# Comparative Quantification of Health Risks 

Global and Regional Burden of Disease Attributable to Selected Major Risk Factors

Volume 1

## Edited by

Majid Ezzati, Alan D. Lopez, Anthony Rodgers and Christopher J.L. Murray


World Health Organization
Geneva

Comparative quantification of health risks : global and regional burden of disease attributable to selected major risk factors / edited by Majid Ezzati . . . [et al.].
$2 \mathrm{v} .+\mathrm{v} .3$ in 1 CD-ROM.
Contents: vol. 1, Childhood and maternal undernutrition-Other nutrition-related risk factors and physical activity-Addictive substancesvol. 2, Sexual and reproductive health-Environmental and occupational risks-Other selected risks-Distribution of risks by poverty-Data analysis and results-Multi-risk assessment.-Annex tables CD-ROM, Population attributable fractions, mortality and disease burden for selected major risk factors.

1. Risk factors 2. Cost of illness 3. Risk assessment
2. Comparative study I. Ezzati, Majid. II. Title: Global and regional burden of disease attributable to selected major risk factors.

ISBN 9241580313
(NLM Classification: WA 105)

## © World Health Organization 2004

All rights reserved. Publications of the World Health Organization can be obtained from Marketing and Dissemination, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland (tel: +41 22791 2476; fax: +41 22791 4857; email: bookorders@who.int). Requests for permission to reproduce or translate WHO publications-whether for sale or for noncommercial distribution-should be addressed to Publications, at the above address (fax: +41 22791 4806; email: permissions@who.int).

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

The World Health Organization does not warrant that the information contained in this publication is complete and correct and shall not be liable for any damages incurred as a result of its use.

## Contents

Volume 1
List of authors ..... vii
Foreword ..... xvii
Richard Peto
Preface ..... xix
Majid Ezzati, Alan D. Lopez, Anthony Rodgers and Christopher J.L. Murray

1. Comparative quantification of health risks: conceptual framework and methodological issues ..... 1
Christopher J.L. Murray, Majid Ezzati, Alan D. Lopez, Anthony Rodgers and Stephen Vander Hoorn
Childhood and maternal undernutrition
2. Childhood and maternal underweight ..... 39
Steven M. Fishman, Laura E. Caulfield, Mercedes de Onis, Monika Blössner, Adnan A. Hyder, Luke Mullany and Robert E. Black
3. Iron deficiency anaemia ..... 163
Rebecca J. Stoltzfus, Luke Mullany and Robert E. Black
4. Vitamin A deficiency ..... 211
Amy L. Rice, Keith P. West Jr. and Robert E. Black
5. Zinc deficiency ..... 257
Laura E. Caulfield and Robert E. Black
Other nutrition-related risk factors and physical inactivity
6. High blood pressure ..... 281
Carlene M.M. Lawes, Stephen Vander Hoorn, Malcolm R. Law, Paul Elliott, Stephen MacMahon and Anthony Rodgers
7. High cholesterol ..... 391
Carlene M.M. Lawes, Stephen Vander Hoorn, Malcolm R. Law and Anthony Rodgers
8. Overweight and obesity (high body mass index) ..... 497
W. Philip T. James, Rachel Jackson-Leach, Cliona Ni Mburchu, Eleni Kalamara, Maryam Shayeghi, Neville J. Rigby, Chizuru Nishida and Anthony Rodgers
9. Low fruit and vegetable consumption ..... 597
Karen Lock, Joceline Pomerleau, Louise Causer and Martin McKee
10. Physical inactivity ..... 729
Fiona C. Bull, Timothy P. Armstrong, Tracy Dixon, Sandra Ham, Andrea Neiman and Michael Pratt
Addictive substances
11. Smoking and oral tobacco use ..... 883
Majid Ezzati and Alan D. Lopez
12. Alcohol use ..... 959
Jürgen Rehm, Robin Room, Maristela Monteiro, Gerhard Gmel, Kathryn Graham, Nina Rehn, Christopher T. Sempos, Ulrich Frick and David Jernigan
13. Illicit drug use ..... 1109
Louisa Degenhardt, Wayne Hall, Matthew Warner-Smith and Michael Lynskey
Volume 2
Sexual and reproductive health
14. Unsafe sex ..... 1177
Emma Slaymaker, Neff Walker, Basia Zaba and Martine Collumbien
15. Non-use and use of ineffective methods of contraception ..... 1255
Martine Collumbien, Makeda Gerressu and John Cleland
Environmental and occupational risk factors
16. Unsafe water, sanitation and hygiene ..... 1321
Annette Prüss-Üstün, David Kay, Lorna Fewtrell and Jamie Bartram
17. Urban air pollution ..... 1353
Aaron J. Cohen, H. Ross Anderson, Bart Ostro, Kiran Dev Pandey, Michal Krzyzanowski, Nino Künzli, Kersten Gutschmidt, C. Arden Pope III, Isabelle Romieu, Jonathan M. Samet and Kirk R. Smith
18. Indoor air pollution from household use of solid fuels ..... 1435
Kirk R. Smith, Sumi Mehta and Mirjam Maeusezabl-Feuz
19. Lead exposure ..... 1495
Annette Prüss-Üstün, Lorna Fewtrell, Philip J. Landrigan and José Luis Ayuso-Mateos
20. Global climate change ..... 1543
Anthony McMichael, Diarmid Campbell-Lendrum, Sari Kovats, Sally Edwards, Paul Wilkinson, Theresa Wilson, Robert Nicholls, Simon Hales, Frank Tanser, David Le Sueur, Michael Schlesinger and Natasha Andronova
21. Selected occupational risk factors ..... 1651
Marisol Concha-Barrientos, Deborah Imel Nelson, Timothy Driscoll, N. Kyle Steenland, Laura Punnett, Marilyn Fingerhut, Annette Prüss-Üstün, James Leigh, Sang Woo Tak and Carlos Corvalan
Other selected risk factors
22. Contaminated injections in health care settings ..... 1803
Anja M. Hauri, Gregory L. Armstrong and Yvan J.F. Hutin
23. Child sexual abuse ..... 1851
Gavin Andrews, Justine Corry, Tim Slade, Cathy Issakidis and Heather Swanston
Distribution of risk factors by poverty
24. Distribution of risks by poverty ..... 1941
Tony Blakely, Simon Hales, Charlotte Kieft, Nick Wilson and Alistair Woodward
Data analysis and results
25. Estimating attributable burden of disease from exposure and hazard data ..... 2129
Stephen Vander Hoorn, Majid Ezzati, Anthony Rodgers, Alan D. Lopez and Christopher J.L. Murray
26. Mortality and burden of disease attributable to individual risk factors ..... 2141
Majid Ezzati, Anthony Rodgers, Alan D. Lopez, Stephen Vander Hoorn and Christopher J.L. Murray
Multi-risk factor assessment
27. Potential health gains from reducing multiple risk factors ..... 2167
Majid Ezzati, Stephen Vander Hoorn, Anthony Rodgers, Alan D. Lopez, Colin D. Mathers and Christopher J.L. Murray
28. Effects of multiple interventions ..... 2191
James Robins, Miguel Hernan and Uwe Siebert
29. Conclusions and directions for future research ..... 2231
Alan D. Lopez, Majid Ezzati, Anthony Rodgers, Stephen Vander Hoorn and Christopher J.L. Murray
Index ..... 2235
CD-ROM
Annex tables
Population attributable fractions, mortality and disease burden forselected major risk factors

## List of AUTHORS

H. Ross Anderson, St George's Hospital Medical School, University of London, London, England (randerso@sghms.ac.uk).

Gavin Andrews, WHO Collaborating Centre for Evidence in Mental Health Policy, St Vincent's Hospital, Sydney, Australia (gavina@crufad.unsw.edu.au).

Natasha Andronova, Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign,Urbana, IL, USA (natasha@uiatma.atmos.uiuc.edu).

Gregory L. Armstrong, Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, GA, USA (GArmstrong@cdc.gov).
Timothy P. Armstrong, Chronic Diseases and Health Promotion, World Health Organization, Geneva, Switzerland (armstrongt@who.int).

José Luis Ayuso-Mateos, Servicio de Psiquiatria, Hospital Universitario de la Princesa, Madrid, Spain (joseluis.ayuso@uam.es).

Jamie Bartram, Protection of the Human Environment, World Health Organization, Geneva, Switzerland (bartramj@who.int).

Robert E. Black, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA (rblack@jhsph.edu).

Tony Blakely, Department of Public Health, Wellington School of Medicine and Health Sciences, University of Otago, Wellington, New Zealand (tblakely@wnmeds.ac.nz).

Monika Blössner, Nutrition for Health and Development, World Health Organization, Geneva, Switzerland (bloessnerm@who.int).

Fiona C. Bull, School of Sport \& Exercise Sciences, Loughborough University, Loughborough, Leicestershire, England (F.C.L.Bull@lboro.ac.uk).

Diarmid Campbell-Lendrum, Protection of the Human Environment, World Health Organization, Geneva, Switzerland (campbelllendrumd@who.int).

Laura E. Caulfield, Department of International Health, Center for Human Nutrition, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA (lcaulfie@jhsph.edu).

Louise Causer, European Centre on the Health of Societies in Transition, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, European London, England.

John Cleland, Centre for Population Studies, London School of Hygiene and Tropical Medicine, London, England (john.cleland@lshtm.ac.uk).

Aaron J. Cohen, Health Effects Institute, Boston, MA, USA (acohen@healtheffects.org).

Martine Collumbien, Centre for Population Studies, London School of Hygiene and Tropical Medicine, London, England (martine.collumbien@bigfoot.com).

Marisol Concha-Barrientos, Asociación Chilena de Seguridad, Gerencia de Salud, Santiago, Chile (gsamcb@gw.achs.cl).

Justine Corry, WHO Collaborating Centre for Evidence in Mental Health Policy, St Vincent's Hospital, Sydney, Australia (j.corry@unsw.edu.au).

Carlos Corvalan, Protection of the Human Environment, World Health Organization, Geneva, Switzerland (corvalanc@who.int).

Mercedes de Onis, Nutrition for Health and Development, World Health Organization, Geneva, Switzerland (deonism@who.int).

Louisa Degenhardt, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia (l.degenhardt@unsw.edu.au)

Tracy Dixon, Cardiovascular Disease, Diabetes and Risk Factor Monitoring Unit, Australian Institute of Health and Welfare, Canberra, Australia (tracy.dixon@aihw.gov.au).

Timothy Driscoll, School of Public Health, University of Sydney, Sydney, Australia (elmatom@optushome.com.au).

Sally Edwards, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, England (edwardss@who.int).

Paul Elliot, Department of Epidemiology and Public Health, Imperial College of Science Technology and Medicine, London, England (p.elliot@ic.ac.uk).

Majid Ezzati, Harvard School of Public Health, Boston, MA, USA (mezzati@hsph.harvard.edu).

Lorna Fewtrell, Centre for Research into Environment and Health, University of Wales, Aberystwyth, Wales (lorna@creh.demon.co.uk).
Marilyn A. Fingerhut, National Institute for Occupational Safety and Health, Washington, DC, USA (mfingerhut@cdc.gov).
Steven M. Fishman, Department of International Health, Center for Human Nutrition, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA (sfishman@jhsph.edu).
Ulrich Frick, Addiction Research Institute, Zurich, Switzerland (ulrich.frick@isf.unizh.ch).

Makeda Gerressu, Department of Primary Care and Population Sciences, Centre for Sexual Health and HIV Research, Royal Free and University College Medical School, London, England (mgerressu@gum.ucl.ac.uk).

Gerhard Gmel, Swiss Institute for the Prevention of Alcohol and Drug Problems, Lausanne, Switzerland (ggmel@sfa-ispa.ch).
Kathryn Graham, Centre for Addiction and Mental Health, London, Canada (kgraham@uwo.ca).

Kersten Gutschmidt, Protection of the Human Environment, World Health Organization, Geneva, Switzerland (gutschmidtk@who.int).

Simon Hales, Department of Public Health, Wellington School of Medicine and Health Sciences, University of Otago, Wellington, New Zealand (shales@wnmeds.ac.nz).
Wayne Hall, Institute for Molecular Biosciences, University of Queensland, Brisbane, Queensland, Australia (w.hall@imb.uq.edu.au).

Sandra Ham, Centers for Disease Control and Prevention, Division of Nutrition and Physical Activity, National Center for Chronic Disease Prevention and Health Promotion, Atlanta, GA, USA (svh6@cdc.gov).

Anja M. Hauri, Staatliches Untersuchungsamt Hessen, Dillenburg, Germany (a.hauri@suah-ldk.hessen.de).

Miguel Hernan, Harvard School of Public Health, Boston, MA, USA (mhernan@hsph.harvard.edu).
Yvan J.F. Hutin, Essential Health Technology, World Health Organization, Geneva, Switzerland (hutiny@who.int).
Adnan A. Hyder, Department of International Health, Health Systems and Management Program, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA (ahyder@jhsph.edu).

Cathy Issakidis, WHO Collaborating Centre for Evidence in Mental Health Policy, St Vincent's Hospital, Sydney, Australia (cathyi@crufad.unsw.edu.au).

Rachel Jackson-Leach, International Obesity Task Force, London, England (rleach@iotf.org; rachjacklea@aol.com).
W. Philip T. James, International Obesity Task Force, London, England (jeanhjames@aol.com).

David Jernigan, Health Policy Institute, Georgetown University, Washington, DC, USA (djernigan@aol.com).

Eleni Kalamara, International Obesity Task Force, London, England (obesity@iotf.org).
David Kay, Centre for Research into Environment and Health, University of Wales, Aberystwyth, Wales (Dave@crehkay.demon.co.uk).

Charlotte Kieft, Department of Public Health, Wellington School of Medicine and Health Sciences, University of Otago, Wellington, New Zealand.

Sari Kovats, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, England (Sari.Kovats@lshtm.ac.uk).

Michal Krzyzanowski, European Center for Environment and Health, World Health Organization, Bonn, Germany (mkr@ecehbonn.euro.who.int).
Nino Künzli, Division of Occupational and Environmental Health, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA (e-mail: Kuenzli@usc.edu).
Philip J. Landrigan, Department of Community \& Preventive Medicine, Mount Sinai School of Medicine, New York, NY, USA (phil.landrigan@mssm.edu).
Malcolm R. Law, Department of Environmental and Preventive Medicine, Wolfson Institute of Preventive Medicine, Barts and The London, Queen Mary's School of Medicine and Dentistry, London, England (M.R.Law@qmul.ac.uk).
Carlene M.M. Lawes, Clinical Trials Research Unit, University of Auckland, Auckland, New Zealand (c.lawes@ctru.auckland.ac.nz).
David Le Sueur (Deceased), Medical Research Council, Durban, South Africa. Deceased.

James Leigh, Centre for Occupational and Environmental Health, School of Public Health, University of Sydney, Sydney, Australia (jleigh@bigpond.com).

Karen Lock, European Centre on the Health of Societies in Transition, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, England (Karen.Lock@lshtm.ac.uk).
Alan D. Lopez, School of Population Health, University of Queensland, Brisbane, Queensland, Australia (e-mail: a.lopez@sph.uq.edu.au).

Michael Lynskey, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia.

Stephen MacMahon, Institute for International Health, University of Sydney, Sydney, Australia (smacmahon@iih.usyd.edu.au).

Mirjam Maeusezahl-Feuz, Division of Epidemiology and Infectious Diseases, Swiss Federal Office of Public Health, Bern, Switzerland (mirjam.maeusezahl@bag.admin.ch).

Colin D. Mathers, Measurement and Health Information, World Health Organization, Geneva, Switzerland (mathersc@who.int).

Martin McKee, European Centre on the Health of Societies in Transition, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, England (martin.mckee@lshtm.ac.uk).

Anthony McMichael, National Centre for Epidemiology and Population Health, Australian National University, Canberra, Australia (Tony.McMichael@anu.edu.au).

Sumi Mehta, Health Effects Institute, Boston, MA, USA (smehta@healtheffects.org).

Maristela Monteiro, Pan American Health Organization, Washington, DC, USA (monteirm@paho.org).

Luke Mullany, Department of International Health, Disease Prevention and Control Program, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA (lmullany@jhsph.edu).

Christopher J.L. Murray, Harvard University Initiative for Global Health, Cambridge, MA, USA (christopher_murray@harvard.edu).
Andrea Neiman, Division of Nutrition and Physical Activity, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA (afn1@cdc.gov).

Deborah Imel Nelson, School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK, USA (imel@ou.edu).

Cliona Ni Mhurchu, Clinical Trials Research Unit, University of Auckland, Auckland, New Zealand (c.nimhurchu@ctru.auckland.ac.nz).

Robert Nicholls, Flood Hazards Research Centre, University of Middlesex, Enfield, England (r.nicholls@mdx.ac.uk).

Chizuru Nishida, Nutrition for Health and Development, World Health Organization, Geneva, Switzerland (nishidac@who.int).

Bart Ostro, California Environmental Protection Agency, Office of Environmental Health Hazard Assessment, Oakland, CA, USA (bostro@oehha.ca.gov).

Kiran Dev Pandey, World Bank, DECRG-IE, Washington, DC, USA (kpandey@worldbank.org).

Richard Peto, Clinical Trial Service Unit and Epidemiological Studies Unit, Medical Sciences Division, University of Oxford, Oxford, England (richard.peto@ctsu.ox.ac.uk).

Joceline Pomerleau, European Centre on the Health of Societies in Transition, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, England (Joceline.Pomerleau@lshtm.ac.uk).

Arden Pope, Department of Economics, Brigham Young University, Provo, UT, USA (arden_pope@byu.edu).

Michael Pratt, Division of Nutrition and Physical Activity, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA (mxp4@cdc.gov).

Annette Prüss-Üstün, Protection of the Human Environment, World Health Organization, Geneva, Switzerland (pruessa@who.int).

Laura Punnett, University of Massachusetts, Lowell, MA, USA (Laura_Punnett@UML.edu).

Jürgen Rehm, Centre for Addiction and Mental Health, Toronto, Canada (jtrehm@aol.com).

Nina Rehn, UNODC—United Nations Office on Drugs and Crime, Viet Nam Field Office, Hanoi, Vietnam (nina.rehn@unodc.org).

Amy L. Rice, Department of International Health, Center for Human Nutrition, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA (arice@jhsph.edu).

Neville J. Rigby, International Obesity Task Force, London, England (nrigby@iotf.org).

James Robins, Harvard School of Public Health, Boston, MA, USA (robins@hsph.harvard.edu).

Anthony Rodgers, Clinical Trials Research Unit, University of Auckland, Auckland, New Zealand (a.rodgers@ctru.auckland.ac.nz).
Isabelle Romieu, National Institute for Public Health, Cuernavaca, Mexico (iromieu@correo.insp.mx).

Robin Room, Centre for Social Research on Alcohol and Drugs, Stockholm University, Stockholm, Sweden (Robin.Room@sorad.su.se).

Jonathan M. Samet, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA (jsamet@jhsph.edu).

Michael Schlesinger, Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA (schlesin@atmos.uiuc.edu).

Christopher T. Sempos, University at Buffalo, Buffalo, NY, USA (sempos@buffalo.edu).

Maryam Shayeghi, International Obesity Task Force, London, England (obesity@iotf.org).

Uwe Siebert, Institute for Technology Assessment, Massachusetts General Hospital, Boston, MA, USA (usiebert@hsph.harvard.edu).

Tim Slade, WHO Collaborating Centre for Evidence in Mental Health Policy, St Vincent's Hospital, Sydney, Australia (tims@crufad.unsw.edu.au).

Emma Slaymaker, Centre for Population Studies, London School of Hygiene and Tropical Medicine, London, England (emma.slaymaker@lshtm.ac.uk).

Kirk R. Smith, School of Public Health, University of California, Berkeley, CA, USA (krksmith@berkeley.edu).
N. Kyle Steenland, Rollins School of Public Health, Emory University, Atlanta, GA, USA (nsteenl@sph.emory.edu).

Rebecca J. Stoltzfus, Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA (rjs62@cornell.edu).

Heather Swanston, WHO Collaborating Centre for Evidence in Mental Health Policy, St Vincent's Hospital, Sydney, Australia (heathers@crufad.unsw.edu.au).

Sang Woo Tak, University of Massachusetts, Lowell, MA, USA (SangWoo_Tak@student.uml.edu).

Frank Tanser, Africa Centre for Population Studies and Reproductive Health, Mtubatuba, South Africa (tanserf@mrc.ac.za).

Stephen Vander Hoorn, Clinical Trials Research Unit, University of Auckland, Auckland, New Zealand (svande@ctru.auckland.ac.nz).
Neff Walker, Joint United Nations Programme on HIV/AIDS (UNAIDS), Geneva, Switzerland (walkern@unaids.org).

Matthew Warner-Smith, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia.

Keith P. West Jr, Department of International Health, Center for Human Nutrition, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA (kwest@hsph.edu).

Paul Wilkinson, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, England (Paul.Wilkinson@lshtm.ac.uk).

Nick Wilson Department of Public Health, Wellington School of Medicine and Health Sciences, University of Otago, Wellington, New Zealand (nwilson@actrix.gen.nz).

Theresa Wilson, University of Middlesex, Enfield, England (Theresa7@mdx.ac.uk).

Alistair Woodward, School of Population Health, University of Auckland, Auckland, New Zealand (a.woodward@auckland.ac.nz).

Basia Zaba, Centre for Population Studies, London School of Hygiene and Tropical Medicine London, England (basia.zaba@lshtm.ac.uk).

## Acknowledgements

This book is the product of more than four years of collaborative effort involving a large group of scientists across the world. This collaboration, led by the World Health Organization (WHO) and known as the comparative risk assessment (CRA) project, is one of the largest and most comprehensive research projects ever undertaken by WHO. The successful completion of the project, and this book, would not have been possible without great effort from a number of individuals and organizations.

More than 150 scientists, experts on the various risk factors and methodological aspects, reviewed the chapters in this book with great care, many of them multiple times, to ensure the scientific integrity, completeness and plausibility of the material and the conclusions. Many individuals and organizations donated their unpublished data to the project to fill some of the existing data gaps. Stephen Vander Hoorn played a key role in the design of the conceptual and methodological framework for the analysis. The literally millions of calculations leading to the disease burden estimates reported in these volumes are the result of his invaluable technical and statistical abilities and perseverance. Our colleagues David Evans, Emmanuela Gakidou, Mie Inoue, Carlene Lawes, Rafael Lozano, Doris Ma Fat, Susan Piccolo, Chalapati Rao, Joshua Salomon, Kenji Shibuya and Niels Tomijima, provided substantial motivational and intellectual support and assistance with the Global Burden of Disease databases. We are especially grateful to Marie-Claude von Rulach for her extraordinary efforts in helping to manage the network of contributors and the collaborators meetings, and for secretarial support. Technical discussions on burden of disease methodology with Colin Mathers and Claudia Stein contributed greatly to improving the scientific basis of the estimates.

The final editing and production of a book of this magnitude, with contributions from numerous authors, is a challenging and complex task. Kaarina Klint and Kai Lashley have managed that process with great care and commitment. Editorial assistance from Stanislava Nikolova, Anna Moore and Margaret Squadrani is highly appreciated. Our thanks also to our technical and copy editors Andrew Colborne, Heidi Mattock, Gillian Stanbridge and Frank Theakston for their efforts to ensure consistency of style across chapters originally written with varied disciplinary and personal styles. Colin Mathers, David Evans
and Ties Boerma generously and efficiently coordinated the production process in its final stages.

The initial concept of the cover artwork was developed by Saied Ezzati; the final arrangement of the cover design was done by Mark Forrest and Emmanuela Gakidou.

The research published in this book forms part of the larger Global Burden of Disease 2000 project, funded by a grant from the National Institute on Aging (PO1-AG17625). Their support is gratefully acknowledged.

## Foreword

During the twentieth century reliable cause-specific mortality statistics became available for many countries, culminating in the Global Burden of Disease project which, during the 1990s, provided estimates for different regions of the world (with, obviously, varying degrees of reliability) of the numbers of deaths due to major diseases, and of the amounts of "disability-adjusted" loss of healthy life from those diseases. The present study goes further, and seeks to estimate the amounts of death and disability due to the main avoidable causes of those diseases. Its preliminary conclusions underlay the 250-page World Health Organization report on "Reducing Risks" (2002), the aim of which was to summarize, for the first time, the amount of death and disability in each of 14 subregions of the world that is attributable not to particular diseases, but to particular avoidable risk factors.

Such attributions of causality throw up, of course, many more difficulties than were encountered in the previous studies of the Global Burden of Disease, which merely tried to classify deaths by the one main disease (or type of accident or violence) that underlay them. For, one death may have several avoidable causes. For example, if a poorly nourished child dies of measles, should "the cause" be thought of as exposure to the virus, or as the lack of measles vaccination (in that child or in the community), or as the poor diet (low in protein, energy and certain micronutrients) that prevented recovery from the illness? The most appropriate answer, if we want to prevent such deaths, is that each of these factors should be thought of as "a cause" of a certain proportion of the childhood deaths from measles. That is what the authors of the World bealth report 2002 tried to do, and in the present much more detailed series of monographs they explain to the interested (or disputative) reader much more about their main conclusions, and about how they reached those conclusions. This is important, because over the years some of the conclusions may need to be revised, as more detailed studies are undertaken or as exposure and disease patterns evolve.

For many decades it has been recognized increasingly clearly by those concerned with global health that much can affordably be achieved even in relatively poor countries if resources are directed to the major diseases of childhood and early adult life, and more recently the affordable avoidability of much other adult mortality and morbidity has been recognized
(see The Health of Adults in the Developing World). Ten years ago, the 1993 World Bank report, Investing in health (together with its companion volume, Disease Control Priorities in Developing Countries) was extremely influential in consolidating these ideas and getting them accepted, and acted upon, by the major international economic institutions.

But, any such cost-effectiveness calculations require, among other things, reliable estimates of effectiveness, and the present report goes further than any other in providing estimates of just how much mortality and morbidity could be avoided by addressing particular causes of disease. In many parts of the world (the main exceptions being where political disruption or HIV predominate) the risk of premature death has been reduced by more than half over the past few decades, and premature death can be halved again over the next few decades if the major intervention options are pursued to control disease and injury, and their causes.

This book will greatly facilitate such progress. It is well organized, stimulating and is an important part of a political and scientific process that is already preventing many millions of deaths a year, and will prevent many more millions of deaths a year in the future.

Sir Richard Peto
Professor of Medical Statistics and Epidemiology
University of Oxford, England

## Preface

A clear understanding of the role and relative magnitude of diseases, injuries and their underlying causes-and effective and affordable interventions to reduce them-should guide policies and programmes for health development. Over the centuries, the health of populations has improved because science has helped us understand the main causes of disease affecting large populations, and how technologies or programmes can be delivered to reduce hazards among those affected or at risk.

While the monitoring and analysis of diseases and mortality in populations has been largely undertaken by actuaries and demographers, much of the work on causes of disease has emanated from research in fields such as epidemiology, toxicology and physiology, which focus on micro-level analysis. By its very nature, this research has quantified hazards in the study population, with its specific characteristics. This body of knowledge has had tremendous application in reducing the established causes of disease, from smoking to iodine deficiency, in many populations. The broader, policy-relevant issue of population effects of exposure to risks, however, has remained under-explored relative to our documentation of established diseases. Thus, while there has been decades of epidemiological research into the leading causes of many major diseases, from childhood diarrhoea to ischaemic heart disease, there have been few attempts to estimate the population-level effects of various exposures, either for specific countries and groups of countries, or for the world and its major regions.

During the last quarter of the twentieth century, a number of works have addressed both the methodological and empirical aspects of population-wide effects of major causes of diseases. Examples include the development of cancer risk models and methods to forecast the health of ageing populations based on their causal determinants, and the estimates of mortality due to risk factors such as smoking, asbestos and childhood malnutrition. This gradual establishment of "risk assessment" or "risk quantification" has been driven partly by the academic curiosity of individual researchers and partly by the demands of regulatory agencies and public policy for better quantitative evidence on the health implications of certain risk exposures.

This book provides a comprehensive assessment of the health effects caused by a range of exposures that are known to be hazardous to human
health. Its origins lie in the expressed need by policy and advocacy groups for comparable data on risk factor exposure and effects in populations. The only previous attempt to quantify risk factor burden worldwide, the Global Burden of Disease (GBD) 1990 project, was affected by a lack of conceptual and methodological comparability across risk factors; the analysis of each risk was constrained by its own disciplinary tradition. It nonetheless stimulated debate about the crucial role of risk factor assessment as a cornerstone of the evidence base for public health action: for instance, the leading risk factor in 1990, malnutrition, accounted for substantially more disease burden worldwide than the leading cause of disease at that time, acute lower respiratory infections.

A key concern of the current work on risks to health is to provide a degree of conceptual and methodological consistency and comparability across risk factors. The results reported in this book, therefore, differ in a number of important ways from those of GBD 1990: a new analytical framework and consistent set of definitions on "risk factor exposure" have been used to enhance comparability; the number of exposures assessed has more than doubled; and the analyses have benefited from more recent and thorough research into causality and geographical variations in population exposures and health effects.

The scope of risks to health studied in this book covers many of the most important hazards to health addressed by various fields of scientific enquiry. Arguably, there are hundreds of risk exposures that are harmful to health; and there are important implications for better understanding the disease burden they cause across the world. We have selected only a relatively small number of exposures for quantification in this book, largely determined by the availability of scientific research about their prevalence and health effects in different parts of the world. It was also important to make choices about the definition of each risk factor. Given the close interrelationships among diet, exercise and physiological risks on the one hand, or among water, sanitation and personal hygiene on the other, the exact definition of what a "risk factor" is, itself requires careful attention. That a particular risk factor like dietary fat intake does not appear in this book does not, of course, imply that it is of limited relevance; or that exposure to lead has been assessed separately from urban air pollution does not override their close linkages. Rather, we have limited ourselves to risk factors for which there was good potential for satisfactory quantification of population exposure distributions and health effects using the existing scientific evidence and available data, and for which intervention strategies are available or might be envisioned to modify their impact on disease burden.

The chapters in Volumes 1 and 2 of this book fall into two broad categories: those that address specific risk factors, and those that provide conceptual, methodological or empirical links across risks. The book begins with a description of some of the important conceptual and methodological issues in quantifying risk factor burden in a consistent
and comparable framework. This is followed by twenty-two chapters, organized under six broad sections, each of which present the background and the scientific evidence and empirical findings for individual risks. These are followed by an attempt to quantify the distributions of some risks by poverty levels. While much is known about the relationship between poverty and health, it is undoubtedly too complex and population-specific to be adequately assessed in a single quantification effort. The research reported here is therefore limited to a simple mapping of risks by poverty, based on existing data. Following the risk factor chapters, the calculus of estimating the burden of disease attributable to each risk factor from exposure and hazard data is presented, followed by a chapter that summarizes the results for individual risk factors.

Many policies and programmes affect multiple risks simultaneously, motivating an assessment of the disease burden from multiple risk factors. The focus on joint exposures and hazards is particularly important because diseases and injuries are almost always caused by multiple risk factors, which may act together on disease processes, or have effects mediated through each other. We have therefore included two chapters on the joint effects of multiple risk factor exposures. The final chapter of the book provides conclusions and recommendations for future research, based on the analytical findings presented in the book, as well as the gaps in data and scientific knowledge that increased uncertainty in quantifying risk factor burden reported here.

The specific risk factor chapters have been grouped according to clusters of exposures likely to be of similar scientific or policy interest. Volume 1 begins with four chapters on childhood and maternal undernutrition, which collectively cause a significant proportion of the childhood infectious disease burden worldwide. With substantial reductions in child mortality over the past few decades in many countries, the focus of scientific enquiry has progressively moved to improving our understanding of the causes of disease and injury among adults. The next five chapters address the various distal (e.g. exercise), more proximal (e.g. overweight and obesity), and physiological (e.g. suboptimal cholesterol levels) risks that are clustered together under the label of nutrition and physical activity. The last section in Volume 1 and the first section in Volume 2, addictive substances and sexual and reproductive health, include the major lifestyle and behavioural risks that are widespread in many societies and, despite being the subject of scientific enquiry and public health intervention for decades, present a range of complexities in risk quantification.

The risk factors that are a part of the physical environment of households (e.g. indoor air pollution from household solid fuel use), communities (e.g. urban air pollution), or specific subgroups (e.g. occupational risk factors) are the next group of risks assessed in Volume 2. The next two chapters, childhood sexual abuse and contaminated medical injec-
tions, do not fall into any of the above broad categories and are presented independently. These two chapters, each representing a risk factor that affects multiple important diseases, illustrate the potential for risk assessment as an analytical tool for improving the public health evidence base across a wide spectrum of health concerns.

In each of the specific risk factor chapters, the authors have provided a definition of the risk factor and introduced an "exposure variable" that best reflects the distribution of hazards in the population. The complexity of disease causation mechanisms (e.g. sexual behaviour and sexually transmitted infections), and the limitations posed by available data and epidemiological studies (e.g. physical inactivity or indoor smoke from solid fuels) have been important factors in the choice of exposure variable. Coupled with this is the choice of a "theoretical-minimum-risk population exposure distribution", which can serve as a consistent baseline for assessing attributable disease burden across difference risks. For some risks such as smoking or childhood abuse, the theoretical-minimum-risk population exposure distribution is obviously zero exposure for the whole population; for others the choice of baseline exposure distribution is less obvious, either because zero exposure is not definable (e.g. blood pressure) or because it may not lead to the lowest risk level in some populations (e.g. alcohol). Each chapter includes current estimates of exposure distributions by age and sex for 14 epidemiological subregions. The chapters also examine in detail the evidence for health outcomes, including the evidence for causality and the estimates of hazard (disease-specific) associated with each level of exposure. Each chapter then concludes with summary results of the burden of disease and injury in 2000 attributable to the risk factor, and when possible using existing evidence and knowledge, estimates of projected future exposure to the risk.

The CD-ROM attached contains detailed tables on the various components of disease burden (i.e. deaths, years of life lost [YLL] due to premature mortality, and disability-adjusted life years [DALYs]) attributable to each risk factor by age, sex and the 14 epidemiological subregions of the world used by the World Health Organization (WHO) in the World health report 2002. The 191 Member States of WHO were divided into five mortality strata on the basis of their levels of child mortality (under five years of age) and 15-59-year-old male mortality. When these mortality strata are applied to the six WHO regions, they produce 14 epidemiological subregions, which are used throughout this book (Table 1).

This book is the culmination of over four years of scientific enquiry and data collection, collectively known as the comparative risk assessment (CRA) project, coordinated by WHO and involving over 100 scientists worldwide. The book is also one of the several planned outputs of the GBD 2000 project which includes multiple analytical and empirical perspectives on global population health. The importance of the collaborative effort in the CRA project goes beyond having leading

## Table I The 14 GBD epidemiological subregions

| WHO region | Mortality stratum ${ }^{\text {a }}$ | Countries |
| :---: | :---: | :---: |
| AFR | D | Algeria, Angola, Benin, Burkina Faso, Cameroon, Cape Verde, Chad, Comoros, Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, GuineaBissau, Liberia, Madagascar, Mali, Mauritania, Mauritius, Niger, Nigeria, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Togo |
|  | E | Botswana, Burundi, Central African Republic, Congo, Côte d'lvoire, Democratic Republic of the Congo, Eritrea, Ethiopia, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, South Africa, Swaziland, Uganda, United Republic of Tanzania, Zambia, Zimbabwe |
| AMR | A | Canada, Cuba, United States of America |
|  | B | Antigua and Barbuda, Argentina, Bahamas, Barbados, Belize, Brazil, Chile, Colombia, Costa Rica, Dominica, Dominican Republic, El Salvador, Grenada, Guyana, Honduras, Jamaica, Mexico, Panama, Paraguay, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Suriname, Trinidad and Tobago, Uruguay, Venezuela |
|  | D | Bolivia, Ecuador, Guatemala, Haiti, Nicaragua, Peru |
| EMR | B | Bahrain, Cyprus, Iran (Islamic Republic of), Jordan, Kuwait, Lebanon, Libyan Arab Jamahiriya, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Tunisia, United Arab Emirates |
|  | D | Afghanistan, Djibouti, Egypt, Iraq, Morocco, Pakistan, Somalia, Sudan, Yemen |
| EUR | A | Andorra, Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, San Marino, Slovenia, Spain, Sweden, Switzerland, United Kingdom |
|  | B | Albania, Armenia, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Georgia, Kyrgyzstan, Poland, Romania, Serbia and Montenegro, Slovakia, Tajikistan, The former Yugoslav Republic of Macedonia, Turkey, Turkmenistan, Uzbekistan |
|  | C | Belarus, Estonia, Hungary, Kazakhstan, Latvia, Lithuania, Republic of Moldova, Russian Federation, Ukraine |
| SEAR | B | Indonesia, Sri Lanka, Thailand |
|  | D | Bangladesh, Bhutan, Democratic People's Republic of Korea, India, Maldives, Myanmar, Nepal |
| WPR | A | Australia, Brunei Darussalam, Japan, New Zealand, Singapore |
|  | B | Cambodia, China, Cook Islands, Fiji, Kiribati, Lao People's Democratic Republic, Malaysia, Marshall Islands, Micronesia (Federated States of), Mongolia, Nauru, Niue, Palau, Papua New Guinea, Philippines, Republic of Korea, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu, Viet Nam |

[^0]researchers for multiple risk factors working simultaneously on the same project. Rather, the interactions of these researchers, with a core network of scientists applying a common analytical framework and methods, has ensured greater consistency and comparability in using and evaluating scientific evidence across risks. As a result, our understanding of the comparative extent of disease burden caused by various exposures worldwide has advanced, and key areas of scientific enquiry necessary to better inform policies to reduce risks have been elucidated. Health advocates and those entrusted with policy and programme development to promote better health now have a more comparable empirical assessment of the hazards to health worldwide, and thus a firmer basis for public health action. We hope that the methodological and empirical findings reported in these volumes will indeed serve as the stimulus for global, regional and national policy action to reduce key hazards to health for decades to come.

Majid Ezzati<br>Alan D. Lopez<br>Anthony Rodgers<br>Christopher J.L. Murray

## Chapter I

# Comparative quantification <br> OF HEALTH RISKS: CONCEPTUAL <br> FRAMEWORK AND <br> METHODOLOGICAL ISSUES 

Christopher J.L. Murray, Majid Ezzati, Alan D. Lopez, Anthony Rodgers and<br>Stephen Vander Hoorn

## 1. Introduction

Detailed description of the level and distribution of diseases and injuries, and their causes are important inputs to strategies for improving population health. Data on disease or injury outcomes alone, such as death or hospitalization, tend to focus on the need for palliative or curative services. Reliable and comparable analysis of risks to health, on the other hand, is key for preventing disease and injury. A substantial body of work has focused on the quantification of causes of mortality, and more recently, the burden of disease (Murray and Lopez 1997; Preston 1976). Analysis of morbidity and mortality due to risk factors, however, has frequently been conducted in the context of methodological traditions of individual risk factors and in a limited number of settings (Kunzli et al. 2000; Leigh et al. 1999; McGinnis and Foege 1993; Peto et al. 1992; Single et al. 1999; Smith 2000; Smith et al. 1999; Willet 2002). The principal conclusions of this body of work are as follows:

- Causal attribution of morbidity and mortality to risk factors has been estimated relative to zero or some other constant level of population exposure. This single, constant baseline, although illustrating the total

Previously published: Copyright © 2003, Murray et al.; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL—http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=156894. Population Health Metrics, 2003, 1:1.
magnitude of the risk, does not provide visions of population health under other alternative exposure distribution scenarios.

- Intermediate stages and interactions in the causal process have not been considered in the causal attribution calculations. As a result, attributable burden could be calculated only for those risk factor-disease combinations for which epidemiological studies had been conducted (often limited to individual risks).
- Causal attribution has often taken place using exposure and/or outcome at one point in time or over an arbitrary period of time (for notable exceptions see the works of Manton and colleagues [Manton et al. 1993b, 1994; Yashin et al. 1986] and Robins [Robins 1986, 1987, 1999a, 1999b; Robins and Greenland 1991; Robins et al. 1999]). Such "counting" of adverse events (such as death) has not been able to clearly distinguish between those cases that would not have occurred in the absence of the risk factor and those where occurrence would have been delayed. More generally, this approach is unable to consider the accumulated effects of time-varying exposure to a risk factor-in the form of years of life lost prematurely or lived with disability.
- The outcome has been morbidity or mortality due to specific disease(s) without conversion to a comparable unit, making comparison among different diseases and/or risk factors difficult.

To allow the assessment of risk factors in a unified framework while acknowledging risk-factor specific characteristics, the comparative risk assessment (CRA) module of the Global Burden of Disease (GBD) 2000 study is a systematic evaluation of the changes in population health which would result from modifying the population distribution of exposure to a risk factor or a group of risk factors (Murray and Lopez 1999). This unified framework for describing population exposure to risk factors and their consequences for population health is an important step in linking the growing interest in the causal determinants of health across a variety of public health disciplines from natural, physical, and medical sciences to the social sciences and humanities. In particular, in the CRA framework:

- The burden of disease due to the observed exposure distribution in a population is compared with the burden from a hypothetical distribution or series of distributions, rather than a single reference level such as the non-exposed population.
- Multiple stages in the causal network of interactions among risk factor(s) and disease outcome are considered to allow making inferences about combinations of risk factors for which epidemiological studies have not been conducted, including the joint effects of changes in multiple risk factors.
- The health loss due to risk factor(s) is calculated as a time-indexed "stream" of disease burden due to a time-indexed "stream" of exposure.
- The burden of disease and injury is converted into a summary measure of population health, which allows comparing fatal and non-fatal outcomes, also taking into account severity and duration.

It is important to emphasize that risk assessment, as defined above, is distinct from intervention analysis, whose purpose is to estimate the benefits of a given intervention or group of interventions in a specific population and at a specific time. Rather, risk assessment aims at mapping alternative population health scenarios to changes in distribution of exposure to risk factors over time, irrespective of whether exposure change is achievable using existing interventions. Therefore, while intervention analysis is a valuable input into cost-effectiveness studies, risk assessment contributes to assessing research and policy options for reducing disease burden by changing population exposure to risk factors.

Summary measures of population health (SMPH) and their use in burden of disease analysis are discussed elsewhere (Murray 1996; Murray et al. 2002). The next three sections of this chapter address the conceptual basis and methodological issues for the remaining three points above. We then discuss the sources and quantification of uncertainty.

## 2. Causal attribution of SMPH to

 RISK FACTORSMathers et al. (2002) describe two traditions for causal attribution of health determinants, outcomes, or states: categorical attribution and counterfactual analysis. In categorical attribution, an event such as death is attributed to a single cause (such as a disease or risk factor) or group of causes according to a defined set of rules (hence $100 \%$ of the event is attributed to the single cause or group of causes). The International Classification of Disease system's (ICD) attribution of causes of death (WHO 1992) and attribution of some injuries to alcohol or occupational conditions are examples of categorical attribution. In counterfactual analysis, the contribution of one or a group of diseases, injuries or risk factors to a summary measure of population health is estimated by comparing the current or future levels of the summary measure with the levels that would be expected under some alternative hypothetical scenario, including the absence of or reduction in the disease(s) or risk factor(s) of interest. This hypothetical scenario is referred to as the counterfactual (see Maldonado and Greenland 2002 for a discussion of conceptual and methodological issues in the use of counterfactuals).

In theory, causal attribution of a summary measure to risk factors can be done using both categorical and counterfactual approaches. For
example, categorical attribution has been used in attribution of diseases and injuries to occupational risk factors in occupational health registries (Leigh et al. 1999) and attribution of motor vehicle accidents to alcohol consumption. In general however, categorical attribution of SMPH to risk factors overlooks the fact that many diseases have multiple causes (Rothman 1976). The epidemiological literature has commonly used the counterfactual approach for the attribution of a summary measure to a risk factor, and compared mortality or disability from the current distribution of exposure to the risk factor to that expected under an alternative exposure scenario.

The dominant counterfactual exposure distribution in these studies has been zero exposure for the whole population (or a fixed non-zero level where zero is not possible such as the case of blood pressure when defined as presence or absence of hypertension). The basic statistic obtained in this approach is the population attributable fraction (PAF) defined as the proportional reduction in disease or death that would occur if exposure to the risk factor were reduced to zero, ceteris paribus (Cole and MacMahon 1971; Eide and Heuch 2001; Greenland 1984; Levin 1953; MacMahon and Pugh 1970; Miettinen 1974; Ouellet et al. 1979; Rockhill et al. 1998; Uter and Pfahlberg 2001). ${ }^{1}$ The attributable mortality, incidence or burden of disease due to the risk factor, $A B$, is then given as $A B=P A F \times B$ where $B$ is the total burden of disease from a specific cause or group of causes affected by the risk factor with a relative risk of $R R$ :

$$
\begin{equation*}
P A F=\frac{P(R R-1)}{P(R R-1)+1} \tag{1a}
\end{equation*}
$$

The exposed population may itself be divided into multiple categories based on the level or length of exposure, each with its own relative risk. With multiple ( $n$ ) exposure categories, the PAF is given by the following generalized form:

$$
\begin{equation*}
P A F=\frac{\sum_{i=1}^{n} P_{i}\left(R R_{i}-1\right)}{\sum_{i=1}^{n} P_{i}\left(R R_{i}-1\right)+1} \tag{1b}
\end{equation*}
$$

Although choosing zero as the reference exposure may be useful for some purposes, it is a restricting assumption for others. The contribution of a risk factor to disease or death can alternatively be estimated by comparing the disease burden due to the observed exposure distribution in a population with that from another distribution (rather than a single reference level such as non-exposed) as described by the generalized
"potential impact fraction" equation (Drescher and Becher 1997; Eide and Heuch 2001; Walter 1980).

$$
\begin{equation*}
P I F=\frac{\int_{x=0}^{m} R R(x) P(x) d x-\int_{x=0}^{m} R R(x) P^{\prime}(x) d x}{\int_{x=0}^{m} R R(x) P(x) d x} \tag{2a}
\end{equation*}
$$

where $R R(x)$ is the relative risk at exposure level $x, P(x)$ is the population distribution of exposure, $P^{\prime}(x)$ is the counterfactual distribution of exposure, and $m$ the maximum exposure level. The first and second terms in the numerator of Equation 2a therefore represent the total exposureweighted risk of mortality or disease in the population under current and counterfactual exposure distributions. The corresponding relationship when exposure is described as a discrete variable with $n$ levels is given by:

$$
\begin{equation*}
\text { PIF }=\frac{\sum_{i=1}^{n} P_{i} R R_{i}-\sum_{i=1}^{n} P_{i}^{\prime} R R_{i}}{\sum_{i=1}^{n} P_{i} R R_{i}} \tag{2b}
\end{equation*}
$$

In addition to relaxing the assumption of the no-exposure group as the reference, analysis based on a broader range of distributions has the advantage of allowing multiple comparisons with multiple counterfactual scenarios. Equation 2a can be further generalized to consider counterfactual relative risks (i.e. relative risk may depend on other risks, new technology, medical services, etc.). For example the relative risk of injuries as a result of alcohol consumption may depend on road conditions and traffic law enforcement. Similarly, people employed in the same occupation may have different risks of occupational injuries because of different safety measures. Therefore, a more general form of Equation 2 a is given by:

$$
\begin{equation*}
P I F=\frac{\int_{x=0}^{m} R R(x) P(x) d x-\int_{x=0}^{m} R R^{\prime}(x) P^{\prime}(x) d x}{\int_{x=0}^{m} R R(x) P(x) d x} \tag{2c}
\end{equation*}
$$

### 2.1 Counterfactual exposure distributions

Various criteria may determine the choice of the counterfactual exposure distributions. Greenland (2002) has discussed some of the criteria for the choice of counterfactuals, arguing that the counterfactuals should
be limited to actions that can be implemented (e.g. anti-smoking campaigns), rather than the effects of removing the outcomes targeted by those actions (e.g. smoking cessation) because, in practice, the implementation of counterfactuals for one risk factor or disease may affect other risks. The solution to Greenland's concern, however, is better analytical techniques for estimating joint risk factor effects, rather than abandoning non-intervention-based counterfactuals which, as argued by Mathers et al. (2002), is a limiting view. Estimating the contributions of risk factors to disease burden and the benefits of their removal, even in the absence of known interventions, can provide an understanding of their role in population health and visions of population health under different scenarios of risk factor exposure. This knowledge of risk factor effects can provide valuable input into public health policies and priorities, as well as research and development.

Murray and Lopez (1999) introduced a taxonomy of counterfactual exposure distributions that, in addition to identifying the size of risk, provides a mapping to policy implementation options. These categories include the exposure distributions corresponding to theoretical minimum risk, plausible minimum risk, feasible minimum risk and cost-effective minimum risk. Theoretical minimum risk refers to the exposure distribution that would result in the lowest population risk, irrespective of whether currently attainable in practice. Plausible minimum refers to a distribution which is imaginable, and feasible minimum is one that has been observed in some population. Finally, cost-effective minimum considers the cost of exposure reduction (through the set of known costeffective interventions) as an additional criterion for choosing the alternative exposure scenario.

In addition to illustrating the total magnitude of disease burden due to a risk factor, the theoretical-minimum-risk distribution (or the current difference between theoretical and plausible or feasible risk levels) can guide research and development resources towards those risk factors for which the mechanisms of reduction (i.e. interventions) are currently underdeveloped. For example, if the reduction in the burden of disease due to improved medical injection safety is high and the methods for risk reduction are well-known, so that plausible/feasible and theoretical minima are identical, then current policy may have to be focused on the implementation of such methods. On the other hand, if there are large differences between plausible/feasible and theoretical minima risk levels for blood lipids or body mass index (BMI) (Powles and Day 2002), then research on reduction methods and their implementation should be encouraged. For this reason the total magnitude of the burden of disease due to a risk factor, as illustrated by the theoretical minimum, provides a tool for considering alternative visions of population health and setting research and implementation priorities.

Biological principles as well as considerations of equity would necessitate that, although the exposure distribution for theoretical minimum
risk may depend on age and sex, it should in general be independent of geographical region or population. Exceptions to this are, however, unavoidable. An example would be the case of alcohol consumption, which in limited quantities and when drunk in certain patterns has beneficial effects on cardiovascular mortality, but is always harmful for other diseases such as cancers and accidents (Puddey et al. 1999). In this case, the composition of the causes of death as well as drinking patterns in a region would determine the theoretical-minimum-risk distribution. In a population where cardiovascular diseases are a dominant cause of mortality, the theoretical-minimum-risk exposure distribution may be nonzero with moderate drinking patterns, whereas in a population with binge drinking and a large burden from injuries the theoretical minimum would be zero. Feasible and cost-effective distributions, on the other hand, may vary across populations based on the current distribution of the burden of disease and the resources and institutions available for exposure reduction.

The above categories of counterfactual exposure distributions are based on the burden of disease in the population as a whole. Counterfactual exposure distributions may also be considered based on other criteria. For example, a counterfactual distribution based on equity would be one in which the highest exposure group (or the group with the highest burden of disease) would be shifted towards low exposure values. Further, such equitable counterfactual distributions for each risk factor may themselves be categorized into theoretical (most equitable), plausible, feasible and cost-effective as described above. Similarly, a counterfactual distribution that focuses on the most susceptible groups in the population is one that gives additional weight to lowering the exposure of this group. Therefore, by permitting comparison of disease burden under multiple exposure distributions based on a range of crite-ria-including, but not limited to, implementation and cost, equity and research prioritization-relaxing the assumption of a constant exposure baseline provides an effective policy and planning tool.

### 2.2 EXPOSURE DISTRIBUTION FOR THEORETICAL MINIMUM RISK

In one taxonomy, risk factors such as those in the GBD project (Ezzati et al. 2002; see also the risk factor chapters in this book) can be broadly classified as physiological, behavioural, environmental and socioeconomic. Some general principles that guide the choice of theoretical-minimum-risk exposure distribution for each category are:

1. Physiological risk factors: This group includes those factors that are physiological attributes of humans, such as blood pressure or blood lipids, and at some level result in increased risk. Since these factors are necessary to sustain life, their "exposure-response" relationship is J-shaped or U-shaped, and the theoretical-minimum-risk distribution is non-zero. For such risk factors, the choice of optimal exposure
needs to be based on empirical evidence from different scientific disciplines. For example, epidemiological research on blood pressure and cholesterol have illustrated a monotonically increasing dose-response relationship for mortality even at low levels of these risk factors (Chen et al. 1991; Eastern Stroke and Coronary Heart Disease Collaborative Research Group 1998; MacMahon et al. 1990; Prospective Studies Collaboration 1995). But, given the role of these factors in sustaining life, this relationship must flatten and reverse at some level. In the blood pressure and cholesterol assessment, a theoretical-minimum-risk exposure distribution with a mean of 115 mmHg for systolic blood pressure and $3.8 \mathrm{mmol} / \mathrm{l}$ for total cholesterol (each with a small standard deviation) were used (Ezzati et al. 2002). This distribution corresponds to the lowest levels at which the dose-response relationship has been characterized in meta-analyses of cohort studies (Chen et al. 1991; Eastern Stroke and Coronary Heart Disease Collaborative Research Group 1998; MacMahon et al. 1990; Prospective Studies Collaboration 1995). Further, these levels of blood pressure and cholesterol are consistent with levels seen in populations which have low levels of cardiovascular disease, such as the Yanomamo Indians (Carvalho et al. 1989) and rural populations in China (He et al. 1991a, 1991b), Papua New Guinea (Barnes 1965; Carvalho et al. 1989), and Africa (Mann et al. 1964). Although meta-analyses of randomized clinical trials have indicated that blood pressure and cholesterol levels may be lowered substantially with no adverse effects (LaRosa et al. 1999; Pignone et al. 2000), it is difficult to justify an optimal exposure distribution lower than that measured in population-based studies, since lower levels in individuals may be caused by factors such as pre-existing disease. Arguments from evolutionary biology would also support the choice of a lower bound on the optimal distribution based on historical survival of populations who are not substantially exposed to factors that raise blood pressure or cholesterol.
2. Behavioural risk factors: The exposure-response relationship for this group of risk factors may be monotonically increasing or J-shaped. For risk factors with a monotonic exposure-response relationship, such as smoking, the optimal exposure would be zero unless there are physical constraints that make zero risk unattainable. For example in the case of blood transfusion, there may be a lower bound on the safety of the blood supply process even using the best monitoring technology. With a J-shaped or U-shaped exposure-response relationship, the minimum risk would occur at the turning point of the exposure-response curve. An example of this is alcohol consumption in adult populations with high cardiovascular disease rates, since moderate consumption may result in a reduction in ischaemic heart disease (IHD) in some age groups (Corrao et al. 2000). With a

J-shaped exposure-response curve, similar to physiological risk factors, empirical evidence would have to be used to determine the theoretical minimum risk.

Finally, some behavioural risks are expressed as the absence of protective factors such as physical inactivity or low fruit and vegetable intake. In such cases, optimal exposure would be the level at which the benefits of these factors would no longer continue. With a monotonic exposure-response relationship or without detailed knowledge about a possible turning point, the theoretical-minimumrisk exposure distribution should be chosen based on empirical evidence about the highest theoretically sustainable levels of intake or exposure (for example very active life style or a purely vegetarian diet).
3. Environmental risk factors: The toxicity of most environmental risk factors is best described as a monotonically increasing function of exposure (potentially with some threshold). Therefore, the theoreti-cal-minimum-risk exposure distribution for this group would be the lowest physically achievable level of exposure, such as background particulate matter concentration due to dust.
4. Socioeconomic "risk factors": Socioeconomic status and factorssuch as income (including levels and distribution) and associated levels of poverty and inequality, education, the existence of social support networks, etc.-are important determinants of health, often through their effects on other risk factors. The effects of each of these factors on health are, however, highly dependent on other socioeconomic variables as well as the policy context, including accessibility and effectiveness of health and welfare systems. For this reason, the theoretical-minimum-risk exposure distribution, even if meaningfully defined, is likely to change over time and space depending on a large number of other factors. Given this heterogeneity, the effects of socioeconomic variables are best assessed relative to counterfactual distributions defined based on policy and intervention options in specific times and settings, as discussed by Greenland (2002).

## 3. Risk Quantification models

Prediction implicitly assumes the use of a conceptual model which infers the value of the variable of interest at a point in time or space based on knowledge from a different time, or another location. Predictive models can be divided along a continuum between aggregate and structural categories. A completely aggregate model uses the previous trend of the variable of interest as the basis for predicting its future value. A structural model, on the other hand, identifies the components-and the relationships among them-of the "system" that determines the variable of interest. It then uses the knowledge of the system for predicting the value of
the variable of interest. Most predictive models lie between the two extremes and use a combination of aggregate and structural modelling. ${ }^{2}$

Consider for example predicting the future population of a city or the future ambient concentration of a pollutant. An aggregate model would extrapolate the historical levels to predict future values. Even in this case the model may include some structural elements. For example, the model may use a specific functional form-linear, exponential, quadratic or logarithmic-for extrapolation which involves an assumption about the underlying system. A structural model, in the case of population prediction would consider the age structure of the population, fertility (which itself may be modelled using data on education and family planning programmes), public health variables and rural-urban migration (which itself can be modelled using economic variables). In the case of air pollution, a structural model may consider demographic variables (themselves modelled as above), the structure of the economy (manufacturing, agriculture or service), the current manufacturing and transportation technology and effects of research and development on new technology, the demand for private vehicles, the price of energy and the atmospheric chemistry of pollution. Once again, in both examples the models may include some aggregation of variables by using historical trends to predict the future values of individual variables in the system, such as funding for family planning or research and development of new technologies.

The comparative advantage of structural and aggregate models lies in the balance between theoretical precision and data requirement. Structural models offer the potential for more robust predictions, especially when the underlying system is complex and highly sensitive to one or more of its components. In such cases, a shift in some of the system variables can introduce large changes in the outcome, which may be missed by extrapolation (such as the discovery of antibiotics and infectious disease trends or the change in tuberculosis mortality after the HIV epidemic). Aggregate models, on the other hand, require considerably less knowledge of the system components and the relationships among them. These models can therefore provide more reliable estimates when such information is not available, especially when the system is not very sensitive to inputs.

### 3.1 Models for risk factor-disease RELATIONSHIP

Using the above aggregate, structural taxonomy, it is also possible to classify models that are used to predict changes in death or disease as a result of changes in exposure to underlying risk factors. Murray and Lopez (1999) described a "causal-web" which includes the various distal (such as socioeconomic), proximal (behavioural or environmental) and physiological and patho-physiological causes of disease, as shown in Figure 1.1. While different disciplinary traditions-from social sciences

Figure I.I Simplified schema for a causal-web illustrating various levels of disease causation


Note: Feedback from outcomes to preceding layers may also exist. For example, individuals or societies may modify their risk behaviour based on health outcomes.
and humanities, to the physical, natural and biomedical sciences-have focused on individual components or stages of these relationships, in a single multi-layer causal model with interactions the term "risk factor" can be used for any of the causal determinants of health (Mathers et al. 2002; Yerushalmy and Palmer 1959). ${ }^{3}$ For example, poverty, location of housing, lack of access to clean water and sanitation, and the existence of a specific pathogen in water can all be considered the causes of diarrhoeal diseases, providing a more complete framework for assessment of interventions and policy options. Similarly, education and occupation, diet, smoking, air pollution, physical activity, BMI and blood pressure are some of the risk factors at various levels of causality for cardiovascular diseases.

Compared to a causal-web, Equations 1 and 2 that use relative risk estimates from epidemiological methods (e.g. the Cox proportional hazard or other regression models) lie further towards aggregate modelling. In general, in such methods, relative risks are estimated so that they incorporate the aggregation of the various underlying relationship (ideally, but not always, controlling for the appropriate confounding variables $)^{4}$ without considering intermediate relationships as separate causal stages. On the other hand, if specified and estimated correctly, considering the complete set of causal pathways which include multiple

Figure I.2 A possible causal diagram based on established relationships for estimating the incidence of ischaemic heart disease


DBP Diastolic blood pressure.
Note: Other interactions may also be possible.
risk factors will allow making inferences about combinations of risk factors and risk factor levels for which direct epidemiological studies may not be available.

As discussed earlier, the appropriateness of the two approaches to estimation of attributable burden depends on the specific risk factor(s), outcomes and available data. For example, the relationship between smoking and lung cancer has been shown to be highly dependent on smoking intensity and duration which, with appropriate indicators of past smoking (Peto et al. 1992), can be readily estimated using the relative risk approach of Equations 1 and 2. Consider, on the other hand, the relationship among age, socioeconomic status and occupation, behavioural risk factors (such as smoking, alcohol consumption, diet, physical activity), physiological variables (such as blood pressure and cholesterol level) and IHD shown in Figure 1.2. Given the multiple complex interactions, IHD risk may be best predicted using a structural (causal-web) approach, especially when some risk factors vary simulta-
neously, such as smoking, alcohol and diet, requiring joint counterfactual distributions. Using a multi-risk model would also allow considering situations for which direct epidemiological studies may not have been conducted, such as the effects of physical activity on those people who have diets different from the study group or those who take medicine to lower blood pressure.

The health effects of global climate change provide another example where a structural approach to risk assessment may be appropriate. Economic activities (including manufacturing, agriculture and forest use, transportation and domestic energy use) affect the emissions of greenhouse gases (GHG). Changes in precipitation, temperature and other meteorological variables due to atmospheric GHG accumulation alter regional ecology, which in turn results in changes in agricultural productivity, quantity and quality of water, dynamics of disease vectors and other determinants of disease. All these effects are in turn modulated by local economic activities, land-use patterns and income (Patz et al. 2000; Reiter 2001; Rogers and Randolph 2000). A model based on the atmospheric physics/chemistry of GHG emissions and accumulation, climate models, plant and vector ecology and human activity might provide the optimal basis for the prediction of the health effects of climate change. ${ }^{5}$

## Specifying The CaUsal-WEB

Assuming for the moment no temporal dimension in the relationship between the different variables in the causal system (temporal aspects are discussed below), each layer of a causal-web may be characterized by the equation:

$$
\begin{equation*}
\mathbf{X}^{n}=f\left(\mathbf{B}\left(\mathbf{X}^{n-1}, \mathbf{X}^{n}\right), \mathbf{X}^{n-1}\right) \tag{3a}
\end{equation*}
$$

where $\mathbf{X}^{n}$ is the vector of the variables in the $n$th layer of the causal-web (which can be causal or output such as $\mathrm{D}, \mathrm{P}, \mathrm{PA}$, or O using the notation of Figure 1.1); $f$ is the functional form connecting the $(n-1)$ th layer to the $n$th layer; B is a matrix of coefficients for $f$ which itself may be dependent on the variables in the $(n-1)$ th and $n$th layers $\left(\mathbf{X}^{n-1} \text { and } \mathbf{X}^{n}\right)^{7}$ (as well as time as we discuss below).

The attributable fraction of disease or mortality due to a single risk factor in the causal-web is then obtained by integrating the outcome ( O ) over the current $(P(\mathbf{x}))$ and counterfactual $\left(P^{\prime}(\mathbf{x})\right)$ population distributions of exposure, as for Equation 2.

$$
\begin{equation*}
A F=\frac{\int_{P(x)} \mathrm{O}(\mathbf{x})-\int_{P(x)} \mathrm{O}(\mathbf{x})}{\int_{P(x)} \mathrm{O}(\mathbf{x})} \tag{4}
\end{equation*}
$$

### 3.2 JOINT RISK FACTOR CHANGES

The attributable fraction relationships described in Equations 1 and 2 are based on individual risk factors. Disease and mortality are however often affected by multiple, and at times correlated, risk factors (Rothman 1976; Walter 1980). Estimating the joint effects of multiple distal and proximal risks is particularly important because many factors act through other, intermediate, factors (Murray and Lopez 1999; Yerushalmy and Palmer 1959), or in combination with other risks. It is therefore important to consider how the burden of disease may change with simultaneous variations in multiple risk factors. Analysis of joint risk factor changes implicitly acknowledges that the disease causation mechanism involves multiple factors, and is therefore suited to a causalweb framework, with $P(\mathbf{x})$ and $P^{\prime}(\mathbf{x})$ in Equation 4 being the joint distributions of the vector of risk factors, $\mathbf{x}$. Alternatively, when using Equations 1 or 2, knowledge of the distribution of all relevant risk factors and the relative risk for each risk factor, estimated at the appropriate level of the remaining risk factors, ${ }^{8}$ is required. Therefore, in Equation $2 \mathrm{a}, R R$ and $P$ may represent joint risks and exposure distributions for multiple risk factors (Eide and Heuch 2001). In this case, the estimates from Equations 2a and 4 may in theory be identical.

## ADDITIVITY OF ATTRIBUTABLE FRACTION

Many users of risk assessment desire information characterized by additive decomposition. In other words, users would like to know what fraction of the disease burden is related to any risk factor or group of risk factors, independent of the changes in other risk factors. As discussed by Mathers et al. (2002), additive decomposition is a property of categorical attribution and, in general, not of counterfactual attribution because many diseases are caused by the interaction of multiple risk factors acting simultaneously and therefore can be avoided by eliminating any of these factors (Rothman 1976; Rothman and Greenland 1998; Yerushalmy and Palmer 1959). Consider for example infant and child mortality due to acute respiratory infections (ARI), which are especially high among malnourished children, as a result of exposure to indoor smoke from solid fuels (Rice et al. 2000; Smith et al. 2000). In this case, removal of either risk factor can reduce mortality, some of which can therefore be attributed to both factors. Similarly the risk of mortality due to cardiovascular diseases among some of those who are exposed to smoking, low physical activity and poor diet may be reduced by elimination of any combination of these risk factors. Counterfactual causal attribution of disease and injury to individual risk factors does not normally allow additive decomposition and the sum of attributable fractions or burdens for a single disease due to multiple risk factors is therefore theoretically unbounded.

Although epidemiologically unavoidable and conceptually acceptable, the lack of additivity presents additional policy complexity and implies great caution is necessary when communicating and interpreting the estimates of attributable fraction and burden. With multiple attribution, the reduction of one risk factor would seem to make other, equally important risk factors potentially irrelevant from the perspective with a limited scope on quantification. At the same time multi-causality offers opportunities to tailor prevention based on availability and cost of interventions. It also necessitates the development of methods to quantify the effects of joint counterfactual distributions for multiple risk factors.

## 4. Temporal dimensions of the risk FACTOR-DISEASE RELATIONSHIP

Both exposure to a risk factor and the health outcomes due to exposure include a time dimension. This can be described by a modified version of Equation 3 in which exposure and outcome as well as the model parameters (B) are dependent on time. In the following two sections we consider the temporal characteristics of exposure and health outcomes, respectively.

### 4.1 Temporal characteristics of exposure

With the exception of acute hazards (e.g. injury risk factors) exposure to a risk factor affects disease over a time period. As a result, the distributional transition between any two exposure distributions includes a temporal dimension as illustrated schematically in Figure 1.3. The transition path is of little importance if exposure changes over a short time interval, especially relative to the time required for the effect of exposure on disease. Over long time periods, however, there is sufficient time for contributions from the intermediate exposure values, and the actual path of transition may be as important as the initial and final distributions in determining the disease burden associated with change in exposure. For example, the effects of reducing the prevalence of smoking or exposure to an occupational carcinogen by half in a population would be markedly different if the change takes place immediately, gradually over a twenty-year period or after twenty years. Therefore, the health effects of exposure to many risk factors depend on the complete profile of exposure over time, and may be further accompanied by a time lag from the period of exposure. Also, for some risk factors there may be complete or partial reversibility, with the role of past exposure gradually declining.

To capture the effects of exposure profiles over time, we begin by considering the role of temporal dimensions of exposure at the level of individuals (or groups of individuals with similar exposure) before considering the whole population.

Figure I.3 A (three-dimensional) representation of a time-indexed distributional transition of population exposure to a risk factor, with a decreasing central tendency


Suppose that at time $T$ the relative risk of a disease, $R R$, for individuals exposed to a risk factor (compared to the non-exposed group) depends on the complete profile or stream of exposure between time $T_{0}$ and $T$, denoted by $x(t)$, with some lag, $L$, between exposure and effect. Then, there is some function, $f(x)$, which can be used to describe the contribution of exposure at any point in time between $T_{0}$ and $T$ to the relative risk ( RR ). In mathematical notation:

$$
\begin{equation*}
\left.R R(x(t))\right|_{T_{0}} ^{T}=R R\left(\int_{T_{0}}^{T} f(x(t-L)) d t\right) \tag{5}
\end{equation*}
$$

The quantity $\int_{T_{0}}^{T} f\left(x(t-L) d x\right.$ is an equivalent exposure ${ }^{10}$ between $T_{0}$ and $T$ and is dependent on: i) the profile of exposure (i.e. level of exposure at any point in time) described by $x(t)$; and ii) the contribution of previous exposure to current hazard characterized by $f(x)$, an accumulative risk function. ${ }^{11}$ Some common forms for the accumulative risk function, $f(x)$, are given in Table 1.1.
Table I.I Possible forms for the accumulative risk function, $f(x)$

| Accumulative risk function, $f(x)$ | Interpretation | Relative risk | Potential example |
| :---: | :---: | :---: | :---: |
| I. $f(x)=\left\{\begin{array}{lc}1 & \text { if } t=T \\ 0 & \text { otherwise }\end{array}\right.$ | RR depends only on current exposure, with no contribution from past exposure | $\left.R R(x(t))\right\|_{T_{0}} ^{T}=R R(x(T))$ | Instantaneous poisoning as a result of exposure to high levels of toxic chemicals; injuries or death in accidents due to binge drinking; infection with Hepatitis B or C as a result of an infected injection |
| 2. $f(x)=1$ | RR depends on the accumulated exposure (or average exposure if normalized with respect to exposure time), without any effects from the temporal distribution of exposure | $\left.R R(x(t))\right\|_{T_{0}} ^{T}=R R\left(\int_{T_{0}}^{T} x(t) d t\right)$ | Cancer risk from lifetime exposure to carcinogens which have no threshold level |
| 3. $f(x)=\left\{\begin{array}{cc}\frac{1}{K} & \text { if } t>T-K \\ 0 & \text { otherwise }\end{array}\right.$ | RR depends on current and past exposures. But the role of past exposure lasts for a limited time, $K$, and declines as a linear function of time | $\left.R R(x(t))\right\|_{T_{0}} ^{T}=R R\left(\int_{T_{0}}^{T} \frac{(t-T+K)}{K} x(t) d t\right)$ | Reduction in cardiovascular events after lowering blood pressure ${ }^{\text {a }}$ |
| 4. $f(x)=\mathrm{e}^{\alpha(t-T)}$ | RR depends on current and past exposures. But the role of past exposure decays as an exponential function of time | $\left.R R(x(t))\right\|_{T_{0}} ^{T}=R R\left(\int_{T_{0}}^{T} \mathrm{e}^{\alpha(t-T)} x(t) d t\right)$ | Reduction in cardiovascular events after lowering blood pressure ${ }^{\text {a }}$ |

[^1]The above framework can be extended from individuals to populations, by indexing the exposure profile $(x(t))$ to individuals (i.e. representing the exposure of the $i$ th individual as $x_{i}(t)$ ) and considering how the distribution of exposure in the population evolves over time. ${ }^{12}$ This in turn provides the population distributions of equivalent exposure (current or expected future and counterfactual) which form the basis of calculating attributable fractions (i.e. the terms in the numerator of Equations 2a, 2 b or 4 ).

It is reasonable to assume that if the exposure of one individual is greater than that of another over the whole exposure period (i.e. tracking) (Foulkes and Davis 1981), the equivalent exposure of the former is also greater than the latter. In other words, the accumulative risk function, $f(x)$, has the following property:

$$
\begin{equation*}
\int_{T_{0}}^{T} f\left(x_{i}(t)\right) d t>\int_{T_{0}}^{T} f\left(x_{j}(t)\right) d t \quad \text { if } \quad x_{i}(t)>x_{j}(t) \quad \forall t \in\left[T_{0}, T\right] \tag{6}
\end{equation*}
$$

With this property, if the ordering of individuals in the exposure distribution remains unchanged over time (i.e. the rank-order correlation of individual exposures equals 1 between different points in time), the equivalent exposure will also have a distribution with the same ordering of individuals.

The method used by Peto et al. (1992) for estimating mortality due to smoking implicitly uses such as framework. It is well known that the accumulated hazards of smoking depend on a number of variables including the age at which smoking began, number of cigarettes smoked per day and cigarette type. Such data however are extremely rare. To overcome this problem, Peto et al. (1992) used the smoking impact ratio, $S I R$, which uses population lung cancer rates as a marker for accumulated hazard of smoking, to estimate the relative risk of the accumulated smoking exposure corresponding to the population. In the above notation:

$$
\left.R R(\operatorname{smoking}(t))\right|_{T_{0}} ^{T}=R R\left(\int_{T_{0}}^{T} f(\operatorname{smoking}(t-L)) d t\right)=R R(S I R(T))
$$

The temporal profile of exposure for some risk factors may be more easily available than the range of indicators that are needed to estimate the accumulated hazards of smoking. For example, exposure to indoor smoke from solid fuels is likely to remain unchanged as long as household fuel and housing conditions remain the same. Therefore, estimating the effects of long-term exposure may require only knowledge of household fuel, housing and participation in cooking. Similarly, in the case of blood pressure, it is known that blood pressure follows a predictable age pattern (Tate et al. 1995; Yong et al. 1993), unless severely affected by a changes in social (stress), behavioral (diet or smoking)
or medical circumstances. In this case, the usual blood pressure of an individual reflects the history of the person's exposure. On the other hand, the patterns of fruit and vegetable consumption, smoking, or exposure to urban air pollution may change rapidly in countries with high rates of economic growth and urbanization, requiring more detailed data.

The above discussion is based on two implicit assumptions:

1. It considers the effects of exposure to a single risk factor over time. This approach may be appropriate for some risk factor-disease relationships (e.g. the effects of accumulated exposure to carcinogens with site-specific effects). But the single equivalent exposure cannot characterize other risk factor-disease relationships where risk factor interactions are important over time (e.g. physical activity, BMI, smoking and cardiovascular diseases). Extending this temporal dimension to multiple risk factors requires considering the accumulated effects of the vector of risk factors as well as their interactions. In this case, Equation 5 would be expressed in terms of the vector of risk factors of interest. Few epidemiological studies, however, have gathered the data needed for assessing accumulated interactive effects.
2. It considers exposure to each risk factor as an exogenous variable (i.e. intermediate exposure at any time, $x(t)$, is not affected by disease or other risk factors) whose accumulated effect can be captured in a single value using the risk accumulation function. For some risk factors, this assumption may not be valid since exposure to behavioural as well as environmental risk factors may be affected by knowledge of their current effects-individuals may change their diet or activity levels based on knowledge of their weight or blood pressure and governments may introduce regulations based on the level of various contaminants in air or water. ${ }^{13}$

Manton et al. (1993a, 1994) have relaxed these assumptions using a diffusion model for forecasting cardiovascular disease mortality in the United States of America. In this model, it is the change in the outcome at any time, $t$, that is modelled as a function of all the other variables in the system (i.e. other risk factors as well as outcomes) and their interactions. Using the notation of Equation 3:

$$
\begin{equation*}
d \mathbf{X}(t)=\mathbf{u}(\mathbf{X}(t), t) d t \tag{7}
\end{equation*}
$$

where $\mathbf{u}(\mathbf{X}(t), t)$ is a drift term whose value depends on the current value of all the variables in the system as well as their interactions (and can be described by a functional form similar to that in Equation 3). ${ }^{14}$ Methods for estimation of such models using longitudinal data are discussed by Robins (1997, 1999b).

### 4.2 Temporal characteristics of health outcomes

If the outcome variable used in causal attribution of disease and mortality to a risk factor only involves counting of adverse events (such as disease incidence or death), it is not possible to characterize those cases whose occurrence would have been delayed in the absence of the risk factor (Greenland and Robins 1988; Robins and Greenland 1989, 1991). A major shortcoming with this approach is that it does not take into account the accumulated effects of exposure-in the form of years of life lost prematurely or lived with disability. Parameterizing the above relationships by age (or birth cohort) would allow estimating the effects of exposure to a risk factor, not as an event without time dimension, but as an event at a certain age and time. More broadly, considering the timeindexed stream of health losses due to a risk factor requires using a timebased (and not event-based) SMPH.

Murray and Lopez (1999) have provided an additional temporal distinction for the burden of disease due to a risk factor by introducing the concepts of "attributable" and "avoidable" burden. Attributable burden is defined as the reduction in current and/or future burden of disease if past exposure to a risk factor had been equal to some counterfactual distribution. Avoidable burden is the reduction in the future burden of disease if the current or future exposure to a risk factor is reduced to a counterfactual distribution. Attributable and avoidable burden are shown graphically in Figure 1.4. While attributable burden is easier to measure and more certain, avoidable burden is more useful for policy purposes. The distinction between attributable and avoidable burden becomes less significant as the time between exposure to risk factor and effects on disease burden decreases. In this case, attributable burden is a good predictor of avoidable disease burden.

Figure 1.4 also illustrates a conceptual complexity in defining and estimating avoidable burden. Attributable burden is defined based on the difference between (accumulated) current exposure and a counterfactual. Measuring current exposure, while difficult and uncertain, is conceptually well defined. Avoidable burden, on the other hand, depends on the expectation of future exposure and a counterfactual, with the former being analogous to current exposure. Consider for example a population exposed to rising levels of air pollution or rates of obesity. In this setting, interventions that would maintain pollutant concentrations or BMI at their current levels would result in avoiding disease and mortality; they reduce exposure from what it would be in their absence. Therefore, avoidable burden (i.e. how much of future burden could be prevented) by definition requires estimates of future exposure (i.e. how much of future burden there is). Projecting future exposure in turn raises the need to provide a projection framework. To provide visions of public health under various intervention and policy scenarios we suggest that the future exposure level, with respect to which avoidable burden is

Figure I.4 Attributable and avoidable burden

estimated, be the expectation of exposure if the current policy and technological context were to continue, referred to as the "business-as-usual" exposure trend. Therefore avoidable burden is the burden of disease averted due to reduction in exposure to a risk factor beyond its expected trends. We emphasize that with this definition, avoidable burden is the difference between two exposure scenarios: the expectation of future trends (business-as-usual) and a reduction with respect to this trend towards theoretical minimum.

### 4.3 Cumulative vs period estimates of attributable and AVOIDABLE BURDEN

Although analytically inconsequential, the starting point and the duration of the time interval over which attributable or avoidable burden is reported has policy implications because reductions in various risk factors may provide health benefits that occur after short or long delays and last for different periods. Consider for example the health benefits of reductions in binge alcohol consumption, smoking and GHG emissions. Reducing binge drinking would result in immediate health benefits from a drop in alcohol-related accidents and injuries (as well as medium- and long-term benefits from reduction in other diseases). Lowering smoking will have some short- and medium-term benefits from reduction of acute respiratory diseases and cardiovascular disease as well as longer-term benefits from lowering cancers and chronic obstructive pulmonary disease (COPD). The benefits of policies that reduce climate change as a result of GHG emissions are likely to be heavily concentrated in the future.

Consider these examples of health benefits in terms of duration: the distribution of a drug that lowers blood pressure, or food aid to reduce malnutrition; and programmes that promote and sustain increased physical activity, the introduction of a new agricultural technology which results in higher food yields, or automotive technology which eliminates the use of leaded gasoline. While the benefits of all these interventions may be equally large and important for the current cohort, the first two actions have one-time health benefits (unless repeated) while the latter three are likely to last indefinitely.

The above discussion would suggest reporting the estimates of avoidable burden in multiple ways including both period (e.g. annual) and cumulative estimates, as well as over short and long time frames. The issue of future estimates and their policy relevance is further complicated by the growing uncertainty of estimates with increasing length of the estimation interval. Therefore, while it may be preferable to increase the prediction horizon, it is important to emphasize that long-term predictions are inherently more uncertain.

### 4.4 Discounting future risk and health effects

Individuals may discount consumption or welfare within their own lifespan and exhibit a preference for benefits today over the future. The theoretical and empirical arguments for and against individual discounting with specific emphasis on health, including the possibility of negative discount rates, are summarized elsewhere (Murray 1996; Murray and Acharya 2002) and are directly incorporated in the calculation of a summary measure of population health. In addition to individual discounting and discount rates, policies dealing with risk confront the issue of addressing benefits to different populations across time. As a result, these policies must address ethical and analytical dilemmas related to the valuation of current and future health and welfare, in the form of social discount rates (Kneese 1999). Discounting future risks, benefits and welfare has been a subject of great debate (Howarth 1996; Lind 1982; Portney and Weyant 1999; Schelling 1995; Toman 1999), motivating some economists to conclude that

> maybe the idea of a unitary decision-maker-like an optimising individual or a wise and impartial adviser-is not very helpful when it comes to the choice of policies that will have distant-future effects about which one can now know hardly anything. Serious policy choice may then be a different animal, quite unlike individual saving and investment decisions... "Responsibility" suggests something less personal (Solow 1999).

The arguments for and against discounting of future health and welfare, and their validity, have been discussed in detail elsewhere (Anand and Hansen 2002; Murray and Acharya 2002; Murray and Lopez 1996; Parfit 1984). According to one specific argument, "the disease eradication and health research paradox", not discounting future health would imply investing all of society's health resources in research programmes or programmes for disease eradication, which would result in an infinite stream of benefits, rather than in any programme that improves the health of the current generation. Such an excessive intergenerational "sacrifice" is a particularly powerful argument for discounting of future health (or more precisely for something that resembles discounting as we discuss shortly) (Parfit 1984). It is important to emphasize that this argument does not imply that future welfare or health is less valuable than current, but rather it uses discounting as a tool to avoid excessive sacrifice for the current generation, to the point of investing all resources in an infinite stream of future health. For this reason, Parfit (1984) argues that the issue of intergenerational distribution should be considered as an independent criterion, rather than explicitly discounting future benefits.

Koopmans (1960), Dasgupta and Mäler (1994) and Dasgupta et al. (1999), however, have shown that any preference-ordering defined over
a set of well-being paths over time can be represented by a numerical function with an apparently utilitarian form ${ }^{15}$ and therefore includes what resembles positive discounting of future well-being. We emphasize that this notion is simply a consequence of considering the paths of wellbeing (or temporal distributions), rather than a statement about the value of current or future welfare. With this formulation, Dasgupta and Mäler (1994) and Dasgupta et al. (1999) consider the implications of the choice of discount rate as a "derived notion", as opposed to a value judgement. Dasgupta and Heal (1974) and Solow (1974b) have shown that if wellbeing is a result of consumption of an exhaustible resource, a zero discount rate would imply investing all available resources for the benefit of future generations, and hence no current consumption. This is because each unit sacrificed by the first generation would yield a finite loss to this generation, but an infinite stream of benefits to future generations (Arrow 1999) which, without discounting, would always be larger than the one-time sacrifice. Although the first generation cannot sacrifice everything, ${ }^{16}$ the logical conclusion of this situation would be that "given any investment [for future benefits], short of the entire income, a still greater investment would be preferred" (Arrow 1999), or a potentially excessive intergenerational sacrifice (Murray 1996).

On the other hand, a positive discount rate would imply that in the long run consumption of resources should become zero. In this case, however, the additional requirement that well-being should never fall below a certain threshold would in turn require downward adjustment of the discount rate (Solow 1974a). The stricter requirement of nondeclining consumption and well-being would require a discount rate lower than the productivity of capital (Dasgupta and Mäler 1994). ${ }^{17}$ Based on these arguments, we suggest discounting of future attributable or avoidable disease burden due to risk factors, but with a low discount rate, to include the welfares of both current and future generations as described above.

## 5. Uncertainty

Quantitative risk assessment is always affected by uncertainty about the existence, magnitude and distribution of risk (Graham et al. 1988). Quantitative analysis of uncertainty greatly adds to the usefullness of the results because it shows not only the "best-estimate" of the magnitude and distribution of exposure to a risk factor, and the resulting burden of disease, but also the range of potential outcomes.

### 5.1 Sources of uncertainty in attributable fractions

Population distribution of Exposure
An important source of uncertainty in risk assessment is characterizing population distribution of exposure. Due to complexity and cost, for
most risk factors, exposure is measured only in small samples and in a limited number of settings. As discussed earlier, because a risk factor can be represented in different layers of causality, variables for which data are more readily available can be used as exposure proxies. The use of exposure proxies is sometimes also necessary because epidemiological studies have used such proxies in estimating hazard size. For example, anthropometric variables such as height-for-age or weight-for-age are used as indicators of childhood nutritional status; the presence of clean water sources or sanitary latrines as indicators of faecal-oral transmission of pathogens; concentration of particles as the measure of exposure to the various pollutants in ambient air, and so on. In addition to reduced data requirements, the use of such indirect indicators of exposure (or exposure scenarios [Kay et al. 2000]) may provide direct mapping to existing interventions. At the same time, these indicators, which are often more distal than actual exposure, do not capture the variability of exposure within each scenario, unless combined with other indicators which affect this variability (Ezzati et al. 2000). For example people using the same water source may experience different levels of faecal-oral transmission of pathogens due to different storage and hygiene behaviours. Therefore, the use of indirect exposure proxies results in additional uncertainty in exposure characterization.

Even with the choice of exposure proxies, extrapolation of exposure between different populations or age groups is often necessary. Such extrapolation (or spatial prediction) can be based on models as simple as using the average of subgroups with data for a whole population, or more complex prediction models. For example, urban air quality monitoring systems provide data on particle concentrations in some but not all cities in each region. Models to predict ambient concentrations of particulate matter based on energy consumption, number of vehicles and level of industrialization can be used to predict ambient air pollution levels for cities where data are not available. Similarly, the level of physical (in)activity in a population may be predicted from a model that uses rural-urban population distribution, income, education, distribution across occupational categories and available transportation modes in each geographical region. Each such extrapolation adds to the uncertainty of exposure distributions.

Finally, as we discussed earlier, for many risk factors, hazards are associated with accumulated effects of sustained exposure. Indicators of accumulated hazard for those risk factors with changing exposure such as smoking, urban air pollution or BMI, are needed but not always available. Further, few epidemiological studies have considered the role of a temporal profile of exposure on disease (see Peto 1986 for an example of an exception). Therefore even if longitudinal data on exposure prevalence were available, they could not always be used together with epidemiological studies that consider a single exposure variable-at the beginning or end of the follow-up period, for example. At the same time,
if the ordering of individuals in the exposure distribution remains unchanged over time (see above), risk estimates from epidemiological studies with similar ordering may be applicable, but result in additional uncertainty. ${ }^{18}$

## RISK FACTOR-DISEASE RELATIONSHIP

At the most fundamental level, quantifying the hazards associated with exposure to a risk factor requires identifying the diseases and injuries that are caused by a risk factor. The criteria for establishing disease causality have been the subject of interest and debate for over a century (summarized in Evans 1976, 1978; Hill 1965; National Research Council 1994; Yerushalmy and Palmer 1959). Epidemiological studies have successfully provided the basis for establishing causality between some risk factor-disease pairs. For other risk factor-disease combinations, where the measurement of exposure or disease has been difficult or the delay between exposure and health effects is very long, observational or experimental epidemiology has had less success in establishing causality (Evans 1976; Robins 1999a). For this reason, epidemiological evidence must often be complemented with inferences from other disciplines such as toxicology, physiology, parasitology and, increasingly, biophysics, in establishing disease causation.

Even when causality is established, the magnitude of the hazard due to a risk factor needs to be quantified. Although the statistical issues around establishing causality and estimating the effect size are similar (lack of causality is equivalent to zero excess risk) (Robins 1999a; Robins and Greenland 2000), in practice, with knowledge from multiple disciplines in establishing causality, it is often the latter that is the source of increased uncertainty in risk assessment. For example, the collectivity of scientific knowledge from disciplines such as economics and behavioural sciences, vector biology, physiology and bio-mechanics and epidemiology would confirm the possibility that climate change or socioeconomic inequality would increase disease, or whether the relationships between occupational factors or physical inactivity and lower back pain are causal. At the same time, risk assessment would require estimating the hazard magnitude for each of these relationships. Therefore, the complexity of the causal relationship, or lack of detailed data, would shift the debate from causality to hazard size.

Epidemiological studies that quantify hazards are often conducted in a limited number of settings, with emphasis on estimating the average effect size in the whole study group. While the robustness of relative risk measures has been confirmed for more proximal factors in studies across populations (Eastern Stroke and Coronary Heart Disease Collaborative Research Group 1998; Law et al. 1994), their extrapolation is an important source of uncertainty for more distal risks (e.g. child sexual abuse) or for those whose effects are heterogeneous (e.g. alcohol and injuries vs alcohol and cancer) and has received less attention in the
epidemiological literature (Horton 2000). For some risk factors, it is likely that the magnitude of the hazard may depend on the levels of other variables (i.e. effect modifiers). Therefore, in extrapolating the results of individual epidemiological studies or meta-analyses, the very strength of the original study-applicability to the average person-would be the source of uncertainty if the population to whom the effect size is extrapolated has characteristics which would result in effect modification (Britton et al. 1999; Horwitz et al. 1990, 1996, 1998). The impact of alcohol drinking patterns on cardiovascular disease risk estimates (Britton and McKee 2000; Puddey et al. 1999) is an example of the importance of considering factors that modulate and modify hazards in risk extrapolation.

## RISK Factor and disease correlations

Because multiple risks and disease are correlated (e.g. higher malnutrition, unsafe water, sanitation and hygiene, indoor smoke and childhood mortality in poor rural households in developing countries; higher smoking, BMI and occupational risks in developed countries [Thun et al. 2000]), estimating attributable fractions would require stratified (e.g. by other risk factors) prevalence as well as disease data. Lack of stratified data is another source of uncertainty, in general leading to the underestimation of effects in the presence of positive risk factor correlations (Greenland 1984).

### 5.2 Characterizing and quantifying uncertainty

Various taxonomies of uncertainty have been used in risk assessment (National Research Council 1994) including:

1. classification based on information type such as uncertainty in hazard identification, exposure assessment, exposure-response assessment, as discussed above;
2. classification based on uncertainty type such as randomness, true variability, and bias; and
3. classification based on the approach to handling uncertainty which divides uncertainty into parameter uncertainty and model uncertainty. Parameter uncertainty includes the uncertainty quantifiable using random-variable methods such as the uncertainty due to sampling and measurement error. Model uncertainty is due to gaps in scientific theory, measurement technology and data (National Research Council 1994). It includes uncertainty in the knowledge of causal relationships or of the form of the exposure-response relationship (threshold vs continuous, linear vs non-linear, etc.), the level of bias in measurement, etc. Defined broadly, model uncertainty also includes extrapolation of exposure or hazard from one population to another. Uncertainty in risk assessment is overwhelmingly dominated by model
uncertainty, which arises due to a lack of direct studies on exposure, hazard and background disease burden.
We distinguish between uncertainty, which is due to gaps in knowledge, methods or data, and variability, which is a real property of the world and itself may be known with certainty or with uncertainty. Variability can nonetheless be a source of uncertainty in the absence of population-specific data on exposure or the exposure-response relationship. For many risk factors, data on exposure distributions are available for a limited number of populations or demographic groups. The exposure distribution for other populations are then extrapolated from the available data based on some model. As discussed earlier, the extrapolation model may be as simple as using the population-weighted average of the existing data or more complex (based on a number of predictors). In such cases, the statistical uncertainty of the estimator (e.g. the $95 \%$ confidence interval of the mean or regression coefficients) is an underestimation of true uncertainty in predicted values due to the unexplained variability in the data. More complex models can increase the predictive power and therefore reduce uncertainty but even the most sophisticated models are unlikely to fully explain the variability of the data, resulting in residual uncertainty. Variability can also be a source of uncertainty in the estimation and extrapolation of exposure-response relationships or relative risks that are measured in a limited number of settings. In the presence of multiple estimates of hazard, it is common to use metaanalytical approaches to obtain an overall estimate. At the same time, the differences between various estimates may reflect true variability in effect size, especially if obtained from different populations, resulting in uncertainty in hazard estimates.

Parameter uncertainty can be readily included in quantitative analysis using random-variable statistical methods (Morgan and Henrion 1990). While we have discussed the various sources of uncertainty, the important issue of extrapolation of exposure and hazard using models requires new approaches to quantifying uncertainty in the presence of limited data. Quantitative analysis of model uncertainty, by definition, would require considering the uncertainty of the models and assumptions used (including assumptions about disease mechanism or data/ parameter extrapolation) using the methods of Bayesian statistics.

## 6. Conclusions

We have described a framework for systematic quantification of the burden of disease due to risk factors that attempts to unify the growing interest in health risks across a number of health, physical and social sciences. The key attributes of the framework, along with the corresponding methodological issues that arise in its application, are:

- comparing the burden of disease due to the observed exposure distribution in a population with the burden from a hypothetical distribution or series of distributions, rather than a single reference level such as non-exposed;
- considering the multiple stages of causality and interactions between risk factor(s) and disease outcome to allow inferences about combinations of risk factors for which epidemiological studies have not been conducted, including the joint effects of changes in multiple risk factors;
- calculating the health loss due to risk factor(s) as a time-indexed "stream" of disease burden due to a time-indexed "stream" of exposure, including consideration of discounting; and
- describing the sources of uncertainty in the risk assessment process.

For each of the above aspects, we have outlined the important conceptual and methodological issues and their implications for risk assessment. While this framework provides a means for considering risk factors in different layers of causality, with multiple counterfactuals (Ezzati et al. 2002; see also the risk factor chapters in this book), its application is limited by the availability of data on risk factors and hazards (Powles and Day 2002). The availability and form of data on both exposure and hazard are often determined by disciplinary boundaries as well as measurement difficulties. Analysis of selected risks highlights data and monitoring needs for better quantification and intervention strategies, especially more detailed data on exposure, hazard accumulation over time, and heterogeneity of risk factor-disease relationships.

For more effective and affordable implementation of a prevention paradigm, policies, programmes and scientific research should acknowledge and take advantage of the interactive role of major risks to health, across and within causality layers. Despite the methodological complexity and empirical difficulties, especially in estimating time-based multi-risk exposures, this framework provides a consistent basis for better and more comparable information about the various causes of disease and injury. In the remaining chapters of this book, the burden of disease due to a diverse set of risk factors is assessed according to this conceptual framework, with a detailed description of data sources and risk factor-specific methodological issues.

## Acknowledgements

This work has been sponsored by the National Institute of Aging Grant 1-PO1-AG17625. The participants in the various CRA project collaborators meetings and two reviewers provided valuable comments. J. Powles and J. Barendregt provided extensive comments on an earlier draft.

## Notes

1 As discussed by Greenland and Robins (1988), attributable fractions without a time dimension are not able to characterize those cases whose occurrence would have been delayed in the absence of exposure. The authors recommend the use of etiologic fractions with a time dimension to account for this shortcoming. Time-based measures are discussed in more detail below.
2 At the extreme, a structural model would attempt to use chemical or physical principles as the unit of analysis and modelling. This would of course be currently impossible in studying any system that involves population health.
3 The "driving force, pressure, state, exposure, effect" (DPSEE) model of Corvalan et al. (1999) does consider the multiple layers of causality. This model however focuses on the risk evolution process, which is less suitable for multi-risk factor interaction within and between layers. More complete discussions of causality and multiple causes are provided by Yerushalmy and Palmer $(1959)$, Evans $(1976,1978)$ and Rothman and Greenland (Rothman 1976; Rothman and Greenland 1998).
4 The exception is those risk factors whose effects occur through intermediate variables which are themselves a risk factor for the outcome considered, such as the relationships between physical inactivity or obesity and IHD, which are mediated through blood pressure or cholesterol. In such cases, controlling for the intermediate risk factor would result in a bias (towards the null) in the estimation of total hazard of the distal factor (Greenland 1987).
5 There are no past studies on "climate change" as it is expected to take place in the future. For this reason, the relationship between climate change and health would always be based on a model which relates climate change to meteorological variables (e.g. temperature or rainfall). The relationship between these variables, disease vectors and disease could then be estimated from past data and vector biology (Craig et al. 1999; Martens et al. 1999; Rogers and Randolph 2000).
6 If the variables in the $n$th layer of the causal-web are affected directly by those in the $(n-2)$ th layer in addition to the $(n-1)$ th layer, or by variables within the $n$th layer itself (see Figure 1.2 for an example), Equation 3a can be expanded to include these links as well:

$$
\begin{equation*}
\mathbf{X}^{n}=f\left(\mathbf{B}_{0}\left(\mathbf{X}^{n}\right), \mathbf{X}^{n} ; \mathbf{B}_{1}\left(\mathbf{X}^{n-1}, \mathbf{X}^{n}\right) \mathbf{X}^{n-1} ; \mathbf{B}_{2}\left(\mathbf{X}^{n-2}, \mathbf{X}^{n}\right) \mathbf{X}^{n-2}\right) \tag{3b}
\end{equation*}
$$

This can be extended to interactions across multiple causal layers, and in general any variable in the system can be affected by any other one as the concept of causal layer becomes more flexible.

7 In this case when some of the variables affect not only the other variables in the causal system but also the relationship(s) between variables, they are equivalent to "effect modification" in epidemiological literature (Rothman 1976). Graphically, in Figure 1.1 they would be represented as links (arrows) not between two variables but as links from one of the variables (the effect modifier) to another link in the system.
8 In other words the $R R(x)$ in Equations (2a) and (2b) are functions of the other covariates, referred to as effect modification earlier. Epidemiological
studies that stratify relative risks based on covariates other than age and sex are however rare.

9 The notation $\left.\right|_{T_{0}} ^{T}$ denotes "estimated between $T_{0}$ and $T$ ".
10 Equivalent exposure is an analytical concept and need not be physically realizable. In fact for many risk factors, such as carcinogens, where the effects are from life-long exposure, the equivalent exposure would be so high that its occurrence at a single instant would be impossible.
11 Further, if there is threshold, $M$, below which exposure has no effect:

$$
\left.R R(x(t))\right|_{T_{0}} ^{T}=R R\left(\int_{T_{0}}^{T} f(x(t-L))(\operatorname{TRUE}(x(t-L) \geq M)) d t\right)
$$

where

$$
\operatorname{TRUE}(x(t) \geq M)= \begin{cases}1 & \text { if } x(t) \geq M \\ 0 & \text { if } x(t)<M\end{cases}
$$

This framework can be easily modified to include cases where exposure has different effects below and above threshold by using $\operatorname{TRUE}(x(t) \geq M)$ for the effect above the threshold and $\operatorname{TRUE}(x(t)<M)$ for the effect below the threshold.
12 In this manner, the evolution of the exposure distribution is analytically similar to a "random process" in which a probability density function (PDF) describes a random variable which is a function of time. Exposure to a risk factor is not a random variable in the strict sense. But since a timedependent exposure distribution has an accumulated distribution of 1.0, it has the same representation as a random process.
13 Robins (1999a) discusses this issue in the case of estimating the effects of a dynamic treatment regime whose dose is dependent on symptoms.
14 The diffusion model also includes a stochastic component to account for those interactive effects not described by the drift term.
15 The functional form is $\int_{0}^{\infty} W_{i} \alpha^{t}$ where $0<\alpha<1$ in Koopmans (1960) and $\int_{0}^{\infty} W_{i} \exp (-\delta t)$ where $\delta>0$ in Dasgupta and Mäler (1994) and Dasgupta et al. (1999); $\alpha$ and $\delta$ are the social rate of discount.
16 The last unit of sacrifice will have infinite marginal utility therefore matching the future infinite stream of benefits.

17 These two additional constraints are external to the economic efficiency arguments as defined by maximizing aggregate welfare. In fact, these additional constraints of minimum acceptable or non-declining welfare result in an "inefficient" outcome in order to achieve better distribution across generations (Montgomery 1999). See Weitzman (1999) for another argument for the choice of lower discount rates.
18 If exposure is sustained for a longer time in the risk assessment population than in the study population and if the whole exposure period contributes to hazards, this would result in an underestimation of risk (and vice versa). For example, in many cohorts in current epidemiological studies, BMI
increased when the subjects were in their twenties or thirties. There is however increasing child and adolescent overweight or obesity in many regions of the world. If this continues into adult life, the hazards may be higher than subjects in the current study cohorts.

## References

Anand S, Hansen K (2002) Age weights and discounting in disability-adjusted life years: a critical review. In: Summary measures of population health. Murray CJL, Salomon J, Mathers CD, Lopez AD, Lozano R, eds. World Health Organization, Geneva.
Arrow KJ (1999) Discounting, morality, and gaming. In: Discounting and intergenerational equity. Portney PR, Weyant JP, eds. Resources for the Future, Washington, DC.
Barnes R (1965) Comparisons of blood pressure and blood cholesterol levels in New Guineans and Australians. Medical Journal of Australia, 1:611-617.
Britton A, McKee M (2000) The relationship between alcohol and cardiovascular disease in eastern Europe: explaining the paradox. Journal of Epidemiology and Community Health, 54:328-332.
Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C (1999) Threats to applicability of randomised trials: exclusion and selective participation. Journal of Health Services Research and Policy, 4:112-121.
Carvalho JJ, Baruzzi RG, Howard PF et al. (1989) Blood pressure in four remote populations in the Intersalt study. Hypertension, 14:238-246.
Chen Z, Peto R, Collins R, MacMahon S, Lu J, Li W (1991) Serum cholesterol concentration and coronary heart disease in population with low cholesterol concentrations. British Medical Journal, 303:276-282.
Cole P, MacMahon B (1971) Attributable risk percent in case-control studies. British Journal of Preventive and Social Medicine, 25:242-244.
Corrao G, Rubbiati L, Bagnardi V, Zambon A, Poikolainen K (2000) Alcohol and coronary heart disease: a meta-analysis. Addiction, 95:1505-1523.
Corvalan CF, Kjellstrom T, Smith KR (1999) Health, environment, and sustainable development: identifying links and interactions to promote action. Epidemiology, 10:656-660.
Craig MH, Snow RW, le Sueur D (1999) A climate-based distribution model of malaria transmission in sub-Saharan Africa. Parasitology Today, 15:105-111.

Dasgupta P, Heal GM (1974) The optimal depletion of exhaustible resources. Review of Economic Studies, 41.
Dasgupta P, Mäler K-G (1994) Poverty, institutions, and the environmental resource base (World Bank Environment Paper No. 9.) World Bank, Washington, DC.
Dasgupta P, Mäler K-G, Barrett S (1999) Intergenerational equity, social discount rates, and global warming. In: Discounting and intergenerational equity. Portney PR, Weyant JP, eds. Resources for the Future, Washington, DC.

Drescher K, Becher H (1997) Estimating the generalized impact fraction from case-control data. Biometrics, 53:1170-1176.

Eastern Stroke and Coronary Heart Disease Collaborative Research Group (1998) Blood pressure, cholesterol, and stroke in eastern Asia. The Lancet, 352:1801-1807.

Eide GE, Heuch I (2001) Attributable fractions: fundamental concepts and their visualization. Statistical Methods in Medical Research, 10:159-193.
Evans AS (1976) Causation and disease: the Henle-Koch postulates revisited. Yale Journal of Biology and Medicine, 49:175-195.
Evans AS (1978) Causation and disease: a chronological journey. American Journal of Epidemiology, 108:249-258.
Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJL, Comparative Risk Assessment Collaborative Group (2002) Selected major risk factors and global and regional burden of disease. The Lancet, 360:1347-1360.
Ezzati M, Saleh H, Kammen DM (2000) The contributions of emissions and spatial microenvironments to exposure to indoor air pollution from biomass combustion in Kenya. Environmental Health Perspectives, 108: 833-839.
Foulkes MA, Davis CE (1981) An index of tracking for longitudinal data. Biometrics, 37:439-446.
Graham JD, Green LC, Roberts MJ (1988) In search of safety: chemicals and cancer risk. Harvard University Press, Cambridge.
Greenland S (1984) Bias in methods for deriving standardized morbidity ratio and attributable fraction estimates. Statistics in Medicine, 3:131-141.
Greenland S (1987) Quantitative methods in the review of epidemiologic literature. Epidemiologic Reviews, 9:1-30.
Greenland S (2002) Causality theory for policy uses of epidemiological measures. In: Summary measures of population health: concepts, ethics, measurement and applications. Murray CJL, Salomon J, Mathers CD, Lopez AD, eds. World Health Organization, Geneva.

Greenland S, Robins JM (1988) Conceptual problems in the definition and interpretation of attributable fractions. American Journal of Epidemiology, 128:1185-1197.

He J, Klag MJ, Whelton PK et al. (1991a) Migration, blood pressure pattern, and hypertension: the Yi migrant study. American Journal of Epidemiology, 134:1085-1101.
He J, Tang YC, Tell GS, Mo PS, He GQ (1991b) Effect of migration on blood pressure: the Yi people study. Epidemiology, 2:88-97.
Hill AB (1965) The environment and disease: association or causation? Proceedings of the Royal Society of Medicine, 58:295-300.
Horton R (2000) Common sense and figures: the rhetoric of validity in medicine (Bradford Hill memorial lecture 1999). Statistics in Medicine, 19:3149-3164.

Horwitz RI, Singer BH, Makuch RW, Viscoli CM (1998) Clinical versus statistical considerations in the design and analysis of clinical research. Journal of Clinical Epidemiology, 51:305-307.
Horwitz RI, Singer BH, Viscoli CM, Makuch RW (1996) Can treatment that is helpful on average be harmful to some patients? A study of the conflicting information needs of clinical inquiry and drug regulation. Journal of Clinical Epidemiology, 49:395-400.
Horwitz RI, Viscoli CM, Clemens JD, Sadock RT (1990) Developing improved observational methods for evaluating therapeutic evidence. American Journal of Medicine, 89:630-638.

Howarth RH (1996) Climate change and overlapping generations. Contemporary Economic Policy, 14:100-111.
Kay D, Prüss AM, Corvalan CF (2000) Methodology for assessment of environmental burden of disease. ISEE session on environmental burden of disease. World Health Organization, Geneva.

Kneese AV (1999) The Faustian bargain. In: The RFF reader in environment and resource management. Oates WE, ed. Resources for the Future, Washington, DC.

Koopmans TC (1960) Stationary ordinal utility and impatience. Econometrica, 28:287-309.
Kunzli N, Kaiser R, Medina S et al. (2000) Public-health impact of outdoor and traffic-related air pollution: a European assessment. The Lancet, 356:795-801.
LaRosa JC, He J, Vupputuri S (1999) Effect of statins on risk of coronary disease: a meta-analysis of randomized controlled trials. Journal of American Medical Association, 282:2340-2346.
Law MR, Wald NJ, Thompson SG (1994) By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? British Medical Journal, 308:367-373.
Leigh J, Macaskill P, Kuosma E, Mandryk J (1999) Global burden of disease and injury due to occupational factors. Epidemiology, 10:626-631.
Levin ML (1953) The occurrence of lung cancer in man. Acta Unio Internationalis Contra Cancrum, 9:531-541.
Lind RC, ed. (1982) Discounting for time and risk in energy policy, Johns Hopkins University Press for Resources for the Future, Baltimore, MD.

MacMahon B, Pugh TF (1970) Epidemiology: principles and method. Little Brown and Company, Boston, MA.
MacMahon S, Peto R, Cutler J et al. (1990) Blood pressure, stroke, and coronary heart disease. Part I, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. The Lancet, 335:765-774.
Maldonado G, Greenland S (2002) Estimating causal effects. International Journal of Epidemiology, 31:422-429.

Mann GV, Shaffer RD, Anderson RS, Sandstead HH (1964) Cardiovascular disease in the Masai. Journal of Atherosclerosis Research, 4:289-312.
Manton KG, Dowd JE, Stallard E (1993a) The effects of risk factors on male and female cardiovascular risks in middle and late age. In: Forecasting the bealth of elderly populations. Manton KG, Singer BH, Suzman RM, eds. Springer-Verlag, New York.
Manton KG, Singer BH, Suzman RM (1993b) Forecasting the health of elderly populations. Springer-Verlag, New York.
Manton KG, Stallard E, Singer BH (1994) Methods for projecting the future size and health status of the US elderly populations. In: Studies in the economics of aging. Wise DA, ed. University of Chicago Press, Chicago, IL.

Martens P, Kovats RS, Nijhof S et al. (1999) Climate change and future populations at risk of malaria. Global Environmental Change-Human and Policy Dimensions, 9:S89-S107.
Mathers CD, Ezzati M, Lopez AD, Murray CJL, Rodgers AD (2002) Causal decomposition of summary measures of population health. In: Summary measures of population health: concepts, ethics, measurement and applications. Murray CJL, Salomon J, Mathers CD, Lopez AD, eds. World Health Organization, Geneva.
McGinnis JM, Foege WH (1993) Actual causes of death in the United States. Journal of American Medical Association, 270:2207-2212.
Miettinen OS (1974) Proportion of disease caused or prevented by a given exposure, trait or intervention. American Journal of Epidemiology, 99: 325-332.

Montgomery WD (1999) A market-based discount rate: comments on Bradford. In: Discounting and intergenerational equity. Portney PR, Weyant JP, eds. Resources for the Future, Washington, DC.
Morgan MG, Henrion M (1990) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis, Cambridge University Press, Cambridge.
Murray CJL (1996) Rethinking DALYs. In: The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, Vol. 1. Murray CJL, Lopez AD, eds. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Murray CJL, Acharya A (2002) Age weights and discounting in health gaps reconsidered. In: Summary measures of population health: concepts, ethics, measurement and applications. Murray CJL, Salomon J, Mathers CD, Lopez AD, eds. World Health Organization, Geneva.
Murray CJL, Lopez AD, eds. (1996) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, Vol. 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.
Murray CJL, Lopez AD (1997) Global mortality, disability, and the contribution of risk factors: global burden of disease study. The Lancet, 349:1436-1442.

Murray CJL, Lopez AD (1999) On the comparable quantification of health risks: lessons from the global burden of disease. Epidemiology, 10:594-605.

Murray CJL, Salomon J, Mathers CD, Lopez AD, eds. (2002) Summary measures of population health: concepts, ethics, measurement and applications. World Health Organization, Geneva.
National Research Council (1994) Science and judgment in risk assessment. National Academy Press, Washington, DC.
Ouellet BL, Romeder JM, Lance JM (1979) Premature mortality attributable to smoking and hazardous drinking in Canada. American Journal of Epidemiology, 109:451-463.
Parfit D (1984) Reasons and persons. Oxford University Press, New York.
Patz JA, McGeehin MA, Bernard SM et al. (2000) The potential health impacts of climate variability and change for the United States: executive summary of the report of the health sector of the US national assessment. Environmental Health Perspectives, 108:367-376.
Peto R (1986) Influence of dose and duration of smoking on lung cancer rates. In: Tobacco: a major international health hazard. Zaridze D, Peto R, eds. (IARC Scientific Publication No. 7.) International Agency for Research on Cancer, Lyon.
Peto R, Lopez AD, Boreham J, Thun M, Heath C, Jr. (1992) Mortality from tobacco in developed countries. The Lancet, 339:1268-1278.
Pignone M, Phillips C, Mulrow C (2000) Use of lipid lowering drugs for primary prevention of coronary heart disease: meta-analysis of randomised trials. British Medical Journal, 321:983-986.
Portney PR, Weyant JP, eds. (1999) Discounting and intergenerational equity. Resources for the Future, Washington, DC.

Powles J, Day N (2002) Interpreting the global burden of disease. The Lancet, 360:1342-1343.

Preston SH (1976) Mortality patterns in national populations: with special reference to recorded causes of death. Academic Press, New York.

Prospective Studies Collaboration (1995) Cholesterol, diastolic blood pressure, and stroke: 13000 strokes in 45000 people in 45 prospective cohorts. The Lancet, 346:1647-1653.
Puddey IB, Rakic V, Dimmitt SB, Beilin LJ (1999) Influence of pattern of drinking on cardiovascular disease and cardiovascular risk factors: a review. Addiction, 94:649-663.
Reiter P (2001) Climate change and mosquito-borne diseases. Environmental Health Perspectives, 109:S141-161.
Rice AL, Sacco L, Hyder A, Black RE (2000) Malnutrition as an underlying cause of childhood deaths associated with infectious diseases in developing countries. Bulletin of the World Health Organization, 108:367-376.

Robins JM (1986) A new approach to causal inference in mortality studies with sustained exposure periods-application to control of the healthy worker survival effect. Mathematical Modelling, 7:1393-1512.

Robins JM (1987) Addendum to "a new approach to causal inference in mortality studies with sustained exposure periods-application to control of the healthy worker survival effect". Computers and Mathematics with Applications, 14:917-921.
Robins JM (1997) Causal inference from complex longitudinal data. In: Latent variable modeling and applications to causality. Lecture notes in statistics (120). Berkane M, ed. Springer-Verlag, New York.

Robins JM (1999a) Association, causation, and marginal structural models. Synthese, 121:151-179.
Robins JM (1999b) Marginal structural models versus structural nested models as tools for causal inference. In: Statistical models in epidemiology. Halloran E, ed. Springer-Verlag, New York.

Robins JM, Greenland S (1989) Estimability and estimation of excess and etiological fractions. Statistics in Medicine, 8:845-859.
Robins JM, Greenland S (1991) Estimability and estimation of expected years of life lost due to a hazardous exposure. Statistics in Medicine, 10:79-93.

Robins JM, Greenland S (2000) Comment on "causal inference without counterfactuals" by Dawid AP. Journal of American Statistical AssociationTheory and Methods, 95:477-482.
Robins JM, Greenland S, Fu-Chang H (1999) Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. Journal of American Statistical Association, 94:687-700.
Rockhill B, Newman B, Weinberg C (1998) Use and misuse of population attributable fractions. American Journal of Public Health, 88:15-19.

Rogers DJ, Randolph SE (2000) The global spread of malaria in a future, warmer world. Science, 289:1763-1766.
Rothman KJ (1976) Causes. American Journal of Epidemiology, 104:587-592.
Rothman KJ, Greenland S (1998) Modern epidemiology. Lippincott-Raven, Philadelphia, PA.

Schelling TC (1995) Intergenerational discounting. Energy Policy, 23:395-401.
Single E, Robson L, Rehm J, Xie X (1999) Morbidity and mortality attributable to alcohol, tobacco, and illicit drug use in Canada. American Journal of Public Health, 89:385-390.

Smith KR (2000) The national burden of disease from indoor air pollution in India. Proceedings of the National Academy of Sciences, 97:13286-13293.

Smith KR, Corvalan CF, Kjellstrom T (1999) How much global ill health is attributable to environmental factors. Epidemiology, 10:573-584.
Smith KR, Samet JM, Romieu I, Bruce N (2000) Indoor air pollution in developing countries and acute lower respiratory infections in children. Thorax, 55:518-532.
Solow RM (1974a) The economics of resources or the resources of economics. The American Economic Review, 64:1-14.

Solow RM (1974b) Intergenerational equity and exhaustible resources. The Review of Economic Studies, 41:29-45.
Solow RM (1999) Foreword. In: Discounting and intergenerational equity. Portney PR, Weyant JP, eds. Resources for the Future, Washington, DC.
Tate RB, Manfreda J, Krahn AD, Cuddy TE (1995) Tracking of blood pressure over a 40-year period in the University of Manitoba follow-up study, 1948-1988. American Journal of Epidemiology, 142:946-954.
Thun MJ, Apicella LF, Henley SJ (2000) Smoking vs other risk factors as the cause of smoking-attributable mortality: confounding in the courtroom. Journal of American Medical Association, 284:706-712.
Toman MA (1999) Reconciling philosophy and economics in long-term discounting. In: Discounting and intergenerational equity. Portney PR, Weyant JP, eds. Resources for the Future, Washington, DC.
Uter W, Pfahlberg A (2001) The application of methods to quantify attributable risk in medical practice. Statistical Methods in Medical Research, 10: 231-237.

Walter SD (1980) Prevention of multifactorial disease. American Journal of Epidemiology, 112:409-416.
Weitzman ML (1999) Just keep discounting, but ... In: Discounting and intergenerational equity. Portney PR, Weyant JP, eds. Resources for the Future, Washington, DC.
Willet WC (2002) Balancing life-style and genomics research for disease prevention. Science, 296:695-698.
WHO (1992) International statistical classification of disease and related health problems. Tenth edition. World Health Organization, Geneva.
Yashin AI, Manton KG, Stallard E (1986) Dependent competing risks: a stochastic process model. Journal of Mathematical Biology, 24:119-140.
Yerushalmy J, Palmer CE (1959) On the methodology of investigations of etiologic factors in chronic diseases. Journal of Chronic Disease, 108:27-40.
Yong LC, Kuller LH, Rutan G, Bunker C (1993) Longitudinal study of blood pressure: changes and determinants from adolescence to middle age. The Dormont High School follow-up study, 1957-1963 to 1989-1990. American Journal of Epidemiology, 138:973-983.

## Chapter 2

# Childhood and maternal UNDERWEIGHT 

Steven M. Fishman, Laura E. Caulfield, Mercedes de Onis, Monika Blössner, Adnan A. Hyder, Luke Mullany and
Robert E. Black

## Summary

Undernutrition, as measured by underweight status, has been associated with substantially increased risk of childhood mortality worldwide, but the magnitude of its effect on specific causes of mortality and morbidity has been less well described. We reviewed extant data and conducted statistical meta-analyses in order to determine the relative risk of disease and death attributable to underweight status. We used these estimates in combination with estimates of underweight prevalence among children aged $0-4$ years and women of reproductive age (15-44 years) to calculate the global burden of disease attributable to undernutrition.

Underweight was defined for children aged $0-4$ years as low weight-for-age relative to the National Center for Health Statistics/World Health Organization (NCHS/WHO) reference median. The theoretical-minimum-risk distribution for this population was equivalent to the reference population distribution, wherein $2.3 \%$ of children have a weight-for-age below -2 standard deviations (SDs) (weight-for-age $<-2$ SDs, or weight-for-age $z$-score [WAZ] $<-2$ ) and are classified as underweight. The relationship between low weight-for-age among children and mortality was analysed as a multiple categorical variable having four levels of exposure: WAZ $<-3$, WAZ -3 to -2 , WAZ -2 to -1 and WAZ >-1 (reference category). Morbidity was analysed as a dichotomous variable, comparing WAZ $<-2$ to WAZ $>-2$ (reference category). Underweight was defined for women of reproductive age as prepregnant body mass index (BMI) below $20 \mathrm{~kg} / \mathrm{m}^{2}$. The outcomes chosen for review were major diseases and disabilities commonly associated with nutritional deficiencies. Sufficient prevalence and risk data were available to analyse childhood underweight status as a risk factor for mortality due to diarrhoeal disease, measles, malaria and pneumonia.

Childhood underweight status was also analysed as a risk factor for increased incidence of these infectious diseases. Maternal underweight was considered as a risk for increased risk of conditions arising during the perinatal period. The relationships were considered causal on the basis of consistent evidence observed across different populations and settings, and persistence of the association after adjustment for recognized confounders and effect modifiers such as age, breastfeeding status and prior morbidity. Current prevalence and risk data were not suitable for deriving burden estimates relating to other outcomes or older populations, although this does not imply the lack of risk in those populations. We therefore reviewed and summarized two major areas of undernutrition research without calculating burden estimates: (i) effects of maternal underweight status during pregnancy on maternal mortality and fetal death; and (ii) the association between undernutrition and cognitive function.

Prevalence of undernutrition among children aged 0-4 years was obtained from the WHO Global Database on Child Growth and Malnutrition, derived from a systematic analysis of raw data sets from 310 nationally representative nutritional surveys that collected data on child anthropometry in 112 countries. Categorical underweight estimates for most subregions ${ }^{1}$ were calculated using multilevel modelling that adjusted for variability between regions, countries and surveys. Prevalence of underweight status was highest in SEAR-B (25.8\%), SEAR-D ( $45.9 \%$ ), AFR-D ( $32.2 \%$ ), AFR-E ( $31.1 \%$ ) and EMR-D ( $25.1 \%$ ). Projections based on national trends suggested declining prevalence in most subregions over the next 30 years but increasing prevalence in the African subregions.

Cause-specific mortality was estimated by extending the methods of Pelletier et al. (1994) to data obtained from the investigators of 10 cohort studies in which both weight-for-age category ( $<-3$ SDs; -3 to -2 SDs; -2 to -1 SDs; and $>-1$ SD) and cause-of-death information were available. All studies contributed information on weight-for-age and risk of diarrhoea, pneumonia and all-cause mortality. However, only six studies contributed information on deaths due to measles, and only three studies contributed information on deaths due to malaria or, in some cases, fever. By calculating the logarithm of the mortality rates by cause and anthropometric status in each country and utilizing weighted random effects models, we estimated the relation between weight-for-age and risk of death and calculated the relative risk of dying for each cause and allcause mortality from these models. The relative risks of mortality due to low weight-for-age were elevated for each cause of death. The attributable fractions of mortality associated with weight-for-age below-1 SD were $44.8 \%$ for measles, $57.3 \%$ for malaria, $52.3 \%$ for pneumonia and $60.7 \%$ for diarrhoea.

Estimation of cause-specific morbidity was based on statistical metaanalysis of published data identified through systematic literature
searches of Medline and other databases. We selected longitudinal studies that compared incidence data according to past anthropometric status. Underweight status among preschool-age children was significantly associated with subsequent risk of diarrhoea and pneumonia episodes, but the association with malaria was not statistically significant. There was no evidence that underweight status influenced susceptibility to measles infection. The overall attributable fractions of morbidity associated with weight-for-age below -2 SDs were $16.5 \%$, $5.3 \%$ and $8.2 \%$ for pneumonia, diarrhoea and malaria, respectively.

Mortality due to perinatal conditions was calculated by estimating the attributable fraction of neonatal mortality due to intrauterine growth retardation (IUGR) and then multiplying the value by the estimated attributable fraction of IUGR due to low maternal pre-pregnancy BMI for each subregion. Attributable fractions were based on prevalence and risk estimates from available published and unpublished sources.

Overall, undernutrition in children aged $0-4$ years, as reflected in underweight or low weight-for-age, caused 3599800 deaths, including 815900 diarrhoea deaths, 1042900 pneumonia deaths, 261300 measles deaths and 549200 malaria deaths. There were an additional 148400 low-birth-weight neonatal deaths associated with low maternal prepregnancy BMI. The total number of child deaths associated with maternal or child underweight status, 3748200 , represented $34.7 \%$ of all child deaths in the age group 0-4 years. The loss in disability-adjusted life years (DALYs) associated with these deaths was over 126 million. An additional 879900 DALYs were lost due to increased morbidity from pneumonia, diarrhoea and malaria. Of the total DALYs associated with undernutrition, 44 million ( $32.0 \%$ ) were lost in SEAR-D, 30 million ( $21.7 \%$ ) in AFR-D, 33 million ( $23.9 \%$ ) in AFR-E, 16 million ( $11.5 \%$ ) in EMR-D and 8 million ( $5.8 \%$ ) lost in WPR-B. Altogether, these five subregions accounted for $95 \%$ of the total DALYs lost due to undernutrition.

The true burden of disease associated with undernutrition extends beyond the narrow focus of this analysis. In addition to the effects of poor pre-pregnancy BMI, inadequate weight gain during pregnancy can lead to IUGR. Fetal growth retardation, in turn, may be associated with impaired immunocompetence, as well as increased risk of chronic disease in later life. Chronic undernutrition, especially in conjunction with poor environmental stimulation, is associated with impaired cognitive development, and severe undernutrition during infancy may contribute to lasting intellectual deficits. Chronic undernutrition among adults can contribute to diminished work capacity. Additional research is necessary into the prevalence and impact of undernutrition throughout the life cycle.

## 1. Introduction

Child undernutrition-measured as poor anthropometric status-is internationally recognized as an important public health indicator for monitoring nutritional status and health in populations. Young children are most vulnerable to undernutrition and face the greatest risk of its adverse consequences. Those who suffer from growth retardation as a result of poor diets and/or recurrent infections tend to have more frequent episodes of severe diarrhoea and are more susceptible to several infectious diseases, such as meningitis and pneumonia (Man et al. 1998; Tomkins and Watson 1989; Victora et al. 1994). A number of studies have demonstrated the association between increasing severity of anthropometric deficits and mortality, and undernutrition is thought to be a contributing factor in over half of all child deaths in developing countries (Pelletier 1994; Pelletier et al. 1993; Rice et al. 2000; Schroeder and Brown 1994). There is strong evidence that poor growth is associated with delayed mental development (de Onis 2001; Mendez and Adair 1999; Pollitt et al. 1993; WHO 1999a), and several studies have shown a relationship between impaired growth status and poor school performance as well as reduced intellectual achievement (Martorell et al. 1992; PAHO 1998). In addition, growth retardation in early childhood is associated with significant functional impairment in adult life (Martorell et al. 1992) and reduced work capacity (Spurr et al. 1977), which in turn has an impact on economic productivity.

The purpose of this chapter is to describe our review and analysis of existing data in order to calculate relative risks of specific causes of death and morbidity attributable to undernutrition (defined by low weight-forage or "underweight"), and to use the estimates, in combination with estimates of underweight prevalence from a comprehensive WHO database, to calculate the global burden of disease associated with underweight status.

### 1.1 Determinants of childhood undernutrition

Undernutrition has several levels of determinants. Poverty is a strong underlying determinant that leads to household food insecurity, poor childcare, maternal undernutrition, unhealthy environments and poor health care. These factors then lead to the immediate determinants of childhood undernutrition, that is, low birth weight, inadequate dietary intake of nutrients and frequent infectious diseases (Baqui and Black 2002). Low birth weight, primarily due to IUGR in developing countries, is a consequence of maternal undernutrition prior to and during pregnancy and subsequently contributes to undernutrition in infancy and childhood (Villar and Belizan 1982). This is especially important in areas such as South Asia, where there is a very high prevalence of low birth weight. The diets of many children in developing countries are inadequate, and children in the first two years of life are at particular risk.

During this period, children have a high rate of growth and demand for calories, protein, essential fats, vitamins and minerals. Breastfeeding provides excellent nutritional support for six months (Kramer and Kakuma 2002), but unfortunately many children in such settings are given fluids and other foods before this age, resulting in reduction of breast milk intake and exposure to infectious agents causing diarrhoea (Lutter 2000). After six months of age, even when breastfeeding is continued, children frequently lack sufficient dietary intake, or consume a diet that is of poor nutritional content (WHO 1998). Again, these dietary deficiencies may involve not only the macronutrients, but also the socalled micronutrients, that is, vitamins and minerals. In young children, the frequently contaminated environments and poor childcare practices cause high rates of infectious disease. These infections result in a reduction in nutrient intake, as well as increased utilization and loss of nutrients, such as vitamin A and zinc.

The relative importance of the three immediate determinants of undernutrition varies by setting. The percentage of growth faltering (compared with an international reference population) due to diarrhoea in developing countries in Latin America, Africa and Asia has been reported to range from $10 \%$ to $80 \%$ (Black 1991). A study in Bangladesh simultaneously examined the role of diarrhoea, other febrile illnesses and dietary intake on weight gain (Becker et al. 1991). In a model using data from this study, it was estimated that improving dietary intake to recommended levels would have a slightly greater effect than eliminating diarrhoea and febrile illness; however, doing both at the same time would be necessary to achieve growth equivalent to an international reference population. There are a number of factors specific to particular developing country settings that can moderate the effect of illness on growth (Black 1991). For example, the adverse effect of diarrhoea on growth is less in exclusively breast-fed children than in children after weaning. Furthermore, an adequate diet, such as that provided in supplementation programmes, may prevent the adverse effect of diarrhoea on growth in some (Lutter et al. 1989) but not all settings (Bhandari et al. 2001). Thus, undernutrition is due to a variety of determinants that act along a pathway from poverty to dietary deficiency and frequent infectious diseases of childhood.

### 1.2 OTHER AGE GROUPS

Undernutrition is not limited to young children. Over 815 million people of all ages are considered undernourished (FAO 2001), including roughly 243 million adults in developing countries who are considered severely underweight (ACC/SCN 2000a). Poor anthropometric status among adults and adolescents has been associated with maternal complications, diminished work capacity and increased risk of mortality (Martorell et al. 1992; Rotimi et al. 1999; Spurr et al. 1977; WHO 1995b). However, prevalence data for most adults, school-age children, adolescents and the
elderly are scarce, and understanding of the health effects of undernutrition in these populations is incomplete. There is a need for further research on undernutrition throughout the life cycle, especially as populations in developing countries go through demographic and epidemiological transitions.

### 1.3 Exposure variable

The internationally recommended way to assess undernutrition at the population level is to take body or anthropometric measurements (e.g. weight and height). Based on combinations of these body measurements anthropometric indices are constructed. These indices are essential for the interpretation of measurements, as the value for body weight alone, for example, has no meaning unless it is related to an individual's age or height (WHO 1995b). In children the three most commonly used anthropometric indices are weight-for-height, height-for-age and weight-for-age. These anthropometric indices can be expressed in terms of $z$-scores, percentiles, or percentage of median, which can then be used to compare a child or a group of children with a reference population.

Weight-for-age was chosen as the index of child nutritional status for this analysis because it is the most widely used in developing countries, allowing for the inclusion of the largest number of studies. Although it does not distinguish between wasting and stunting, low weight-for-age (underweight) represents a combination of both aspects and has a high positive predictive value as an indicator for child malnutrition in developing countries (WHO 1995b). Underweight is defined internationally as the proportion of preschool children falling below -2 SDs (weight-for-age $<-2$ SDs, or WAZ <-2) from the NCHS/WHO international reference median value for weight-for-age (de Onis and Blössner 1997; WHO 1995b). The burden of disease estimates for mortality were based on four weight-for-age categories: $<-3$ SDs; between -3 and -2 SDs; between -2 and -1 SDs; and the reference category of $>-1 \mathrm{SD}$. The estimates for morbidity were based on two categories: weight-for-age $<-2$ SDs and the reference category of $>-2$ SDs.

Although the terms are often used synonymously, "underweight" should be distinguished from the terms "malnourished" and "undernourished". "Undernourished" and "malnourished" refer to the internal, physiological state of nutriture. "Underweight" is an observable anthropometric marker for that state, while the internal process can only be inferred.

BMI was selected as the index of adult nutritional status recommended by the International Dietary Energy Consultative Group of the Administrative Committee on Coordination/Sub-Committee on Nutrition (Shetty and James 1994). BMI is less biased by height differences than other adult indicators, such as absolute weight, and correlates with
health-related outcome variables such as overall mortality risk. For evaluation of nutritional status in relation to pregnancy, only pre-pregnant BMI was considered. Unlike the classification of underweight status in children, there is no comparable international reference for BMI among adults. A BMI $<18.5 \mathrm{~kg} / \mathrm{m}^{2}$ represents chronic energy deficiency, and the Institute of Medicine recommends a BMI cut-off of $<19.8 \mathrm{~kg} / \mathrm{m}^{2}$ as the definition for underweight (Institute of Medicine 1990; Shetty and James 1994). The cut-off of $20 \mathrm{~kg} / \mathrm{m}^{2}$ is commonly used in studies of maternal risk, and, therefore, underweight status among women of reproductive age was defined as BMI $<20 \mathrm{~kg} / \mathrm{m}^{2}$ for this analysis. Further details on BMI can be found in the chapter on overweight and obesity (chapter 8).

### 1.4 Considerations in choice of child ANTHROPOMETRIC INDICATOR

Nutritional status can be assessed using clinical signs of malnutrition, biochemical indicators and anthropometry. Inadequacies in nutritional intake eventually alter functional capacity and result in many adverse health outcomes that are distinct expressions of malnutrition's different levels of severity. Initially, children adapt to inadequate diets through reduced physical activity and slowed rates of growth. At moderate degrees of malnutrition, activity and growth rates are affected to a greater degree, and in addition signs of wasting and some biochemical abnormalities (e.g. reduction in serum albumin), begin to show. At advanced stages of severity, all linear growth ceases, physical activity is severely curtailed, body wasting is marked and clinical signs (e.g. oedema, hair and skin changes) are noticeable. Anthropometry thus has an important advantage over other nutritional indicators: body measurements are sensitive over the full spectrum of malnutrition, whereas biochemical and clinical indicators are useful only at the extremes. In addition, anthropometric measurements are non-invasive, inexpensive and relatively easy to obtain. The main disadvantage of anthropometry is its lack of specificity, as changes in body measurements are also sensitive to several other factors, including infection, altitude, stress and genetic background (de Onis 2000).

A child's body responds to malnutrition in two ways that can be measured by anthropometry: (i) a deceleration or cessation of growth, which over the long-term results in low height-for-age or stunting; and (ii) body wasting, which is a short-term response to inadequate intakes, commonly assessed by weight relative to height. Height-for-age and weight-for-height thus discriminate between different biological processes, unlike weight-for-age, which could be low because of stunting (short stature) and/or wasting (recent weight loss). The current estimates relied exclusively on weight-for-age, which reflects body mass relative to chronological age. It is influenced by both the height of the child (height-for-age) and her weight (weight-for-height). Hence weight-
for-age cannot discriminate between short- and long-term forms of malnutrition given that children classified on its basis are a mixed group in terms of their nutritional status. In the 1990s weight-for-height emerged as a very important indicator (WHO 1995b) and, in fact, several authors have identified low weight-for-height as the indicator of choice for screening malnourished children who are at increased risk of dying (de Onis 2000).

### 1.5 Theoretical-minimum-exposure distribution

The prevalence of underweight status is defined as the proportion of preschool-age children falling below - 2 SDs indicated from the NCHS/ WHO international reference median value (de Onis and Blössner 1997; WHO 1995b). The theoretical-minimum-risk distribution is equivalent to this distribution, wherein $13.6 \%$ of children aged 0-4 years have weight-for-age between -1 to -2 SDs and $2.3 \%$ have a weight-for-age <-2SDs.

Anthropometric values are compared across individuals or populations in relation to a set of reference values based upon a healthy population. The choice of reference population to assess nutritional status has a significant impact on the proportion of children identified as being malnourished and, in turn, important implications for establishing relationships between nutritional status and functional outcomes (WHO 1995a). Much has been written about growth references, but unanswered questions remain about the many factors that determine human growth and indeed what constitutes "normal" growth. WHO has been recommending since the late 1970s the use of the NCHS/WHO international reference population $z$-scores for the comparison and presentation of child malnutrition data (Waterlow et al. 1977). A detailed account of the different growth references used prior to the current international reference is provided elsewhere (de Onis and Yip 1996). In the mid-1990s the NCHS/WHO international reference was found to have important technical and biological drawbacks (de Onis and Habicht 1996; de Onis and Yip 1996; WHO 1995a). Consequently, an international effort coordinated by WHO is under way to develop a new international growth reference for infants and young children (de Onis et al. 1997). The new international reference-which will be constructed from primary data collected for this purpose-includes a number of features that, taken together, will result in a reference population substantially different from the existing ones (de Onis et al. 2001). An important characteristic of this new reference is that it will be based on a truly international sample. Six countries, representing the major global geographic regions, are participating in this effort. Another notable feature is that it takes the breast-fed infant as the biological "norm", recognizing the health and nutritional benefits of breastfeeding. The extent to which the new curves differ from the current ones in shape and the spread of values around the mean will affect the relationship-established using the NCHS/WHO
reference values-between child anthropometry and functional outcomes such as morbidity and mortality.

Once a reference population has been selected, it is necessary to determine the limits of "normality". Current practice is to use a defined cutoff point using one of the three available classification systems for comparing a child, or a group of children, to the reference population: $z$-scores (SD scores), percentiles and percentage of median. The use of a statistically defined cut-off point (e.g. $-2 \mathrm{SD} z$-score) is not unique to anthropometry; indeed, it is widely applied in many fields of biology and medicine. Nevertheless, it is important to bear in mind that using a cut-off-based criterion to define what is "abnormal" is arbitrary. In reality, there are not two distinct populations-one well nourished and the other malnourished-but rather a continuous gradation of nutritional status (de Onis 2000). The risk of undesirable health outcomes such as mortality does not change dramatically by simply crossing the cut-off line; risks are continuous within the "normal" range. For many purposes, the best descriptor of a population's nutritional status is the mean $z$-score, which in less developed environments is usually shifted to the left. This concept is based on data from many different countries which showed a high consistency in the SD of weight-for-height among young children (de Onis and Blössner 1997). Even under extreme conditions, such as during famines, where the mean $z$-score is two or three units below the reference, the value of the SD of $z$-scores is very close to unity. This shows that the entire distribution is shifted so that all individuals, not only those below a given cut-off point, are affected (Rose 1985). Taking such a populationapproach resolves the problem of focusing solely on the severely malnourished subpopulation falling below a certain cut-off. In most instances, the mild and moderately malnourished subpopulations will be of greater importance from a public health perspective because there are many more children in these categories than in the severely malnourished one. For this reason, estimates are presented using low weight-for-age as both a dichotomous and multi-categorical variable when possible.

## 2. <br> Prevalence of underweight AMONG CHILDREN

Over the years the WHO Department of Nutrition has been monitoring trends in child malnutrition using anthropometric data. A major difficulty has been the lack of comparability of survey results. Although many nutritional surveys have been conducted since the 1970s, many of them had used distinct definitions of malnutrition (i.e. different anthropometric indicators, reporting systems, cut-off points and reference values) making comparison among studies difficult. This lack of comparable data prompted the beginning of WHO's systematic collection and standardization of data on the nutritional status of the world's population of children <5 years. Initial results of this effort, published in 1993 (de

Onis et al. 1993), were updated in 1997 and presented together with estimates of trends in child growth retardation in developing countries (de Onis and Blössner 1997). A more recent analysis using multilevel modelling updated these earlier estimates, describing trends in child malnutrition from 1980-2005 (de Onis et al. 2000).

### 2.1 DATA SOURCES AND QUALITY CONTROL

Cross-sectional data on the prevalence of underweight were obtained from nationally representative nutritional surveys that collected data on child anthropometry and are included in the WHO Global Database on Child Growth and Malnutrition. This database was initiated in 1986 to compile, standardize and disseminate the results of nutritional surveys performed in both developing and developed countries (de Onis and Blössner 1997). A distinct feature of this database is the systematic analysis of raw data sets in a standard format to produce comparable results. Only nationally representative data have been used for the present analysis. Three hundred and ten national nutrition surveys from 112 countries carried out from 1965 onwards were analysed to estimate the prevalence of underweight in children. The majority of surveys were conducted by either the relevant national ministry of the country concerned, by the Demographic and Health Surveys (DHS) programme of Macro International, or by the Multiple Indicators Cluster Surveys (MICS) carried out by the United Nations Children's Fund (UNICEF). Appendix A lists all national surveys included in the present analysis by country in alphabetical order.

Survey data for inclusion in the WHO Global Database on Child Growth and Malnutrition are identified by various ways and mechanisms.

- An automated Medline search provides results according to the established search history following the weekly online Medline updates. Selected abstracts are reviewed and relevant surveys with data searched for in the library. Data are extracted and frequently the principal investigators and/or data holders are contacted to complement the data provided in the article and/or give clarification on methodological issues.
- A wide network of collaborators, including international organizations (e.g. UNICEF, Food and Agriculture Organization of the United Nations [FAO]), nongovernmental organizations (NGOs) (e.g. Helen Keller International), Ministries of Health and National Institutes of Nutrition, as well as independent institutions (e.g. Macro International) and individual researchers provide their data directly to the WHO Global Database. In case of queries on the data the same procedure as above takes place to clarify any unclear issues.
- Principal investigators contact the WHO Global Database to enquire about possible inclusion of their data into the Global Database.
- Other WHO database managers within WHO headquarters share data sets and survey documents with the Global Database.

The sampling methods for each of the survey data used in this analysis were carefully reviewed to ensure national representativeness. Multistage random sampling methods were applied for sample selection in the majority of countries. Only a few countries, such as Argentina, Chile, Croatia, Uruguay and Venezuela based their estimates on well-established national nutritional surveillance systems with high population coverage (country-specific details on sampling procedures are available from the authors on request). Surveys generally followed standard procedures of measuring length up to 24 months of age and height from 24 months onwards. The anthropometric measurement techniques used in each survey are described in the comprehensive survey reports, which are available on request. Some surveys included information on reliability of the measurements while others did not.

Survey results were checked for inconsistencies between the estimates based on height-for-age, weight-for-age and weight-for-height. The observed SDs of the $z$-score distribution were used to assess the quality of the survey results. With accurate age estimates and anthropometric measurements, the SDs of the observed height-for-age, weight-for-age and weight-for-height $z$-score distributions should be relatively constant and close to the expected value of 1.0 for the reference distribution (ranging within approximately 0.2 units). This nearly constant SD in height-based and weight-based $z$-score distributions provides an opportunity to assess data quality (WHO 1995b). Surveys with a SD outside the expected ranges were subjected to closer examination because of possible problems related to age assessment and anthropometric measurements. Surveys with inaccurate data resulting from measurement error or incorrect age reporting were excluded from this analysis.

## $2.2 \quad$ Statistical analysis

The analysis followed three separate steps, which are described below. The number of countries and percentage of population aged $<5$ years covered by data for each subregion are listed in Appendix B. The population aged <5 years derived from the 1998 revision of the UN population estimates (UN 1999) for the particular survey year were added as population weights. As there were no significant differences observed between male and female child underweight prevalence, the model exercise included data only for sexes combined, and estimates were assumed to be the same for males and females.

Primary analysis of raw data sets
As an essential first step in producing subregional estimates, primary analysis of the 310 national raw data sets was conducted to produce
standardized estimates of prevalence of underweight as defined in section 1.1. This was necessary because many nutritional surveys used distinct definitions of malnutrition (i.e. different anthropometric indicators, reporting systems, cut-off points and reference values) making impossible the pooling of reported prevalences. There were three steps involved in this analysis: (i) identifying the data holder for each of the individual surveys; (ii) requesting re-analysis of the original data sets following the standardized definition, or otherwise, requesting a copy of the raw data set for standard analysis by the WHO Department of Nutrition; and (iii) initiating quality control procedures for each individual survey. Countryspecific prevalences of child underweight by age, sex and subregion derived for each of the national surveys listed in Appendix A can be found on the web site of the WHO Global Database on Child Growth and Malnutrition at http://www.who.int/nutgrowthdb.

## Subregional estimates: Underweight as a dichotomous variable

The methodology applied to derive the underweight estimates depended on the availability of data points for the different subregions. In detail these were as follows: for AFR-D, AFR-E, AMR-B, AMR-D, EMR-B, EMR-D, SEAR-B, SEAR-D and WPR-B the underweight estimates were calculated using a multilevel model. Multilevel modelling allows for more than one component of variation, which in this case includes between subregions, between countries within subregions and between surveys over time within countries. This multilevel model is an extended form of regression, which separates estimates for the three levels of variation, and the total variation is obtained by combining over the three levels. To adjust for skewness in the distribution of prevalence, the model used the logit transform of the underweight prevalence, with the additional advantage of stabilizing the variance. Because estimates were calculated on the logistic scale, estimated prevalence and their CIs were derived by back-transformation. The nature of the logistic function ensured that all estimated prevalence and their CIs would lie above zero. CIs for prevalences close to zero were asymmetric, and were narrower than for values close to $50 \%$-these are intrinsic properties of the logistic function (Armitage and Berry 1994).

Multilevel modelling was implemented in the Statistical Analysis System (SAS) Proc Mixed program, a procedure that accommodates both categorical and continuous covariates and incomplete series of time measurements, and allows for the added variability introduced by the multiple levels of analysis. Analyses were conducted separately for each subregion, given that the trend of underweight between subregions could differ substantially.

Separate analysis for each subregion required that two rather than three levels be modelled, thus reducing complexity of the model. An assumption of the analyses was that the extent of available data for countries was not related to the prevalence of underweight. In other
words, it was assumed that the trend of the prevalence over time for a subregion could be estimated in an unbiased manner from the data available. Under this assumption the model accommodated the variable patterns of available surveys for the countries. Countries with only one survey contributed information to the estimation of the overall intercept, whereas countries with more than one survey also contributed to the estimation of the regression coefficient(s) relating the trend to the survey year. The models specified a linear trend relationship between prevalence of underweight and survey year. Thus, these models assumed that the rate of change in the prevalence of underweight was constant.

Possible non-linear trends were examined by including quadratic and cubic polynomial terms. There was some evidence for non-linear relationship in only one of the subregions (SEAR-D), showing slightly larger differences between countries and indicating a better fit using the quadratic and cubic functions. However, due to unrealistic drops in prevalence the model outcome was discarded. Further explanation on the multilevel modelling approach to derive estimates has been described elsewhere (ACC/SCN 2000b; de Onis et al. 2000).

No modelling was possible in AMR-A, EUR-A, EUR-B, EUR-C and WPR-A due to lack of sufficient national survey data. To derive estimates for these subregions we proceeded as follows:

- In AMR-A and EUR-A, which are composed of developed countries, we assumed that children have a similar nutritional status as the reference population. Therefore, these two subregions were assigned the prevalence values derived from the international reference population, i.e. $2.3 \%$ (theoretical minimum). For these subregions no CIs were available and uncertainty was assumed as zero.
- In EUR-B, EUR-C and WPR-A, we calculated the weighted mean prevalence of underweight following the method of de Onis and Blössner (2000) for estimating prevalence and trends of overweight among preschool-age children in developing countries. Based on the available survey data the CIs fell within a $95 \%$ CI of: $0.1-15.5$ in EUR-B; 0.1-5.1 in EUR-C; and 0.7-6.9 in WPR-A.

Subregional estimates: Underweight as a multi-Categorical variable
Growth in populations of children is normally distributed. In a standard normal distribution the SD is 1 and the mean is located at 0 . As there is a relationship between the prevalence falling below the cut-off ( $<-2$ SDs) and the mean $z$-score, to convert the mean $z$-score to a prevalence based on $<-2$ SDs, the probit function, which converts $z$-score values to cumulative percentiles, was used as recommended by an Expert Committee (WHO 1995b). To derive probability that the $z$-score of a child is between two cut-off points (<-3SDs; -3 SDs to -2 SDs; -2 SDs to -1 SD; and $>-1$ SD), we calculated the difference between the two cumulative
probabilities obtained from the probit function. Prevalence data below -3 SDs were derived by subtracting the calculated values for the interval -3 to -2 SDs from the total of the prevalence below $<-2$ SDs provided by the categorical analysis (see Table 2.1 and section 4.2), which included all ranges below this cut-off.

### 2.3 Exposure estimates

The prevalence of low weight-for-age among children aged 0-4 years by subregion is listed as a dichotomous variable in Table 2.1 (a) and as a multi-categorical variable in Table 2.1 (b), below.

## 3. Prevalence of underweight among women of reproductive age

Subregional BMI estimates listed in Table 2.2 are from chapter 8. Underweight status is widespread among women of reproductive age in developing countries, especially in sub-Saharan Africa and South Asia. Of 26 sub-Saharan African countries with recent ( $<10$ years old) survey estimates, 23 have greater than $30 \%$ prevalence of $\mathrm{BMI}<20 \mathrm{~kg} / \mathrm{m}^{2}$-five of which (Chad, Eritrea, Ethiopia, Madagascar and the Niger) have prevalence $>50 \%$. Among South Asian women, $70-76 \%$ in Bangladesh and India have a BMI $<20 \mathrm{~kg} / \mathrm{m}^{2}$, along with $54-61 \%$ in Cambodia and Nepal. Low BMI is less common in the Americas, where prevalence of

Table 2.1(a) Underweight prevalence by subregion ${ }^{\text {a }}$

| Subregion | Prevalence of underweight (\% below -2 SDs) | $95 \% \mathrm{Cl}$ |
| :--- | :---: | :---: |
| AFR-D | 32.2 | $(26.7-38.2)$ |
| AFR-E | 31.0 | $(24.7-38.0)$ |
| AMR-A | 2.3 | $(2.0-2.6)$ |
| AMR-B | 5.0 | $(3.8-6.6)$ |
| AMR-D | 12.4 | $(7.7-19.6)$ |
| EMR-B | 8.1 | $(5.4-12.1)$ |
| EMR-D | 25.1 | $(14.5-39.7)$ |
| EUR-A | 2.3 | $(2.0-2.6)$ |
| EUR-B | 7.6 | $(0.1-15.5)$ |
| EUR-C | 2.6 | $(0.1-5.1)$ |
| SEAR-B | 25.8 | $(16.3-38.4)$ |
| SEAR-D | 45.9 | $(40.7-51.2)$ |
| WPR-A | 3.8 | $(0.7-6.9)$ |
| WPR-B | 16.0 | $(12.7-19.9)$ |

[^2]Table 2.1(b) Mean z-scores and underweight prevalence by weight-forage category and subregion ${ }^{\text {a }}$

| Subregion | Mean z-score | Percentage of children in weight-for-age category |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | $<-3$ SDs | $>-3$ to <-2 SDs | $>-2$ to $<-I S D$ | $>-1$ SD to $<0$ |
| AFR-D | -1.54 | 7.1 (4.0-10.4) | 25.1 (19.7-30.3) | 38.3 (32.3-44.2) | 23.3 (18.1-28.5) |
| AFR-E | -I. 5 | 6.8 (3.1-10.3) | 24.2 (18.1-30.4) | 38.3 (31.3-45.3) | 24.2 (17.9-30.2) |
| AMR-A | 0 | 0.1 (0.1-0.2) | 2.1 (1.9-2.4) | 13.6 (13.0-14.3) | 34.1 (33.3-35.1) |
| AMR-B | -0.35 | 0.5 (0.0-0.8) | 4.5 (3.3-6.0) | 20.8 (18.4-23.7) | 37.9 (34.8-41.1) |
| AMR-D | -0.84 | 1.6 (0.0-3.8) | 10.8 (5.3-16.5) | 31.3 (23.1-39.8) | 36.3 (27.6-44.9) |
| EMR-B | -0.6 | 0.8 (0.0-1.9) | 7.3 (4.1-10.5) | 26.3 (21.1-31.9) | 38.1 (32.2-44.1) |
| EMR-D | -1.33 | 4.7 (0.0-10.9) | 20.4 (8.7-32.1) | 37.8 (23.7-51.9) | 27.9 (14.9-4I.0) |
| EUR-A | 0 | 0.1 (0.1-0.2) | 2.1 (1.9-2.4) | 13.6 (13.0-14.3) | 34.1 (33.3-35.1) |
| EUR-B | -0.57 | 0.7 (0.0-3.3) | 6.9 (0.0-14.2) | 25.7 (13.0-38.4) | 38.2 (24.1-52.3) |
| EUR-C | -0.05 | 0.2 (0.0-0.8) | 2.4 (0.0-4.9) | 14.5 (9.1-20.2) | 34.9 (27.5-42.5) |
| SEAR-B | -1.35 | 5.0 (0.0-10.4) | 20.8 (10.6-31.1) | 37.9 (25.7-50.2) | 27.5 (16.2-38.7) |
| SEAR-D | -1.9 | 13.4 (9.9-17.1) | 32.5 (27.5-37.3) | 35.8 (30.6-40.7) | 15.5 (11.8-19.4) |
| WPR-A | -0.22 | 0.3 (0.0-1.1) | 3.5 (0.5-6.5) | 18.0 (11.9-24.4) | 36.9 (29.2-44.8) |
| WPR-B | -I | 2.3 (0.8-3.8) | 13.6 (10.3-17.1) | 34.1 (29.6-38.9) | 34.1 (29.4-38.7) |

[^3]Table 2.2 BMI distribution of women aged I5-44 years, by subregion

| Subregion | Mean BMI (kg/m ${ }^{2}$ ) (SD) |  | Percentage of women with BMI $\leq 20 \mathrm{~kg} / \mathrm{m}^{2}$ |  |
| :---: | :---: | :---: | :---: | :---: |
|  | 15-29 years | 30-44 years | 15-29 years | 30-44 years |
| AFR-D | 20.6 (3.5) | 22.1 (3.9) | 43.3 | 29.5 |
| AFR-E | 21.2 (3.9) | 22.9 (4.8) | 37.8 | 27.4 |
| AMR-A | 23.8 (5.7) | 26.1 (7.1) | 25.1 | 19.5 |
| AMR-B | 23.5 (4.7) | 26.2 (5.1) | 22.7 | 11.1 |
| AMR-D | 23.7 (3.1) | 25.8 (4.4) | 11.7 | 9.3 |
| EMR-B | 22.5 (4.4) | 25.8 (5.1) | 28.4 | 12.7 |
| EMR-D | 21.6 (6.3) | 23.8 (8.5) | 40.1 | 32.6 |
| EUR-A | 23.3 (4.1) | 25.1 (4.7) | 20.9 | 13.8 |
| EUR-B | 22.7 (3.9) | 25.6 (5.2) | 24.5 | 14.0 |
| EUR-C | 22.7 (3.9) | 26.5 (5.0) | 24.5 | 9.7 |
| SEAR-B | 20.5 (3.5) | 22.7 (2.3) | 44.4 | 12.1 |
| SEAR-D | 19.5 (3.0) | 20.8 (4.0) | 56.7 | 42.1 |
| WPR-A | 20.8 (3.5) | 22.5 (4.1) | 40.9 | 27.1 |
| WPR-B | 21.9 (4.2) | 22.8 (4.I) | 32.6 | 24.8 |

$\mathrm{BMI}<20 \mathrm{~kg} / \mathrm{m}^{2}$ among women of reproductive age is generally below 20\% (ACC/SCN 2000b; DHS web site, http://www.measuredhs.com).

## 4. Risk factor-Disease Relationships

A synergistic relationship between malnutrition and infection has been recognized for decades (Scrimshaw et al. 1968), providing a foundation for the choice of health outcomes in this analysis, but the bi-directional nature of the relationship complicates attributing causality. The changes in body size used as a marker for undernutrition in this analysis are commonly the result, in part, of previous infections, raising the possibility that co-existing infection is responsible for some of the observed risk relationship. Age, sex, socioeconomic status, season, breastfeeding status and behavioural factors will also influence both anthropometric status and risk of infection (Pelletier 1994).

Undernutrition is most prevalent where poverty persists, and it is often difficult to isolate the effects of undernutrition from the complex web of environmental and socioeconomic factors that contribute to mortality and morbidity in such surroundings. In reviewing the various risk relationships, the likelihood of causality was considered according to Hill's standards (Hill 1965). In each case, there was judged to be sufficient biological plausibility and experimental evidence to support the relationship. We chose risk relationships that have been reported by multiple researchers in different populations under different circumstances. The associations generally, albeit not always, persisted after adjustment for prior morbidity and known confounders. Further, we selected for analysis only those studies in which anthropometric assessment clearly preceded the observed health outcome. In previous analyses (Pelletier 1994) and in our analysis, the consistency of the risk relationship between weight-for-age and mortality in different settings with quite different mortality rates and environmental conditions strongly suggests a causal rather than confounding relationship.

The relationship between undernutrition and disease may be mediated through many biological mechanisms. The lack of adequate protein and energy is associated with numerous immunological effects, including impairment of antibody and complement formation; atrophy of the thymus and other lymphoid tissues; reduction in T-lymphocytes; depressed lymphocyte activation; decreased secretory immunoglobin A (IgA) concentrations; and delayed cutaneous hypersensitivity reaction (Rivera and Martorell 1988; Scrimshaw and SanGiovanni 1997). Undernutrition is associated with impaired turnover and maturation of intestinal epithelial cells and compromised epithelial tissue integrity (Patwari 1999). Other anatomic barriers and secretory forms of protection such as lysozymes and mucus are also altered by undernutrition (Scrimshaw and SanGiovanni 1997).

In the course of the review, several health outcomes were dropped from further consideration toward the final burden of disease estimate. In many instances, there was insufficient published evidence within a developing country setting to support a causal relationship between low weight-for-age or BMI and the outcome, although the relationship might be more widely supported given a different choice of anthropometric indicator. Ultimately, risk estimates were derived for eight outcomes for childhood underweight:

- malaria mortality;
- malaria incidence;
- pneumonia (acute lower respiratory infection [ALRI]) mortality;
- pneumonia/ALRI incidence;
- diarrhoea mortality;
- diarrhoea incidence;
- measles mortality; and
- all-cause child mortality.

All of the estimates apply to children aged $0-4$ years. In addition, mortality due to perinatal causes among neonates was attributed to maternal underweight status. There was insufficient evidence to evaluate relationships among older age groups. Although some individual studies might suggest sex differences, the summary risk estimates here were assumed to be the same for males and females. Further, although risk relationships might vary by setting, the risk estimates were regarded in the analysis as constant across all subregions of the world.

Other outcomes such as cognitive impairment and additional pregnancy effects were reviewed below, but were not included toward the burden of disease estimate. The lack of current epidemiological evidence suitable to calculate a burden estimate does not necessarily imply the lack of an effect. For example, there is a large body of research into the influence of undernutrition on cognitive function. However, it was felt that there was too much heterogeneity in the types of studies and interventions, exposure variables, and outcome measurements to adequately quantify the undernutrition relationship through meta-analysis. There is also ample evidence describing the role of insufficient weight gain during pregnancy on the risk of IUGR, but prevalence data on pregnancy weight gain in developing countries are scarce.

### 4.1 Risk of mortality due to childhood infections

It has been demonstrated that a child's risk of dying due to undernutrition is not limited to only those children with the most severe undernutrition (Pelletier et al. 1994). Rather, there exists a spectrum of risk
associated with all degrees of undernutrition-mild, moderate and severe. Although the risk of dying is highest among those most undernourished, when one considers the small but significantly elevated risk of mortality associated with mild and moderate undernutrition and the high prevalence of mild undernutrition worldwide, much of the burden of childhood deaths due to undernutrition may be attributable to mild and moderate, rather than to severe undernutrition.

Pelletier et al. (1994) estimated the relations between weight-for-age and risk of mortality from all causes, and used this information to calculate the burden of child deaths attributable to undernutrition. These relations however are likely to vary depending on specific causes of death. For example, the relative risk of dying from malaria associated with varying degrees of undernutrition may not be similar to the risk for diarrhoea. Despite problems in defining cause of death, particularly in developing countries, in which one may often need to rely on verbal autopsy methods, it is worthwhile to consolidate the evidence to-date and consider the implications of analyses relating variation in anthropometric status and risk of dying across the principal causes of death for children in developing countries: diarrhoea, pneumonia, measles and malaria.

## METHODS FOR ESTIMATING RISK OF MORTALITY

Initially, we surveyed the published literature to gather data to estimate the relationship between anthropometric status and cause-specific mortality (Rice et al. 2000). Because insufficient data were available, we contacted investigators with relevant data (published or unpublished) and asked them to contribute specific study results for this analysis. We sought community-based prospective cohort studies in which anthropometric status was assessed to determine weight-for-age, vital status monitored and cause of death determined by documented methods. Data sets ultimately contributing to the mortality analysis are listed in Table 2.3.

Each investigator was asked to provide the following information: (i) a description of their study; (ii) the number of children in the study or child-years of study with WAZ $<-3,-2$ to $-3,-1$ to $-2,0$ to -1 and $>0$; and (iii) deaths in each category attributable to diarrhoea, pneumonia, measles, malaria, or other cause and total (all-cause). During analysis, we collapsed the last two categories of anthropometric status, so that the reference group would be composed of those children with WAZ >-1.

## Analytic methods used to combine mortality data

We followed the procedures of Pelletier et al. $(1993,1994,1995)$ and used the SAS Proc Mixed program for all procedures. Because this has been described elsewhere, we describe the steps only briefly here. First, we calculated the mortality rate (per 1000) by anthropometric status and each cause of death for each study. These rates are listed in Table 2.4
Table 2.3 Data sets with information contributing to the childhood mortality analysis

| Country | Study | Age range (months) | Follow-up (months) | Number of children |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | $<-3$ SDs | $>-3$ to $<-2$ SDs | $>-2$ to $<-1$ SD | >-I SD |
| Sudan | Fawzi et al. (1997) | 6-72 | 6 | 365 | 985 | 1123 | 592 |
| Senegal | Garenne et al. (1987) | 0-59 | 6 | 1663 | 2186 | 2035 | 1249 |
| Guinea-Bissau | Andersen (1997) | 0-59 | 6-12 | 1155 | 4366 | 7341 | 2731 |
| Ghana | WHO/CHD (1998) | 0-12 | 12 | 46 | 183 | 506 | 1410 |
| Nepal | West et al. (1991, 1997) | 0-72 | 12 | 1030 | 2082 | 2002 | 892 |
| Bangladesh | Arifeen et al. (2001) | $0-11$ | 3 | 79 | 250 | 475 | 478 |
| Pakistan | Khan et al. (1993); <br> Jalil et al. (1993) | 0-24 | 6 | 309 | 778 | 1182 | 1165 |
| India | WHO/CHD (1998) | 0-12 | 12 | 172 | 639 | 1259 | 1295 |
| Indonesia | Sutrisna et al. (1993) | 0-60 | 18 | 207 | 240 | 797 | 4856 |
| Philippines | Ricci and Becker (1996) | 0-59 | 3 | 800 | 3144 | 4834 | 5115 |



Figure 2.1 Underweight and all-cause mortality: (a) deaths per I000; (b) logarithm of deaths per 1000

and presented graphically in Figures 2.1(a)-2.5(a). We then calculated the logarithm of the mortality rate by anthropometric status and each cause of death for each study, and re-graphed the data (Figures $2.1[\mathrm{~b}]-2.5[\mathrm{~b}])$. Second, we regressed the logarithm of mortality on WAZ and compared these results with regressions of the simple mortality rates by WAZ. Third, we compared the goodness-of-fit characteristics for the two models ( $\mathrm{R}^{2}$ and MSE). The models with $\log$ mortality rate as the outcome had better goodness-of-fit. For these analyses, we utilized a weighted regression with the weighting scheme of Pelletier and col-

Figure 2.2 Underweight and diarrhoea mortality: (a) deaths per 1000; (b) logarithm of deaths per 1000

leagues: (1/deaths) + (1/children). Fourth, we utilized weighted random effects models (Proc Mixed in SAS—see section 2.2 under subregional estimates), to provide combined estimates of the relation between weight-for-age and risk of mortality. The midpoint of the anthropometric category was used for estimation, and -0.5 SD was considered the value of the reference category. The results of the regression analyses are provided in Table 2.5. From these coefficients, estimated relative risk and $95 \%$ CIs were calculated. These steps were performed for all-cause mortality as well as by cause of death.

Figure 2.3 Underweight and pneumonia mortality: (a) deaths per I000; (b) logarithm of deaths per 1000

(b)


## Results

In all, we identified 12 potential studies, and have utilized data from 10 studies in the current analysis. We excluded two studies from the analysis: data from Peru were excluded because the study was too small and provided insufficient deaths for the analysis; data from Brazil were excluded because they resulted from a case-control study and were therefore not appropriate for our analysis.

The 10 studies were conducted in sub-Saharan Africa and Asia (Table 2.3). Each study contributed data on deaths due to diarrhoea and pneu-

Figure 2.4 Underweight and measles mortality: (a) deaths per 1000; (b) logarithm of deaths per 1000

monia (see Figures 2.2 and 2.3); however, exposure to infectious diseases such as malaria and measles depends on the ecology of the study setting and health care utilization (i.e. measles vaccination rates), and thus, not all studies contributed data on these causes of death. In all, six studies from Ghana, Guinea-Bissau, Indonesia, Nepal, the Philippines and Senegal contributed data on deaths due to measles (Figure 2.4). Three studies, from Ghana, Guinea-Bissau and Senegal contributed data on malaria-related deaths (Figure 2.5).

From the regression analyses we calculated the relative risk (95\% CI) of death by cause of death for each category of weight-for-age as

Figure 2.5 Underweight and malaria mortality: (a) deaths per I000; (b) logarithm of deaths per 1000


Table 2.5 Regression coefficients for models relating low weight-forage with mortality, by cause

| Outcome | Intercept (SE) | Weight-for-age (SE) |
| :--- | :--- | :---: |
| All-cause mortality | $+2.26^{\mathrm{a}}(0.30)$ | $-0.722^{\mathrm{a}}(0.077)$ |
| Diarrhoea mortality | $+0.490(0.338)$ | $-0.842^{\mathrm{a}}(0.094)$ |
| Pneumonia mortality | $+0.459(0.36 \mathrm{I})$ | $-0.697^{\mathrm{a}}(0.105)$ |
| Malaria mortality | $+0.058(0.454)$ | $-0.750^{\mathrm{a}}(0.182)$ |
| Measles mortality | $-0.263(0.370)$ | $-0.55 \mathrm{I}^{\mathrm{a}}(0.140)$ |

a $\quad P<0.05$.

Table 2.6 RR of mortality associated with low weight-for-age estimated from regression analysis, by cause of death ${ }^{\text {a }}$

| Cause of death | $<-3 S D s$ <br> $(95 \% \mathrm{Cl})$ | -2 to -3 SDs <br> $(95 \% \mathrm{Cl})$ | -1 to -2 SDs <br> $(95 \% \mathrm{Cl})$ | $>-$ I SD |
| :--- | :---: | :---: | :---: | :---: |
| Diarrhoea | $12.50(7.19-21.73)$ | $5.39(3.73-7.79)$ | $2.32(1.93-2.79)$ | 1.0 |
| Pneumonia | $8.09(4.36-15.0 \mathrm{I})$ | $4.03(2.67-6.08)$ | $2.01(1.63-2.47)$ | 1.0 |
| Malaria | $9.49(3.25-27.66)$ | $4.48(2.20-9.15)$ | $2.12(1.48-3.02)$ | 1.0 |
| Measles | $5.22(2.29-11.88)$ | $3.01(1.74-5.21)$ | $1.73(1.32-2.28)$ | 1.0 |
| All-cause | $8.72(5.55-13.72)$ | $4.24(3.13-5.73)$ | $2.06(1.77-2.39)$ | 1.0 |

Calculated at $-3.5,-2.5,-1.5$ vs 0.5 SD weight-for-age.
compared to the reference category of weight-for-age >-1 SD (Table 2.6). As shown, there are significant risks of death associated with low weight-for-age for overall mortality as well as for each cause of death examined.

Several results on the relationships between anthropometric status and specific causes of mortality should be interpreted with caution. First, studies on the influence of low weight-for-age on measles mortality have yielded equivocal results. For example, a community-based longitudinal study by Chen et al. (1980) in Bangladesh found that underweight children were at over twice the risk for measles mortality (risk ratio 2.37, $95 \%$ CI 0.96-5.86), while Aaby et al. (1988) observed little effect attributable to pre-existing nutritional status in Guinea-Bissau (estimated risk ratio $0.83,95 \%$ CI $0.25-2.78$ ). Other factors, such as intra-household crowding and the different levels of exposure experienced by primary and secondary cases, appeared to have greater influence on clinical outcome. Second, in the study by Andersen (1997) in Guinea-Bissau, deaths due to malaria were reported as due to fever, and thus deaths due to febrile illnesses other than malaria were necessarily included. However, as shown in Figure 2.5 (b), the results are quite consistent with those from Senegal. The results from Ghana suggest a different pattern of relationship with mortality risk, but it should be noted that the number of deaths due to malaria is quite small $(n=8)$. Further research is needed to refine these estimates. Finally, as described earlier, poor nutritional status is associated with poverty and poor environment. Some of the effects of such factors on child health are mediated through modifying anthropometry and some through other mechanisms. Therefore, it is likely that the hazard estimates presented here are affected by confounding due to a number of such factors.

### 4.2 RISK OF MORBIDITY FROM CHILDHOOD INFECTIONS

The incidence of infection and the risk of other morbidities associated with underweight status were estimated through a literature review and meta-analysis of published data.

## Search Strategy for identifying studies

Articles relating to undernutrition were identified for review based upon database searches of literature published between 1966 and 2001 in English, or having an English abstract. Searches of Medline, Popline and PsychInfo databases were conducted through PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) and Internet Grateful Med (http://igm.nlm.nih.gov, now discontinued) using numerous keywords, author names and "related articles" links. After reviewing the abstracts, English-language articles that examined the relationship between anthropometric status and health outcomes among human populations were selected.

Once copies of articles were obtained, additional publications were identified from the reference lists of those articles. Several nutrition researchers were consulted as to their awareness of any ongoing studies or additional published or unpublished data. However, beyond this and the search strategy described above, there was no systematic attempt to identify unpublished studies or dissertations.

## INCLUSION/EXCLUSION CRITERIA

The studies considered for review were restricted to: (i) original research reports of community-based or facility-based controlled trials, cohort studies, case-control studies, cross-sectional studies and retrospective analyses of records; (ii) critical review articles that contained original research results not available elsewhere; and (iii) meta-analyses of original research results. Studies that were excluded from the review and analysis were: (i) animal studies; (ii) case reports; (iii) ecological analyses; (iv) studies of special populations (e.g. dialysis patients, acquired immunodeficiency syndrome [AIDS] patients); (v) studies without a control or comparison group; (vi) studies where nutritional status was defined only by birth weight; (vii) studies where the term "malnutrition" was used without further descriptive information or reference to a source providing such information; and (viii) studies where clear risk estimates were not presented or could not be calculated from the reported data.

The analysis was limited to studies conducted among populations in developing countries, among whom poor anthropometric status was considered more likely to reflect undernutrition. Studies of protein/energy supplementation or fortification were excluded from the analysis because the exposure variable (i.e. supplemented vs not supplemented) was not compatible with the anthropometric definition of undernutrition used here. The analysis of studies was limited to those reporting weight-for-age according to either $z$-score or percentage of median relative to the Harvard (Jelliffe 1966) or NCHS/WHO (Hamill 1977; WHO 1978) references.

## ANALYTIC METHODS USED TO COMBINE INDIVIDUAL STUDY RISK ESTIMATES

Meta-analyses were performed using unadjusted data from studies that satisfied all of the inclusion/exclusion criteria and provided sufficient statistical information (e.g. sample sizes per group, means and SDs or significance test values). Original source data were not obtained for these meta-analyses. Each meta-analysis was based on data exactly as reported in the publication, whenever possible. In some cases, the data had to be adapted from reported rates and statistical information in order to convert them into comparable units of measurement or to conform to comparable nutritional categories. For example, a multiplier would be used to convert diarrhoea episodes per month into diarrhoea episodes per year, or categorical data corresponding to two or more degrees of malnutrition would be collapsed into a single category. Estimates were made of data plotted in figure form if the figure provided sufficient clarity and precision.

All modelling and testing was conducted using Intercooled STATA 6.0. Studies reviewing malaria risk associated with underweight status were combined using Mantel-Haenszel fixed-effect models and tests of heterogeneity. Relative risk of diarrhoea associated with low weight-for-age was calculated from summary data and estimates of lower and upper confidence limits were calculated using the incidence-density binomial approximation to the normal distribution. Individual study estimates were then combined using Mantel-Haenszel combination techniques and tests of heterogeneity. Random effects models were used to combine individual study estimates for risk of pneumonia associated with low weight-for-age. For each outcome, sensitivity analyses were conducted to determine if any studies were significantly driving the estimate.

Although $z$-score is now preferable, most studies reported anthropometric status in terms of percentage of median, and cut-off points to classify malnutrition varied among different studies. As a result, the definition of underweight was not absolutely consistent among the studies combined in each summary estimate. The differences in definition in such cases were distinguished in this review through the use of multiple analysis "levels", wherein studies were separated by their nutritional categorization scheme during the analysis process in order to compare and contrast the results. Analysis levels are described further in the sections on pneumonia and diarrhoea results.

Weight-for-age was analysed according to multiple categories of severity, when possible, in order to describe any potential dose-response effect and better characterize possible health risks among the large proportion of mild-to-moderate malnutrition cases. The ability to examine multiple weight-for-age levels, however, was highly limited by the available published data; the majority of studies analysed anthropometric status as a dichotomous variable only. Therefore, for all health outcomes, underweight was analysed as a dichotomous variable, comparing individuals
falling below the designated weight-for-age cut-off point to those above the cut-off point (the reference group). A drawback to the dichotomous approach is that if a dose-response trend exists, combining intermedi-ate-risk and low-risk groups into a single category for a dichotomous analysis could have the effect of underestimating the true risk. An additional analysis was conducted when data on multiple weight-for-age categories were available, in which case data were conflated into three anthropometric categories: moderately to severely underweight (e.g. weight-for-age below -2 SDs), mildly underweight (e.g. weight-for-age between -1 and -2 SDs) and normal (e.g. weight-for-age greater than $-1 \mathrm{SD})$. Risk ratios were then estimated for both categories in relation to the reference category of weight-for-age >-1 SD. When weight-for-age was defined by percentage of median, the overall analysis was subdivided into separate analysis levels according to the cut-off points used.

## MEASLES RESULTS

There were an inadequate number of eligible studies for a summary risk analysis of weight-for-age and measles incidence. Evidence from an observational study in the Gambia (Lamb 1988) and a controlled supplementation trial in India (Gopalan et al. 1973) suggest, however, that pre-morbid underweight status has no significant effect on the incidence of measles infection. In India, Sinha et al. (1977) observed a reduced incidence of measles rash among children in the lowest weight percentiles, but this may represent a lowered immune response rather than a reduced incidence of infection. Beau et al. (1987) in Senegal noted a higher prevalence of measles among hospitalized children who were wasted at admission, but reverse causality would explain this observation.

## Malaria results

Research and clinical observation over the past several decades have generated debate as to how undernutrition may influence susceptibility to malaria. Conflicting results occur, in part, because susceptibility to malaria attack and/or infection is highly dependent on the individual's immune status, which, in turn, is influenced by factors such as age, pregnancy status and the level of endemicity within a population (Shankar 2000). The relationship between malaria and undernutrition has been examined according to several outcome variables, including presence or density of malaria parasites in the blood, episodes of malarial illness, incidence of complicated malaria (cerebral malaria or severe malarial anaemia) among malaria cases and case-fatality. This comparative risk analysis focused only on the incidence of malaria attacks, defined clinically as episodes of fever with slide-confirmation of parasites in the blood, or, in some studies, less specifically as fever with chills. Most of the data related to infection by Plasmodium falciparum, although some studies included data on Plasmodium vivax, as well.

Table 2.7 Risk of malaria attack associated with low weight-for-age from five studies

| Location | Study | Study type | Cut-off point | Sample <br> size | Risk estimate (95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Congo | Tonglet et al. (1999) | Community-based cohort | WA <25th percentile | 842 | $R \mathrm{R}=0.92$ (0.72-1.17) |
| Gambia | Snow et al. (1991) | Community-based cohort | WA < $70 \%$ | 138 | $R \mathrm{R}=1.52$ (0.59-3.93) |
| Nigeria | Adelekan et al. (1997) | Facility-based case-control | WAZ <-2 | 65 | $\mathrm{OR}=1.50$ (0.5I-4.4I) |
| Sudan | el Samani et al. (1987) | Community-based cross-sectional | WA < $75 \%$ | 445 | $R \mathrm{R}=1.63$ (1.18-2.27) |
| Vanuatu | Williams et al. (1997) | Community-based cohort | WAZ <-2 | 911 | $R \mathrm{R}=\mathrm{I} .28$ (0.87-I.89) |

WA Weight-for-age.

It is difficult to adequately estimate any effect of low weight-for-age on risk of malaria morbidity given the scarcity of studies reporting both weight-for-age and malaria data. Five studies were identified that fulfilled that initial requirement (Table 2.7). These studies examined nutritional status as a dichotomous variable, comparing underweight children (defined as WAZ $<-2$, weight-for-age $<70 \%$, or weight-for-age $<75 \%$ of the NCHS reference median) to adequate-weight children; there were insufficient data from eligible prospective studies to examine any potential risk associated with milder degrees of malnutrition (i.e. WAZ between -1 and -2 ). Of the five studies, cohort studies from the Gambia (Snow et al. 1991), the Congo (Tonglet et al. 1999) and Vanuatu (Williams et al. 1997) among children aged $1-4$ years, $<2$ years and $<10$ years, respectively, reported statistically non-significant differences in malaria morbidity between underweight and normal children. In a facility-based case-control study in Nigeria (Adelekan et al. 1997) malaria cases in children aged $<8$ years had slightly, albeit not significantly, lower weight-for-age than non-infected controls, but this study was dropped from the final analysis because nutritional status was assessed after the onset of illness, leaving causality ambiguous. A cross-sectional study in the Sudan (el Samani et al. 1987) observed significantly greater history of malaria attacks (defined as fever with rigors and joint pain) among underweight children aged $<5$ years, but the causality was unclear in this case, as well.

Of the remaining three prospective studies, the Congo study (Tonglet et al. 1999), which reported a relative risk of 0.92 ( $95 \%$ CI $0.72-1.17$ ) for a malaria episode, was dropped from the summary analysis because nutritional categories were defined according to local reference data. Following meta-analysis of the two remaining cohort studies, the most

Table 2.8 Combined estimate of risk for malaria attack associated with low weight-for-age from two cohort studies

| Location | Study | Risk estimate (95\% Cl) |
| :--- | :--- | :---: |
| Gambia | Snow et al. (1991) | $1.52(0.59-3.93)$ |
| Vanuatu | Williams et al. (1997) | $1.28(0.87-\mathrm{I} .89)$ |
| Combined estimate |  | $1.31(0.92-1.88)$ |

appropriate for this assessment, the combined relative risk for malaria among children having weight-for-age $<-2$ SDs was statistically nonsignificant (RR1.31, 95\% CI 0.92-1.88) (Table 2.8). This overall estimate was more heavily influenced by the Vanuatu cohort study, which reported a slightly but not significantly higher incidence of $P$. falciparum malaria (defined as fever with parasitaemia $>1000 / \mu \mathrm{l}$ ) among children who were classified as underweight prior to the period of morbidity follow-up.

## PNEUMONIA/ALRI RESULTS

Undernutrition may increase the risk of pneumonia through a number of potential mechanisms. In general, cell-mediated immunity is depressed in undernourished children (Rivera and Martorell 1988; Scrimshaw and SanGiovanni 1997). Respiratory muscles weaken, impairing the cough reflex and reducing a child's ability to adequately clear secretions from the respiratory tract or respond to hypoxia through tachypnea (Wilson 1985). Salivary IgA is decreased, compromising the integrity of the respiratory tract mucosa and leaving the mucosa susceptible to invasion by microorganisms (Lehmann et al. 1988).

Over 60 studies evaluating the relationship between acute respiratory infections and malnutrition were initially identified. These included a collaborative data group from the United States National Academy of Sciences Research Program of the Board on Science and Technology for International Development (BOSTID). We immediately excluded studies that examined upper respiratory tract infections only; did not distinguish between upper and lower respiratory tract infections in their data collection and/or analysis; limited the report to single etiological agents (e.g. only pneumococcus); or did not report anthropometric status in terms of weight-for-age. The remaining studies are briefly described below, although not all fulfilled the criteria for inclusion in a summary analysis, usually for reasons of insufficient data or statistical information. Studies that provided estimates of risk are listed in Tables 2.9 and 2.10.

Studies varied in their case definitions of ALRI and pneumonia, classifying them on the basis of mothers' recall of the physician's diagnosis (Victora et al. 1990); radiologically confirmed pneumonia (Alam et al. 1984; Fonseca et al. 1996; Victora et al. 1994); or cough and fever with

Table 2.9 Risk of pneumonia/ALRI incidence associated with low weight-for-age from studies included in summary estimate

| Analysis level | Study | Location | Cut-off point | Sample <br> size | Crude RR (95\% CI) | Adjusted RR ${ }^{\text {a }}$ (95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| I | Victora et al. (1990) | Brazil | WAZ <-2 | 4486 | 1.32 (0.70-2.47) | $1.75{ }^{\text {b }}$ |
| I | Zaman et al. (1996) | Bangladesh | WAZ <-2 | 696 | 1.25 (0.76-2.05) | - |
| 3 | James (1972) | Costa Rica | WA $<75 \%$ | 137 | 12.37 (3.11-49.25) | - |
| 4 | Ballard and Neumann (1995) | Kenya | WA $<80 \%$ | 109 | 0.90 (0.37-2.17) | $\begin{gathered} 1.8^{c} \\ (0.52-6.4) \end{gathered}$ |
| 4 | Deb (1998) | India | Unclear | 800 | 2.53 (1.72-3.73) | - |

- No data.
a All crude relative risks reported in Table 2.9 represent risk associated with weight-for-age below the cut-off point compared to risk associated with weight-for-age above the cut-off point. Adjusted relative risks, however, varied in their referent group and in the variables adjusted as indicated in the table footnotes.
b Adjusted for income. Adjusted risk ratio was based on WAZ $<-2$ relative to $W A Z \geq 0$. Crude risk estimate of 1.32 was based on WAZ $<-2$ relative to WAZ $>-2$. (Crude risk estimate based on WAZ $<-2$ relative to WAZ $\geq 0$ is 2.38 [ $95 \% \mathrm{Cl}$ I.18-4.78].)
c Adjusted for season, socioeconomic status and geographic region. Adjusted RR was based on weight-for-age $<80 \%$ relative to weight-for-age $>90 \%$; crude RR was based on weight-for-age $<80 \%$ relative to weight-for-age $>80 \%$. (Crude RR based on weight-for-age $<80 \%$ relative to weight-for-age $>90 \%$ is 2.03 [ $95 \% \mathrm{Cl} 0.44-9.25$ ].)
rapid respiratory rate (Cunha 2000; Deb 1998; Smith et al. 1991) and/or chest indrawing (Ballard and Neumann 1995; Spooner et al. 1989; Zaman et al. 1996). The case definition for pneumonia recommended by WHO is cough or difficult breathing with rapid respiration $(>50$ breaths/min for infants aged 2 months to under 1 year; $>40$ breaths $/ \mathrm{min}$ for children aged 1 to 5 years) or chest indrawing, stridor or general danger signs such as vomiting, convulsions, lethargy, unconsciousness or inability to drink/breastfeed (WHO 1999b). The coordinated BOSTID data group defined ALRI as the presence of at least one of the following: rales or crepitations, wheezing, stridor, cyanosis, rapid respiratory rate ( $>50$ breaths $/ \mathrm{min}$ ), or chest indrawing (Selwyn 1990). Studies were considered for our summary analysis if the case definition for ALRI was consistent with the recommended WHO definition. With the exception of a Bangladesh study among diarrhoea patients aged $<12$ years (Alam et al. 1984), all of the studies examined children aged $\leq 5$ years. Incidence was reported predominantly as the number of ALRI episodes over a given follow-up period (i.e. multiple episodes per individual child counted as separate events), although a few studies reported the proportion of children with one or more ALRI episodes (i.e. multiple episodes per individual child counted as a single event).
Table 2.10 Risk of pneumonia/ALRI incidence associated with low weight-for-age from excluded studies

| Location | Study | Study type | Cut-off point | Crude risk estimate (95\% Cl) | Adjusted risk estimate ${ }^{\text {a }}$ (95\% CI) | Reason for exclusion |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Brazil | Cunha et al. (2000) | Community-based cohort | WAZ -3 to -2 | - | $\begin{gathered} \mathrm{OR}=1.59^{\mathrm{b}} \\ (1.09-2.33) \end{gathered}$ | Insufficient incidence data |
| Brazil | Fonseca et al. (1996) | Community-based cohort | WAZ <-2 | - | $\begin{gathered} \mathrm{OR}=4.57^{\mathrm{c}} \\ (2.93-7.13) \end{gathered}$ | Reverse causality |
| Brazil | Victora et al. (1994) | Facility-based case-control | WAZ <-2 | $\begin{gathered} \mathrm{OR}=5.87 \\ (3.30-10.44) \end{gathered}$ | $\begin{gathered} \mathrm{OR}=4.77^{\mathrm{d}} \\ (2.46-9.06) \end{gathered}$ | Reverse causality |
| China | Liu et al. (1991) | Community-based case-control | Not reported | $\begin{gathered} \mathrm{OR}=5.79 \\ (3.70-9.14) \end{gathered}$ | - | Reverse causality; no PEM definition |
| India | Shah et al. (1994) | Facility-based case-control | WA < $75 \%$ | $\begin{gathered} \mathrm{OR}=2.36 \\ (1.53-3.65) \end{gathered}$ | - | Reverse causality |
| Papua New Guinea | Binns (1976) | Community-based cohort | WA <80\% | $R \mathrm{R}=1.59$ | - | No categorical sample size data |
| Papua New Guinea | Smith et al. (1991) | Community-based cohort | WAZ <-2 | - | $\begin{aligned} & \mathrm{RR}=2.1^{\mathrm{e}} \\ & (1.3-3.4) \end{aligned}$ | Data reported in figure form |
| South Africa | Wesley and Loening (1996) | Community-based case-control | WA <10th percentile | $\begin{gathered} \mathrm{OR}=1.47 \\ (0.38-5.88) \end{gathered}$ | - | Reverse causality; cut-off point |

[^4]
## Excluded studies

The collaborative BOSTID study reported community-based longitudinal data from Guatemala, Papua New Guinea, the Philippines and Uruguay (Selwyn 1990). In each population, researchers observed higher incidence of ALRI among underweight children aged 18-59 months, with relative risks ranging between 1.2 (Guatemala) and 2.7 (Uruguay). An increased risk was observed among the younger group of 0-17-month-old children only in the Philippines ( $\mathrm{RR}=1.3$ ). While the methodology used in these four countries was standardized for better internal comparability among study sites, the results were not included in our summary analysis because nutritional status was evaluated using an anthropometric cut-off point (10th percentile) higher than WAZ of -2 (equivalent to approximately the third percentile-Pelletier 1994) and the case definitions used in Guatemala and Uruguay could have classified asthma as ALRI.

Ten ALRI investigations were case-control studies or record reviews and were excluded because of the possibility of reverse-causality. In Colombia (Berman et al. 1983) and India (Shah et al. 1994), children with pneumonia were significantly more likely to be moderately or severely malnourished than were non-ALRI paediatric patients or outpatients. In South Africa, in contrast, odds of being malnourished were not significantly different between pneumonia patients and a control group of upper respiratory infection patients (Wesley and Loening 1996). In a case history review in Bangladesh, malnourished diarrhoea patients aged 1-5 years had a higher incidence of pneumonia during hospitalization than diarrhoea patients considered well-nourished, although the relationship was not observed in infants (Alam et al. 1984).

Case-control studies in Brazil, China and Papua New Guinea (Fonseca et al. 1996; Victora et al. 1994) compared cases to healthy controls from the surrounding community. Children having lower respiratory infections in China (Liu et al. 1991) and Papua New Guinea (Spooner et al. 1989) were at significantly greater odds of being underweight compared to healthy controls, and underweight children in Papua New Guinea were four times as likely to be admitted to the hospital with pneumonia as non-malnourished children (Barker et al. as cited in Lehmann et al. 1988). Similarly, in an urban area of southern Brazil, children aged $<2$ years hospitalized with pneumonia had a reported OR of 5.87 ( $95 \%$ CI 3.30-10.44) for weight-for-age $<-2$ SDs compared to healthy, age-matched neighbourhood controls; adjusted for age, sex, crowding and socioeconomic factors, the odds ratio decreased slightly to 4.77 (95\% CI 2.46-9.06) (Victora et al. 1994). Fonseca et al. (1996) conducted a similarly designed study in northern Brazil and found undernutrition to be the most important risk factor for pneumonia among children aged $<2$ years attending an outpatient clinic; the odds ratio for weight-for-age $<-2$ SDs among cases was 4.57 ( $95 \%$ CI 2.93-7.13) adjusted for income, parents' education and previous pneumonia.

In a community-based survey in Brazil, Cunha et al. (2000) compared anthropometric status of children aged $<5$ years according to mothers' report of respiratory illness in the previous week (Cunha et al. 2000). Adjusted for age, sex and household crowding, children with weight-forage below -2 SDs had significantly greater odds for an ALRI episode compared to children with weight-for-age >-1 SD.

Prospective cohort studies that assessed anthropometric status prior to illness were more appropriate for evaluating malnutrition as a causal factor in ALRI incidence. In rural Papua New Guinea, Smith et al. (1991) assessed nutritional status and subsequent incidence of ALRI through twice-weekly home visits, reporting an age-adjusted incidence rate ratio of approximately 2.1 ( $95 \%$ CI $1.25-3.4$ ) associated with WAZ $<-2$ relative to weight-for-age $>0 \mathrm{SD}$. There was no evidence of confounding by socioeconomic status (defined by education, income and household crowding) or previous episodes of illness. The data, however, were reported only in figure form and lacked sufficient statistical information for inclusion in the meta-analysis. An earlier cohort study in Papua New Guinea by Binns (1976) observed a relative risk of 1.59 for pneumonia among children having weight-for-age $<80 \%$, but, likewise, did not report sufficient statistical data for inclusion (such as sample sizes per weight-for-age category).

## Studies included in the risk estimate

James (1972) in urban Costa Rica reported the highest risk estimate among cohort studies. Based on weekly physician visits and mothers' recall, underweight children aged $<5$ years had a similar number of overall respiratory tract infections as normal-weight children, but experienced a relative risk of 12.4 ( $95 \%$ CI $3.11-49.25$ ) for bronchopneumonia, with no apparent differences between groups in breastfeeding duration, age distribution, levels of household crowding, incomes or sanitation. In urban Brazil, Victora et al. (1990) followed a cohort of infants for subsequent hospital admissions, reporting a significant relative risk (adjusted for family income) of 1.75 for pneumonia among children with pre-morbid WAZ $<-2$ compared to $\geq 0$. The risk estimate remained statistically significant after adjusting for previous pneumonia hospitalizations (data not reported). Incidence referred to hospital admissions, which represented only the more severe cases of ALRI, and there is potential bias if underweight ALRI cases are more likely to be admitted than the normal weight cases. In a community-based cohort study in Bangladesh, Zaman et al. (1996) followed up morbidity every four days and observed an increased, but not statistically significant, risk of 1.25 between low weight-for-age and subsequent incidence of ALRI among children aged $\leq 5$ years; adjusted for age, sex and socio-demographic variables, the relative risk became statistically significant, with a one-unit decrease in WAZ being associated with a $55 \%$ increase in the incidence of ALRI. A smaller cohort study in Kenya reported a slightly higher,
but statistically non-significant, relative risk for ALRI of 1.8 (95\% CI $0.52-6.4$ ) among children aged 18-25 months with weight-for-age $<80 \%$ relative to weight-for-age $>90 \%$, adjusted for season, socioeconomic status and geographic region (Ballard and Neumann 1995). In an urban and rural community-based cohort study in India (Deb 1998), underweight children aged $<5$ years experienced a significantly higher rate of pneumonia over the course of 18 months relative to normal-weight children $(\mathrm{RR}=2.53)$, although it was not reported how comparable malnourished and normal groups were in relation to potentially confounding variables.

## Relative risk estimate

There were an insufficient number of studies reporting multiple categories of weight to compare the risk of ALRI associated with mild-to-moderate degrees of underweight, and, therefore, only a dichotomous analysis was performed. Two cohort studies reported anthropometric status in terms of $z$-score and were treated as "analysis level one" (Table 2.11). Analysis levels three and four used cut-off points in terms of percentage of median. (Analysis level two was reserved for studies using a cut-off point of $70 \%$, but no eligible studies using that cut-off point were included here.)

Relative risks according to cut-off point are represented graphically in Figure 2.6 and summary risk estimates are listed in Table 2.12 according to analysis level. Overall, community-based cohort studies that

Table 2.1I Analysis levels according to anthropometric classification used

| Analysis level | Comparison |
| :--- | :---: |
| 1 | WAZ $<-2$ vs WAZ $>-2$ |
| 2 | WA $<70 \%$ vs WA $>70 \%$ |
| 3 | WA $<75 \%$ vs WA $>75 \%$ |
| 4 | WA $<80 \%$ vs WA $>80 \%$ |

Table 2.12 Combined estimate of risk for pneumonia/ALRI incidence associated with low weight-for-age

| Study type | Analysis level | Number of studies | Combined risk <br> estimate (95\% Cl) |
| :--- | :---: | :---: | :---: |
| Cohort | I | 2 | $1.28(0.86-\mathrm{I} .89)$ |
| Cohort | 3 | I | $12.37(2.98-5 \mathrm{I} .3 \mathrm{I})$ |
| Cohort | 4 | 2 | $2.22(1.55-3.18)$ |
| Cohort | $\mathrm{I}+4$ | 4 | $1.72(1.32-2.25)$ |
| Combined estimate | $\mathrm{I}+3+4$ | 5 | $1.86(1.06-3.28)$ |

Figure 2.6 Relative risk of pneumonia/ALRI according to weight-for-age cut-off point used in study

provided sufficient results and statistical data yielded a combined risk estimate for ALRI of 1.86 associated with low weight-for-age (Table 2.12). Even excluding the high estimate observed in the study by James (1972), the risk ( $R R=1.72$ ) remained statistically significant.

## DIARRHOEA RESULTS

There is a large body of literature examining the relationship between malnutrition and diarrhoea, and cross-sectional studies among children have consistently shown a trend of increasing diarrhoea prevalence associated with decreases in anthropometric status. Malnutrition and concomitant growth faltering and weight decline are not uncommon following severe bouts of diarrhoea among children in developing coun-
tries, but the reverse relationship-that pre-morbid undernutrition increases susceptibility to subsequent diarrhoea attack-is more uncertain. We selected prospective studies that specifically examined premorbid anthropometric status as a risk factor for subsequent diarrhoea, rather than the reverse relationship. Studies that simultaneously assessed weight-for-age and diarrhoea were not considered because the direction of the relationship could not be determined-i.e. there was potential reverse causality with regard to the risk relationship. All of the studies related to children aged $\leq 5$ years; little if any data were available for older age groups. In some studies, morbidity was compared to the initial assessment of nutritional status only, whereas other studies reclassified children, as necessary, according to periodic follow-up measurements. An episode of diarrhoea was typically defined as three or more loose, liquid or watery stools, or at least one bloody stool in a 24 -hour period. An episode was generally considered terminated when normal stool patterns returned for at least three days. Some of the outcome variables relating to diarrhoea included incidence of general diarrhoea, incidence of dysentery and incidence of persistent diarrhoea, defined as diarrhoea lasting 14 days or longer. We restricted our analysis to the incidence of general diarrhoea, and excluded studies that focused solely on a single pathogenic agent (e.g. Cryptosporidium, Vibrio cholerae), watery diarrhoea alone or bloody diarrhoea alone.

Literature supports an increased risk of diarrhoea mortality associated with low weight-for-age (Bhan et al. 1986; Chen et al. 1980; Fawzi et al. 1997; Rice et al. 2000; Yoon et al. 1997), but the effect on diarrhoea incidence has been less clear. Cohort studies in Bangladesh (Baqui et al. 1993; Chowdhury et al. 1990), Brazil (Schorling et al. 1990), Costa Rica (James 1972), the Congo (Tonglet et al. 1999), Ethiopia (Lindtjorn et al. 1993), the Gambia (Tomkins et al. 1989), Ghana (Biritwum et al. 1986), Guatemala (Gordon et al. 1964), India (Ghai and Jaiswal 1970; Walia et al. 1989), Mexico (Sepulveda et al. 1988), Papua New Guinea (Binns 1976) and the Sudan (el Samani et al. 1988; Kossman et al. 2000), reported higher diarrhoea incidence rates among underweight children (rate ratios of 1.1 to 2.4 ), while other studies observed no statistically significant effect (Anand et al. 1994; Bairagi et al. 1987; Bhan et al. 1986; Black et al. 1984; Chen et al. 1981; Henry et al. 1987; Khan and Yunus 1990; Mathur et al. 1985; Molbak et al. 1997b; Thongkrajai et al. 1990; Tomkins 1981; Victora et al. 1990). Factors such as age, socioeconomic status, season, breastfeeding and recent history of diarrhoea, among others, can be important confounders in the relationship between anthropometric status and diarrhoeal illness. When statistical adjustment had been made for such variables, several studies (e.g. el Samani et al. 1988; Kossman et al. 2000; Schorling et al. 1990; Sepulveda et al. 1988; Tonglet et al. 1999; Wierzba et al. 2001) continued to show a significant relationship, while others (Chowdhury et al. 1990; Lindtjorn et al. 1993) did not.

Table 2.13 Diarrhoea cohort studies included in dichotomous analysis

| Analysis <br> level | Study | Location | Cut-off point | Sample size | Crude RR (95\% CI) | Adjusted $R R^{2}$ <br> (95\% Cl) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | Wierzba <br> et al. (200I) | Egypt | WAZ <-2 | 143 | 1.83 (1.36-2.47) | $1.7(1.2-2.3)^{\text {b }}$ |
| 1 | Baqui et al. (1993) | Bangladesh | WAZ <-2 | 512 | 1.11 (1.00-1.23) | 1.22 (1.09-1.35) ${ }^{\text {c }}$ |
| 1 | Schorling et al. (1990) | Brazil | WAZ <-2 | 61 | 1.18 (1.06-1.34) | $\begin{aligned} & \mathrm{OR}=3.4 \\ & (1.0-11.9)^{\mathrm{d}} \end{aligned}$ |
| 1 | Victora et al. (1990) | Brazil | WAZ <-2 | 4486 | 0.78 (-0.26-1.95) | $0.58{ }^{\text {e }}$ |
| 2 | Bhan et al. (1986) | India | WA < $70 \%$ | 1467 | 0.90 (0.80-1.01) | - |
| 2 | Walia et al. (1989) | India | WA < $70 \%$ | 838 | 1.33 (1.07-1.62) | - |
| 2 | Anand et al. (1994) | India | WA < $70 \%$ | 250 | 1.08 (0.92-1.26) | - |
| 2 | Sepulveda et al. (1988) | Mexico | WA < $70 \%$ | 284 | 1.86 (1.54-2.21) | $1.7{ }^{\text {f }}$ |
| 3 | Mathur <br> et al. (1985) | India | WA < $75 \%$ | 687 | 1.40 (1.24-1.58) | - |
| 3 | Chen et al. (1981) | Bangladesh | WA < $75 \%$ | 207 | 1.04 (0.91-1.20) | - |
| 3 | Black et al. (1984) | Bangladesh | WA < $75 \%$ | 125 | 1.10 (0.88-1.33) | - |
| 3 | Bairagi et al. (1987) | Bangladesh | WA < $75 \%$ | 1454 | 1.10 (0.96-1.26) | - |
| 3 | Henry et al. (1987) | Bangladesh | WA < $75 \%$ | 300 | 0.93 (0.82-I.06) | - |
| 3 | James (1972) | Costa Rica | WA < $75 \%$ | 137 | 1.08 (0.93-1.26) | - |
| 3 | Gordon et al. (1964) | Guatemala | WA $<75 \%$ | 179 | 1.73 (1.45-2.07) | - |
| 3 | Tomkins et al. (198।) | Nigeria | WA < $75 \%$ | 343 | 1.22 (1.00-1.45) | - |
| 3 | el Samani et al. (1988) | Sudan | WA < $75 \%$ | 403 | 1.22 (1.00-1.42) | $\begin{aligned} & \mathrm{OR}=1.3 \\ & (1.0-1.7)^{\mathrm{g}} \end{aligned}$ |
| 4 | Khan and <br> Yunus (1990) | India | WA <80\% | 183 | 1.39 (0.75-2.58) | - |
| 4 | Biritwum et al. (1986) | Ghana | WA <80\% | 250 | 1.53 (1.1-2.01) | - |
| 5 | Chowdhury et al. (1990) | Bangladesh | WA <85\% | 753 | 1.54 (0.87-2.72) | - |

- No data.
a All crude risk estimates in Table 2.13 represent incidence below cut-off relative to incidence above cut-off. Adjusted risk estimates, however, vary in their referent category and adjusted variables as shown in the following table footnotes.
b Estimate adjusted for age, sex, socioeconomic status, breastfeeding and previous morbidity. WAZ <-2 vs WAZ >-2.
c Estimate adjusted for age. WAZ $<-2$ vs WAZ $>-2$. When stratified by previous diarrhoea experience, age-adjusted rate ratios were $I .00(0.75-I .33)$ and 1.26 (I.07-I.49) for underweight children without recent morbidity and with recent morbidity, respectively.
d Estimate adjusted for previous diarrhoea morbidity, age, sex and household crowding. Estimate is an OR comparing WAZ $<-3$ to WAZ >-3.
e Estimate adjusted for income. WAZ $<-2$ vs WAZ $\geq 0$.
f Estimate adjusted for age. Weight-for-age $<75 \%$ vs weight-for-age $\geq 90 \%$.
g Estimate adjusted for age, sex, season, socioeconomic status and previous morbidity. Weight-for-age $<90 \%$ vs weight-forage $\geq 90 \%$.

Table 2.14 Analysis levels for dichotomous analysis according to anthropometric definition used in study

| Analysis level | Comparison |
| :--- | :--- |
| 1 | WAZ $<-2$ vs WAZ $>-2$ |
| 2 | WA $<70 \%$ vs WA $>70 \%$ |
| 3 | WA $<75 \%$ vs $W A>75 \%$ |
| 4 | WA $<80 \%$ vs WA $>80 \%$ |
| 5 | WA $<85 \%$ vs WA $>85 \%$ |

Table 2.13 summarizes prospective studies used to estimate diarrhoea incidence according to weight-for-age. The majority of diarrhoea studies reported anthropometric status in terms of percentage of median rather than $z$-score and the cut-off points varied by study. In India, for example, most studies conformed to the Indian Academy of Pediatrics classification in which weight-for-age $>80 \%$ represents normal status and weight-for-age $<70 \%$ represents second-degree, third-degree or fourth-degree malnutrition (Indian Academy of Pediatrics 1972). Many others used the Gomez classification, defining second-degree malnutrition and worse as weight-for-age $<75 \%$ (Gomez et al. 1956). For comparison, studies were grouped into different analysis levels according to the cut-off points used to classify malnutrition (Table 2.14).

## Dichotomous estimate

In developing a summary risk estimate, nutritional status was first examined as a dichotomous variable comparing all children below the cut-off point ("malnourished") to children above that point (reference group). The dichotomous approach allowed for the inclusion of the greatest number of studies, since many studies did not further subdivide the malnourished group into discrete anthropometric categories.

## Excluded studies

Ten studies that were identified were excluded from the risk analysis (Table 2.15). Seven prospective studies were excluded due to insufficient statistical data (e.g. sample sizes per anthropometric category), data presented as a figure only, or the use of a different growth reference. Six of the seven reported an increased incidence of diarrhoea episodes associated with low weight-for-age, including studies in the Sudan (Kossman et al. 2000) and the Congo (Tonglet et al. 1999) that adjusted for previous diarrhoea morbidity. An additional three cross-sectional studies in Ecuador (Brussow et al. 1993), India (Luwang and Datta 1982) and El Salvador (Stetler et al. 1981) reported risk estimates on the order of 1.45 to 1.57 , but reverse causality disqualified the studies.

## Studies included in the risk estimate

Figure 2.7 illustrates the distribution of individual study risk estimates according to the cut-off point used. Twenty cohort studies provided sufficient results and statistical data for inclusion in a summary estimate (Table 2.13). The evidence regarding risk of diarrhoea was equivocal

Figure 2.7 Relative risk for diarrhoea according to weight-for-age cut-off point used in study


Table 2.15 Diarrhoea studies excluded from dichotomous analysis

| Location | Study | Cut-off point | Sample size | Crude RR $(95 \% C l)^{a}$ | Adjusted RR (95\% CI) | Cause for exclusion |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Congo | Tonglet et al. (1999) | WA <25th percentile | 842 | $\begin{aligned} & 1.22^{\mathrm{b}} \\ & (1.02-1.46) \end{aligned}$ | $\begin{gathered} \mathrm{OR}= \\ 1.49(0.94-2.05) ; \\ 1.50(1.10-1.89) \end{gathered}$ | Local growth reference |
| Ecuador | Brussow et al. (1993) | WAZ <-2 | 321 | 1.45 (0.99-2.12) | - | Cross-sectional; reverse causality |
| El Salvador | Stetler et al. (1981) | WA < $75 \%$ | 3705 | 1.46 (1.24-1.71) | - | Cross-sectional; reverse causality |
| Ethiopia | Lindtjorn et al. (1993) | WAZ <-2 | 425 | $2.2{ }^{\text {c }}$ | - | Data reported in figure form; no categorical sample size data |
| Gambia | Tomkins et al. (1989) | WAZ <-2 | 211 | $1.18 ; 1.33^{\text {d }}$ | - | Data reported in figure form; no categorical sample size data |
| India | Ghai and Jaiswal (1970) | WA $<85 \%$ | 925 | 2.35 (1.97-2.81) | - | Uncertain growth reference |
| India | Luwang and Datta (1982) | WA $<70 \%$ | 508 | 1.57 (1.26-1.94) | - | Cross-sectional; reverse causality |
| Papua New Guinea | Binns (1976) | WA <80\% | 630 | 1.75 | - | No categorical sample size data |
| Sudan | Kossman et al. (2000) | WAZ <-2 | 28753 | 1.32; $1.84{ }^{\text {e }}$ | $\mathrm{OR}=$ <br> I. 13 (I.2I-I.45); <br> 1.75 (1.56-1.96) | No categorical sample size data |
| Thailand | Thongkrajai et al. (1990) | WA $<75 \%$ | 1339 | 0.18 (0.06-0.43) | - | Local growth reference |
| No data. |  |  |  |  |  |  |
| Crude risk estimates represent incidence below cut-off point relative to incidence above cut-off point except where indicated as follows. |  |  |  |  |  |  |
| b Weight-for-age $<25$ th percentile vs weight-for-age $>75$ th percentile. Crude RR is a Mantel-Haenszel weighted risk ratio. Adjusted ORs are an months, respectively. Adjusted for age, sex and previous morbidity. |  |  |  |  |  |  |
| c $W$ AZ $<-2$ vs WAZ -1 to 0 . Estimate of data reported in figure form. |  |  |  |  |  |  |
| WAZ<-2 vs WAZ -I to 0 . Relative risks of diarrhoea on day of monthly visit during the dry and rainy seasons, respectively. Estimates of data reported in figure form. |  |  |  |  |  |  |
| WAZ -3 to -2 vs $W A Z \geq-I$, and WAZ -4 to -3 vs $W A Z \geq-I$, respectively. OR was adjusted for age, sex, socioeconomic status, season, bre morbidity. |  |  |  |  |  |  |

Table 2.16 Combined estimate for diarrhoea incidence from dichotomous analysis

| Study type | Analysis level | Number of studies | Combined risk estimate (95\% Cl) |
| :--- | :--- | :---: | :---: |
| Cohort | 1 | 4 | $1.25(1.04-1.50)$ |
| Cohort | 2 | 4 | $1.24(0.90-1.70)$ |
| Cohort | 3 | 9 | $1.18(1.05-1.33)$ |
| Cohort | $4+5$ | 3 | $1.51(1.20-1.90)$ |
| Combined estimate | $1+2+3+4+5$ | 20 | $1.23(1.12-1.35)$ |

among these studies, with half reporting no significant effect. Individual studies reported crude risk estimates ranging between 0.78 in Brazil (Victora et al. 1990) and 1.86 in Mexico (Sepulveda et al. 1988). Three of four studies that adjusted for previous diarrhoea morbidity found a significantly increased risk of diarrhoea incidence associated with low weight-for-age (el Samani et al. 1988; Schorling et al. 1990; Wierzba et al. 2001). The fourth, in Bangladesh, observed an increased risk of diarrhoea among underweight children $(R R=1.26)$ who had already suffered diarrhoea within the preceding three months, but the risk was not significant for those children without recent diarrhoea experience ( $\mathrm{RR}=1.00$ ) (Baqui et al. 1993). In the most recently published investigation, Wierzba et al. (2001) in Egypt observed that WAZ $<-2$ predisposed toward increased incidence of diarrhoea among children aged $<3$ years, yielding a relative risk of 1.7 ( $95 \%$ CI1.2-2.3) compared to $z$-score $>-2$, after adjusting for age, sex, socioeconomic status, breastfeeding and previous episodes of diarrhoea.

## Dichotomous relative risk estimate

Based on all cohort studies, a combined estimate suggested underweight children were at a significantly increased relative risk of 1.23 (95\% CI 1.12-1.35) for developing an episode of diarrhoea (Table 2.16). The estimate based on a cut-off point of $z$-score $<-2(R R=1.25)$ was similar to the estimate using percentage of median $<70 \%(R R=1.24)$ and percentage of median $<75 \% ~(R R=1.18)$. Higher cut-off points of $80 \%$ or $85 \%$ yielded a higher relative risk of 1.51 , possibly due to the small number of studies or the comparatively better nutritional status of the referent group.

## Multi-categorical estimate

We also attempted to examine the risk of diarrhoea incidence according to multiple anthropometric categories, comparing incidence among moderately to severely underweight (e.g. WAZ $<-2$ ), mildly underweight (e.g. WAZ between -2 and -1 ) and "normal"(e.g. WAZ >-1) children. The

Table 2.17 Analysis levels for multiple categorical analysis according to anthropometric classification used in study

| Analysis level | Underweight categories | Cut-off point |
| :--- | :--- | :--- |
| I | Moderate-severe | WAZ $<-2$ |
|  | Mild | WAZ -2 to -I |
|  | Normal (referent) | WAZ >-I |
| 2 | Moderate-severe | WA $<70 \%$ |
|  | Mild | WA $70-80 \%$ |
|  | Normal (referent) | WA $>80 \%$ |
| 3 | Moderate-severe | WA $<75 \%$ |
|  | Mild | WA 75-90\% |
|  | Normal (referent) | WA $>90 \%$ |
|  | Moderate-severe | WA $<60$ |
|  | Mild | WA $60-85 \%$ |
|  | Normal (referent) | WA $>85 \%$ |

studies used different cut-off points based on percentage of median and $z$-score and were subdivided into analysis levels according to the anthropometric classification scheme used (Table 2.17). A subset of 10 cohort studies reporting multiple anthropometric categories provided sufficient statistical data for inclusion in a summary analysis (Table 2.18).

The different classification schemes complicated combining studies to a greater extent than in the dichotomous analysis. Eight of 10 studies, however, fell into the second or third analysis levels, which produced relative risks of 1.23 and 1.28 , respectively, associated with moderate to severely low weight-for-age (Table 2.19). The corresponding risk estimates associated with mildly low weight-for-age were 1.17 and 0.95 for these analysis levels (Table 2.20). The overall risk estimates for diarrhoea based on all analysis levels were 1.25 (statistically significant) for moderate and severe underweight and 1.12 (not statistically significant) for mild underweight. Since the risk relationship was not significant for mild underweight, only the risk for moderate-severe underweight was considered toward the burden of disease estimate. This value (1.25) was similar to the value calculated in the dichotomous analysis (1.23). As there was no significant difference between the two estimates and the dichotomous value was based on a larger number of studies, the dichotomous value of 1.23 was selected for calculating burden of disease.

## Summary of morbidity risk estimates

Three morbidity outcomes contributed to the overall calculation of disease burden: diarrhoea incidence, pneumonia/ALRI incidence and malaria incidence. Table 2.21 summarizes the risk of disease incidence associated with low weight-for-age for each outcome. There was no evidence that undernutrition contributed to incidence of measles and other disease relationships cannot be adequately quantified at this time.

Table 2.18 Cohort studies of diarrhoea incidence included in categorical analysis

| Analysis <br> level | Study | Location | Sample <br> size |  | Crude risk estimate (95\% CI) |
| :--- | :--- | :--- | :---: | :--- | :--- |

Table 2.19 Combined estimates of diarrhoea incidence associated with severe to moderate underweight status relative to WAZ >-I

| Analysis level | Category | Number of studies | Combined risk estimate (95\% CI) |
| :--- | :---: | :---: | :---: |
| I | WAZ $<-2$ | 1 | $0.94(0.22-3.94)$ |
| 2 | WA $<70 \%$ | 4 | $1.23(0.95-1.60)$ |
| 3 | WA $<75 \%$ | 4 | $1.28(0.85-1.94)$ |
| 4 | WA $<60 \%$ | 1 | $1.47(0.65-3.34)$ |
| $I+2+3$ |  | 9 | $1.25(1.02-1.53)$ |
| $I+2+3+4$ |  | 10 | $1.25(1.03-1.52)$ |

### 4.3 RISK OF MORTALITY DUE TO PERINATAL CONDITIONS

"Perinatal conditions" are responsible for approximately $18 \%$ of all deaths among children aged $<5$ years in developing countries (de Onis et al. 1998). These deaths, concentrated in the neonatal period ( $\leq 28$ days postpartum), result mostly from low birth weight, birth asphyxia and trauma, neonatal infections (e.g. tetanus and syphilis) and congenital

Table 2.20 Combined estimates of diarrhoea incidence associated with mild underweight status relative to WAZ >-I

| Analysis level | Category | Number of studies | Combined risk estimate (95\% Cl) |
| :--- | :--- | :---: | :---: |
| I | WAZ -2 to -I | 1 | $1.93(1.05-3.56)$ |
| 2 | WA 70-80\% | 4 | $1.17(1.05-1.31)$ |
| 3 | WA 75-90\% | 4 | $0.95(0.68-1.34)$ |
| 4 | WA 60-85\% | 1 | $2.15(0.99-4.64)$ |
| $1+2+3$ |  | 9 | $1.09(0.92-1.29)$ |
| $1+2+3+4$ |  | 10 | $1.12(0.95-1.32)$ |

Table 2.21 Relative risk of morbidity associated with weight-for-age <-2SDs

| Morbidity outcome | Relative risk (95\% CI) |
| :--- | :---: |
| Diarrhoea incidence | $1.23(1.12-\mathrm{I} .35)$ |
| Pneumonia incidence | $1.86(1.06-3.28)$ |
| Malaria incidence | $1.31(0.92-1.88)$ |

anomalies (Anonymous 1999). In developed regions such as North America and western Europe, $23 \%$ of neonatal deaths are due to congenital anomalies and an additional $65 \%$ are the result of other perinatal conditions (C. Stein, personal communication, 2001). Of the deaths from perinatal conditions, the largest share ( $32-65 \%$ ) is attributed to low birth weight.

The Global Burden of Disease (GBD) study categorizes low birth weight, birth asphyxia, birth trauma and other conditions (such as neonatal sepsis, maternal and placental complications, respiratory distress, fetal blood loss, fetal haematological disorders, anaemia, perinatal infections and maternal diabetes) as "perinatal conditions", while deaths due to congenital anomalies, neonatal tetanus and syphilis are addressed separately (WHO 1992a). Among these conditions, low birth weightspecifically that due to IUGR-is the most strongly linked to undernutrition and is the only perinatal condition considered in our analysis.

Low birth weight is defined as birth weight below 2500 grams. It is a product of IUGR, preterm birth, or both in combination (Kramer 1987). In developing countries, the majority of low-birth-weight births are due to IUGR (usually defined as birth weight less than the tenth percentile of weight-for-gestational-age) whereas preterm birth ( $<37$ weeks gestation) is the predominant cause in most developed countries (Ashworth 1998). Preterm low-birth-weight infants tend to have higher neonatal mortality rates than full-term, growth-retarded infants; at
highest risk are low-birth-weight infants who are both growth-retarded and preterm (Barros et al. 1992; Behrman et al. 1971; Cogswell and Yip 1995; Gray et al. 1991; Sappenfield et al. 1987).

The etiology of IUGR is complex and multifactorial. In developing countries maternal undernutrition is the major determinant of IUGR, and evidence across populations has demonstrated a greater incidence of IUGR births among women who are underweight or stunted prior to conceiving, or who fail to gain sufficient weight during pregnancy (Bakketeig et al. 1998; King and Weininger 1989; Kramer 1987; WHO 1997). Poor maternal nutrition during pregnancy is thought to account for $14 \%$ of IUGR in developing countries; maternal stunting may account for $18.5 \%$ (ACC/SCN 2000b). Malaria, other acute and chronic infections and cigarette smoking are also important etiologic factors for IUGR in developing countries. In developed countries, smoking is the most important determinant of IUGR, followed by factors such as maternal nutrition, pre-eclampsia, genetic factors and alcohol or drug use (Bakketeig et al. 1998). IUGR is also associated with multiple births and primiparity (Cogswell and Yip 1995).

Fetal growth retardation takes different forms, which may have different implications for neonatal and infant health. IUGR can be subdivided into asymmetric (wasted) IUGR, characterized by adequate length and head circumference, but reduced weight and low ponderal index; and symmetric (stunted) IUGR, in which the ponderal index is normal but weight, length and head circumference are all reduced (Bakketeig 1998). Symmetric IUGR generally reflects early onset or chronic undernutrition in utero, while asymmetric IUGR is thought to result from undernutrition of later onset (Ashworth 1998). The difference may be clinically important, as asymmetrically growth-retarded infants have demonstrated higher risks of asphyxia, hypoglycaemia and other morbidities, and higher mortality rates in the early neonatal period (Ashworth 1998; Caulfield et al. 1991; Villar et al. 1990). Stunted infants, on the other hand, have greater risks of mortality in later infancy (Ashworth 1998; Balcazar and Haas 1990; Cheung et al. 2001).

The evidence relating low maternal BMI to preterm birth is more ambiguous. A 1995 WHO meta-analysis based on data sets from 20 developed and developing countries calculated a combined OR of 1.3 ( $95 \%$ CI 1.1-1.4) for preterm birth associated with pre-pregnancy BMI $<20 \mathrm{~kg} / \mathrm{m}^{2}$ (WHO 1995b). Individual studies, however, have been inconsistent. In Papua New Guinea, BMI was found to be a significant predictor of preterm delivery; a one-unit increase in BMI was associated with a reduced risk of preterm delivery ( $\mathrm{OR}=0.79$, $95 \%$ CI $0.66-0.94$ ), and remained significant after being adjusted for haemoglobin concentration, smoking, gravidity and anti-malarial use (Allen et al. 1998). A case-control study in India observed an increased risk of preterm birth with underweight status as well (Mavalankar et al. 1994). Other studies in Malawi (Pelletier et al. 1995) and Indonesia (Husaini et al. 1995)
reported little or no association between pre-pregnancy BMI and preterm delivery. In developed settings, studies from Canada (Kramer et al. 1995), Italy (Spinillo et al. 1998), the United Kingdom (Sebire et al. 2001) and the United States (Edwards et al. 1979; Naeye 1990; Siega-Riz et al. 1996) observed higher rates of preterm labour or delivery among underweight women, while other studies in Canada (Kramer et al. 1992), Finland (Rantakallio et al. 1995) and Sweden (Cnattingius et al. 1998) did not. Recently, in a study in the United States of America based on the National Maternal and Infant Health Survey found that underweight women had an increased risk for preterm delivery only when their pregnancy weight gain was inadequate (Schieve et al. 2000). Because of the uncertain risk relationship, low birth weight due to preterm birth was excluded from our analysis.

Although many studies conducted in developing countries examined the influence of maternal body size on infant birth weight and others compared neonatal mortality rates according to birth weight, few studies directly compared maternal weight to neonatal or perinatal mortality. These studies used postpartum weight as a proxy for pre-pregnant weight, and therefore, possible effects of pregnancy weight gain cannot be excluded. Among these, a slightly elevated risk of early neonatal death (adjusted $\mathrm{OR}=1.4,95 \%$ CI 1.0-1.9) was associated with maternal weight below 50 kg in a nested case-control study in Brazil (Gray et al. 1991). The authors adjusted for low birth weight in one regression model to distinguish between risk factors that might operate in part through low birth weight and factors that might influence neonatal death independently. Low maternal weight remained significantly associated with neonatal death only when low birth weight was excluded from the model, suggesting maternal weight influences neonatal death through low birth weight.

In other studies, maternal weight $<40 \mathrm{~kg}$ and low weight-to-height ratio were significantly associated with perinatal death in Ahmedabad, India ( $\mathrm{OR}=2.9,95 \%$ CI $1.8-4.7$ and $\mathrm{OR}=3.0,95 \% 1.9-4.4$, respectively), adjusted for maternal age, parity, obstetric history and antenatal care (Mavalankar et al. 1994). In the Sudan, maternal weight below 50 kg was significantly associated with perinatal death, after adjustment for birth interval, prior fetal loss and antenatal care, among other factors (adjusted OR=2.3, 95\% CI 1.1-4.8) (Taha et al. 1994). In a commu-nity-based case-control study in Punjab, perinatal death was significantly associated with weight $<40 \mathrm{~kg}$ and height $<152 \mathrm{~cm}$, but not BMI $<20 \mathrm{~kg} / \mathrm{m}^{2}$, after controlling for factors such as birth interval and length of gestation (Sachar and Soni 2000).

In developed countries, there is a consistent association between maternal underweight status and low birth weight, but the evidence relating maternal underweight status to perinatal death has been mixed. In New York City, low BMI prior to pregnancy was associated with slightly, but not significantly, higher perinatal mortality rates in one hospital-
based study (Bracero and Byrne 1998). Another analysis using data from the USA Collaborative Perinatal Study reported reduced perinatal mortality for infants of underweight mothers; among mothers with low pregnancy weight gain ( $<0.8 \mathrm{~kg} /$ month $)$, however, underweight status was associated with higher perinatal mortality (Naeye 1990). In Australia, the odds ratios for stillbirth/neonatal death (adjusted for gestational age, parity, smoking and maternal age) were 1.65 and 2.64 among women with low postpartum BMI (defined for postpartum women as $20-$ $\left.24.4 \mathrm{~kg} / \mathrm{m}^{2}\right)$ and very low postpartum BMI ( $<20 \mathrm{~kg} / \mathrm{m}^{2}$ ), respectively (Cattanach et al. 1993). In Finland, perinatal and childhood mortality rates were similar between women having low and normal BMI (Rantakallio et al. 1995), and in Sweden low maternal BMI was associated with reduced risk of late fetal death and a slightly, but not significantly, reduced risk of early neonatal death relative to maternal BMI of $20-25 \mathrm{~kg} / \mathrm{m}^{2}$ (Cnattingius et al. 1998).

Such studies have generated uncertainty about the significance of low birth weight, per se, as an intermediate variable in the causal pathway between prenatal factors and perinatal mortality (Rush 2001; Wilcox 2001). As with other anthropometric indicators, birth weight is regarded as a proxy for underlying biological processes that must be inferred. All prenatal factors that influence birth weight may not affect infant health equally. For example, high altitude is associated with lower birth weight, but not necessarily higher mortality (Cogswell and Yip 1995; Wilcox 2001). Therefore, caution must be taken when generalizing across populations.

## Estimate of mortality due to perinatal conditions

The relationship between maternal underweight status and neonatal mortality was estimated by considering two stages of the conceptual pathway mediated through IUGR. First, what is the proportion of IUGR attributable to poor maternal pre-pregnancy anthropometric status? And second, what proportion of neonatal mortality can be attributed to IUGR? There are four components to this equation: the proportion of IUGR births among live births in each subregion, an estimate of infant mortality risk associated with IUGR, subregional prevalence of underweight status among women of reproductive age and an estimate of the risk of IUGR associated with pre-pregnant underweight status.

By estimating the risk of IUGR in relation to pre-pregnancy BMI only, we are not considering the fraction of IUGR attributable to inadequate weight gain during pregnancy, and are potentially underestimating the overall influence of maternal weight on infant birth weight. Gains in pre-pregnancy and pregnancy weight have independent, additive effects on birth weight (Krasovec and Anderson 1991). Across diverse populations, low pregnancy weight gain, particularly in the second and third trimesters, has been associated with risk ratios of 1.7 to 2.0 for IUGR; among mothers who are below average height or weight before preg-
nancy, the risk ratios are 3.1 and 5.5, respectively (Strauss and Dietz 1999; WHO 1997). The prevalence data on pregnancy weight gain in developing countries is scarce, but there is evidence that poor weight gain is common, especially among women who have low pre-pregnancy BMI and must gain more to compensate. In rural Indonesia, for example, $79 \%$ of women in a cohort study failed to reach their recommended total weight gain, including $82.4 \%$ of the women with a pre-pregnancy BMI below $20 \mathrm{~kg} / \mathrm{m}^{2}$ (Winkvist et al. 2002).

## Methods

The overall attributable fraction of neonatal mortality due to maternal underweight status was calculated as the product of the attributable fraction of neonatal mortality due to IUGR and the attributable fraction of IUGR due to maternal underweight status. Attributable fractions were calculated according to the formula $[(\mathrm{P})(\mathrm{RR}-1)] /[1+(\mathrm{P})(\mathrm{RR}-1)]$, where P is the prevalence of the risk factor and RR is the associated risk ratio. Prevalence and risk ratio estimates were derived from available published and unpublished sources.

## Incidence of IUGR

Classification of IUGR in developing countries can be problematic because gestational age cannot be accurately determined in most cases, and reference curves that are adjusted for gestational age are not widely used (ACC/SCN 2000b). Therefore, the incidence of low birth weight at term (referred to as "term-LBW" or "IUGR-LBW") is used as the best proxy for IUGR incidence, although this will underestimate the actual incidence of IUGR because it excludes growth-retarded infants born preterm and those born over 2500 grams but below the tenth percentile for their gestational age (de Onis et al. 1998).

Incidence rates of IUGR-LBW for most developing countries were previously estimated by de Onis et al. (1998) using low birth weight estimates from the WHO database on Low Birth Weight, compiled by the Maternal Health and Safe Motherhood Programme (WHO 1992b). The authors applied a linear regression model $[\mathrm{Y}=-3.2452+0.8528 \mathrm{X}]$ derived from numerous studies in developing countries where birth weight and valid gestational age assessments were recorded. The dependent variable in the model was IUGR-LBW and the independent variable was the total incidence rate of low birth weight ( $\beta=0.8528$; SE $=0.0282 ; P=0.0001 ; r=0.96$ ) (de Onis et al. 1998; Villar et al. 1994). The linear regression model was found to be unsuitable as a predictor of IUGR-LBW in developed countries, however.

Incidence rates of IUGR-LBW in developed countries were based on Medline (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi), Popline (http://db.jhuccp.org/popinform/index.stm) and Google (http://www.google.com/) searches for country-level data sets and nationally-representative study samples that reported rates of IUGR-

LBW or sufficient statistical information (including numbers of births, mean birth weights and SDs or percentiles according to week of gestation) for calculating the rate of IUGR-LBW among total live births. The studies used in estimating rates were published between 1993 and 2001. The data reported in most studies were collected between 1992 and 2001, although in a few instances the data reflect populations from as early as 1984. In general, the reported data were limited to singleton births and excluded births having major congenital anomalies. Where data were unavailable-as was the case for the majority of European countries-the rate of IUGR-LBW was assumed for these purposes to be equivalent to the rate in AMR-A, or $2.5 \%$ of total live births (Kramer et al. 2001; Ventura et al. 1999). Estimates of small-for-gestational-age incidence were not appropriate for this analysis because they can include births weighing over 2500 grams. The estimated incidence rate of IUGRLBW for each country was multiplied by the estimated total number of live births in that country for the year 2000; the country estimates were then collapsed into the 14 subregions in order to calculate subregional rates (Table 2.22). Estimates of live births per year for each country were taken from the United Nations Population Division (UN 2001).

## IUGR and risk of neonatal mortality

Rice et al. (unpublished) calculated the risk of neonatal mortality associated with IUGR-LBW in an initial review and meta-analysis of the rela-

Table 2.22 Incidence of IUGR-LBW, by subregion

|  | Total live births <br> per year (000s) | IUGR-LBW births <br> as percentage of <br> total live births | Estimated number <br> of IUGR-LBW births <br> per year (000s) |
| :--- | :---: | :---: | :---: |
| Subregion | 11185 | 10.5 | 1176 |
| AFR-D | 13240 | 9.3 | 1229 |
| AFR-E | 4426 | 2.5 | 112 |
| AMR-A | 9303 | 6.7 | 621 |
| AMR-B | 2011 | 6.6 | 133 |
| AMR-D | 3410 | 4.1 | 141 |
| EMR-B | 12003 | 12.1 | 1455 |
| EMR-D | 4233 | 2.5 | 107 |
| EUR-A | 3743 | 2.6 | 96 |
| EUR-B | 2527 | 2.5 | 63 |
| EUR-C | 5933 | 4.2 | 250 |
| SEAR-B | 42147 | 22.7 | 6866 |
| SEAR-D | 358 | 2.3 | 8 |
| WPR-A | 26840 | 2.9 | 769 |
| WPR-B | 129493 | 10.1 | 13027 |
| World |  |  |  |

tionship between malnutrition and cause-specific mortality. The authors conducted a Medline search for English-language research reports published between 1966 and 1999 on child mortality, nutritional status, low birth weight, perinatal causes of death and neonatal mortality, excluding studies conducted in developed countries. They selected community and hospital-based studies that reported sufficient statistical and followup data on infants who were low birth weight at term, excluding data on very low birth weight $(<1000 \mathrm{~g})$ infants. Follow-up times and cut-off points varied among the studies as summarized in Table 2.23. Unadjusted data from studies were combined according to the random-effects method described by Morris (Everson and Morris 1983; Morris 1983). The combined risk estimate from all studies was 5.53 (3.74-8.17) and the risk estimate from the subset of studies using a 2500 g cut-off was 6.00 (3.63-9.90). Because the overall combined estimate of 5.53 was heavily weighted toward studies with short post-natal follow-up periods and higher birth weight cut-off points, we selected the 6.00 risk estimate for our calculations.

An elevated mortality risk has been observed among IUGR-LBW infants in developed countries, as well. Ashworth (1998) estimated the risk of neonatal mortality according to birth weight, combining 10 data sets from developing and developed countries (predominantly weighted toward births in the United States) and excluding most preterm births. Individual studies showed a consistent dose-response effect of increasing mortality risk with decreasing birth weight. Compared to birth weight of $2500-2999 \mathrm{~g}$, the overall relative risk associated with birth weight between 2000 g and 2499 g was 4.0 and the risk associated with birth weight $1500-1999 \mathrm{~g}$ was 18.0 . An overall dichotomous risk estimate comparing birth weight $<2500 \mathrm{~g}$ to $>2500 \mathrm{~g}$ was not reported, but after pooling published data from four (Behrman et al. 1971; Binkin et al. 1985; Lubchenco et al. 1972; Sappenfield et al. 1987) of six United States studies included in the estimate, the combined risk for neonatal mortality associated with birth weight $<2500 \mathrm{~g}$ at term was $10.64(95 \%$ CI 9.94-11.38). This crude estimate is based on mostly white, singleton births between 38 and 42 weeks gestation and excludes births below 1000 g .

## Prevalence of low maternal BMI

Refer to section 3: prevalence of underweight among women of reproductive age.

## Maternal pre-pregnancy BMI and risk of IUGR

In $1995, \mathrm{WHO}$ published a meta-analysis based on data sets from 25 studies that related maternal anthropometry to pregnancy outcomes (WHO 1995b). These sets represented over 111000 births in 20 developing and developed countries throughout the world. Countries were grouped together for analysis according to similarities in the population
Table 2.23 Risk of neonatal ( $\leq 28$ days) or early neonatal ( $\leq 7$ days) death among term infants according to birth weight ${ }^{\text {a }}$

| Location | Study | Study type | Neonatal period studied | Sample size | Birth weight comparison | Crude relative risk (95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Bolivia | Haas et al. (1987) | Retrospective cohort | $\leq 2$ days | 12280 | $<2900 \mathrm{~g}$ vs $\geq 2900 \mathrm{~g}$ | 3.63 (1.93-6.82) |
| Brazil | Barros et al. (1987) | Prospective cohort | $\leq 7$ days | 5356 | $<2500 \mathrm{~g}$ vs $\geq 2500 \mathrm{~g}$ | 6.28 (2.57-15.33) |
| Brazil | de Almeida and Jorge (1998) | Retrospective cohort | $\leq 28$ days | 2014 | $<10$ th percentile vs $\geq 10$ th percentile | 10.61 (3.34-33.72) |
| Guatemala | Mata (1978) | Community-based cohort | $\leq 28$ days | 385 | $<2500 \mathrm{~g}$ vs $\geq 2500 \mathrm{~g}$ | 3.41 (0.63-18.40) |
| India | Arora et al. (1987) | Prospective cohort | $\leq 28$ days | 200 | $<10$ th percentile vs $\geq 10$ th percentile | 3.03 (0.32-28.64) |
| India | Bhargava et al. (1985) | Prospective cohort | $\leq 7$ days | 13806 | $<-2$ SDs vs $\geq-2$ SDs ${ }^{\text {b }}$ | 11.08 (7.34-16.73) |
| India | Ghosh et al. (1979) | Community-based cohort | $\leq 28$ days | 3650 | $\leq 2500 \mathrm{~g}$ vs $>2500 \mathrm{~g}$ | 6.25 (3.14-12.42) |
| Mexico | Balcazar and Hass (1991) | Retrospective cohort | $\leq 3$ days | 8526 | $<10$ th percentile vs $\geq 10$ th percentile | 5.35 (2.69-10.65) |
| Mexico | Haas et al. (1987) | Retrospective cohort | $\leq 2$ days | 9228 | $<2900 \mathrm{~g}$ vs $\geq 2900 \mathrm{~g}$ | 2.13 (1.12-4.06) |
| Combined estimate (all studies) |  |  |  |  |  | 5.53 (3.74-8.17) |
| Combined estimate (Barros et al. 1987; Mata et al. 1978; Ghosh et al. 1979) |  |  |  |  |  | 6.00 (3.63-9.90) |

[^5]Table 2.24 Analysis groups used by WHO Collaborative Study on Maternal Anthropometry and Pregnancy Outcomes

| Group | Countries/data sets included in group | 25th quartile <br> cut-off (BMI) | 75th quartile <br> cut-off (BMI) |
| :--- | :--- | :---: | :---: |
| I | India (Pune), Sri Lanka | 17.3 | 20.1 |
| 2 | China, Gambia, India (Hyderabad), | 18.4 | 21.0 |
|  | Indonesia, Myanmar, Nepal (Rural), |  |  |
|  | Viet Nam |  | 22.7 |
| 3 | Guatemala, Malawi, Thailand | 19.4 | 25.0 |
| 4 | Argentina, Cuba, United Kingdom, USA | 20.1 | 26.7 |

Source: WHO 1995a.

Table 2.25 Odds ratio for IUGR according to pre-pregnancy BMI

| Group | BMI cut-off | BMI referent | OR for IUGR (95\% CI) |
| :--- | :---: | :---: | :---: |
| $I$ | $\leq 17.3$ | $\geq 20.1$ | $0.7(0.1-2.6)$ |
| 2 | $\leq 18.4$ | $\geq 21.0$ | $1.8(1.5-2.3)$ |
| 3 | $\leq 19.4$ | $\geq 22.7$ | $1.8(1.3-2.5)$ |
| 4 | $\leq 20.1$ | $\geq 25.0$ | $2.0(1.8-2.2)$ |
| 5 | $\leq 21.0$ | $\geq 26.7$ | $1.5(1.3-1.7)$ |
| Combined | $\leq 19.7^{\mathrm{a}}$ | $\geq 24.2^{\mathrm{a}}$ | $1.8(1.7-2.0)$ |

a Combined BMI cut-offs were not reported. Values here are the weighted means of the 25th and 75th quartile values from each data set.
Source: WHO 1995a.
distributions for various anthropometric indicators, including prepregnancy weight, height, BMI and arm circumference. Table 2.24 describes the composition of analysis groups and the corresponding BMI cut-offs. Unadjusted odds ratios for pregnancy outcomes were then calculated by comparing the lowest quartile to the highest quartile within each analysis group.

Results are presented in Table 2.25. The meta-analysis reported significantly greater risk for IUGR associated with pre-pregnancy BMI below $19.7 \mathrm{~kg} / \mathrm{m}^{2}$ relative to BMI above $24.2 \mathrm{~kg} / \mathrm{m}^{2}$, with an overall OR of 1.8 ( $95 \%$ CI 1.7-2.0). This estimate was consistent across the analysis groups (with the exception of group one studies) despite the differences in their cut-off and reference points.

## Results

Subregional estimates of the attributable fraction of IUGR due to low pre-pregnancy BMI are listed in Table 2.26 for women aged 15-29 and

Table 2.26 Attributable fraction of IUGR due to low BMI among women aged 15-44 years, by subregion

| Subregion | Prevalence of $\mathrm{BMI} \leq 20 \mathrm{~kg} / \mathrm{m}^{2}$ (\%) |  | Fraction of IUGR attributable to $\mathrm{BMI} \leq 20 \mathrm{~kg} / \mathrm{m}^{2}$ (\%) |  |
| :---: | :---: | :---: | :---: | :---: |
|  | 15-29 years | 30-44 years | 15-29 years | 30-44 years |
| AFR-D | 43.3 | 29.5 | 25.7 | 19.1 |
| AFR-E | 37.8 | 27.4 | 23.2 | 18.0 |
| AMR-A | 25.1 | 19.5 | 16.7 | 13.5 |
| AMR-B | 22.7 | 11.1 | 15.4 | 8.2 |
| AMR-D | 11.7 | 9.3 | 8.6 | 6.9 |
| EMR-B | 28.4 | 12.7 | 18.5 | 9.2 |
| EMR-D | 40.1 | 32.6 | 24.3 | 20.7 |
| EUR-A | 20.9 | 13.8 | 14.3 | 9.9 |
| EUR-B | 24.5 | 14.0 | 16.4 | 10.1 |
| EUR-C | 24.5 | 9.7 | 16.4 | 7.2 |
| SEAR-B | 44.4 | 12.1 | 26.2 | 8.8 |
| SEAR-D | 56.7 | 42.1 | 31.2 | 25.2 |
| WPR-A | 40.9 | 27.1 | 24.7 | 17.8 |
| WPR-B | 32.6 | 24.8 | 20.7 | 16.6 |

30-44 years based on the prevalence of low BMI and the risk ratio of 1.8 reported by the WHO meta-analysis. Table 2.27 lists the attributable fraction of neonatal deaths due to IUGR-LBW for each subregion based on the prevalence of IUGR-LBW and the risk ratio of 6.0 calculated by Rice et al. (unpublished). The results of both tables are summarized in Table 2.28 with the overall attributable fraction of neonatal deaths due to low pre-pregnancy BMI. The fraction of neonatal deaths attributed to maternal BMI ranged between $0.8 \%$ (women aged 30-44 years in EUR-C) and 16.6 \% (women aged 15-29 years in SEAR-D).

### 4.4 Other health outcomes

Other important health outcomes related to underweight status were reviewed, but ultimately not used for burden of disease estimates. The adverse effects of undernutrition on risk of dysentery and persistent diarrhoea, pregnancy outcome, cognitive function and chronic diseases in later life are described below.

## Dysentery and persistent diarrhoea

Dysentery, defined by diarrhoea with blood, was specifically examined in two community-based prospective studies, both among children aged $<24$ months in Bangladesh. Henry et al. (1987) observed no significant difference in incidence of dysentery between children having weight-for-

Table 2.27 Attributable fraction of neonatal mortality due to IUGR-LBW, by subregion

| Subregion | Incidence of IUGR-LBW <br> (\% of total live births per year) | Attributable fraction of neonatal <br> deaths due to IUGR-LBW (\%) |
| :--- | :---: | :---: |
| AFR-D | 10.5 | 34.4 |
| AFR-E | 9.3 | 31.7 |
| AMR-A | 2.5 | 11.1 |
| AMR-B | 6.7 | 25.1 |
| AMR-D | 6.6 | 24.8 |
| EMR-B | 4.1 | 17.0 |
| EMR-D | 12.1 | 37.7 |
| EUR-A | 2.5 | 11.1 |
| EUR-B | 2.6 | 11.5 |
| EUR-C | 2.5 | 11.1 |
| SEAR-B | 4.2 | 17.4 |
| SEAR-D | 22.7 | 53.2 |
| WPR-A | 2.3 | 10.3 |
| WPR-B | 2.9 | 12.7 |

Table 2.28 Attributable fraction of neonatal mortality due to low maternal BMI, by subregion

| Subregion | Attributable fraction of IUGR due to $\mathrm{BMI} \leq 20 \mathrm{~kg} / \mathrm{m}^{2}$ (\%) |  | Attributable fraction of neonatal deaths due to IUGR (\%) | Attributable fraction of neonatal deaths due to low BMI (\%) |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | 15-29 years | 30-44 years |  | 15-29 years | 30-44 years |
| AFR-D | 25.7 | 19.1 | 34.4 | 8.9 | 6.6 |
| AFR-E | 23.2 | 18.0 | 31.7 | 7.4 | 5.7 |
| AMR-A | 16.7 | 13.5 | 11.1 | 1.9 | 1.5 |
| AMR-B | 15.4 | 8.2 | 25.1 | 3.9 | 2.1 |
| AMR-D | 8.6 | 6.9 | 24.8 | 2.1 | 1.7 |
| EMR-B | 18.5 | 9.2 | 17.0 | 3.2 | 1.6 |
| EMR-D | 24.3 | 20.7 | 37.7 | 9.2 | 7.8 |
| EUR-A | 14.3 | 9.9 | 11.1 | 1.6 | 1.1 |
| EUR-B | 16.4 | 10.1 | 11.5 | 1.9 | 1.2 |
| EUR-C | 16.4 | 7.2 | 11.1 | 1.8 | 0.8 |
| SEAR-B | 26.2 | 8.8 | 17.4 | 4.6 | 1.5 |
| SEAR-D | 31.2 | 25.2 | 53.2 | 16.6 | 13.4 |
| WPR-A | 24.7 | 17.8 | 10.3 | 2.5 | 1.8 |
| WPR-B | 20.7 | 16.6 | 12.7 | 2.6 | 2.1 |

age $<75 \%$ and those above $75 \%$. Black et al. (1984) found that incidence of Shigella diarrhoea, specifically, did not vary significantly according to weight-for-age.

The incidence of persistent diarrhoea, defined as an episode lasting 14 or more days, was greater among underweight children in communitybased studies in Brazil (Guerrant et al. 1992; Schorling et al. 1990), India (Bhandari et al. 1989) and Bangladesh (Baqui 1990). Children aged <5 years having weight-for-age at or below $75 \%$ in the Brazil cohort study were at a relative risk of 1.59 for persistent diarrhoea. In Bangladesh, the relative risk for persistent diarrhoea among children having $z$-score $<-2$ was 1.29 . Combining unadjusted figures from the two cohort studies yielded an overall relative risk of 1.35 ( $95 \%$ CI $0.97-1.90$ ). In a case-control study in India, pre-morbid weight-for-age below $70 \%$ was significantly more common among persistent diarrhoea cases compared to healthy age-matched controls, producing an OR of 3.25 (95\% CI 1.46-7.29). A community-based cohort study in Guinea-Bissau (Molbak et al. 1997) assessed risk according to stature and observed a statistically non-significant relative risk of 1.17 ( $95 \%$ CI $0.68-2.01$ ) associated with height-for-age $z$-score $<-2$.

## Other pregnancy outcomes

There is extensive literature supporting the influence of maternal undernutrition on fetal growth and risk of low birth weight as described in previous sections, but the relationship between low pre-pregnancy BMI and other adverse reproductive outcomes such as fetal death and maternal mortality is less clear. Research on maternal undernutrition as a risk factor for congenital anomalies, other than those attributable to micronutrient deficiencies (e.g. folic acid) is lacking.

## Obstructed labour and maternal mortality

Few studies from developing countries have been published relating low BMI to maternal mortality, although studies have described an increased risk of obstructed or prolonged labour associated with maternal height of 160 cm and below (Adadevoh et al. 1989; Anonymous 1984; Bhatt et al. 1967; Essex and Everett 1977; Konje and Ladipo 2000; Kwawukume et al. 1993; Mati 1983; Sokal et al. 1991; Tsu 1992), and one study in India reported an increased risk of caesarean section associated with low weight-height product index (weight times height over weight times height of reference population median) (Thilothammal et al. 1992). This risk may result from cephalopelvic disproportion (CPD) between mother and fetus (Rush 2000), increasing the likelihood of intrapartum caesarean delivery (Adadevoh et al. 1989; Merchant et al. 2001; Sokal et al. 1991; Tsu 1992) and perinatal distress (Merchant et al. 2001). In Nigeria, primigravida women below 150 cm were at a relative risk of 10.34 for CPD relative to women 160 cm or taller (Harrison et al. 1985). In Malawi, the odds ratio for CPD associated

Table 2.29 Study of early fetal death (<28 weeks)

| Study | Location | Study type | Comparison | Crude risk estimate ( $95 \% \mathrm{Cl}$ ) | Adjusted risk estimate (95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Agarwal et al. (1998) | India | Communitybased cohort | Weight $<42.5 \mathrm{~kg}$ vs $>42.5 \mathrm{~kg}$ | $\begin{aligned} & \mathrm{RR}=1.21 \\ & (1.05-1.38) \end{aligned}$ | - |
| Same as above | Same as above | Same as above | Height $<147.5 \mathrm{~cm}$ vs $>147.5 \mathrm{~cm}$ | $\begin{aligned} & \mathrm{RR}=1.21 \\ & (1.06-1.38) \end{aligned}$ | - |

- No data.
with height at or below 154 cm was 3.8 , adjusted for birth weight and parity (Brabin et al. 2002).

The 1995 WHO meta-analysis examined assisted delivery, defined as non-spontaneous delivery covering a range of complications (but not prolonged labour), and estimated an odds ratio between 1.0 and 2.1 (1.6, overall) associated with maternal height below the 25 th percentile (WHO 1995b). The odds ratio for assisted delivery according to prepregnancy BMI was 0.7 , explained as the result of constrained fetal growth reducing the risk of CPD; however, numerous studies have found that an association between maternal anthropometric status and assisted delivery remains after adjusting for multiple parameters of fetal size (Witter et al. 1995).

## Fetal death

Fetal death may be more common among underweight mothers (Tables 2.29 and 2.30). A community-based cohort study in India reported significant risks of spontaneous abortion (defined in the study as death <28 weeks' gestation) and late fetal death (>28 weeks) associated with both short stature and low absolute weight (Agarwal et al. 1998). A case-control study in the Sudan reported an adjusted odds ratio for stillbirth of 2.3 associated with absolute weight below 50 kg (Taha et al. 1994). On the other hand, Conde-Agudelo et al. (2000) analysed over 800000 births recorded in the Perinatal Information System database, a system used by over 700 hospitals throughout Latin American and the Caribbean, and found no significant risk of fetal death associated with BMI $<19.8 \mathrm{~kg} / \mathrm{m}^{2}$ relative to BMI between 19.8 and $26.0 \mathrm{~kg} / \mathrm{m}^{2}$.

## Post-neonatal consequences of IUGR

Growth-retarded neonates may experience partial catch-up growth in the first two years of life relative to non-IUGR controls, but they remain shorter and lighter than controls and are more likely to be classified as underweight (Martorell et al. 1998). They are at increased risk of infection in infancy as a result of multiple immunological abnormalities, including reduced T - and B -lymphocyte numbers and activity, lower

Table 2.30 Studies of stillbirth

| Study | Location | Study type | Comparison | Crude risk estimate (95\% CI) | Adjusted risk estimate (95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Agarwal et al. (1998) | India | Communitybased cohort | $\begin{aligned} & \text { Weight }<45 \mathrm{~kg} \\ & \text { vs }>45 \mathrm{~kg} \end{aligned}$ | $\begin{aligned} & \mathrm{RR}=5.07 \\ & (2.94-8.73) \end{aligned}$ | - |
| Same as above | Same as above | Same as above | $\begin{aligned} & \text { Height } \\ & <147.5 \mathrm{~cm} \\ & \text { vs }>147.5 \mathrm{~cm} \end{aligned}$ | $\begin{aligned} & \mathrm{RR}=4.18 \\ & (2.88-6.05) \end{aligned}$ | - |
| Taha et al. (1994) | Sudan | Facility-based case-control | Weight $<50 \mathrm{~kg}$ vs $\geq 70 \mathrm{~kg}$ | $\begin{aligned} & \mathrm{OR}=2.0 \\ & (1.0-4.1) \end{aligned}$ | $\begin{gathered} \mathrm{OR}=2.3 \\ (\mathrm{I} . \mathrm{I}-4.8)^{\mathrm{a}} \end{gathered}$ |
| Conde-Agudelo et al. (2000) | Latin <br> America, Caribbean | Facility-based cohort | $\begin{aligned} & \mathrm{BMI}<19.8 \mathrm{~kg} / \mathrm{m}^{2} \text { vs } \\ & 19.8-26.0 \mathrm{~kg} / \mathrm{m}^{2} \end{aligned}$ | - | $\begin{gathered} \mathrm{RR}=0.98 \\ (0.88-1.08)^{\mathrm{b}} \end{gathered}$ |

- No data.
a OR adjusted for prior fetal loss, number of antenatal care visits and malaria in early pregnancy. Several socioeconomic status indicators were assessed, but none was significantly associated with stillbirth and therefore were not included in the final regression model.
b Relative risk adjusted for antenatal care, age, literacy, cigarette smoking and prior fetal loss. Study addressed fetal death, defined as death $>20$ weeks' gestation.
levels of IgG and impaired bactericidal polymorphonuclear neutrophil function (Ferro-Luzzi et al. 1998; Xanthou 1985). Depending on the severity of the growth retardation, this immune deficiency may persist into later childhood (Chandra 1977). In addition to higher neonatal mortality rates, IUGR infants are at increased risk of mortality in the postneonatal period, including an increased risk of sudden infant death syndrome (Ferro-Luzzi et al. 1998).

IUGR may increase the risk of neurological dysfunction and mild cognitive impairment. IUGR infants show higher rates of hyperactivity, attention deficit and impaired motor coordination, depending on the degree of growth retardation, but the association of IUGR with low socioeconomic status and hypoxia makes it difficult to interpret the impact of IUGR (Goldenberg et al. 1998). Similarly, IUGR children have manifested small, but statistically significant IQ and developmental deficits, aggravated by impoverished environmental surroundings and poor psychosocial stimulation, but few studies have been conducted in developing countries, and socioeconomic factors may be confounding the relationship (Grantham-McGregor et al. 1998).

IUGR has been associated with increased susceptibility to chronic diseases in later life. The "fetal origins of disease" hypothesis posits that fetal undernutrition causes permanent structural and metabolic changes that potentiate subsequent risk of cardiovascular and endocrine disease (Barker 1995). IUGR infants have demonstrated insulin resistance and
higher blood pressure in childhood, and increased rates of ischaemic heart disease and non-insulin-dependent diabetes mellitus have been observed among adults who were born growth-retarded (Barker 1995; Chatelain et al. 1998; Law et al. 2001; Stein et al. 1996). Studies are inconsistent and causality is uncertain, but the possible risk relationship has important health implications for developing countries as they undergo epidemiological transition.

## Cognitive function

Cognitive function has been studied in terms of global measures of development and intelligence such as IQ, along with school performance and more narrowly defined intellectual, psychomotor, and behavioural skills such as attention, memory, verbal reasoning, motivation, visual-spatial abilities and social interaction. An array of psychometric tests and scales have been used to evaluate these functions, including modified versions of Stanford-Binet, Weschler Intelligence Scale for Children (WISC), Bayley Scales of Infant Development, Griffiths Mental Development Scale, Goodenough Drawing Test, Raven's Progressive Matrices, Piagetian tests of conservation and the Bender Visual Motor Gestalt Test. A major portion of research into malnutrition and cognition has addressed whether early childhood presents a critical period of vulnerability during which severe undernutrition causes lasting cognitive deficits in later life, and if potential deficits are amenable to subsequent nutritional and psychosocial interventions.

The immediate response to acute malnutrition is irritability, lethargy and apathy (Grantham-McGregor 1984). During this stage, children demonstrate lower activity levels and reduced exploratory behaviour and developmental quotients tend to be extremely low (Grantham-McGregor 1984). Electroencephalograms performed on children with acute malnutrition show nonspecific abnormalities (e.g. diffuse slowing of background rhythm) that improve with recovery (Chopra and Sharma 1992). During recovery and rehabilitation, behaviour improves to normal or near-normal levels and developmental quotients generally improve, but the long-term developmental implications remain uncertain (GranthamMcGregor 1995).

Numerous short-term and long-term follow-up studies have reported lasting developmental deficits associated with an early history of marasmus and/or kwashiorkor. Persistent deficits have been described in shortterm memory (Nwuga 1977), visual-spatial perception (Champakam et al. 1968; Cravioto et al. 1971; Ghai 1975; Hoorweg and Stanfield 1976; Reyes et al. 1990; Stoch and Smythe 1976), motivation (Stoch and Smythe 1967), Piagetian conservation tasks (Galler and Ramsey 1987) and perceptual-motor function (Grantham-McGregor et al. 1997). Poor academic performance and behavioural learning disabilities have also been recognized (Galler et al. 1984, 1990; Richardson et al. 1973).

Studies that assessed general measures of cognitive function among school-age children have observed significantly lower overall developmental and intelligence scores associated with early malnutrition (Berkman et al. 2002; Bhat et al. 1973; Birch et al. 1971; Botha-Antoun et al. 1968; Cabak and Najdanvic 1965; Cravioto et al. 1971; Fisher 1972; Galler et al. 1983, 1987a, 1987b; Hertzig et al. 1972; Ivanovic et al. 2000; McLaren et al. 1973; Mehta et al. 1975; Mendez and Adair 1999; Parekh et al. 1974; Pek et al. 1967; Sigman et al. 1991; Srikantia and Sastri 1971; Stoch and Smythe 1967, 1976; Udani et al. 1976). IQ differentials among these studies were predominantly in the order of 8 to 18 points. Studies in South Africa (Evans et al. 1971, 1980), however, observed no significant IQ difference between early kwashiorkor cases and non-hospitalized siblings. In Jamaica, Richardson et al. (1978) observed that height-for-age or weight-for-height among severely malnourished infants had no significant effect on IQ scores at school age after adjusting for social and environmental factors.

Among most of the studies showing long-term deficits, it could not be distinguished whether the observed impairment was related to the severe episode of acute malnutrition or to underlying or subsequent chronic undernutrition. In Jamaica, Grantham-McGregor (1982) reported that initial height-for-age, but not weight-for-height or presence of oedema, significantly predicted developmental quotients one month after hospitalization, suggesting the presence of chronic malnutrition was more closely associated with cognitive function than acute malnutrition (Grantham-McGregor 1982; Grantham-McGregor et al. 1989a, 1989b). In a long-term longitudinal study by Walker et al. (2000) height and head circumference in the first 24 months of life were found to be more significantly predictive of IQ at age 11 years than were anthropometric measurements taken at or near the time of cognitive testing, even after controlling for age, sex and socioeconomic factors, suggesting chronic malnutrition at an early age could have enduring effects on intelligence despite subsequent improvements in growth (Grantham-McGregor et al. 2000). In Peru, school-age children who were severely stunted in the second year of life scored 10.0 points lower than those moderately stunted or not stunted in the second year of life, after adjusting for socioeconomic status and schooling (Berkman et al. 2002). In the Philippines, severe stunting at 2 years of age was significantly associated with cognitive deficits at age 8 years, adjusting for schooling, socioeconomic status, sex and other covariates, but the differences were not significant by age 11 years (Mendez and Adair 1999). In Jamaica, Richardson (1976), observed that the long-term cognitive effect of early acute malnutrition was more heavily determined by chronic undernutrition and social background factors such as caretaker capability, presence of electricity and appliances, and child's access to toys, radio or stories. For the children having adequate growth and an advantageous social back-
ground, malnutrition during infancy was associated with an average IQ only 2 points lower than those not malnourished during infancy; for those experiencing both an unfavourable background and inadequate growth, the difference in IQ between children with and without malnutrition during infancy was 9 points by age $6-10$ years (Richardson 1976).

A large number of correlational studies have examined the association between current anthropometric status and cognitive function. Low height-for-age was significantly associated with low IQ or poor school achievement in Brazil (Paine et al. 1992), China (Jamison 1977), Guatemala (Johnston et al. 1987), India (Agarwal et al. 1987), Nepal (Moock and Leslie 1986) and the Philippines (Florencio 1988), but not Chile (Colombo et al. 1988). Deficits in other developmental scores were observed among stunted children in Chile (Monckeberg 1972), Guatemala (Lasky et al. 1981), India (Agarwal et al. 1989), Jamaica (Powell and Grantham-McGregor 1985) and Nigeria (Ashem and Janes 1978). Low weight-for-age has been associated with delayed motor development (Agarwal et al. 1992; Groos 1991; Heywood et al. 1991; Sathy et al. 1991; Vazir et al. 1998), delayed language development (Agarwal et al. 1992; Vazir et al. 1998) and poor performance on conservation tasks (Agarwal et al. 1989). Lower scores on aggregate measures such as IQ have been observed among underweight children in Ethiopia (Aboud and Alemu 1995), India (Agarwal et al. 1992; Gupta et al. 1975; Kalra et al. 1980; Lahiri et al. 1994; Sathy et al. 1991; Singh and Sidhu 1987; Upadhyay et al. 1989), Indonesia (Pek 1967), Kenya (Sigman et al. 1989) and Thailand (Rajatasilpin et al. 1970), with evidence of progressively increasing cognitive impairment associated with decreasing weight-for-age (Agarwal et al. 1992; Kalra et al. 1980; Lahiri et al. 1994; Sathy et al. 1991; Singh et al. 1976; Upadhyay et al. 1989). The differences in mean IQ between normal and underweight children in these correlational studies ranged between 7 and 31 points.

Among the more recent of these studies to report IQ, Agarwal et al. (1992) found Indian children aged 36 months with weight-for-age below $70 \%$ had a mean IQ approximately 8.7 points lower than adequately nourished children with weight-for-age of $80 \%$ or above; the relative risk for IQ below 80 points associated with underweight status (weight-forage $<80 \%$ ) was 2.33 ( $95 \%$ CI 1.46-3.74). Upadhyay et al. (1989) reported a difference of 7.2 points between children below $75 \%$ and those above $90 \%$. Lahiri et al. (1994) observed a significant trend toward lower IQ scores with decreasing weight-for-age (chi-square $=6.78$, $P<0.05$ ) among children aged 3-6 years from low socioeconomic levels in rural India, but the effect was not statistically significant when adjusted for age and educational exposure.

Neuroanatomical changes observed in animal models of PEM offer a theoretical foundation for the possibility that severe early malnutrition, particularly during the period of rapid brain growth and myelination in the first two years of life, presents a permanent structural insult to
brain function, leading to irreversible intellectual impairment (Strupp and Levitsky 1995). PEM has also generated abnormalities in neurotransmitter activity and receptor number (Levitsky and Strupp 1995). However, it is widely held now that neurobiological changes, on their own, are not sufficient to explain the complicated, multifactorial relationship between undernutrition and cognition, particularly among mild-to-moderately undernourished cases (Pollitt 1987). Increasing attention has been paid to environment, social context and experiential factors as potential modifiers or confounders of the nutrition-cognition relationship. Lack of energy, impaired psychomotor function and poor social interaction may prevent a child from exploring her surroundings as fully as other children, leading to less stimulation and slower acquisition of skills (Grantham-McGregor et al. 1989a, 1989b). The poverty associated with malnutrition is an environment of scarce resources, overcrowding, poor sanitation, illiteracy, few adequate educational opportunities and high morbidity, lacking in constructive forms of cognitive stimulation. Households tend to have few toys or books, social contacts are limited, and parents may offer less care and attention because of poor health, low income, low intelligence and education levels or large family sizes (Grantham-McGregor 1995). Returning to such an environment following severe malnutrition in infancy, for example, may explain persisting cognitive deficits observed in follow-up studies of infant malnutrition. Saco-Pollitt et al. (1985) offered a cumulative deficit hypothesis that suggests cognitive deficits increase as children are continuously exposed to environments that fail to meet their physiological, emotional and educational requirements (Pollitt 1987). Experimental studies have utilized interventions based on some of these neurobiological and psychosocial concepts.

Intervention studies to prevent or remedy impairments have focused on two general types of treatment: nutritional supplements provided to very young children and, in some cases, pregnant women; and psychosocial stimulation with or without nutritional supplementation provided to very young children. Evidence from nutritional interventions among high-risk or undernourished children has suggested early ( $<2$ years of age) supplementary feeding improves their developmental scores, with some indication of long-term benefits. Short-term ( 90 days) and long-term (2-3 years) supplementation programmes in Colombia (Waber et al. 1981), Guatemala (Pollitt et al. 1993), Indonesia (Hussaini et al. 1991) and Jamaica (Grantham-McGregor et al. 1991) have improved motor development among infants. Concurrent improvements in other developmental outcomes have been less consistent, but Waber et al. (1981) in Colombia observed significantly higher scores in per-sonal-social, speech and language, and performance subscales in addition to locomotor and eye-hand coordination subscales among infants supplemented until 3 years of age. Long-term follow-up identified improvements in certain achievement-related abilities, but not basic cog-
nitive skills by 6 years of age (Gorman et al. 1995). In India, Elizabeth and Sathy (1997) reported larger gains in developmental quotient and higher overall IQ (+2.6 points) among malnourished infants by the end of a two-year nutritional management and supplementation intervention. In Guatemala, children exposed to prenatal and early postnatal supplementation demonstrated long-term cognitive benefits, performing significantly better at ages 13-19 years on general intelligence tests and achievement-related subtests of numeracy, general knowledge, reading and vocabulary, even after adjusting for socioeconomic factors and educational experience; some information-processing tasks showed improvement as well (Pollitt et al. 1995). In Indonesia, infants who began 90 days of supplementation before 18 weeks of age demonstrated a positive effect on memory skills at 8 years of age, but differences in other cognitive processes were not significant (Pollitt et al. 1997).

Intervention programmes based on early psychosocial stimulation have improved developmental scores among moderately to severely malnourished infants and preschool-age children. Various studies have employed educational day care programmes (McKay et al. 1978), clinicbased therapy and education (Elizabeth and Sathy 1997) and weekly home visits with mothers and infants (Grantham-McGregor et al. 1994; Waber et al. 1981). In Jamaica, demonstrating play techniques to mothers and severely malnourished infants at home over three years was associated with both immediate and long-term improvements in global intelligence and development scores relative to a non-intervened control group (Grantham-McGregor et al. 1994). At 14 years follow up, mean scores on WISC for the intervened group were 8.6 points higher than the non-intervened group, but both intervened and non-intervened groups still performed significantly worse than an adequately nourished, generally higher socioeconomic status control group (9.7 and 18.3 IQ points below, respectively). Similar results were observed in a facilitybased study in India among malnourished infants provided nutritional management with or without cognitive stimulation (Elizabeth and Sathy 1997). At the end of two years, both interventions were associated with significant improvements in developmental quotients, with the final IQ of stimulated infants 8.3 points higher than those receiving nutritional management alone, but both treatment groups ultimately performed below an adequately nourished/higher mean socioeconomic status control group by 5.4 to 13.7 IQ points. In a community-based component of the study, stimulated infants achieved IQ scores 6.5 points higher than an untreated, comparable control group and 3.8 points higher than infants receiving nutritional management alone; it is too early to know long-term effects in this case.

In Colombia, McKay et al. (1978) assigned underweight children aged 3 years to varying durations of an integrated health, nutrition and education programme and found that earlier initiation and longer duration of intervention were associated with higher general cognitive scores.

However, by school age, the intervened groups continued to score below an adequately nourished, higher socioeconomic status control group, and the absence of a comparable malnourished/low socioeconomic status control group made the absolute effect of the intervention unclear. Waber et al. (1981) evaluated cognitive development of nutritionally at-risk Colombian infants according to their exposure to prenatal and postnatal nutritional supplementation and a maternal-and-child educational programme. Nutritional supplementation was associated with higher scores on tests of primarily motor skills, and maternal-child education was associated with better language performance.

In a randomized controlled trial among stunted infants in Jamaica, Grantham-McGregor et al. $(1991,1997)$ assessed the immediate and long-term developmental effects of nutritional supplementation with or without psychosocial stimulation. By the end of the two-year intervention, supplementation and stimulation each had a significant benefit on developmental quotients (by 7 and 8 points, respectively), with the combination of treatments having an additive effect. Follow-up at ages 7-8 years (Grantham-McGregor et al. 1997) and 11-12 years (Walker et al. 2000), indicated persistent benefits in overall IQ, vocabulary and reasoning ability associated with the early psychosocial stimulation, but less of an effect from nutritional supplementation. Supplementation alone was associated with a benefit of 4.2 IQ points (not statistically significant), stimulation alone was associated with 6.3 points, and both treatments combined were associated with 6.1 points relative to a control group of untreated, stunted children; all three groups performed below a non-stunted, untreated control group.

Unfortunately, few studies (and no individually randomized controlled trials) assessed the cognitive benefit of sustained nutritional supplementation among school-age children in developing countries. In Guatemala, a year of supplementation had no consistent effect on children aged 5-7 years (Pollitt et al. 1993). A two-year school-based supplementation programme in India observed marginal but significant effects on IQ scores and Piagetian tasks, but treatment-related differences in school attendance may have influenced the results (Agarwal et al. 1989). Several studies have examined the short-term effects of breakfast programmes (Chandler et al. 1995) or breakfast omission (Lopez et al. 1993; Simeon and Grantham-McGregor 1989) on cognitive performance among schoolchildren. In Jamaica, controlled breakfast omission was associated with impaired performances in short-term memory, idea generation and problem-solving ability among stunted or previously malnourished children, but not among adequately nourished children (Simeon and Grantham-McGregor 1989). A subsequent randomized controlled trial found that only the undernourished children responded to breakfast supplementation, showing improvement in idea generation capacity but not memory or problem-solving skills (Chandler et al. 1995). In Chile, controlled breakfast omission had no significant effect on short-term visual
memory, problem-solving capacity, or attention, regardless of preexisting nutritional status (Lopez et al. 1993).

As with other health outcomes, the close relationship between poverty and malnutrition raises the possibility of confounding. While many of the correlational and matched case-control studies attempted to control or adjust for potential socioeconomic and environmental differences, those variables were frequently defined in broad terms and may have overlooked subtler, as-yet-unidentified, non-nutritional determinants of cognitive function. Educational status could have played a role in some findings, as well; for example, initial school enrolment may be delayed among stunted children because they appear or act younger than their chronological age (Brown and Pollitt 1996). Undernutrition could influence cognition indirectly through higher morbidity; illness could reduce activity and social interaction, or contribute to higher school absentee and drop-out rates (Mendez and Adair 1999; Neumann et al. 1991; Pollitt 1983).

Another important consideration in estimating the effect of low weight-for-age on cognitive function is possible confounding by deficiencies of iron and other micronutrients (Pollitt 1995). Iron deficiency has been associated with developmental delays (Lozoff et al. 1991; Soewondo et al. 1989; Walter 1993) and iron supplementation studies have demonstrated improvements in cognitive function among irondeficient subjects (Pollitt 1995; Pollitt and Metallinos-Katsaras 1990; Seshadri and Gopaldas 1989). Further, studies of iron deficiency and growth have observed significant covariance between the two (PeragalloGuarda 1984; Pollitt 1995), with evidence that iron supplementation improves both growth velocity (Chwang et al. 1988) and weight (Lawless et al. 1994) among anaemic children. Zinc deficiency may also play a role in cognitive ability (Golub et al. 1995). Several intervention studies described above (Elizabeth and Sathy 1997; Husaini et al. 1991; McKay et al. 1978; Waber et al. 1981) provided micronutrient-rich indigenous foods or supplements between experimental groups differentially and most observational studies did not control for other nutritional factors.

### 4.5 RISK REVERSIBILITY

Risk reversibility refers to the extent to which an increased relative risk persists among a previously exposed group after that exposure is removed. There has been some investigation of the lasting effect of fetal malnutrition on immune function of children and adults in Africa (Moore et al. 1999, 2001). However, there is insufficient evidence at this time to state that past underweight status during childhood increases the mortality or morbidity risk of malaria, measles, pneumonia or diarrhoea among those no longer underweight. Therefore, the increased risk of mortality and morbidity associated with low weight-for-age is considered completely reversible. Likewise, the risk of neonatal death attribut-
able to low maternal BMI is considered completely reversible once underweight status is corrected.

## 5. Estimates of ATtRibutable burden

Estimates of the attributable fraction, attributable mortality and attributable burden of underweight status in the age group 0-4 years were calculated for mortality due to measles, malaria, diarrhoea, pneumonia and perinatal causes based upon the above estimates of underweight prevalence, relative risk and reversibility. Attributable fractions and attributable burdens were also calculated for morbidity due to malaria, pneumonia and diarrhoea. Because we were unable to calculate relative risks of morbidity for the proportion of children falling in the weight-for-age category of -1 SD to -2 SDs , the mortality and morbidity estimates are based on different anthropometric thresholds. The mortality estimates represent the cause-specific deaths that are attributable to weight-for-age less than -1 SD, and the morbidity estimates refer to the cause-specific episodes of illness attributable to weight-for-age less than -2 SDs. The burden estimates for mortality from perinatal causes are based on maternal pre-pregnancy $\mathrm{BMI} \leq 20 \mathrm{~kg} / \mathrm{m}^{2}$ and infant deaths attributed to low birth weight. Results are listed in Tables 2.31 through 2.35

Table 2.3I Attributable burden of low weight-for-age on measles infection among children aged 0-4 years, by subregion

| Subregion | Mortality (WA <-I SD) |  |  | Incidence (WA <-2 SDs) |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Attributable fraction (\%) | Attributable mortality (000s) | Attributable burden DALYs (000s) | Attributable fraction (\%) | Attributable burden DALYs (000s) |
| AFR-D | 45.2 | 94.7 | 3302.1 | 0.0 | 0.0 |
| AFR-E | 44.2 | 64.2 | 2238.0 | 0.0 | 0.0 |
| AMR-A | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| AMR-B | 9.3 | 0.0 | 0.0 | 0.0 | 0.0 |
| AMR-D | 24.3 | 0.0 | 0.0 | 0.0 | 0.0 |
| EMR-B | 16.5 | 0.0 | 0.6 | 0.0 | 0.0 |
| EMR-D | 39.2 | 26.3 | 915.8 | 0.0 | 0.0 |
| EUR-A | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| EUR-B | 15.5 | 0.7 | 25.4 | 0.0 | 0.0 |
| EUR-C | 1.5 | 0.0 | 0.0 | 0.0 | 0.0 |
| SEAR-B | 39.9 | 9.2 | 320.4 | 0.0 | 0.0 |
| SEAR-D | 53.8 | 59.6 | 2073.9 | 0.0 | 0.0 |
| WPR-A | 5.7 | 0.0 | 0.1 | 0.0 | 0.0 |
| WPR-B | 29.3 | 6.5 | 226.0 | 0.0 | 0.0 |
| World | 44.8 | 261.3 | 9102.1 | 0.0 | 0.0 |

Table 2.32 Attributable burden of low weight-for-age on malaria infection among children aged 0-4 years, by subregion

| Subregion | Mortality (WA <-I SD) |  |  | Morbidity/incidence$(W A<-2 S D s)$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Attributable fraction (\%) | Attributable mortality (000s) | Attributable burden DALYs (000s) | Attributable fraction (\%) | Attributable burden DALYs (000s) |
| AFR-D | 57.7 | 251.5 | 8480.9 | 8.4 | 115.9 |
| AFR-E | 56.7 | 235.6 | 7982.9 | 8.1 | 115.6 |
| AMR-A | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| AMR-B | 13.8 | 0.1 | 1.7 | 0.8 | 0.1 |
| AMR-D | 33.7 | 0.0 | 0.9 | 3.0 | 0.1 |
| EMR-B | 23.7 | 0.0 | 0.0 | 1.8 | 0.1 |
| EMR-D | 51.3 | 23.3 | 787.3 | 6.6 | 7.4 |
| EUR-A | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| EUR-B | 22.3 | 0.0 | 0.0 | 1.6 | 0.0 |
| EUR-C | 2.3 | 0.0 | 0.0 | 0.1 | 0.0 |
| SEAR-B | 52.0 | 1.7 | 56.0 | 6.8 | 1.8 |
| SEAR-D | 66.4 | 35.9 | I 223.1 | 11.8 | 11.4 |
| WPR-A | 8.5 | 0.0 | 0.0 | 0.5 | 0.0 |
| WPR-B | 39.8 | 1.2 | 40.1 | 4.1 | 0.5 |
| World | 57.3 | 549.2 | 18572.7 | 8.2 | 253.0 |

Table 2.33 Attributable burden of low weight-for-age on pneumonia/ALRI among children aged $0-4$ years, by subregion

| Subregion | Mortality (WA <-I SD) |  |  | Morbidity/incidence$(W A<-2 S D s)$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Attributable fraction (\%) | Attributable mortality (000s) | Attributable burden DALYs (000s) | Attributable fraction (\%) | Attributable burden DALYs (000s) |
| AFR-D | 54.6 | 171.1 | 5729.9 | 20.1 | 38.2 |
| AFR-E | 53.6 | 200.9 | 6739.8 | 19.5 | 42.4 |
| AMR-A | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| AMR-B | 12.6 | 3.4 | 113.6 | 2.2 | 6.7 |
| AMR-D | 31.2 | 6.6 | 221.5 | 7.9 | 4.8 |
| EMR-B | 21.7 | 4.5 | 151.2 | 4.7 | 5.0 |
| EMR-D | 48.2 | 108.2 | 3661.4 | 16.1 | 51.6 |
| EUR-A | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| EUR-B | 20.4 | 9.8 | 328.2 | 4.3 | 1.9 |
| EUR-C | 2.1 | 0.2 | 5.2 | 0.3 | 0.0 |
| SEAR-B | 49.0 | 20.5 | 692.7 | 16.5 | 37.7 |
| SEAR-D | 63.4 | 432.0 | 14606.4 | 26.9 | 262.3 |
| WPR-A | 7.7 | 0.0 | 0.9 | 1.3 | 0.0 |
| WPR-B | 37.1 | 85.9 | 2884.4 | 10.4 | 77.8 |
| World | 52.3 | 1042.9 | 35135.0 | 16.5 | 528.3 |

Table 2.34 Attributable burden of low weight-for-age on diarrhoea infection among children aged 0-4 years, by subregion

| Subregion | Mortality (WA <-I SD) |  |  | Morbidity/incidence$(W A<-2 S D s)$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Attributable fraction (\%) | Attributable mortality (000s) | Attributable burden DALYs (000s) | Attributable fraction (\%) | Attributable burden DALYs (000s) |
| AFR-D | 62.7 | 118.0 | 3960.1 | 6.4 | 12.4 |
| AFR-E | 61.7 | 198.6 | 6698.3 | 6.2 | 14.0 |
| AMR-A | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| AMR-B | 16.1 | 4.2 | 140.0 | 0.6 | 0.9 |
| AMR-D | 38.1 | 7.3 | 247.1 | 2.3 | 0.7 |
| EMR-B | 27.1 | 4.0 | 133.8 | 1.3 | 0.7 |
| EMR-D | 56.3 | 138.1 | 4657.9 | 5.0 | 9.0 |
| EUR-A | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| EUR-B | 25.5 | 3.8 | 128.0 | 1.2 | 0.5 |
| EUR-C | 2.8 | 0.1 | 1.7 | 0.1 | 0.0 |
| SEAR-B | 57.1 | 15.2 | 507.2 | 5.1 | 4.8 |
| SEAR-D | 71.1 | 295.5 | 9975.4 | 9.1 | 47.6 |
| WPR-A | 10.0 | 0.0 | 0.1 | 0.3 | 0.0 |
| WPR-B | 44.5 | 31.3 | 1050.5 | 3.0 | 8.0 |
| World | 60.7 | 815.9 | 27500.1 | 5.3 | 98.6 |

Table 2.35 Attributable burden of low pre-pregnancy BMI on mortality due perinatal conditions, by subregion

|  | Mortality $\left(\mathrm{BMI} \leq 20 \mathrm{~kg} / \mathrm{m}^{2}\right)$ |  |  |
| :--- | :---: | :---: | :---: |
| Subregion | Attributable fraction (\%) | Attributable mortality <br> $(000 \mathrm{~s})$ | Attributable burden <br> DALYs (000s) |
| AFR-D | 8.0 | 10.1 | 358.5 |
| AFR-E | 6.7 | 9.0 | 318.2 |
| AMR-A | 1.7 | 0.1 | 4.8 |
| AMR-B | 3.2 | 2.5 | 87.6 |
| AMR-D | 1.9 | 0.3 | 10.0 |
| EMR-B | 2.6 | 0.3 | 11.9 |
| EMR-D | 8.7 | 17.5 | 637.8 |
| EUR-A | 1.4 | 0.1 | 2.7 |
| EUR-B | 1.6 | 0.3 | 10.6 |
| EUR-C | 1.5 | 0.1 | 3.3 |
| SEAR-B | 3.5 | 1.3 | 46.2 |
| SEAR-D | 15.5 | 104.5 | 3792.0 |
| WPR-A | 2.5 | 0.0 | 0.9 |
| WPR-B | 2.4 | 2.4 | 85.8 |
| World | 10.5 | 148.4 | 5370.1 |

Table 2.36 Total burden of underweight status among children aged $0-4$ years

| Disease | Mortality (WA <-I SD) |  |  | Morbidity/incidence$(W A<-2 S D s)$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Attributable fraction (\%) | Attributable mortality (000s) | Attributable burden DALYs (000s) | Attributable fraction (\%) | Attributable burden DALYs (000s) |
| Protein-energy malnutrition | 100.0 | 153.6 | 5252.7 | 100 | 9632.5 |
| Perinatal conditions ${ }^{\text {a }}$ | 10.5 | 148.4 | 4967.0 | 0.0 | 0.0 |
| Pneumonia/ ALRI | 52.3 | 1042.9 | 35135.0 | 16.5 | 528.3 |
| Diarrhoea | 60.7 | 815.9 | 27500.1 | 5.3 | 98.6 |
| Malaria | 57.3 | 549.2 | 18572.7 | 8.2 | 253.0 |
| Measles | 44.8 | 261.3 | 9102.1 | 0.0 | 0.0 |
| Other | 53.1 | 776.9 | 26355.8 | - | - |
| Total | 34.7 | 3748.2 | 126885.4 | - | 10512.2 |
| No data. |  |  |  |  |  |
| "Perinatal conditions" include low birth weight, birth asphyxia and trauma, neonatal sepsis, maternal and placental complications, respiratory distress, fetal blood loss, fetal haematological disorders, anaemia, perinatal infections, maternal diabetes and various other conditions. The category does not include congenital anomalies, neonatal tetanus or syphilis. The burden estimates represent infant deaths due to low birth weight only. |  |  |  |  |  |

for both sexes combined according to outcome and subregion. Table 2.36 summarizes the total burden estimates associated with underweight status, including estimates of mortality directly attributable to PEM and "other" indeterminate deaths in the residual category. The total attributable fraction of mortality among children aged $<5$ years due to maternal and child undernutrition was $34.7 \%$ (Table 2.36). This was calculated from the sum of the cause-specific attributable mortality estimates divided by the estimate of total deaths among children aged 0-59 months. This fraction includes deaths directly and indirectly attributable to undernutrition. In determining the fraction of deaths indirectly caused by undernutrition, the analysis differentiated postneonatal from neonatal deaths, with maternal undernutrition considered as the risk factor for neonatal deaths.

## 6. Discussion

The results of these analyses indicate that undernutrition in young children contributes significantly toward the global burden of disease. We
estimate that undernutrition in children aged <5 years, as reflected in underweight or low weight-for-age, causes 3599800 deaths, including 815900 diarrhoea deaths, 1042900 pneumonia deaths, 261300 measles deaths and 549200 malaria deaths. The estimates also illustrate that the burden of undernutrition in children begins in utero, with approximately 148400 neonatal deaths from low birth weight attributable to maternal underweight status. The 3748200 total deaths attributable to maternal and child undernutrition constitute $34.7 \%$ of all deaths among children aged $<5$ years. The loss in DALYs associated with these deaths is stag-gering-over 126 million. Among the principal causes of death in young children, $60.7 \%$ of deaths due to diarrhoea, $52.3 \%$ of deaths due to pneumonia, $44.8 \%$ of deaths due to measles and $57.3 \%$ of deaths due to malaria are attributable to undernutrition. These attributable fractions are large because undernutrition substantially increases a child's risk of dying from common childhood illnesses, and because undernutrition is still highly prevalent in many regions of the world. In particular, the subregions SEAR-D, AFR-D and AFR-E respectively account for $32 \%, 22 \%$ and $24 \%$ of the total DALYs lost from undernutrition, followed by EMR-D (12\%) and WPR-B (6\%).

Undernutrition contributes to the morbidity burden among children as well. Our analyses indicate that having a weight-for-age less than -2 SDs places a child at increased risk of developing pneumonia, diarrhoea or malaria. We estimate that $16.5 \%$ of pneumonia, $5.3 \%$ of diarrhoea illness and $8.2 \%$ of malarial attacks are attributable to low weight-for-age, with an associated loss of DALYs amounting to 528300 for pneumonia, 98600 for diarrhoea and 253000 for malaria. Available evidence suggests that undernutrition does not influence a child's risk of contracting measles and thus none of the measles morbidity burden is attributed to this risk factor.

The relatively greater burden of death as compared to illness attributable to undernutrition is understandable, given the known synergy between illness and malnutrition. This synergy was first described by Scrimshaw et al. (1968): "The simultaneous presence of malnutrition and infection results in an interaction that is more serious for the host than would be expected from the combined effect of the two working independently". Undernutrition potentiates the risk of mortality by increasing the likelihood that the illness will be prolonged or become severe; and more prolonged or severe illness is more likely to negatively affect the nutritional status of the child, placing her at ever-increasing risk of future and more prolonged or severe illness episodes. This has been most clearly demonstrated in the literature for diarrhoeal illnesses (Black et al. 1984), and our analysis makes the extension to three other principal causes of death among young children: pneumonia, malaria and measles. Such findings underscore the need to prioritize the improvement of the nutritional status of children, including within disease control programmes (Becker et al. 1991).

In attempting to quantify the relationship between underweight status and disease, it is important to consider that undernutrition, as assessed by the underweight indicator (or a stunting indicator), is commonly associated with deficiencies of micronutrients (Bhan et al. 2001). In fact, some of these deficiencies, such as zinc, may themselves contribute to poor growth (Brown 2002). These deficiency conditions, especially of vitamin A, iron and zinc, put the child at risk of adverse outcomes, including morbidity and mortality from infectious disease and cognitive impairment. The consequences of these deficiency states can be estimated separately, as has been done in other chapters in this book. The association of underweight condition with these deficiencies means that some of the risk attributed to being underweight may be, in fact, due to specific micronutrient deficiencies. On the other hand, these other deficiencies can occur in children who are not considered underweight, so the risk of adverse outcomes is not entirely encompassed in the subset of the population who are underweight. In fact, the effect of vitamin A supplementation on reducing mortality in vitamin A-deficient populations has been found to be similar in children who were more or less well nourished at the beginning of the trial (Sommer 1986; West 1991). Likewise, zinc supplementation in populations that are presumably zinc deficient has had similar effects on reducing infectious disease morbidity in children who were classified as both "wasted" or "not wasted" (Zinc Investigators' Collaborative Group 1999). Therefore, the effects of undernutrition as identified by being underweight and those of deficiencies of vitamin A, iron and zinc cannot be simply added to determine the overall burden of disease due to nutritional risk factors, nor is it appropriate to assume that all of the consequences of micronutrient deficiencies are subsumed in the underweight calculation. Additional work is needed to determine the joint prevalence distributions of being underweight and being deficient in each of these micronutrients and to assess the joint effects of these aspects of undernutrition.

The outcomes described here represent some of the fundamental global health challenges facing populations at greatest risk for undernutrition. However, the burden of undernutrition on human health and well-being no doubt extends beyond the relatively narrow focus of this analysis. We presented findings to suggest a role for maternal malnutrition as a risk factor for maternal mortality, as well as a role for undernutrition in disability due to poor cognitive development, but the research base thus far does not allow for the calculation of burden estimates. Other considerations relating to undernutrition may include the short-term or long-term effects of underweight status on work capacity or the effect of IUGR on childhood development and morbidity. Many areas are still largely unknown or speculative, such as the role of early or ongoing undernutrition in potentiating chronic disease (Barker 1995;

Law et al. 2001; Stein et al. 1996) or the effect of undernutrition on susceptibility to infection among adult and elderly populations-questions that will only gain in importance as developing countries continue to go through demographic and epidemiological transitions.

Several reports suggest that through current programmatic efforts, rates of undernutrition among children are declining at $1 \%$ per year. In contrast, our projections indicate that over the coming decades we can expect variable changes in these prevalences. Because our analyses clearly indicate the overwhelming magnitude of the disease burden associated with child malnutrition, effective strategies to reduce child undernutrition are urgently needed. Energy supplementation for pregnant women, counselling and promotion of breastfeeding and adequate complementary food intake, and child growth monitoring are some of the more effective and affordable interventions for preventing low birth weight and improving child growth (ACC/SCN 2001; WHO 2002). Micronutrient supplementation or fortification, the use of oral rehydration therapy and childhood immunizations, have also been key elements to improving nutritional status and preventing severe childhood disease. Further innovations will be necessary in order to meet the challenge of undernutrition in all populations.

## 7. Projections of exposure: TRENDS IN CHILD UNDERWEIGHT STATUS FROM 2000 TO 2030

The estimates for 2010 and 2020 are directly derived from the model as outlined in section 2.2 , based on estimates of underweight prevalence in individual countries at multiple points in time. To extend these model estimates to 2030, we applied the lowest delta between last available trend intervals, i.e. for AFR-D and AFR-E, we used 2010-2015 and for the other subregions the deltas 2015-2020; the respective difference was multiplied by 2 (as they refer to a five-year interval) and then subtracted from or added to the estimate of 2020 to derive the value for 2030 (Table 2.37).

Table 2.37 Estimated prevalence of underweight among children aged $0-4$ years from 2000 to 2030, by subregion

|  | \% having weight-for-age below -2 SDs |  |  |  |
| :--- | ---: | :---: | :---: | :---: |
| Subregion | 2000 | 2010 | 2020 | 2030 |
| AFR-D | 32.2 | 34.2 | 36.2 | 38.2 |
| AFR-E | 31.0 | 37.3 | 44.1 | 50.7 |
| AMR-A | 2.3 | 1.5 | 0.7 | $0.0^{\mathrm{a}}$ |
| AMR-B | 5.0 | 3.3 | 2.2 | 1.1 |
| AMR-D | 12.4 | 9.5 | 7.2 | 5.0 |
| EMR-B | 8.1 | 4.3 | 2.3 | 0.5 |
| EMR-D | 25.1 | 20.4 | 16.4 | 12.6 |
| EUR-A | 2.3 | 1.5 | 0.7 | $0.0^{\mathrm{a}}$ |
| EUR-B | 7.6 |  | Overall stagnation |  |
| EUR-C | 2.6 | 18.7 | Overall declining trend |  |
| SEAR-B | 25.8 | 37.6 | 3.2 | 8.0 |
| SEAR-D | 45.9 | 2.5 | 1.2 | 22.6 |
| WPR-A | 3.8 | 16.8 | 8.6 | $0.0^{\mathrm{b}}$ |
| WPR-B |  |  | 5.6 |  |

[^6](i) EUR-B

The available data for this subregion show a diverse pattern. There are countries with decreasing rates in child underweight and parallel there are others where increase in the national prevalence of child underweight can be observed. With the information available to date it is very difficult to estimate a trend in child underweight from 2000 to 2030 . The trend pattern is expected to be driven by Turkey, which has $40 \%$ of the total $<5$ population in this subregion. However, given the available trend data for other countries in this subregion and the unfavourable conditions (demographic and economic transition associated in parts with political instability) in some of them an overall stagnation of the underweight prevalence from 2000 to 2030 is to be expected. Appendix $\mathrm{E}(\mathrm{a})$ lists references by country in alphabetical order to document the forecasted trend in this subregion.
(ii) EUR-C

Given that any trend in this subregion is likely to be dominated by Russia (which has $56 \%$ of the total population of children aged $<5$ years in this subregion) and that current evidence shows raising rates in overweight in this country we would estimate an overall declining trend of underweight. However, given the enduring unstable situation in some of the countries in EUR-C and the uncertainty about support and investments from developed countries, a quantification of the expected decline cannot be made. Appendix $\mathrm{E}(\mathrm{b})$ lists references by country in alphabetical order derived from the WHO Global Database, Medline and the Internet which contain additional information to document the forecasted trend in this subregion.

## Acknowledgements

We thank Majid Ezzati for guidance in global burden of disease comparative risk assessment methodology; Steve Vander Hoorn for comparative risk assessment analysis and burden of disease calculations; Amy Rice, Lisa Sacco and Constantine Frangakis for their analysis of malnutrition and perinatal causes of death; Andres Lescano for additional statistical analysis of morbidity data; Jeffrey McGuckin for his assistance in coordinating comparative risk assessment efforts; and Julie West and Seema Rai for their assistance with the literature search and compilation of risk data.

We are grateful to Wafaie Fawzi, Michel Garenne, Marc Andersen, Keith West, Shams el Arifeen, Fehmida Jalil, S.R. Khan, Bambang Sutrisna, and Stan Becker, for providing original data on cause-specific mortality; and K. Zaman, for providing data on pneumonia morbidity.

We are grateful for the statistical support received from Allen Shoemaker for conducting primary data analysis of raw data sets relating to underweight prevalence; Richard Morris for the modelling of the subregional estimates; and Gilda Piaggio for transforming the categorical underweight prevalences into continuous categories.

This manuscript was supported in part by the Family Health and Child Survival (FHACS) Cooperative Agreement between the United States Agency for International Development, Office of Health and Nutrition and the Johns Hopkins Bloomberg School of Public Health, Department of International Health; and by the World Health Organization, Evidence and Information for Policy Department, Department of Child and Adolescent Health and Development, and Department of Nutrition for Health and Development.

## Notes

1 See preface for an explanation of this term.

## References

Aaby P, Bukh J, Lisse IM, da Silva MC (1988) Decline in measles mortality: nutrition, age at infection, or exposure? British Medical Journal, 296:1225-1228.

Aboud FE, Alemu T (1995) Nutrition, maternal responsiveness and mental development of Ethiopian children. Social Science and Medicine, 41:725-732.
ACC/SCN (2000a) Ending malnutrition by 2020: An agenda for change in the millennium. Food and Nutrition Bulletin, 21:S3-88.
ACC/SCN (2000b) Fourth report on the world nutrition situation. United Nations Administrative Committee on Coordination/Sub-Committee on Nutrition, Geneva.

ACC/SCN (2001) What works? A review of the efficacy and effectiveness of nutrition interventions. Allen LH, Gillespie SR, eds. United Nations Administrative Committee on Coordination/Sub-Committee on Nutrition, Geneva in collaboration with the Asian Development Bank, Manila.
Adadevoh SW, Hobbs C, Elkins TE (1989) The relation of the true conjugate to maternal height and obstetric performance in Ghanaians. International Journal of Gynaecology and Obstetrics, 28:243-251.
Adelekan DA, Adeodu OO, Thurnham DI (1997) Comparative effects of malaria and malnutrition on plasma concentrations of antioxidant micronutrients in children. Annals of Tropical Paediatrics, 17:223-227.
Agarwal DK, Agarwal A, Singh M, Satya K, Agarwal S, Agarwal KN (1998) Pregnancy wastage in rural Varanasi: relationship with maternal nutrition and sociodemographic characteristics. Indian Pediatrics, 35:1071-1079.
Agarwal DK, Agarwal KN, Upadhyay SK (1989) Effect of mid-day meal programme on physical growth and mental function. Indian Journal of Medical Research, 90:163-174.
Agarwal DK, Awasthy A, Upadhyay SK, Singh P, Kumar J, Agarwal KN (1992) Growth, behavior, development and intelligence in rural children between 1-3 years of life. Indian Pediatrics, 29:467-480.
Agarwal DK, Upadhyay SK, Agarwal KN (1989) Influence of malnutrition on cognitive development assessed by Piagetian tasks. Acta Paediatrica Scandinavica, 78:115-122.

Agarwal DK, Upadhyay SK, Tripathi AM, Agarwal KN (1987) Nutritional status, physical work capacity and mental function in school children. (Scientific Report No. 6.) Nutritional Foundation of India, New Delhi.

Alam AN, Chowdhury AA, Kabir IA, Sarker SA, Rahaman MM (1984) Association of pneumonia with undernutrition and shigellosis. Indian Pediatrics, 21:609-613.
Allen SJ, Raiko A, O’Donnell A, Alexander ND, Clegg JB (1998) Causes of preterm delivery and intrauterine growth retardation in a malaria endemic region of Papua New Guinea. Archives of Disease in Childhood, Fetal and Neonatal Edition, 79:F135-140.

Anand K, Sundaram KR, Lobo J, Kapoor SK (1994) Are diarrheal incidence and malnutrition related in under five children? A longitudinal study in an area of poor sanitary conditions. Indian Pediatrics, 31:943-948.

Andersen M (1997) Anthropometric measurements in health programmes: Epidemiological and statistical aspects [Dissertation]. University of Copenhagen, Copenhagen.
Anonymous (1984) Antenatal screening for fetopelvic dystocias. A costeffectiveness approach to the choice of simple indicators for use by auxiliary personnel. The Kasongo Project Team. Journal of Tropical Medicine and Hygiene, 87:173-183.

Anonymous (1999) Reducing perinatal and neonatal mortality. (Child Health Research Project Special Report. Vol. 3, No. 1.) Johns Hopkins University, Baltimore, MD.

Arifeen S, Black RE, Antelman G, Baqui A, Caulfield L, Becker S (2001) Exclusive breastfeeding reduces acute respiratory infection. Pediatrics, 108:E67.

Armitage P, Berry G (1994) Statistical methods in medical research. 3rd edn. Blackwell Science, Oxford.

Arora NK, Paul VK, Singh M (1987) Morbidity and mortality in term infants with intrauterine growth retardation. Journal of Tropical Pediatrics, 33: 186-189.
Ashem B, Janes MD (1978) Deleterious effects of chronic undernutrition on cognitive abilities. Journal of Child Psychology and Psychiatry, 19:23-31.
Ashworth A (1998) Effects of intrauterine growth retardation on mortality and morbidity in infants and young children. European Journal of Clinical Nutrition, 52:S34-41.

Bairagi R, Chowdhury MK, Kim YJ, Curlin GT, Gray RH (1987) The association between malnutrition and diarrhoea in rural Bangladesh. International Journal of Epidemiology, 16:477-481.
Bakketeig LS (1998) Current growth standards, definitions, diagnosis and classification of fetal growth retardation. European Journal of Clinical Nutrition, 52:S1-4.

Bakketeig LS, Butte N, de Onis M, et al. (1998) Report of the IDECG Working Group on definitions, classifications, causes, mechanisms and prevention of IUGR. European Journal of Clinical Nutrition, 52:S94-96.
Balcazar H, Haas J (1990) Classification schemes of small-for-gestational age and type of intrauterine growth retardation and its implications to early neonatal mortality. Early Human Development, 24:219-230.

Balcazar H, Haas JD (1991) Retarded fetal growth patterns and early neonatal mortality in a Mexico City population. Bulletin of the Pan American Health Organization, 25:55-63.

Ballard TJ, Neumann CG (1995) The effects of malnutrition, parental literacy and household crowding on acute lower respiratory infections in young Kenyan children. Journal of Tropical Pediatrics, 41:8-13.
Baqui AH (1990) Epidemiology of persistent diarrhea in Bangladeshi children. [Dissertation]. Johns Hopkins University School of Hygiene and Public Health, Baltimore, MD.
Baqui AH, Black RE (2002) Childhood infectious diseases and their contribution to undernutrition. In: Public health issues in infant and child nutrition. Black RE, Michaelson KF, eds. (Nestlé Nutrition Workshop Series, Pediatric Program, Volume 48.) Lippincott Williams \& Wilkins, Philadelphia, PA.

Baqui AH, Black RE, Sack RB, Chowdhury HR, Yunus M, Siddique AK (1993) Malnutrition, cell-mediated immune deficiency, and diarrhea: a communitybased longitudinal study in rural Bangladeshi children. American Journal of Epidemiology, 137:355-365.
Barker DJ (1995) The fetal and infant origins of disease. European Journal of Clinical Investigation, 25:457-463.

Barros FC, Huttly SR, Victora CG, Kirkwood BR, Vaughan JP (1992) Comparison of the causes and consequences of prematurity and intrauterine growth retardation: a longitudinal study in southern Brazil. Pediatrics, 90:238244.

Barros FC, Victora CG, Vaughan JP, Estanislau HJ (1987) Perinatal mortality in southern Brazil: a population-based study of 7392 births. Bulletin of the World Health Organization, 65:95-104.
Beau JP, Garenne M, Diop B, Briend A, Diop MI (1987) Diarrhoea and nutritional status as risk factors of child mortality in a Dakar hospital (Senegal). Journal of Tropical Pediatrics, 33:4-9.
Becker S, Black RE, Brown KH (1991) Relative effects of diarrhea, fever, and dietary energy intake on weight gain in rural Bangladeshi children. American Journal of Clinical Nutrition, 53:1499-1503.

Behrman RE, Babson GS, Lessel R (1971) Fetal and neonatal mortality in white middle class infants. Mortality risks by gestational age and weight. American Journal of Diseases of Children, 121:486-489.
Berkman DS, Lescano AG, Gilman RH, Lopez SL, Black MM (2002) Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. The Lancet, 359:564571.

Berman S, Duenas A, Bedoya A, et al. (1983) Acute lower respiratory tract illnesses in Cali, Colombia: a two-year ambulatory study. Pediatrics, 71: 210-218.
Bhan MK, Arora NK, Ghai OP, Ramachandran K, Khoshoo V, Bhandari N (1986) Major factors in diarrhoea related mortality among rural children. Indian Journal of Medical Research, 83:9-12.

Bhan MK, Sommerfelt H, Strand T (2001) Micronutrient deficiency in children. British Journal of Nutrition, 85:S199-203.
Bhandari N, Bahl R, Nayyar B, Khokhar P, Rohde JE, Bhan MK (2001) Food supplementation with encouragement to feed it to infants from 4 to 12 months of age has a small impact on weight gain. Journal of Nutrition, 131:1946-1951.

Bhandari N, Bhan MK, Sazawal S, Clemens JD, Bhatnagar S, Khoshoo V (1989) Association of antecedent malnutrition with persistent diarrhoea: a casecontrol study. British Medical Journal, 298:1284-1287.

Bhat US, Udani PM, Shah BP, Naik PA, Ezhuthachan SG (1973) Nutritional growth failure and mental development. Indian Pediatrics, 10:657-660.
Bhargava SK, Sachdev HP, Ghosh S (1985) Distribution of live births and early neonatal mortality in relation to gestation and intrauterine growth. Indian Journal of Medical Research, 82:95-97.
Bhatt RV, Modi NS, Acharya PT (1967) Height and reproductive performance. Journal of Obstetrics and Gynaecology of India, 17:75-79.
Binkin NJ, Williams RL, Hogue CJ, Chen PM (1985) Reducing black neonatal mortality. Will improvement in birth weight be enough? Journal of the American Medical Association, 253:372-375.

Binns CW (1976) Food, sickness and death in children of the highlands of Papua New Guinea. Journal of Tropical Pediatrics and Environmental Child Health, 22:9-11.

Birch HG, Pineiro E, Alcalde E, Toca T, Cravioto J (1971) Relation of kwashiorkor in early childhood and intelligence at school age. Pediatric Research, 5:579-585.

Biritwum RB, Isomura S, Assoku A, Torigoe S (1986) Growth and diarrhoeal disease surveillance in a rural Ghanaian pre-school child population. Transactions of the Royal Society of Tropical Medicine and Hygiene, 80:208-213.

Black RE (1991) Would control of childhood infectious diseases reduce malnutrition? Acta Paediatrica Scandinavica, 374:S133-140.

Black RE, Brown KH, Becker S (1984) Malnutrition is a determining factor in diarrheal duration, but not incidence, among young children in a longitudinal study in rural Bangladesh. American Journal of Clinical Nutrition, 39:87-94.

Botha-Antoun E, Babayan S, Harfouche JK (1968) Intellectual development related to nutritional status. Journal of Tropical Pediatrics, 14:112-115.
Brabin L, Verhoeff F, Brabin BJ (2002) Maternal height, birthweight and cephalo pelvic disproportion in urban Nigeria and rural Malawi. Acta Obstetricia et Gynecologica Scandinavica, 81:502-507.
Bracero LA, Byrne DW (1998) Optimal maternal weight gain during singleton pregnancy. Gynecologic and Obstetric Investigation, 46:9-16.
Brown JL, Pollitt E (1996) Malnutrition, poverty and intellectual development. Scientific American, 274:38-43.

Brown KH, Peerson JM, Rivera JA, Allen LH (2002) Effect of supplemental zinc on the growth and serum zinc concentrations of prepubertal children: a metaanalysis of randomized controlled trials. American Journal of Clinical Nutrition, 75:1062-1071.
Brussow H, Rahim H, Barclay D, Freire WB, Dirren H (1993) Nutritional and environmental risk factors for diarrhoeal diseases in Ecuadorian children. Journal of Diarrhoeal Diseases Research, 11:137-142.
Cabak V, Najdanvic R (1965) Effect of undernutrition in early life on physical and mental development. Archives of Disease in Childhood, 40:532534.

Cattanach S, Morrison J, Andersen MJ, Williams GM, Sheahan T (1993) Pregnancy hazards associated with low maternal body mass indices. Australian and New Zealand Journal of Obstetrics and Gynaecology, 33: 45-47.

Caulfield LE, Haas JD, Belizan JM, Rasmussen KM, Edmonston B (1991) Differences in early postnatal morbidity risk by pattern of fetal growth in Argentina. Paediatric and Perinatal Epidemiology, 5:263-275.
Champakam S, Srikantia SG, Gopalan C (1968) Kwashiorkor and mental development. American Journal of Clinical Nutrition, 21:844-852.

Chandler AM, Walker SP, Connolly K, Grantham-McGregor SM (1995) School breakfast improves verbal fluency in undernourished Jamaican children. Journal of Nutrition, 125:894-900.

Chandra RK (1977) Lymphocyte subpopulations in human malnutrition: cytotoxic and suppressor cells. Pediatrics, 59:423-427.
Chatelain PG, Nicolino M, Claris O, Salle B, Chaussain J (1998) Multiple hormone resistance in short children born with intrauterine growth retardation? Hormone Research, 49:S20-22.
Chen LC, Chowdhury A, Huffman SL (1980) Anthropometric assessment of energy-protein malnutrition and subsequent risk of mortality among preschool aged children. American Journal of Clinical Nutrition, 33: 1836-1845.

Chen LC, Huq E, Huffman SL (1981) A prospective study of the risk of diarrheal diseases according to the nutritional status of children. American Journal of Epidemiology, 114:284-292.
Cheung YB, Yip PS, Karlberg JP (2001) Parametric modelling of neonatal mortality in relation to size at birth. Statistics in Medicine, 20:2455-2466.
Chopra JS, Sharma A (1992) Protein energy malnutrition and the nervous system. Journal of the Neurological Sciences, 110:8-20.
Chowdhury MK, Gupta VM, Bairagi R, Bhattacharya BN (1990) Does malnutrition predispose to diarrhoea during childhood? Evidence from a longitudinal study in Matlab, Bangladesh. European Journal of Clinical Nutrition, 44:515-525.

Chwang LC, Soemantri AG, Pollitt E (1988) Iron supplementation and physical growth of rural Indonesian children. American Journal of Clinical Nutrition, 47:496-501.

Cnattingius S, Bergstrom R, Lipworth L, Kramer MS (1998) Prepregnancy weight and the risk of adverse pregnancy outcomes. New England Journal of Medicine, 338:147-152.
Cogswell ME, Yip R (1995) The influence of fetal and maternal factors on the distribution of birthweight. Seminars in Perinatology, 19:222-240.
Colombo M, de Andraca I, Lopez I (1988) Mental development and stunting. In: Linear growth retardation in less developed countries. Nestlé Nutrition Workshop. Waterlow JC, ed. Raven Press, New York.
Conde-Agudelo A, Belizan JM, Diaz-Rossello JL (2000) Epidemiology of fetal death in Latin America. Acta Obstetricia et Gynecologica Scandinavica, 79:371-378.

Cravioto J, DeLicardie ER, Pinero C, Lindoro M, Arroyo M, Alcade E (1971) Mental development and malnutrition. Neurointegrative development and intelligence in school children recovered from malnutrition in infancy. Indian Journal of Medical Research, 59:S195-215.
Cunha AL (2000) Relationship between acute respiratory infection and malnutrition in children under 5 years of age. Acta Paediatrica, 89:608-609.
de Almeida MF, Jorge MH (1998) [Small for gestational age: risk factor for neonatal mortality]. Revista de Saude Publica, 32:217-224.
de Onis M (2000) Measuring nutritional status in relation to mortality. Bulletin of the World Health Organization, 78:1271-1274.
de Onis M (2001) Child growth and development. In: Nutrition and health in developing countries. Semba RD, Bloem MW, eds. Humana Press, Totowa, NJ.
de Onis M, Blössner M (1997) WHO Global Database on Child Growth and Malnutrition. Available at http://www.who.int/nutgrowthdb.
de Onis M, Blössner M (2000) Prevalence and trends of overweight among preschool children in developing countries. American Journal of Clinical Nutrition, 72:1032-1039.
de Onis M, Blössner M, Villar J (1998) Levels and patterns of intrauterine growth retardation in developing countries. European Journal of Clinical Nutrition, 52:S5-15.
de Onis M, Frongillo EA, Blössner M (2000) Is malnutrition declining? An analysis of changes in levels of child malnutrition since 1980. Bulletin of the World Health Organization, 78:1222-1233.
de Onis M, Garza C, Habicht JP (1997) Time for a new growth reference. Pediatrics, 100:E8.
de Onis M, Habicht JP (1996) Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. American Journal of Clinical Nutrition, 64:650-658.
de Onis M, Monteiro C, Akre J, Glugston G (1993) The worldwide magnitude of protein-energy malnutrition: an overview from the WHO global database on child growth. Bulletin of the World Health Organization, 71:703-712.
de Onis M, Victora CG, Garza C, Frongillo EA, Cole TJ, for the WHO Working Group on the Growth Reference Protocol (2001) A new international growth reference for young children. In: Perspectives in human growth, development and maturation. Dasgupta P, Hauspie R, eds. Kluwer Academic Publishers, Dordrecht.
de Onis M, Yip R (1996) The WHO growth chart: historical considerations and current scientific issues. Biblotheca Nutritio et Dieta, 53:74-89.
Deb SK (1998) Acute respiratory disease survey in Tripura in case of children below five years of age. Journal of the Indian Medical Association, 96: 111-116.

Edwards LE, Alton IR, Barrada MI, Hakanson EY (1979) Pregnancy in the underweight woman. Course, outcome, and growth patterns of the infant. American Journal of Obstetrics and Gynecology, 135:297-302.
el Samani EF, Willett WC, Ware JH (1988) Association of malnutrition and diarrhea in children aged under five years. A prospective follow-up study in a rural Sudanese community. American Journal of Epidemiology, 128: 93-105.
el Samani FZ, Willett WC, Ware JH (1987) Nutritional and socio-demographic risk indicators of malaria in children under five: a cross-sectional study in a Sudanese rural community. Journal of Tropical Medicine and Hygiene, 90:69-78.

Elizabeth KE, Sathy N (1997) The role of developmental stimulation in nutritional rehabilitation. Indian Pediatrics, 34:681-695.

Essex BJ, Everett VJ (1977) Use of an action-orientated record for antenatal screening. Tropical Doctor, 7:134-138.
Evans D, Hansen JD, Moodie AD, van der Spuy HI (1980) Intellectual development and nutrition. Journal of Pediatrics, 97:358-363.

Evans DE, Moodie AD, Hansen JD (1971) Kwashiorkor and intellectual development. South African Medical Journal, 45:1413-1426.
Everson PJ, Morris CN (1983) Hierarchical normal regression model. (Technical Report.) Harvard University, Boston, MA.
FAO (2001) The state of food insecurity in the world. The Food and Agriculture Organization of the United Nations, Rome.
Fawzi WW, Herrera MG, Spiegelman DL, el Amin A, Nestel P, Mohamed KA (1997) A prospective study of malnutrition in relation to child mortality in the Sudan. American Journal of Clinical Nutrition, 65:1062-1069.

Ferro-Luzzi A, Ashworth A, Martorell R, Scrimshaw NS (1998) Report of the IDECG Working Group on effects of IUGR on infants, children and adolescents: immunocompetence, mortality, morbidity, body size, body composition, and physical performance. European Journal of Clinical Nutrition, 52:S97-99.

Fisher MM (1972) Malnutrition and reasoning ability in Zambian schoolchildren. Transactions of the Royal Society of Tropical Medicine and Hygiene, 66:471-478.
Florencio CA (1988) Nutrition, health and other determinants of academic achievement and school related behavior of grade one to grade six pupils. University of the Philippines, Quezon City.

Fonseca W, Kirkwood BR, Victora CG, Fuchs SR, Flores JA, Misago C (1996) Risk factors for childhood pneumonia among the urban poor in Fortaleza, Brazil: a case-control study. Bulletin of the World Health Organization, 74:199-208.

Galler JR, Ramsey F (1987) A follow-up study of the influence of early malnutrition on development: V. Delayed development of conservation (Piaget). Journal of the American Academy of Child and Adolescent Psychiatry, 26:23-27.
Galler JR, Ramsey F, Solimano G (1984) The influence of early malnutrition on subsequent behavioral development III. Learning disabilities as a sequel to malnutrition. Pediatric Research, 18:309-313.

Galler JR, Ramsey F, Solimano G, Lowell WE, Mason E (1983) The influence of early malnutrition on subsequent behavioral development. I. Degree of impairment in intellectual performance. Journal of the American Academy of Child Psychiatry, 22:8-15.
Galler JR, Ramsey FC, Forde V, Salt P, Archer E (1987a) Long-term effects of early kwashiorkor compared with marasmus. II. Intellectual performance. Journal of Pediatric Gastroenterology and Nutrition, 6:847-854.

Galler JR, Ramsey FC, Morley DS, Archer E, Salt P (1990) The long-term effects of early kwashiorkor compared with marasmus. IV. Performance on the national high school entrance examination. Pediatric Research, 28: 235-239.

Galler JR, Ramsey FC, Salt P, Archer E (1987b) Long-term effects of early kwashiorkor compared with marasmus. III. Fine motor skills. Journal of Pediatric Gastroenterology and Nutrition, 6:855-859.
Garenne M, Maire B, Fontaine O, Dieng K, Briend A (1987) Risques de décès associés à différents états nutritionnels chez l'enfant d'âge préscolaire [Risks of dying associated with different nutritional status in pre-school aged children]. Rapport final. ORSTOM, Dakar, September 1987. Reprinted 2000 by the Centre Population et Développement (CEPED), Paris.
Ghai OP (1975) Effect of early marasmic malnutrition on subsequent mental development. Indian Pediatrics, 12:85.
Ghai OP, Jaiswal VN (1970) Relationship of undernutrition to diarrhoea in infants and children. Indian Journal of Medical Research, 58:789-795.

Ghosh S, Ramanujacharyulu TK, Hooja V, Madhavan S (1979) Mortality pattern in an urban birth cohort. Indian Journal of Medical Research, 69:616-623.

Goldenberg R, Hack M, Grantham M, Schurch B (1998) Report of the IDECG/IUNS Working Group on IUGR effects on neurological, sensory, cognitive, and behavioral function. European Journal of Clinical Nutrition, 52:S100-101.

Golub MS, Keen CL, Gershwin ME, Hendrickx AG (1995) Developmental zinc deficiency and behavior. Journal of Nutrition, 125:S2263-2271.
Gomez F, Galvan R, Frenk S, Cravioto Munoz J, Chavez R, Vazquez L (1956) Mortality in second and third degree malnutrition. Journal of Tropical Pediatrics, 2:77-83.

Gopalan C, Swaninathan MC, Kumari VK, Rao DH, Vijayaraghavan K (1973) Effect of calorie supplementation on growth of undernourished children. American Journal of Clinical Nutrition, 26:563-566.
Gordon JE, Guzman MA, Ascoli W, Scrimshaw NS (1964) Acute diarrhoeal disease in less developed countries. 2. Patterns of epidemiological behaviour in rural Guatemalan villages. Bulletin of the World Health Organization, 31:9-20.

Gorman KS (1995) Malnutrition and cognitive development: evidence from experimental/quasi-experimental studies among the mild-to-moderately malnourished. Journal of Nutrition, 12:S2239-2244.
Grantham-McGregor S (1982) The relationship between developmental level and different types of malnutrition in children. Human Nutrition: Clinical Nutrition, 36:319-320.
Grantham-McGregor S (1984) Chronic undernutrition and cognitive abilities. Human Nutrition. Clinical Nutrition, 38:83-94.

Grantham-McGregor S (1995) A review of studies of the effect of severe malnutrition on mental development. Journal of Nutrition, 125:S2233-2238.

Grantham-McGregor S, Meeks Gardner JM, Walker S, Powell C (1989a) The relationship between undernutrition, activity levels and development in young children. In: Activity, energy expenditure and energy requirements of infants and children. (Proceedings of an IDECG workshop held in Cambridge, Massachusetts, USA, November 14 to 17, 1989.) Schurch B, Scrimshaw NS, eds. United Nations University Press, New York.
Grantham-McGregor S, Powell C, Fletcher P (1989b) Stunting, severe malnutrition and mental development in young children. European Journal of Clinical Nutrition, 43:403-409.

Grantham-McGregor S, Powell C, Walker S, Chang S, Fletcher P (1994) The long-term follow-up of severely malnourished children who participated in an intervention program. Child Development, 65:428-439.
Grantham-McGregor SM, Lira PI, Ashworth A, Morris SS, Assuncao AM (1998) The development of low birth weight term infants and the effects of the environment in northeast Brazil. Journal of Pediatrics, 132:661666.

Grantham-McGregor SM, Powell CA, Walker SP, Himes JH (1991) Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study. The Lancet, 338:1-5.

Grantham-McGregor SM, Walker SP, Chang S (2000) Nutritional deficiencies and later behavioural development. Proceedings of the Nutrition Society, 59:47-54.

Grantham-McGregor SM, Walker SP, Chang SM, Powell CA (1997) Effects of early childhood supplementation with and without stimulation on later development in stunted Jamaican children. American Journal of Clinical Nutrition, 66:247-253.

Gray RH, Ferraz EM, Amorim MS, de Melo LF (1991) Levels and determinants of early neonatal mortality in Natal, northeastern Brazil: results of a surveillance and case-control study. International Journal of Epidemiology, 20:467-473.

Groos AD (1991) Delayed motor development in relation to nutritional status among children under two years of age in two districts of Simbu Province. Papua New Guinea Medical Journal, 34:238-245.

Guerrant RL, Schorling JB, McAuliffe JF, de Souza MA (1992) Diarrhea as a cause and an effect of malnutrition: diarrhea prevents catch-up growth and malnutrition increases diarrhea frequency and duration. American Journal of Tropical Medicine and Hygiene, 47:28-35.
Gupta S, Dhingra DC, Singh MV, Anand NK (1975) Impact of nutrition on intelligence. Indian Pediatrics, 12:1079-1082.

Haas JD, Balcazar H, Caulfield L (1987) Variation in early neonatal mortality for different types of fetal growth retardation. American Journal of Physical Anthropology, 73:467-473.

Hamill PV, Drizd TA, Johnson CL, Reed RB, Roche AF (1977) NCHS growth curves for children birth-18 years. United States. Vital Health Statistics, 11:i-iv, 1-74.

Harrison KA, Rossiter CE, Chong H (1985) Relations between maternal height, fetal birth weight and cephalopelvic disproportion suggest that young Nigerian primigravidae grow during pregnancy. British Journal of Obstetrics and Gynaecology, 92:S40-48.
Henry FJ, Alam N, Aziz KM, Rahaman MM (1987) Dysentery, not watery diarrhoea, is associated with stunting in Bangladeshi children. Human Nutrition. Clinical Nutrition, 41:243-249.
Hertzig ME, Birch HG, Richardson SA, Tizard J (1972) Intellectual levels of school children severely malnourished during the first two years of life. Pediatrics, 49:814-824.

Heywood AH, Marshall T, Heywood PF (1991) Motor development and nutritional status of young children in Madang, Papua New Guinea. Pариа New Guinea Medical Journal, 34:109-116.

Hill AB (1965) The environment and disease: association or causation? Proceedings of the Royal Society of Medicine, 58:295-300.
Hoorweg J, Stanfield JP (1976) The effects of protein energy malnutrition in early childhood on intellectual and motor abilities in later childhood and adolescence. Developmental Medicine and Child Neurology, 18:330-350.
Husaini MA, Husaini YK, Sandjaja S et al. (1995) Maternal anthropometry and pregnancy outcomes in Indonesia. Bulletin of the World Health Organization, 73:77-79.

Husaini MA, Karyadi L, Husaini YK, Sandjaja S, Karyadi D, Pollitt E (1991) Developmental effects of short-term supplementary feeding in nutritionally-at-risk Indonesian infants. American Journal of Clinical Nutrition, 54:799-804.

IDECG Working Group (1987) Pregnancy, lactation and childhood: report of working group 1. In: Chronic energy deficiency: consequences and related issues. Schurch B, Scrimshaw NS, eds. International Dietary Energy Consultancy Group, Lausanne.
Indian Academy of Pediatrics (1972) Nutrition sub committee of IAP: report of the convenor. Indian Pediatrics, 9:360

Institute of Medicine (1990) Nutrition during pregnancy. National Academy Press, Washington, DC.

Ivanovic DM, Leiva BP, Perez HT et al. (2000) Long-term effects of severe undernutrition during the first year of life on brain development and learning in Chilean high-school graduates. Nutrition, 16:1056-1063.

Jalil F, Lindblad BS, Hanson LA et al. (1993) Early child health in Lahore, Pakistan: I. Study design. Acta Paediatrica, Supplement, 82:S3-16.
James JW (1972) Longitudinal study of the morbidity of diarrheal and respiratory infections in malnourished children. American Journal of Clinical Nutrition, 25:690-694.
Jamison DT (1977) Child malnutrition and school performance in China. Journal of Development Economics, 20:299-309.
Jelliffe DB (1966) The assessment of the nutritional status of the community. (WHO Monograph 53). World Health Organization, Geneva.

Johnson FE, Low SM, de Baessa Y, MacVean RB (1987) Interaction of nutritional and socioeconomic status as determinants of cognitive development in disadvantaged urban Guatemalan children. American Journal of Physical Anthropology, 73:501-506.
Kalra A, Mishra MN, Kumar R, Prasad R, Dayal RS (1980) The impact of malnutrition on intelligence. Indian Pediatrics, 17:109-116.
Khan MA, Yunus M (1990) Host factors in childhood diarrhoea. Journal of the Royal Society of Health, 110:94-95.
Khan SR, Jalil F, Zaman S, Lindblad BS, Karlberg J (1993) Early child health in Lahore, Pakistan: X. Mortality. Acta Paediatrica, Supplement, 82: S109-117.

King JC, Weininger J (1989) Nutrition during pregnancy. Seminars in Perinatology, 13:162-168.

Konje JC, Ladipo OA (2000) Nutrition and obstructed labor. American Journal of Clinical Nutrition, 72:S291-297.
Kossmann J, Nestel P, Herrera MG, el Amin A, Fawzi WW (2000) Undernutrition in relation to childhood infections: a prospective study in the Sudan. European Journal of Clinical Nutrition, 54:463-472.
Kramer MS (1987) Determinants of low birth weight: methodological assessment and meta-analysis. Bulletin of the World Health Organization, 65:663-737.

Kramer MS, Coates AL, Michoud MC, Dagenais S, Hamilton EF, Papageorgiou A (1995) Maternal anthropometry and idiopathic preterm labor. Obstetrics and Gynecology, 86:744-748.
Kramer MS, Kakuma R (2002) Optimal duration of exclusive breastfeeding. Cochrane Database of Systematic Reviews, 1:CD003517

Kramer MS, McLean FH, Eason EL, Usher RH (1992) Maternal nutrition and spontaneous preterm birth. American Journal of Epidemiology, 136:574583.

Kramer MS, Platt RW, Wen SW et al. (2001) A new and improved populationbased Canadian reference for birth weight for gestational age. Pediatrics, 108:E35.

Krasovec K, Anderson MA (1991) Maternal anthropometry for prediction of pregnancy outcomes. (Memorandum from a USAID/WHO/PAHO/ MotherCare meeting.) Bulletin of the World Health Organization, 69: 523-532.

Kwawukume EY, Ghosh TS, Wilson JB (1993) Maternal height as a predictor of vaginal delivery. International Journal of Gynaecology and Obstetrics, 41:27-30.

Lahiri SK, Mukhopadhyay SP, Das KK, Ray SK, Biswas D (1994) Study of the impact of epidemiological factors on intelligence of rural children of 3 to 6 years age group belonging to low socioeconomic status. Indian Journal of Public Health, 38:133-142.

Lamb WH (1988) Epidemic measles in a highly immunized rural West African (Gambian) village. Reviews of Infectious Diseases, 10:457-462.

Lasky RE, Klein RE, Yarbrough C, Engle PL, Lechtig A, Martorell R (1981) The relationship between physical growth and infant behavioral development in rural Guatemala. Child Development, 52:219-226.

Law CM, Egger P, Dada O et al. (2001) Body size at birth and blood pressure among children in developing countries. International Journal of Epidemiology, 30:52-57.
Lawless JW, Latham MC, Stephenson LS, Kinoti SN, Pertet AM (1994) Iron supplementation improves appetite and growth in anemic Kenyan primary school children. Journal of Nutrition, 124:645-654.

Lehmann D, Howard P, Heywood P (1988) Nutrition and morbidity: acute lower respiratory tract infections, diarrhoea and malaria. Papua New Guinea Medical Journal, 31:109-116.

Levitsky DA, Strupp BJ (1995) Malnutrition and the brain: changing concepts, changing concerns. Journal of Nutrition, 125:S2212-2220.

Lindtjorn B, Alemu T, Bjorvatn B (1993) Nutritional status and risk of infection among Ethiopian children. Journal of Tropical Pediatrics, 39: 76-82.

Liu YL, Lin LM, Zhu J, Zhang ZJ, Li Z (1991) Case-control study on pathogenesis of pneumonia in children aged 0-2 years. Chinese Medical Journal (Engl.), 104:22-26.
Lopez I, de Andraca I, Perales CG, Heresi E, Castillo M, Colombo M (1993) Breakfast omission and cognitive performance of normal, wasted and stunted schoolchildren. European Journal of Clinical Nutrition, 47:533-542.

Lozoff B, Jimenez E, Wolf AW (1991) Long-term developmental outcome of infants with iron deficiency. New England Journal of Medicine, 325:687-694.

Lubchenco LO, Searls DT, Brazie JV (1972) Neonatal mortality rate: relationship to birth weight and gestational age. Journal of Pediatrics, 81:814822.

Lutter C (2000) Length of exclusive breastfeeding: linking biology and scientific evidence to a public health recommendation. Journal of Nutrition, 130:1335-1338.

Lutter CK, Mora JO, Habicht JP et al. (1989) Nutritional supplementation: effects on child stunting because of diarrhea. American Journal of Clinical Nutrition, 50:1-8.

Luwang NC, Datta SN (1982) Association between diarrhoeal disease and undernutrition in pre-school children. Indian Journal of Medical Sciences, 36:29-30.

Man WD, Weber M, Palmer A et al. (1998) Nutritional status of children admitted to hospital with different diseases and its relationship to outcome in the Gambia, West Africa. Tropical Medicine and International Health, 3: 678-686.

Martorell R, Ramakrishnan U, Schroeder DG, Melgar P, Neufeld L (1998) Intrauterine growth retardation, body size, body composition and physical performance in adolescence. European Journal of Clinical Nutrition, 52:S43-52.

Martorell R, Rivera JA, Kaplowitz H, Pollitt E (1992) Longterm consequences of growth retardation during early childhood. In: Human growth: basic and clinical aspects. Hernandez M, Argente J, eds. Elsevier Science Publishers B.V., Amsterdam.
Mata LJ (1978) The children of Santa Maria Cauque: a prospective field study of health and growth. MIT Press, Cambridge, MA.

Mathur R, Reddy V, Naidu AN, Ravikumar, Krishnamachari KA (1985) Nutritional status and diarrhoeal morbidity: a longitudinal study in rural Indian preschool children. Human Nutrition. Clinical Nutrition, 39:447-454.
Mati JK (1983) The Nairobi birth survey III. Labor and delivery. Journal of Obstetrics and Gynaecology of Eastern and Central Africa, 2:47-56.
Mavalankar DV, Trivedi CC, Gray RH (1994) Maternal weight, height and risk of poor pregnancy outcome in Ahmedabad, India. Indian Pediatrics, 31: 1205-1212.
McKay H, Sinisterra L, McKay A, Gomez H, Lloreda P (1978) Improving cognitive ability in chronically deprived children. Science, 200:270278.

McLaren DS, Yaktin US, Kanawati AA, Sabbagh S, Kadi Z (1973) The subsequent mental and physical development of rehabilitated marasmic infants. Journal of Mental Deficiency Research, 17:273-281.
Mehta S, Teja JS, Wig NN, Thukral S, Pasricha S (1975) Protein calorie malnutrition-a follow-up study. Indian Journal of Medical Research, 63: 576-582.

Mendez MA, Adair LS (1999) Severity and timing of stunting in the first two years of life affect performance on cognitive tests in late childhood. Journal of Nutrition, 129:1555-1562.
Merchant KM, Villar J, Kestler E (2001) Maternal height and newborn size relative to risk of intrapartum caesarean delivery and perinatal distress. British Journal of Obstetrics and Gynaecology, 108:689-696.
Molbak K, Jensen H, Ingholt L, Aaby P (1997) Risk factors for diarrheal disease incidence in early childhood: a community cohort study from Guinea-Bissau. American Journal of Epidemiology, 146:273-282.

Monckeberg F (1972) Malnutrition and mental capacity. In: Nutrition, the nervous system and behaviour. (PAHO Scientific Publication, No. 251.) Pan American Health Organization, Washington, DC.
Moock PR, Leslie J (1986) Childhood malnutrition and schooling in the Teri region of Nepal. Journal of Development Economics, 20:33-52.
Moore SE, Cole TJ, Collinson AC, Poskitt EM, McGregor IA, Prentice AM (1999) Prenatal or early postnatal events predict infectious deaths in young adulthood in rural Africa. International Journal of Epidemiology, 28: 1088-1095.

Moore SE, Collinson AC, Prentice AM (2001) Immune function in rural Gambian children is not related to season of birth, birth size, or maternal supplementation status. American Journal of Clinical Nutrition, 74: 840-847.

Morris CN (1983) Parametric empirical Bayes inference: theory and applications. Journal of the American Statistical Association, 78:47-65.
Naeye RL (1990) Maternal body weight and pregnancy outcome. American Journal of Clinical Nutrition, 52:273-279.
Neumann C, McDonald MA, Sigman M, Bwibo N, Marquardt M (1991) Relationships between morbidity and development in mildly to moderately malnourished Kenyan toddlers. Pediatrics, 88:934-942.

Nwuga VC (1977) Effect of severe kwashiorkor on intellectual development among Nigerian children. American Journal of Clinical Nutrition, 30: 1423-1430.
PAHO (1998) Nutrition, health and child development. (Scientific Publication No. 566.) Pan American Health Organization, Washington, DC.

Paine P, Dorea JG, Pasquali J, Monteior AM (1992) Growth and cognition in Brazilian school children: a spontaneously occurring intervention study. International Journal of Behavioral Development, 15:160-183.
Parekh UC, Udani PM, Naik PA, Shah BP (1974) Mental development of children with severe protein calorie malnutrition in the past and present. Indian Pediatrics, 11:465-469.
Patwari AK (1999) Diarrhoea and malnutrition interaction. Indian Journal of Pediatrics, 66: S124-134.

Pek HL, Tjiook TH, Oey HJ, Lauw TG (1967) Evaluation of mental development in relation to early malnutrition. American Journal of Clinical Nutrition, 20:1290-1294.

Pelletier DL (1994) The relationship between child anthropometry and mortality in developing countries: implications for policy, programs and future research. Journal of Nutrition, 124:S2047-2081.
Pelletier DL, Arimond M, Johnson FC, et al. (1995) Maternal anthropometry predictors of intrauterine growth retardation and prematurity in the Malawi Maternal and Child Nutrition Study. Bulletin of the World Health Organization, 73:80-81.
Pelletier DL, Frongillo EA, Jr, Habicht JP (1993) Epidemiologic evidence for a potentiating effect of malnutrition on child mortality. American Journal of Public Health, 83:1130-1133.

Pelletier DL, Frongillo EA, Jr, Schroeder DG, Habicht JP (1994) A methodology for estimating the contribution of malnutrition to child mortality in developing countries. Journal of Nutrition, 124:S2106-2122.

Pelletier DL, Frongillo EA, Jr, Schroeder DG, Habicht JP (1995) The effects of malnutrition on child mortality in developing countries. Bulletin of the World Health Organization, 73:443-448.

Peragallo-Guarda N (1984) Severity of iron deficiency anemia and its relationship to growth and morbidity in a population of pre-schoolers in rural Guatemala [Dissertation]. University of Texas, Houston, TX.
Pollitt E (1983) Morbidity and infant development. International Journal of Behavioral Development, 6:461-475.
Pollitt E (1987) A critical view of three decades of research on the effects of chronic energy malnutrition on behavioral development. In: Chronic energy deficiency: consequences and related issues. Schurch B, Scrimshaw NS, eds. International Dietary Energy Consultancy Group, Lausanne.
Pollitt E (1995) Functional significance of the covariance between protein energy malnutrition and iron deficiency anemia. Journal of Nutrition, 125:S22722277.

Pollitt E, Gorman KS, Engle PL, Martorell R, Rivera JA (1993) Early supplementary feeding and cognition: effects over two decades. Monographs of the Society for Research in Child Development, 58:1-99.

Pollitt E, Gorman KS, Engle PL, Rivera JA, Martorell R (1995) Nutrition in early life and the fulfillment of intellectual potential. Journal of Nutrition, 125:S1111-1118.

Pollitt E, Metallinos-Katsaras E (1990) Iron deficiency and behavior: constructs, methods, and validity of the findings. In: Nutrition and the brain. Wurtman RJ, Wurtman JJ, eds. Raven Press, New York.
Pollitt E, Oh S (1994) Early supplementary feeding, child development, and health policy. Food and Nutrition Bulletin, 15(3).

Pollitt E, Watkins WE, Husaini MA (1997) Three-month nutritional supplementation in Indonesian infants and toddlers benefits memory function 8 years later. American Journal of Clinical Nutrition, 66:1357-1363.
Powell CA, Grantham-McGregor S (1985) The ecology of nutritional status and development in young children in Kingston, Jamaica. American Journal of Clinical Nutrition, 41:1322-1331.
Rajatasilpin A, Suepsaman B, Yamarat V (1970) Intellectual development and its relationship to the nutritional status among school children. Journal of the Medical Association of Thailand, 53:788-793.
Rantakallio P, Laara E, Koiranen M, Sarpola A (1995) Maternal build and pregnancy outcome. Journal of Clinical Epidemiology, 48:199-207.
Reyes MR, Valdecanas CM, Reyes OL, Reyes TM (1990) The effects of malnutrition on the motor, perceptual, and cognitive functions of Filipino children. International Disability Studies, 12:131-136.

Ricci JA, Becker S (1996) Risk factors for wasting and stunting among children in Metro Cebu, Philippines. American Journal of Clinical Nutrition, 63:966-975.

Rice AL, Sacco L, Hyder A, Black RE (2000) Malnutrition as an underlying cause of childhood deaths associated with infectious diseases in developing countries. Bulletin of the World Health Organization, 78:1207-1221.
Rice AL, Sacco L, Hyder A, Frangakis C, Black RE (2001) Malnutrition as an associated cause of death in children: evidence for an impact on cause-
specific mortality. Report by the Department of International Health, Bloomberg School of Public Health, Johns Hopkins University [unpublished].

Richardson SA (1976) The relation of severe malnutrition in infancy to the intelligence of school children with differing life histories. Pediatric Research, 10:57-61.

Richardson SA, Birch HG (1973) School performance of children who were severely malnourished in infancy. American Journal of Mental Deficiency, 77:623-632.
Richardson SA, Koller H, Katz M, Albert K (1978) The contributions of differing degrees of acute and chronic malnutrition to the intellectual development of Jamaican boys. Early Human Development, 2:163-170.

Rivera JA, Martorell R (1988) Nutrition, infection and growth. Part II: effects of malnutrition on infection and general conclusions. Clinical Nutrition, 7:163-167.

Rose G (1985) Sick individuals and sick populations. International Journal of Epidemiology, 14:32-38.
Rotimi C, Okosun I, Johnson L et al. (1999) The distribution and mortality impact of chronic energy deficiency among adult Nigerian men and women. European Journal of Clinical Nutrition, 53:734-739.
Rush D (2000) Nutrition and maternal mortality in the developing world. American Journal of Clinical Nutrition, 72:S212-240.
Rush D (2001) Maternal nutrition and perinatal survival. Journal of Health, Population, and Nutrition, 19:S217-264.
Sachar RK, Soni RK (2000) Perinatal mortality in rural Punjab-a populationbased study. Journal of Tropical Pediatrics, 46:43-45.
Saco-Pollitt C, Pollitt E, Greenfield D (1985) The cumulative deficit hypothesis in the light of cross-cultural evidence. International Journal of Behavioral Development, 8:75-97.

Sappenfield WM, Buehler JW, Binkin NJ, Hogue CJ, Strauss LT, Smith JC (1987) Differences in neonatal and postneonatal mortality by race, birth weight, and gestational age. Public Health Reports, 102:182-192.
Sathy N, Elizabeth KE, Nair MK, Bai NS (1991) Growth faltering and developmental delay in children with PEM. Indian Pediatrics, 28:255-258.
Schieve LA, Cogswell ME, Scanlon KS et al. (2000) Prepregnancy body mass index and pregnancy weight gain: associations with preterm delivery. Obstetrics and Gynecology, 96:194-200.
Schorling JB, McAuliffe JF, de Souza MA, Guerrant RL (1990) Malnutrition is associated with increased diarrhoea incidence and duration among children in an urban Brazilian slum. International Journal of Epidemiology, 19: 728-735.

Schroeder DG, Brown KH (1994) Nutritional status as a predictor of child survival: summarizing the association and quantifying its global impact. Bulletin of the World Health Organization, 72:569-579.

Scrimshaw NS, SanGiovanni JP (1997) Synergism of nutrition, infection, and immunity: an overview. American Journal of Clinical Nutrition, 66: S464-477.

Scrimshaw NS, Taylor CE, Gordon JE (1968) Interaction of nutrition and infection. (Monograph Series 57.) World Health Organization, Geneva.
Sebire NJ, Jolly M, Harris J, Regan L, Robinson S (2001) Is maternal underweight really a risk factor for adverse pregnancy outcome? A populationbased study in London. British Journal of Obstetrics and Gynaecology, 108:61-66.

Selwyn BJ (1990) The epidemiology of acute respiratory tract infection in young children: comparison of findings from several developing countries. Coordinated Data Group of BOSTID Researchers. Reviews of Infectious Diseases, 12:S870-888.

Sepulveda J, Willett W, Munoz A (1988) Malnutrition and diarrhea. A longitudinal study among urban Mexican children. American Journal of Epidemiology, 127:365-376.
Seshadri S, Gopaldas T (1989) Impact of iron supplementation on cognitive functions in preschool and school-aged children: the Indian experience. American Journal of Clinical Nutrition, 50:S675-684.
Shah N, Ramankutty V, Premila PG, Sathy N (1994) Risk factors for severe pneumonia in children in south Kerala: a hospital-based case-control study. Journal of Tropical Pediatrics, 40:201-206.

Shankar AH (2000) Nutritional modulation of malaria morbidity and mortality. Journal of Infectious Diseases, 182:S37-53.

Shetty PS, James WP (1994) Body mass index. A measure of chronic energy deficiency in adults. FAO Food and Nutrition Paper, 56:1-57.

Siega-Riz AM, Adair LS, Hobel CJ (1996) Maternal underweight status and inadequate rate of weight gain during the third trimester of pregnancy increases the risk of preterm delivery. Journal of Nutrition, 126:146-153.
Sigman M, McDonald MA, Neumann C, Bwibo N (1991) Prediction of cognitive competence in Kenyan children from Toddler nutrition, family characteristics and abilities. Journal of Child Psychology and Psychiatry, 32:307320.

Sigman M, Neumann C, Baksh M, Bwibo N, McDonald MA (1989) Relationship between nutrition and development in Kenyan toddlers. Journal of Pediatrics, 115:357-364.

Simeon DT, Grantham-McGregor S (1989) Effects of missing breakfast on the cognitive functions of school children of differing nutritional status. American Journal of Clinical Nutrition, 49:646-653.
Singh MB, Sidhu K (1987) Effect of age, sex and birth order on the intelligence of well-nourished and malnourished slum children. Indian Pediatrics, 24:713-718.
Singh MV, Anand NK, Gupta S, Dhingra DC (1976) Intelligence in relation to degree of malnutrition. Indian Journal of Clinical Psychology, 3:117120.

Sinha DP (1977) Measles and malnutrition in a West Bengal village. Tropical and Geographical Medicine, 29:125-134.

Smith TA, Lehmann D, Coakley C, Spooner V, Alpers MP (1991) Relationships between growth and acute lower-respiratory infections in children aged less than 5 years in a highland population of Papua New Guinea. American Journal of Clinical Nutrition, 53:963-970.
Snow RW, Byass P, Shenton FC, Greenwood BM (1991) The relationship between anthropometric measurements and measurements of iron status and susceptibility to malaria in Gambian children. Transactions of the Royal Society of Tropical Medicine and Hygiene, 85:584-589.

Soewondo S, Husaini M, Pollitt E (1989) Effects of iron deficiency on attention and learning processes in preschool children: Bandung, Indonesia. American Journal of Clinical Nutrition, 50:S667-673.

Sokal D, Sawadogo L, Adjibade A (1991) Short stature and cephalopelvic disproportion in Burkina Faso, West Africa. Operations Research Team. International Journal of Gynaecology and Obstetrics, 35:347-350.
Sommer A, Tarwotjo I, Djunaedi E et al. (1986) Impact of vitamin A supplementation on childhood mortality: a randomized controlled community trial. The Lancet 1:1169-1173.

Spinillo A, Capuzzo E, Piazzi G, Ferrari A, Morales V, Di Mario M (1998) Risk for spontaneous preterm delivery by combined body mass index and gestational weight gain patterns. Acta Obstetricia et Gynecologica Scandinavica, 77:32-36.

Spooner V, Barker J, Tulloch S et al. (1989) Clinical signs and risk factors associated with pneumonia in children admitted to Goroka Hospital, Papua New Guinea. Journal of Tropical Pediatrics, 35:295-300.

Spurr GB, Barac-Nieto M, Maksud MG (1977) Productivity and maximal oxygen consumption in sugar cane cutters. American Journal of Clinical Nutrition, 30:316-321.
Srikantia SG, Sastri CY (1971) Observations on malnutrition and mental development. Indian Journal of Medical Research, 59:S216-220.
Stein CE, Fall CH, Kumaran K, Osmond C, Cox V, Barker DJ (1996) Fetal growth and coronary heart disease in South India. The Lancet, 348: 1269-1273.

Stetler HC, Trowbridge FL, Huong AY (1981) Anthropometric nutrition status and diarrhea prevalence in children in El Salvador. American Journal of Tropical Medicine and Hygiene, 30:888-893.

Stoch MB, Smythe PM (1967) The effect of undernutrition during infancy on subsequent brain growth and intellectual development. South Africa Medical Journal, 41:1027-1030.
Stoch MB, Smythe PM (1976) 15-year developmental study on effects of severe undernutrition during infancy on subsequent physical growth and intellectual functioning. Archives of Disease in Childhood, 51:327-336.

Strauss RS, Dietz WH (1999) Low maternal weight gain in the second or third trimester increases the risk for intrauterine growth retardation. Journal of Nutrition, 129:988-993.

Strupp BJ, Levitsky DA (1995) Enduring cognitive effects of early malnutrition: a theoretical reappraisal. Journal of Nutrition, 125:S2221-2232.
Taha TE, Gray RH, Abdelwahab MM, Abdelhafeez AR, Abdelsalam AB (1994) Levels and determinants of perinatal mortality in central Sudan. International Journal of Gynaecology and Obstetrics, 45:109-115.
Thilothammal N, Kamala KG, Ashok TP et al. (1992) Influence of maternal nutritional status on mode of delivery and asphyxia neonatorum. Indian Journal of Pediatrics, 59:325-329.
Thongkrajai E, Stoeckel J, Thongkrajai P (1990) Nutritional status and the incidence and duration of diarrhoeal disease among children in northeast Thailand. Social Science and Medicine, 30:773-776.

Tomkins A (1981) Nutritional status and severity of diarrhoea among pre-school children in rural Nigeria. The Lancet, 1:860-862.

Tomkins A, Watson F (1989) Malnutrition and infection: a review. (ACC/SCN State-of-the-Art Series, Nutrition Policy Discussion Paper No. 5.) United Nations Administrative Committee on Coordination/Subcommittee on Nutrition, Geneva.
Tomkins AM, Dunn DT, Hayes RJ (1989) Nutritional status and risk of morbidity among young Gambian children allowing for social and environmental factors. Transactions of the Royal Society of Tropical Medicine and Hygiene, 83:282-287.

Tonglet R, Mahangaiko LE, Zihindula PM, Wodon A, Dramaix M, Hennart P (1999) How useful are anthropometric, clinical and dietary measurements of nutritional status as predictors of morbidity of young children in central Africa? Tropical Medicine and International Health, 4: 120-130.

Tsu VD (1992) Maternal height and age: risk factors for cephalopelvic disproportion in Zimbabwe. International Journal of Epidemiology, 21: 941-946.
Udani PM, Bhat US, Shah BP (1976) Mental development in severe protein calorie malnutrition (a study of 252 children and 25 mothers). Indian Pediatrics, 13:507-516.

UN (1999) UN world population prospects. The 1998 revision. United Nations, Department of International Economic and Social Affairs, New York.
UN (2001) UN world population prospects. The 2000 revision. United Nations, Department of International Economic and Social Affairs, New York.
Upadhyay SK, Agarwal DK, Agarwal KN (1989) Influence of malnutrition on intellectual development. Indian Journal of Medical Research, 90:430441.

Vazir S, Naidu AN, Vidyasagar P (1998) Nutritional status, psychosocial development and the home environment of Indian rural children. Indian Pediatrics, 35:959-966.

Ventura SJ, Martin JA, Curtin SC, Mathews TJ (1999) Births: final data for 1997. National Vital Statistics Report, 47:1-96.

Victora CG, Barros FC, Kirkwood BR, Vaughan JP (1990) Pneumonia, diarrhea, and growth in the first 4 years of life: a longitudinal study of 5914 urban Brazilian children. American Journal of Clinical Nutrition, 52:391-396.

Victora CG, Fuchs SC, Flores JA, Fonseca W, Kirkwood B (1994) Risk factors for pneumonia among children in a Brazilian metropolitan area. Pediatrics, 93:977-985.

Villar J, Belizan JM (1982) The relative contribution of prematurity and fetal growth retardation to low birth weight in developing and developed societies. American Journal of Obstetrics and Gynecology, 143:793-798.

Villar J, de Onis M, Kestler E, Bolanos F, Cerezo R, Bernedes H (1990) The differential neonatal morbidity of the intrauterine growth retardation syndrome. American Journal of Obstetrics and Gynecology, 163:151-157.
Villar J, Ezcurra EJ, Gurtner de la Fuente V, Campodonico L (1994) Preterm delivery syndrome: the unmet need. Research and Clinical Forums, 16:938.

Waber DP, Vuori-Christiansen L, Ortiz N et al. (1981) Nutritional supplementation, maternal education, and cognitive development of infants at risk of malnutrition. American Journal of Clinical Nutrition, 34:S807-813.
Walia BN, Singhi S, Gambhir SK, Sroa SR (1989) Impact of acute diarrhoea and oral rehydration solution on nutritional status of preschool children. Indian Journal of Medical Research, 90:415-425.

Walker SP, Grantham-McGregor SM, Powell CA, Chang SM (2000) Effects of growth restriction in early childhood on growth, IQ, and cognition at age 11 to 12 years and the benefits of nutritional supplementation and psychosocial stimulation. Journal of Pediatrics, 137:36-41.
Walter T (1993) Impact of iron deficiency on cognition in infancy and childhood. European Journal of Clinical Nutrition, 47:307-316.
Waterlow JC, Buzina R, Keller W, Lane JM, Nichaman MZ, Tanner JM (1977) The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years. Bulletin of the World Health Organization, 55:489-498.
Wesley AG, Loening WE (1996) Assessment and 2-year follow-up of some factors associated with severity of respiratory infections in early childhood. South Africa Medical Journal, 86:365-368.

West KP, LeClerq SC, Shrestha SR et al. (1997) Effects of vitamin A on growth of vitamin A-deficient children: field studies in Nepal. Journal of Nutrition, 127:1957-1965.

West KP, Jr, Pokhrel RP, Katz J et al. (1991) Efficacy of vitamin A in reducing preschool child mortality in Nepal. The Lancet, 338:67-71.
WHO (1978) A growth chart for international use in maternal and child health care: guidelines for primary health care personnel. World Health Organization, Geneva.

WHO (1992a) International statistical classification of diseases and related bealth problems, 10th revision (ICD-10). World Health Organization, Geneva.

WHO (1992b) Low birth weight: a tabulation of available information. Maternal Health and Safe Motherhood Programme. World Health Organization, Geneva.

WHO (1995a) An evaluation of infant growth: the use and interpretation of anthropometry in infants. WHO Working Group on Infant Growth. Bulletin of the World Health Organization, 73:165-174.
WHO (1995b) Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. World Health Organization Technical Report Series, 854:1-452.
WHO (1997) WHO Global Database on Child Growth and Malnutrition. World Health Organization, Geneva.

WHO (1998) Complementary feeding of young children in developing countries: a review of current scientific knowledge. World Health Organization, Geneva.

WHO (1999a) A critical link. Interventions for physical growth and psychological development. A review. (WHO/CHS/CAH/99.3.) World Health Organization, Geneva.
WHO (1999b) WHO recommended surveillance standards. 2nd edn. (WHO/CDS/CSR/ISR/99.2.) World Health Organization, Geneva.
WHO (2002) The world health report 2002: Reducing risks, promoting healthy life. World Health Organization, Geneva.
WHO/CHD Immunisation-linked Vitamin A Supplementation Study Group (1998) Randomised trial to assess benefits and safety of vitamin A supplementation linked to immunisation in early infancy. WHO/CHD Immunisation-Linked Vitamin A Supplementation Study Group. The Lancet, 352:1257-1263.

Wierzba TF, el Yazeed RA, Savarino SJ et al. (2001) The interrelationship of malnutrition and diarrhea in a periurban area outside Alexandria, Egypt. Journal of Pediatric Gastroenterology and Nutrition, 32:189-196.
Wilcox AJ (2001) On the importance—and the unimportance-of birthweight. International Journal of Epidemiology, 30:1233-1241.
Williams TN, Maitland K, Phelps L et al. (1997) Plasmodium vivax: a cause of malnutrition in young children. Quarterly Journal of Medicine, 90: 751-757.

Wilson DO, Rogers RM, Hoffman RM (1985) Nutrition and chronic lung disease. American Review of Respiratory Disease, 132:1347-1365.

Winkvist A, Stenlund H, Hakimi M, Nurdiati DS, Dibley MJ (2002) Weightgain patterns from prepregnancy until delivery among women in Central Java, Indonesia. American Journal of Clinical Nutrition, 75:1072-1077.

Witter FR, Caulfield LE, Stoltzfus RJ (1995) Influence of maternal anthropometric status and birth weight on the risk of cesarean delivery. Obstetrics and Gynecology, 85:947-951.

Xanthou M (1985) Immunologic deficiencies in small-for-dates neonates. Acta Paediatrica Scandinavica, Supplement, 319:143-149.

Yoon PW, Black RE, Moulton LH, Becker S (1997) The effect of malnutrition on the risk of diarrheal and respiratory mortality in children $<2$ years of age in Cebu, Philippines. American Journal of Clinical Nutrition, 65: 1070-1077.

Zaman K, Baqui AH, Yunus M et al. (1996) Association between nutritional status, cell-mediated immune status and acute lower respiratory infections in Bangladeshi children. European Journal of Clinical Nutrition, 50:309-314.
Zinc Investigators' Collaborative Group (1999) Prevention of diarrhea and pneumonia by zinc supplementation in children in developing countries: pooled analysis of randomized controlled trials. Journal of Pediatrics 135:689-697.
Appendix A
National surveys included in prevalence estimates

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :---: | :---: | :---: |
| Afghanistan |  | (1998) Afghanistan 1997 multiple indicator baseline (MICS). Report to UNICEF. Centro de Investigacion de Enfermedades Tropicales (CIET), Acapulco. ${ }^{\text {b }}$ |
| Algeria | Kellou K | (1987) Etat nutritionnel des enfants algériens de 0 à 10 ans et niveaux d'urbanisation d'après les résultats préliminaires de l'enquête épidémiologique sur la malnutrition protéino-énérgétique en 1987. Institut National de Santé Publique, Algiers. |
| Algeria | République Algérienne Démocratique et Populaire | (1992) Enquête algérienne sur la santé de la mère et de l'enfant. Office national des statistiques. (PAPCHILD Surveys.) The League of Arab States, Cairo. |
| Algeria | Ministère de la Santé et de la Population | (1996) Enquête nationale sur les objectifs de la mi-décennie, "MDG Algérie", 1995. Algiers.' |
| Angola |  | (I999) Inquerito de indicadores multiplos (MICS) 1996. Instituto Nacional de Estatistica, Gabinete de Monitorização das Condições de Vida da População. Luanda. ${ }^{\text {b }}$ |
| Argentina Armenia | Lejarraga H, Krupitzky S, Gimenez E et al. | (1997) The organisation of a national survey for evaluating child psychomotor development in Argentina. Paediatric and Perinatal Epidemiology, II:359-373.b <br> (I998) The health and nutritional status of children and women in Armenia. National Institute of Nutrition, Italy. ${ }^{\text {b }}$ |
| Azerbaijan | Branca F, Burkholder B, Hamel M, Parvanta I, Robertson A | (1996) Health and nutrition survey of internally displaced and resident population of Azerbaijan-April 1996. Baku.b |
| Bahrain | Ministry of Health | (1992) Bahrain child health survey 1989. Manama. |
| Bangladesh | Helen Keller International, Institute of Public Health | (1985) Bangladesh nutritional blindness study, 1982-83: nutritional findings. Dhaka. ${ }^{\text {b }}$ |
| Bangladesh | Government of the People's Republic of Bangladesh | (1987) Report of the child nutrition status module, Bangladesh household expenditure survey 1985-86. Bangladesh Bureau of Statistics, Dhaka. |


| Bangladesh | Government of the People's Republic of Bangladesh | (1991) Report of the child nutrition status survey 1989-90. Bangladesh Bureau of Statistics, Dhaka. |
| :---: | :---: | :---: |
| Bangladesh | Ministry of Planning | (1994) Child nutrition survey of Bangladesh 1992. Bangladesh Bureau of Statistics, Dhaka. |
| Bangladesh Barbados |  | (1997) Bangladesh demographic and health survey 1996-97. (Demographic and Health Surveys.) National Institute for Population Research and Training, Dhaka. ${ }^{\text {b }}$ <br> (1986) National nutrition survey of Barbados, I98I. Caribbean Food and Nutrition Institute, Jamaica. ${ }^{\text {b }}$ |
| Belize | Ministry of Health | (1992) Assessment of the food, nutrition and health situation of Belize. (INCAP Publication DCI/002.) Institute of Nutrition of Central America and Panama, Kingston. |
| Benin | Kodjogbé N, Mboup G, Tossou J et al. | (1997) Enquête démographique et de santé 1996. (Demographic and Health Surveys.) Ministère du Plan, de la Restructuration Economique et de la Promotion de l'Emploi, Cotonou. |
| Bhutan |  | (1989) Bhutan Directorate of Health Services. Report on the national nutrition survey. Bhutan. |
| Bhutan | Ministry of Health and Education | (1999) National anthropometric survey of under five children in Bhutan. Division of Health Services, Thimphu. ${ }^{\text {b }}$ |
| Bolivia | Government of Bolivia | (1982) Bolivia national nutritional status survey, 1981: summary report. National Institute for Food and Nutrition, La Paz. |
| Bolivia | Ministerio de Planeamiento y Coordinacion | (1990) Encuesta nacional de demografia y salud 1989. (Demographic and Health Surveys.) La Paz. ${ }^{\text {b }}$ |
| Bolivia | Ministerio de Planeamiento y Coordinacion | (1992) Situacion alimentaria y nutricional de Bolivia 1992. Instituto Nacional de Alimentacion y Nutricion, La Paz. |
| Bolivia | Ministerio de Desarrollo <br> Sostenible y Medio <br> Ambiente | (1994) Encuesta nacional de demografia y salud 1994. (Demographic and Health Surveys.) La Paz. ${ }^{\text {b }}$ |
| Bolivia | Ministerio de Desarrollo Humano | (1994) Bolivia: mapa de la desnutricion 1990-1992. La Paz. |
| Bolivia | Gutierrez Sardan M | (1997) Encuesta nacional de multiples indicadores 1996 (MICS). Ministerio de Desarrollo Humano, La Paz. ${ }^{\text {b }}$ |

National surveys included in prevalence estimates (continued)

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :--- | :--- | :--- |
| Bolivia | Ministerio de Haciendo, <br> Instituto Nacional <br> de Estadistica | (1998) Encuesta nacional de demografia y salud 1998. (Demographic and Health Surveys.) La Paz. ${ }^{\text {b }}$ |

(1988) Anthropometric assessment of the nutritional status of preschool children in Cape Verde. Bulletin of the World Health Organization, 66:375-386.
(1996) A saúde das crianças menores de cinco anos em Cabo Verde. Ministério de Saúde e Promoção Social and UNICEF, Cape Verde.
(1995) Etat nutritionnel de la population. Rapport préliminaire de l'enquête de nutrition mai-juillet 1995. Bangui.
(1995) Enquête démographique et de santé, République Centrafricaine 1994-95. (Demographic and Health Surveys.) Bangui. ${ }^{\text {b }}$
(1998) Enquête démographique et de santé, Tchad 1996-97. (Demographic and Health Surveys.) Bureau Central du Recensement, Direction de la Statistique, des Etudes Economiques et Démographiques, N'Djamena.b
(1986) Estado nutricional de la poblacion en control de salud. Sistema de vigilancia alimentaria y nutricional. Departamento de Control y Evaluacion, Santiago.
(1988) Situacion alimentaria nutricional de Chile, periodo 1984-87. Universidad de Chile, Santiago.
(1987) Estado nutricional de la poblacion infantil, 1986. Santiago.
(1987) Infant and pre-school nutrition: economical development versus intervention strategies-the case of Chile. Nutrition Research, 7:327-342.
(1994) National health service system. Nutrition Unit, Santiago.
(1996) Alimentacion del menor de 18 meses: relacion con el estado nutricional. Revista Chilena de Pediatria, 67:22-28.
(1995) National health service system. Santiago.
(1997) National health service system. Santiago.
(1999) Boletin anual de vigilancia nutricional, año 1998. Departamento Coordinacion e Informatica, Santiago.
(1995) The dietary and nutritional status of Chinese population (1992 national nutrition survey). Institute of Nutrition and Food Hygiene, Beijing. ${ }^{\text {b }}$
National surveys included in prevalence estimates (continued)

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :---: | :---: | :---: |
| Colombia | Mora JO | (1982) Situacion nutricional de la poblacion Colombiana en 1977-80. Volumen I: Resultados antropometricos y de laboratorio; Comparacion con 1965/66, Ministerio de Salud, Instituto Nacional de Salud, Bogota. |
| Colombia | Mora JO, de Paredes B, de Navarro L, Rodriguez E | (1992) Consistent improvement in the nutritional status of Colombian children between 1965 and 1989. Bulletin of PAHO, 26: I-13. |
| Colombia | Ministerio de Salud | (1988) Tercera encuesta nacional de prevalencia del uso de anticonceptivos y primera de demografia y salud, 1986. (Demographic and Health Surveys.) Institute for Resource Development, Bogota. ${ }^{\text {b }}$ |
| Colombia |  | (1995) Encuesta nacional de demografia y salud 1995. (Demographic and Health Surveys.) Bogota. ${ }^{\text {b }}$ |
| Comoros | Ministère de la Santé Publique et de la Population | (I995) Rapport sur l'état nutritionnel et les facteurs impliqués chez les enfants de moins de deux ans en République Fédérale Islamique des Comores 1991. Direction de la Santé Familiale, Comoros. |
| Comoros | Mondoha KA, Schoemaker J, Barrère M | (1997) Enquête démographique et de santé, Comores 1996. (Demographic and Health Surveys.) Centre National de Documentation et de Recherche Scientifique, Moroni. |
| Congo | Cornu A, Delpeuch F, Simondon F et al. | (1990) Enquête nationale sur l'état nutritionnel des enfants d'age préscolaire au Congo. (Collection Etudes et Thèses.) ORSTOM, Institut Francais de Recherche Scientifique pour le Développement en Coopération, Paris. |
| Costa Rica | Ministerio de Salud | (1982) Encuesta nacional de nutrition 1982. Departamento de Nutricion, San José. |
| Costa Rica | Ministerio de Salud | (1994) Analisis del estado nutricional de la poblacion Costarricense 1992. Departamento de Nutricion y Atencion Integral, Seccion Vigilancia Nutricional, San José. |
| Costa Rica | Ministerio de Salud | (1996) Estado nutricional de preescolares atendidos por el programa de atencion primaria. Departamento de Nutricion, Seccion de Vigilancia Nutricional, San José. |
| Costa Rica | Ministerio de Salud | (1996) Encuesta nacional de nutricion: I fasciculo antropometria. San José. |
| Côte d'lvoire | Sahn DE | (1990) Malnutrition in Côte d'Ivoire, prevalence and determinants. (Working Paper No. 4.) World Bank, Washington, DC. ${ }^{\text {b }}$ |
| Côte d'lvoire | Sombo N'Cho, Kouassi L, Kouamé Koffi A et al. | (1995) Enquête démographique et de santé, Côte d'lvoire 1994. (Demographic and Health Surveys.) Abidjan. ${ }^{\text {b }}$ |

Democratic
Ministère du Plan et
Reconstruction Nationale
Ministère du Plan et
Reconstruction Nationale
Ministère de la Santé
Publique et des Affaires Sociales
Freire W, Dirren
Mora J et al.
(1989) Egypt demographic and health survey 1988. (Demographic and Health Surveys.) Egypt National Population Council, Cairo. ${ }^{\text {b }}$ Statistics, Cairo. ${ }^{\text {b }}$
(1992) Egyptian maternal and child health survey. (PAPCHILD Surveys.) Agency for Public Mobilisation and
(I993) Egypt demographic and health survey 1992. (Demographic and Health Surveys.) Cairo.b
(1995) National survey for assessment of vitamin "A" status in Egypt. Nutrition Institute, Cairo.
(I996) Egypt demographic and health survey 1995. (Demographic and Health Surveys.) National Population
(1996) Enquête nationale sur la situation des enfants et des femmes au Zaire en 1995. Kinshasa.
(1990) Enquête couverture vaccinale malnutrition. République de Djibouti. Djibouti. Tourisme, Direction Nationale de la Statistique, Djibouti.
(1997) Enquête djiboutienne auprès des ménages indicateurs sociaux (EDAM-IS 1996). Ministère du Commerce et du
(1987) Encuesta demografica y de salud 1986. (Demographic and Health Surveys.) Consejo Nacional de Poblacion y Familia, Santo Domingo. ${ }^{\text {b }}$
(1992) Encuesta demografica y de salud 1991. (Demographic and Health Surveys.) Santo Domingo.b
(1997) Encuesta demografica y de salud 1996. (Demographic and Health Surveys.) Centro de Estudios Sociales y Demograficos, Asociacion Dominicana Pro Bienestar de la Familia, Oficina Nacional de Planificacion, Santo Domingo. ${ }^{\text {b }}$
(1988) Diagnostico de la situacion alimentaria, nutricional y de salud de la poblacion ecuatoriana menor de cinco años. Ministerio de Salud Publica y Consejo Nacional de Desarrollo, Quito.
(1978) National nutrition survey, 1978. Cairo. ${ }^{\text {b }}$


> Council, Cairo. ${ }^{\text {b }}$
continued
National surveys included in prevalence estimates (continued)

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :---: | :---: | :---: |
| Egypt |  | (1998) Egypt demographic and health survey 1997. (Demographic and Health Surveys.) El-Zanaty and Associates, Cairo. |
| El Salvador | Trowbridge FL | (1984) Center for Disease Control Survey (CDC): Data on El Salvador, 1975. Atlanta, GA. |
| El Salvador |  | (1990) Evaluacion de la situacion alimentaria nutricional en EI Salvador (ESANES-88). Ministerio de Salud Publica y Asistencia Social, San Salvador. |
| El Salvador | Salvadoran Demographic Association | (1994) National family health survey 1993 (FESAL-93). Government of El Salvador, San Salvador. ${ }^{\text {b }}$ |
| Eritrea | Ministry of Finance and Development | (1994) Children and women in Eritrea: situation analysis, 1994. Government of the State of Eritrea and UNICEF, Asmara. |
| Eritrea |  | (1997) Eritrea demographic and health survey 1995. (Demographic and Health Surveys.) Asmara. |
| Ethiopia |  | (1993) Report on the national rural nutrition survey, core module. (Statistical Bulletin No II3.) Transitional Government of Ethiopia, Central Statistical Authority, Addis Ababa. |
| Fiji |  | (1995) 1993 National nutrition survey-main report. National Food and Nutrition Committee, Suva. |
| Gambia |  | (I997) Report of the progress of the mid-decade goals in the Gambia (MICS), 1996 (draft). Central Statistics Department, Banjul. |
| Ghana | Alderman H | (1989) Nutritional status in Ghana and its determinants. (Working Paper No. 3.) World Bank, Washington, DC. (And additional analysis on the Living Standards Survey, Ghana 1987-88.) |
| Ghana |  | (1989) Ghana demographic and health survey 1988. (Demographic and Health Surveys.) Ghana Statistical Service, Accra. ${ }^{\text {b }}$ |
| Ghana |  | (I994) Ghana demographic and health survey 1993. (Demographic and Health Surveys.) Ghana Statistical Service, Accra. ${ }^{\text {b }}$ |
| Ghana |  | (I999) Ghana demographic and health survey 1998. (Demographic and Health Surveys.) Ghana Statistical Service, Accra. ${ }^{\text {b }}$ |
| Guatemala | Delgado H, Hidalgo E, Giron EM, Pareja G | (1989) Encuesta nacional de salud materno infantil 1987. (Demographic and Health Surveys.) Ministerio de Salud Publica y Asistencia Social, Guatemala.b |

Guatemala


N
$\frac{\text { 프N }}{\text { 포N }}$
$\frac{: \overline{ }}{\mathbb{T}}$
Haiti
Honduras
Honduras
Honduras
 믐
India
National surveys included in prevalence estimates (continued)

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :--- | :--- | :--- |
| India | Vijayaraghavan K, <br> Hanumantha Rao D <br> International Institute for <br> Population Sciences | (I998) Diet and nutrition situation in rural India. Indian Journal of Medical Research, I 08:243-253. ${ }^{\text {b }}$ |
| India |  |  |
| (I995) National family health survey, India I992-93. (Demographic and Health Surveys.) Bombay. |  |  |

(1998) Jamaica survey of living conditions, 1997. The Planning Institute and the Statistical Institute of Jamaica, Kingston. ${ }^{\text {b }}$
(1992) Jordan population and family health survey 1990. (Demographic and Health Surveys.) Ministry of Health, Amman. ${ }^{\text {b }}$
(1998) Jordan population and family health survey 1997. (Demographic and Health Surveys.) Department of Statistics, Amman. ${ }^{\text {b }}$
(1996) Kazakhstan demographic and health survey 1995. (Demographic and Health Surveys.) Almaty.b
(1991) Fourth rural child nutrition survey, 1987. Central Bureau of Statistics, Ministry of Planning and National Development, Nairobi.
(1994) Kenya demographic and health survey 1993. (Demographic and Health Surveys.) Central Bureau of Statistics, Nairobi. ${ }^{\text {b }}$
(1995) Fifth child nutrition survey, 1994. Welfare monitoring survey. Nairobi.
(1999) Kenya demographic and health survey 1998. (Demographic and Health Surveys.) Central Bureau of Statistics, Nairobi.
(I990) National nutrition survey, 1985 (draft version). Government of Kiribati, South Tarawa.
(1985) Kuwait nutritional survey: comparison of the nutritional status of Kuwaiti children, 0-5 years with the NCHS/CDC reference. World Health Organization Bulletin, 63:521-526.
(1996) Nutritional status survey of preschool children in Kuwait. Eastern Mediterranean Health Journal, 2:386-394. (1998) Kyrgyz Republic demographic and health survey 1997. (Demographic and Health Surveys.) Bishkek City. ${ }^{\text {b }}$ (1995) Diagnostic de la situation nutritionnelle et consommation alimentaire au Laos. Rapport complet de l'étude sur l'état nutritionnel de la population laotienne. (ESNA: TCP/LAO/2354.) Food and Agriculture Organization of the United Nations, Rome.
(1994) Women and children in the Lao People's Democratic Republic. Results from the Lao social indicator survey (LSIS). Mother and Child Institute, Vientiane.
Zou'bi A, Poedjastoeti S,
Ayad M
 National Council for Population and Development Ministry of Health and Family Planning Bayoumi A, Moussa MAA
Amine EK, Al-Awadi FA
Ministry of Public Health Nutrition
Jamaica
Jordan
Jordan
Kazakhstan
Kenya
Kenya
Kenya
Kenya
Kiribati
Kuwait
Kuwait
Kyrgyzstan
Lao People's
Democratic
Rep Democratic
Republic
National surveys included in prevalence estimates (continued)

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :---: | :---: | :---: |
| Lebanon |  | (1996) Lebanon maternal and child health survey (preliminary report). (PAPCHILD Surveys.) The League of Arab States, Cairo. |
| Lesotho |  | (1977) Lesotho national nutrition survey. The Government of Lesotho, Maseru. |
| Lesotho | Ministries of Health and Agriculture | (1992) National nutrition survey report, May-June 1992. Maseru. |
| Lesotho | Ministry of Health | (1994) National survey on iodine, vitamin A and iron status of women and children in Lesotho. Maseru. |
| Lesotho | Spring CA | (I996) Mid-decade goals: progress towards the world summit, May 1996 (MICS). Bureau of Statistics and UNICEF, Maseru. |
| Liberia | Ministry of Health and Social Welfare | (1978) Liberia national nutrition survey 1975-1976. Monrovia. |
| Libyan Arab Jamahiriya |  | (1997) Libyan maternal and child health survey. (PAPCHILD Surveys.) The League of Arab States, Cairo. |
| Madagascar | Adrianasolo R | (1986) Etude des aspects épidémiologiques de l'allaitement maternel à Madagascar. Laboratoire Central de Nutrition MINSAN, Antananarivo. |
| Madagascar | Refeno G, Rabeza V, <br> Mboup G, Schoemaker J | (1994) Enquête nationale démographique et sanitaire 1992. (Demographic and Health Surveys.) Centre National de Recherches sur l'Environnement, Antananarivo. ${ }^{\text {b }}$ |
| Madagascar | Institut National de la Statistique | (1995) Enquête permanente auprès des ménages—rapport principal, décembre 1995. Antananarivo. |
| Madagascar | Institut National de la Statistique | (1996) Enquête par grappes à indicateurs multiples (résultats préliminaires). Multiple Indicators Cluster Survey. Antananarivo. |
| Madagascar |  | (1998) Enquête démographique et de santé, Madagascar 1997. (Demographic and Health Surveys.) Institut National de la Statistique, Antananarivo. ${ }^{\text {b }}$ |
| Malawi | Malawi Government | (1984) National sample survey of Agriculture 1980/8I. Vol. III. National Statistics Office, Zomba. |
| Malawi |  | (1993) Malawi demographic and health survey 1992. (Demographic and Health Surveys.) National Statistics Office, Zomba. ${ }^{\text {b }}$ |

Malawi
Malaysia
Malaysia
Maldives
Maldives Maldives
Mauritania
Mauritania Mauritania Ministry of Planning Ministère du Plan, Direction des Ressources Humaines

## Elder JA

 Traoré B, Konaté M,Stanton C
Coulibaly S, Dicko F,
Traoré SM, Sidibé O,
Seroussi M, Barrère B Planification et de Statistique, Bamako. ${ }^{\text {b }}$
(1996) Enquête démographique et de santé Mali 1995-1996. (Demographic and Health Surveys.) Cellule de
(I996) Enquête à indicateurs multiples au Mali (EIM) 1996 (MICS). Rapport d'analyse. Direction Nationale de la Statistique et de l'Informatique (DNSI), Bamako.
(1990) The socioeconomic determinants of nutritional status among children under five in Mauritania. (Social dimensions of adjustment surveys.) World Bank, Washington, DC.
(1992) Mauritania maternal and child health survey 1990-91. (PAPCHILD surveys.) Nouakchott. ${ }^{\text {b }}$
(1996) Enquête nationale sur les indicateurs des objectifs à mi-terme en Mauritanie (MICS). Nouakchott.
(1988) Mauritius national nutrition survey 1985: summary report. Evaluation and Nutrition Unit, Port Louis.
(1996) A survey on nutrition in Mauritius and Rodrigues, 1995 (final report). Port Louis.
National surveys included in prevalence estimates (continued)

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :---: | :---: | :---: |
| Mexico |  | (1990) Encuesta nacional de alimentacion en el medio rural, 1974. Instituto de la Nutricion 'Salvador Zubiran', Comision Nacional de Alimentacion, Division de Nutricion, INNSZ, Tlalpan, Mexico City. |
| Mexico |  | (1990) Encuesta nacional de alimentacion en el medio rural, 1979. Instituto de la Nutricion 'Salvador Zubiran', Comision Nacional de Alimentacion, Division de Nutricion, INNSZ, Tlalpan, Mexico City. |
| Mexico | Sepulveda AJ, Lezana MA, <br> Tapia Conyer R, <br> Valdespino IL, Madrigal H, <br> Kumate J | (I990) Estado nutricional de preescolares y las mujeres en Mexico: resultados de una encuesta probabilistica nacional. Gaceta Medica de Mexico, 126:207-226. ${ }^{\text {b }}$ |
| Mexico |  | (1990) Encuesta nacional de alimentacion en el medio rural, 1989 (ENAL). Instituto de la Nutricion 'Salvador Zubiran', Comision Nacional de Alimentacion, Division de Nutricion, INNSZ, Tlalpan, Mexico City. ${ }^{\text {b }}$ |
| Mongolia | Kachondham Y | (1992) Report of a consultancy on the Mongolian child nutrition survey. Institute of Nutrition, Nakornpathom. |
| Mongolia | Kachondham Y | (2000) Report on the 2nd national child and nutrition survey, Mongolia 1999. Institute of Nutrition and Faculty of Medicine, Ramathibodi Hospital, Mahidol University. ${ }^{\text {b }}$ |
| Morocco | Azelmat M, Ayad M, Belhachmi H | (1989) Enquête nationale sur la planification familiale, la fécondité et la santé de la population au Maroc (ENPS) 1987. (Demographic and Health Surveys.) Ministère de la Santé Publique, Rabat. ${ }^{\text {b }}$ |
| Morocco | Azelmat M, Ayad M, Housni El A | (1993) Enquête nationale sur la population et la santé (ENPS-II) 1992. (Demographic and Health Surveys.) Rabat. ${ }^{\text {b }}$ |
| Mozambique | Government of Mozambique, UNICEF | (1996) Multiple indicator cluster survey Mozambique-1995. MICS surveys. Ministry of Planning and Finance, Maputo. |
| Mozambique |  | (1998) Moçambique inquérito demográfico e de saúde 1997. (Demographic and Health Surveys.) Instituto Nacional de Estatística, Maputo. ${ }^{\text {b }}$ |
| Myanmar | Daw Cho Nwe Oo | (1981) Feeding practices in infants and young children in Rangoon Division. 1980-1981. Rangoon. |
| Myanmar | Daw Cho Nwe Oo | (1986) Feeding practices in young children and infants. Department of Medical Research, Rangoon. |
| Myanmar | Ministry of Health | (1991) Nutrition situation of Myanmar children. Preliminary report of the national nutrition survey 1990. Rangoon. |

Myanmar
Myanmar
Myanmar
Myanmar帚
Z
Z $\qquad$ Nepal
Nepal Aryal, Gokarna Regmi,

Bharat Baan, Pavalavalli Govindasamy

Ministerio de Salud
Ministry of Health
Ministerio de Salud
Ministère de la Santé Publique et des Affaires Sociales

Kourguéni IA, Garba B, Barrère B
(1995) National nutrition survey, 1994. National Nutrition Centre, Yangon.
(1995) Monitoring progress toward the goals of the World Summit for Children through multiple indicator cluster survey (MICS). Yangon.
(2000) National nutrition survey 1997. National Nutrition Centre, Yangon. ${ }^{\text {b }}$
(1993) Namibia demographic and health survey 1992. (Demographic and Health Surveys.) Windhoek. ${ }^{\text {b }}$
(1975) Nepal nutrition status survey, January-May 1975. Kathmandu. ${ }^{\text {b }}$
(1996) Nepal multiple indicator surveillance: cycle I, January-March 1995 health and nutrition-final report (MICS). Kathmandu. ${ }^{\text {b }}$
(I997) Nepal family health survey, 1996. (Demographic and Health Surveys.) Ministry of Health, Kathmandu.b
(1988) Enfoque de riesgo y estado nutricional de los ninos menores de 5 anos en la region III, 1988. Centro de investigaciones y estudios de la salud, Managua.
(I997) Nicaragua 1993 living standards measurement survey (LSMS). World Bank, Washington, DC.b (1999) Encuesta nicaraguense de demografia y salud 1998. (Demographic and Health Surveys.) Instituto Nacional de Estadisticas y Censos, Managua. ${ }^{\text {b }}$

> (I985) Enquête nationale sur la morbidité et la mortalité, rapport No I. Cellule de Planification, Niamey.
(1993) Enquête démographique et de santé, Niger 1992. (Demographic and Health Surveys.) Niamey. ${ }^{\text {b }}$
National surveys included in prevalence estimates (continued)

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :--- | :--- | :--- |
| Niger | Attama S, Seroussi M, <br> Kourguéni Al, Koché H,, | (1998) Enquête démographique et de santé, Niger 1998. (Demographic and Health Surveys.) Care International, <br> Niger and Macro International Inc., Calverton, MD. ${ }^{\text {b }}$ |
|  | Barrère B |  |

Papua New
Jenkins C, Zemel B
Carron JM, Loaiza E,
Ministerio de Salud
Padilla A, Ochoa LH,
Marckwardt AM
Food and Nutrition Research Institute
National Economics and Statistics Section
Department of Science and Technology
Department of Science and Technology
Department of Science and Technology Amine EK
Meheus A, Butera S,
Dindinian O, Eylenbosch
W (1990) Ancient diversity and contemporary change in the growth patterns of Papua New Guinea children. (59th Annual Meeting of the American Association of Physical Anthropologists.) Miami, FL.
(1991) Encuesta nacional de demografia y salud 1990. (Demographic and Health Surveys.) Asuncion. ${ }^{\text {b }}$
(1988) Situacion nutricional en el Peru. Encuesta nacional de nutricion y salud (ENSSA), 1984; and Evaluacion nacional de nutricion por antropometria (ENPPE), 1975. Ministerio de Salud, Lima.
(1992) Encuesta demografica y de salud familiar 1991/1992. (Demographic and Health Surveys.) Lima. ${ }^{\text {b }}$
(1997) Peru demographic and health survey 1996. (Demographic and Health Surveys.) Instituto Nacional de Estadistica e Informacion, Lima. ${ }^{\text {b }}$
(1975) 197I-75 unpublished data. Department of Physiological Hygiene and Nutrition, Institute of Public Health, Manila.
(1982) Second nationwide nutrition survey, Philippines, 1982. National Science and Technology Authority, Manila.
(I99I) Regional updating of nutritional status of Filipino children, 1989-90. Food and Nutrition Research Institute,
(1991) Third national nutrition survey Philippines, 1987. Food and Nutrition Research Institute. Manila. ${ }^{\text {b }}$
(1994) The 1992 regional nutrition survey. Food and Nutrition Research Institute, Manila. ${ }^{\text {b }}$
(1995) The fourth national nutrition survey: Philippines 1993. Food and Nutrition Institute, Manila. ${ }^{\text {b }}$
(1996) Nutritional assessment in Qatar. (Assignment report EM/NUTI69/E/R/0I.96/27.) World Health Organization Regional Office, Alexandria.
(1977) Evaluation de l'état nutritionnel des enfants de 0 à 5 ans dans la République Rwandaise, 1976. Université Instelling, Antwerpen. ${ }^{\text {b }}$
National surveys included in prevalence estimates (continued)

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :--- | :--- | :--- |
| Rwanda | Barrère B, Schoemaker J, <br> Barrère M, Habiyakare T, <br> Kabagwira A, | (I994) Enquête démographique et de santé, Rwanda 1992. (Demographic and Health Surveys.) Kigali. |

(1976) Sri Lanka nutrition status survey, 1976. Colombo. ${ }^{\text {b }}$
(1987) Sri Lanka demographic and health survey 1987. (Demographic and Health Surveys.) Colombo.b
(1979) National nutritional survey 1977-78 (resurvey). Colombo. ${ }^{\text {b }}$
(1995) Sri Lanka demographic and health survey 1993. (Demographic and Health Surveys.) Colombo.
(1997) Preliminary report on the fourth national $n$ of Medical Science, 40:13-24.
(1994) Sudanese maternal and child health survey, 1993. (PAPCHILD surveys.) The League of Arab States, Cairo.
(1987) Acute and chronic undernutrition in Swaziland. Journal of Tropical Pediatrics, 33:35-42.b
(1994) Syrian maternal and child health survey. (PAPCHILD surveys.) The League of Arab States, Cairo.
(1996) Multiple indicator cluster survey in the Syrian Arab Republic (MICS). Central Bureau of Statistics, Damascus.
(I988) Thailand demographic and health survey 1987. (Demographic and Health Surveys.) Institute of Population Studies, Chulalongkorn University, Bangkok. ${ }^{\text {b }}$
(1996) Random survey on nutritional status of children of ages under five. Thailand Journal of Health Promotion and Environmental Health, 19:57-66. ${ }^{\text {b }}$
(1978) Togo nutrition status survey, 1977. Lomé. ${ }^{\text {b }}$
(1989) Enquête démographique et de santé au Togo 1988. (Demographic and Health Surveys.) Direction Générale de la Santé, Lomé. ${ }^{\text {b }}$
(I996) Enquête nationale sur la situation des enfants au Togo en 1995 (MICS) (version préliminaire). Ministère du Plan et de l'Aménagement du Territoire, Lomé.
National surveys included in prevalence estimates (continued)

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :---: | :---: | :---: |
| Togo | Anipah K, Mboup G, Ouro-Gnao AM, Boukpessi <br> B, Messan PA, Salami-Odjo R | (1999) Enquête démographique et de santé, Togo 1998. (Demographic and Health Surveys.) Ministère de la Planification et du Développement Economique, Direction de la Statistique, Lomé. ${ }^{\text {b }}$ |
| Trinidad and Tobago | Gueri M, Andrews N, Jutsum P, Rawlins R | (1980) Nutritional status of young children in Trinidad and Tobago. Journal of Tropical Pediatrics, 26:11-15.b |
| Trinidad and Tobago | Heath K, Da Costa-Martinez D, Sheon AR | (1988) Trinidad and Tobago demographic and health survey 1987. (Demographic and Health Surveys.) Family Planning Association of Trinidad and Tobago, Port-of-Spain. ${ }^{\text {b }}$ |
| Tunisia | Forbes AL, Pelletier O, Lowenstein FW, Lane M, Keller W | (1976) Preliminary report of the 1973-1975 Tunisian national nutrition survey. Tunisian National Institute of Nutrition and Food Technology, Tunis. ${ }^{\text {b }}$ |
| Tunisia | Aloui T, Ayad M, Fourati H | (1989) Enquête démographique et de santé en Tunisie 1988. (Demographic and Health Surveys.) Office National de la Famille et de la Population, Tunis. ${ }^{\text {b }}$ |
| Tunisia | Ministère de la Santé Publique, Institut National de Nutrition et de Technologie Alimentaire | (I998) Enquête nationale 1996-1997. Evaluation de l'état nutritionnel de la population Tunisienne: Rapport national. Sotepa Grafic, Tunis. ${ }^{\text {b }}$ |
| Turkey |  | (1994) Turkey demographic and health survey, 1993. (Demographic and Health Surveys.) Ankara. ${ }^{\text {b }}$ |
| Turkey | Ministry of Health | (1996) Multiple indicator cluster survey in Turkey 1995. National Bureau of Statistics, Ankara. |
| Turkey |  | (I999) Turkish demographic and health survey 1998. (Demographic and Health Surveys.) Hacettepe University, Institute of Population Studies, Ankara. ${ }^{\text {b }}$ |
| Uganda | Kaijuka EM, Kaija EZA, Cross AR, Loaiza E | (1989) Uganda demographic and health survey 1988/89. (Demographic and Health Surveys.) Ministry of Health, Entebbe. ${ }^{\text {b }}$ |
| Uganda |  | (1996) Uganda demographic and health survey 1995. (Demographic and Health Surveys.) Entebbe. ${ }^{\text {b }}$ |


| United Republic of Tanzania | Ngallaba S, Kapiga SH, Ruyobya I, Boerma JT | (1992) Tanzania demographic and health survey 1991/92. (Demographic and Health Surveys.) Dar es Salaam. ${ }^{\text {b }}$ |
| :---: | :---: | :---: |
| United Republic of Tanzania | Bureau of Statistics | (1997) Tanzania demographic and health survey 1996. (Demographic and Health Surveys.) Dar es Salaam. ${ }^{\text {b }}$ |
| Uruguay | Bove MI | (1987) Uruguay: Situacion alimentario-nutricional, algunos factores condicionantes 1970-87. (Document prepared for the ACC/SCN.) Montevideo. |
| Uruguay | Ministerio de Salud Publica | (1996) Sistema de vigilancia epidemiologica nutricional (SISVEN). Departamento de nutricion, Montevideo |
| Uzbekistan | Ministry of Health | (I997) Uzbekistan demographic and health survey 1996. (Demographic and Health Surveys.) Institute of Obstetrics and Gynecology, Tashkent City. ${ }^{\text {b }}$ |
| Vanuatu | Hung MM | (1983) National nutrition survey report. Department of Health, Port Vila. |
| Venezuela | Instituto Nacional de Nutricion | (1986) Encuesta nacional de nutricion 1981-82. Caracas. |
| Venezuela |  | (I995) Proyecto Venezuela 1987. Centro de estudios sobre crecimiento y desarrollo de la poblacion venezolana, Caracas. ${ }^{\text {b }}$ |
| Venezuela | Oficina SISVAN | (1996) Resultados de la evalucion antropometrica del componente menores de 5 años del Sistema de Vigilancia Alimentaria y Nutricional (SISVAN) segun los puntos de corte de la OMS: Venezuela 1990-1993. Instituto Nacional de Nutricion, Caracas. |
| Venezuela | Oficina SISVAN | (1996) Clasificacion antropometrica nutricional, Venezuela 1994. INN-SISVAN componente menores de 15 años. Instituto Nacional de Nutricion, Caracas. |
| Venezuela | Oficina SISVAN | (1998) Evaluacion antropometrica nutricional de los menores de cinco años, para comparacion internacional:Venezuela 1995-1997. Instituto Nacional de Nutricion, Caracas. |
| Venezuela | Oficina SISVAN | (1999) Evaluacion antropometrica nutricional de los menores de cinco años, para comparacion internacional. Venezuela 1990-1998. Instituto Nacional de Nutricion, Caracas. |
| Viet Nam | Tu Giay, Ha Huy Khoi, Luong Tan Thanh | (1986) Some characteristics about protein energy nutritional status in Vietnamese children. In: Applied nutrition, proceedings of the international conference on applied nutrition, Hanoi 25/26 April 1986:201-208. |
| Viet Nam | Ministry of Health | (1991) Report on reanalysed data collected by the General Nutrition Survey 1987-89. Department of Planning, Hanoi. |
| Viet Nam | Bloem MW, Gorstein J | (I995) Viet Nam: Xerophthalmia free; 1994 national Vitamin A deficiency and protein-energy malnutrition prevalence survey. Consultancy report 5-17 March 1995. National Institute of Nutrition, Hanoi. |

National surveys included in prevalence estimates (continued)

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :---: | :---: | :---: |
| Viet Nam |  | (1998) Viet Nam living standards survey 1992-93 (VNLSS). World Bank, Washington, DC. ${ }^{\text {b }}$ |
| Viet Nam | Dibley MJ, Khoi HH, Khan NC et al. | (I999) National protein energy malnutrition survey, Viet Nam 1998. National Institute of Nutrition, Hanoi; and Centre for Clinical Epidemiology and Biostatistics, Newcastle, Australia. ${ }^{\text {b }}$ |
| Viet Nam | Khoi HH, Khan NC, Tuyen LD, Ngu T, Xuan TT | (2000) 1999 Viet Nam —child nutrition situation. The national goal for child malnutrition control. Medical Publishing House, Hanoi. ${ }^{\text {b }}$ |
| Yemen | Ministry of Health | (1992) Yemen maternal and child health survey. (PAPCHILD Surveys.) Sana'a. ${ }^{\text {b }}$ |
| Yemen |  | (1996)Yemen multiple indicator cluster survey (March 1996): Final results. Ministry of Planning and Development, Sana'a. ${ }^{\text {b }}$ |
| Yemen |  | (1998) Yemen democratic and maternal and child health survey 1997. (Demographic and Health Surveys.) Central Statistical Organization, Sana'a. |
| Zambia | Wenlock RW | (1980) Nutritional risk and the family environment in Zambia. Ecology of Food and Nutrition, 10:79-86. ${ }^{\text {b }}$ |
| Zambia | Gaisie K, Cross AR, Nsemukila G | (1993) Zambia demographic and health survey 1992. (Demographic and Health Surveys.) Central Statistical Office, Lusaka. ${ }^{\text {b }}$ |
| Zambia | Cogill B, Zaza M | (1990) Report of the nutrition module as part of the crop forecasting survey—Rural Zambia 1990. Ministry of Health, Lusaka. |
| Zambia |  | (1997) Zambia demographic and health survey 1996. (Demographic and Health Surveys.) Central Statistical Office, Lusaka. ${ }^{\text {b }}$ |
| Zimbabwe | Ministry of Health | (1987) Report of the nutrition component of the national health information system. Harare. |
| Zimbabwe | Ministry of Finance, Economic Planning and Development | (1989) Zimbabwe demographic and health survey, 1988. (Demographic and Health Surveys.) Harare. ${ }^{\text {b }}$ |
| Zimbabwe |  | (1995) Zimbabwe demographic and health survey 1994. (Demographic and Health Surveys.) Harare. ${ }^{\text {b }}$ |

[^7]
## Appendix B

Numbers of countries and population coverage of children aged < 5 Years for underweight, by subregion

| Subregion | No. of countries/total | \% population aged $<5$ years covered |
| :--- | :---: | :---: |
| AFR-D | $23 / 26$ | 99.1 |
| AFR-E | $20 / 20$ | 100 |
| AMR-A | $1 / 3$ | 87.9 |
| AMR-B | $19 / 26$ | 99.8 |
| AMR-D | $6 / 6$ | 100 |
| EMR-B | $10 / 13$ | 81.6 |
| EMR-D | $8 / 9$ | 96.7 |
| EUR-A | $2 / 26$ | 20.5 |
| EUR-B | $5 / 16$ | 58.9 |
| EUR-C | $3 / 9$ | 69.8 |
| SEAR-B | $3 / 3$ | 100 |
| SEAR-D | $6 / 7$ | 98.4 |
| WPR-A | $2 / 5$ | 92.3 |
| WPR-B | $13 / 22$ | 97.4 |

Appendix C
Trends in underweight status; references for AMR-A, EUR-A, and WPR-A

| Country | Author(s) | Reference |
| :---: | :---: | :---: |
| Australia | Margarey AM, Daniels LA, Boulton TJ | (2001) Prevalence of overweight and obesity in Australian children and adolescents: reassessment of 1985 and 1995 data against new standard international definitions. Medical Journal of Australia, 174:561-564. |
| Australia | Booth ML, Wake M, Amstrong T, Chey T, Hesketh K, Mathur S | (2001) The epidemiology of overweight and obesity among Australian children and adolescents, 1995-97. Australian and New Zealand Journal of Public Health, 25:162-169. |
| Canada | Hanley AJG, Harris SB, Gittelsohn J, Wolever TMS, Saksvig, Zinman B | (2000) Overweight among children and adolescents in a native Canadian community: prevalence and associated factors. American Journal of Clinical Nutrition, 71:693-700. |
| France | Deheeger M, Rolland-Cachera MF, Labadie MD, Rossignol C | (1994) Etude longitudinale de la croissance et de l'alimentation d'enfants examinés de l'âge de 10 mois à 8 ans. Cahiers de Nutrition et de Diététique, XXIX:16-23 |
| France | Lehingue Y, Miginiac M, Locard E, Mamelle N | (1993) Birth weight and obesity at the age of 6 . Study from the growth curves of a population of schoolchildren. Pediatrie, 48:623-632. |
| Germany | Schaefer F, Georgi M, Wuhl E, Scharer K | (1998) Body mass index and percentage fat mass in healthy German school children and adolescents. International Journal of Obesity Related Disorders, 22:46I-469. |
| Germany | Kromeyer-Hauschild K, Zellner K, Jaeger U, Hoyer H | (1999) Prevalence of overweight and obesity among school children in Jena (Germany). International Journal of Obesity Related Disorders, 23:1143-1150. |
| Greece | Mamalakis G, Kafatos A, Manios Y, Anagnosto-poulou T, Apostolaki I, | (2000) Obesity indices in a cohort of primary school children in Crete: a six year prospective study. International Journal of Obesity \& Related Metabolic Disorders, 24:765-77I. |
| Japan | Mitsunori Murata | (2000) Secular trends in growth and changes in eating patterns of Japanese children. American Journal of Clinical Nutrition, 72:SI379-I383. |
| Netherlands | Cole TJ, Roede MJ | (1999) Centiles of body mass index for Dutch children aged 0-20 years in 1980—a baseline to assess recent trends in obesity. Annals of Human Biology, 26:303-308. |
| Netherlands | Fredriks AM, van Buuren S, Burgmeijer RJF et al. | (2000) Continuing positive secular growth change in the Netherlands 1955-1997. Pediatric Research, 47:316-323. |

Netherlands
(1984) Paidos'84, estudio epidemiologico sobre nutricion y obesidad infantil. Graficas Jomagar, Mostoles, Madrid.
(2000) Trends in body mass index and overweight prevalence among children and adolescents in the region of Aragon (Spain) from 1985 to 1995. International Journal of Obesity and Related Metabolic Disorders, 24:925-93I.
(1999) Prevalence of childhood overweight in northwestern Spain: a comparative study of two periods with a ten year interval. International Journal of Obesity and Related Metabolic Disorders, 23:1095-1098. (2001) Sociodemographic factors and trends on overweight prevalence in children and adolescents in Aragon (Spain) from 1985 to 1995. Journal of Clinical Epidemiology, 54:921-927.
(1998) Trends in growth and obesity in ethnic groups in Britain. Archives of Disease in Childhood,
(1999) Epidemic of obesity in UK children. The Lancet, 354:1874-1875.
(I999) Prevalence of overweight and obesity in British children: cohort study. British Medical Journal, 319:1039.
(2001) Prevalence and trends in overweight and obesity in three cross-sectional studies of British
children, 1974-94. British Medical Journal, 322:24-26.
(2001) Prevalence of overweight and obese children between 1989 and 1998: population based series of cross-sectional studies. British Medical Journal, 322:313-314.
(200I) Increasing prevalence of obesity in primary school children: cohort study. British Medical Journal, 322:1094-1095.
(2000) Stature, weight, and body mass among young US children born at term with appropriate birth $\qquad$ (2001) Prevalences of overweight in a triethnic pediatric population of San Antonio, Texas. International Journal of Obesity and Related Metabolic Disorders, 25:409-416.
(1998) Increasing prevalence of overweight among US low-income preschool children: the Centers for Disease Control and Prevention pediatric nutrition surveillance, 1983 to 1995. Pediatrics, 101:EI2

Proyecto Universitario
Moreno LA, Sarria A, Fleta J,
Rodriguez G, Bueno M
 Garcia-Mayor EG, Garcia-Mayor RV Moreno LA, Sarria A, Fleta J, Rodriguez G, Gonzalez JM, Bueno M Chinn S, Hughes JM, Rona RJ Reilly JJ, Dorosty AR.
 Chinn S, Rona RJ Bundred P, Kitchine

Bundred P, Kitchiner D, Buchan I Rudolf MC, Sahota P, Barth JH, Walker J.
Overpeck MD, Hediger ML, Ruan Overpeck MD, Hediger ML, Ruan
WJ et al.

Park MK, Menard SW, Schoolfield J Mei Z, Scanlon KS, Grummer-Strawn LM, Freedman DS, Yip R, United Kingdom

United Kingdom United Kingdom

## United Kingdom

United Kingdom
United Kingdom


USA

## Appendix D

National survey data on trends of underweight (<-2 SD weight-for-age) in children aged < 5 years in EUR-B and EUR-C COUNTRIES

| Country | Year of survey | $\begin{gathered} \%<-2 \text { SD } \\ \text { weight-for-age } \end{gathered}$ | Overall trend ${ }^{\text {a }}$ | pp/yr ${ }^{\text {b }}$ | Pop. estimate ${ }^{\text {© }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| EUR-B |  |  |  |  |  |
| Albania | 1996-1998, 2000 ${ }^{\text {d }}$ | $8.1,14.3{ }^{\text {d }}$ | $\uparrow$ | 2.07 | 309 |
| Armenia | 1998 | 3.3 |  |  | 207 |
| Azerbaijan | 1996, 2000 ${ }^{\text {d }}$ | 10.1, 16.8 ${ }^{\text {d }}$ | $\uparrow$ | 1.68 | 609 |
| Bosnia and Herzegovina | $2000^{\text {d }}$ | $4.1{ }^{\text {d }}$ |  |  | 205 |
| Bulgaria | - | - |  |  | 317 |
| Georgia | 1999 | 3.1 |  |  | 299 |
| Kyrgyzstan | 1997 | 11.0 |  |  | 524 |
| Poland | - | - |  |  | 1994 |
| Romania | 1991 | 5.7 |  |  | 1137 |
| Slovakia | - | - |  |  | 289 |
| Tajikistan | - | - |  |  | 773 |
| The former Yugoslav Republic of Macedonia | 1999 | 5.9 |  |  | 145 |
| Turkey | 1993, 1995, 1998 | 10.4, 10.3, 8.3 | $\downarrow$ | -0.42 | 7108 |
| Turkmenistan | - | - |  |  | 602 |
| Uzbekistan | 1996 | 18.8 |  |  | 2761 |
| Former Yugoslavia | 1996, 2000 | 1.6, 1.9 | $\leftrightarrow$ | -0.08 | 640 |
| EUR-C |  |  |  |  |  |
| Belarus | - | - |  |  | 468 |
| Estonia | - | - |  |  | 61 |
| Hungary | 1980-1988 | 2.2 |  |  | 490 |
| Kazakhstan | 1995, 1999 | 8.3, 4.2 | $\downarrow$ | -1.03 | 1273 |
| Latvia | - | - |  |  | 93 |
| Lithuania | - | - |  |  | 186 |
| Republic of Moldova | - | - |  |  | 258 |
| Russian Federation | 1993, 1995 | 4.2, 3.0 | $\downarrow$ | -0.60 | 6362 |
| Ukraine | $2000^{\text {d }}$ | $3.0{ }^{\text {d }}$ |  |  | 2190 |

Key: $\uparrow$, Rising: $\geq 0.30$ percentage points per year; $\leftrightarrow$, Static: $<0.30$ or $>-0.30$ percentage points per year; $\downarrow$, Falling; $\geq-0.30$ percentage points per year.

- No data.
a For countries with no arrow, a trend could not be established.
b Percentage point change per year calculated by dividing the difference between the earliest and the last data points by the number of years between the two surveys. Trends are classified as rising, static or falling according to the cut-offs listed above.
c World population prospects: the 2000 revision; estimates refer to total number ( 000 s ) of children aged <5 years in 2000, sexes combined.
d MICS end-decade draft results.
Appendix E
Trends in underweight status; references for EUR-B and EUR-C


## (a) References for EUR-B

| Country | Author(s) ${ }^{\text {a }}$ | Reference |
| :---: | :---: | :---: |
| Georgia |  | (2000) Georgia multiple indicator cluster survey 1999 (MICS). Tibilisi. ${ }^{\text {b }}$ |
| Hungary | Nemeth A, Eiben OG | (1997) Secular growth changes in Budapest in the 20th century. Acta Medica Auxologica, 29:5-12. |
| a Where <br> b Survey | listed, documents have be reanalysed either by resp | tiple authors such as organizations, institutions and governments. horities or by WHO. |

(b) References for EUR-C
Country
nutrition transition Journal of Nutition, 126:3009-3016.
Russia longitudinal monitoring survey, http://www.cpc.unc.edu/projects/rlms/
(1997) Dynamics of growth and development of rural children in the Republic of Sakha (Yakutia) over a 70 year period. Gigiena i Sanitariia, 4:30-31.
 Voprosy Pitaniia, 69:3-7.

## Chapter 3

# Iron Deficiency anaemia 

Rebecca J. Stoltzfus, Luke Mullany and<br>Robert E. Black

## Summary

Iron deficiency is a highly prevalent form of undernutrition, affecting around one-fourth of the world's women and children, and is one of the most common causes of anaemia.

We conducted comprehensive reviews of published literature linking iron deficiency to disability and death for four potential outcomes: child mortality, perinatal mortality, maternal mortality and mild mental retardation. For all of these outcomes, the best available data were prospective observational studies in which anaemia or haemoglobin concentration was the risk factor. Data on child mortality were not adequate for this task, although a true risk cannot be precluded by the data. Summary relative risks for perinatal mortality ( 10 studies), maternal mortality (six studies) and mental retardation (five studies) were estimated using random effects models (both mortality outcomes) or a fixed-effects model (retardation) and weighting individual estimates by the inverse of their within-study variance. For mortality outcomes, the bivariate relations between haemoglobin and death were used. In two studies of perinatal mortality, unadjusted and multivariate adjusted odds ratios were compared to assess the potential degree of bias in the unadjusted associations. For mental retardation, published multivariate adjusted relations between haemoglobin and IQ were used. Global anaemia prevalence data were supplied by the World Health Organization (WHO), and converted to mean haemoglobin concentrations, assuming normal distribution and observed standard deviations from a large number of studies. To estimate the haemoglobin distribution if iron deficiency were corrected, we assumed the prevalence of anaemia in women and children would be reduced by $50 \%$. On average, for the world, this would increase haemoglobin concentration by about $0.45 \mathrm{~g} / \mathrm{dl}$ (range: $0.0 \mathrm{~g} / \mathrm{dl}$ to $1.28 \mathrm{~g} / \mathrm{dl}$ ).

The relative risks associated with a $1 \mathrm{~g} / \mathrm{dl}$ increase in population mean haemoglobin were 0.75 ( $95 \%$ CI $0.62-0.89$ ) for maternal mortality, 0.72 ( $0.65-0.81$ ) for perinatal mortality and 0.78 ( $0.70-0.86$ ) for mental retardation. Subgroup analyses suggested that the relative risk for perinatal mortality in malaria-endemic regions, 0.65 ( $0.56-0.75$ ), was lower than that in regions without endemic Plasmodium falciparum malaria, 0.80 (0.73-0.87). These estimates were attenuated by $20 \%$ to account for probable bias leading to overestimation of the true relationship. Based on these estimates of iron deficiency anaemia as a risk factor for mortality, iron deficiency is estimated to cause 591000 perinatal deaths and 115000 maternal deaths globally. The associated loss of healthy life years amounts to more than 19 million disability-adjusted life years (DALYs) from perinatal causes and more than 3 million from maternal causes. When the direct sequelae of iron deficiency anaemia are added, the total global burden attributed to iron deficiency anaemia amounts to 841000 deaths and 35057000 DALYs.

The available evidence suggests that iron deficiency anaemia contributes substantially to death and disability in the world. The great majority of this disease burden derives from anaemia in pregnancy and early childhood and is borne by women and children in Asia and Africa. The high global prevalence of anaemia and its potentially associated disease burden, as reflected in these estimates, constitute an urgent research agenda. Because these estimates are uncertain in many respects, their most important use may be to motivate public health scientists to provide definitive evidence.

## 1. Introduction

Iron deficiency is one of the most prevalent nutrient deficiencies in the world, affecting an estimated two billion people (Stoltzfus and Dreyfuss 1998). Young children and pregnant and postpartum women are the most commonly and severely affected because of the high iron demands of infant growth and pregnancy. However, where diets are based mostly on staple foods with little meat intake, or people are exposed to infections that cause blood loss (primarily hookworms and urinary schistosomiasis), iron deficiency may occur throughout the life span. Current WHO/International Nutritional Anemia Consultative Group/United Nations Children's fund/United Nations Children's fund (WHO/INACG/UNICEF) guidelines recommend universal iron and folic acid supplementation of young children and pregnant women where anaemia is highly prevalent (Stoltzfus and Dreyfuss 1998).

Although much is known about iron metabolism, the health consequences of iron deficiency continue to be a subject of research and debate. This is partly because in many regions of the world iron supplements are the standard of care for individuals with anaemia. Most trials of iron supplementation have measured haemoglobin concentra-
tion as the primary outcome. There is a relatively small body of clinical trials of iron repletion to humans with functional iron deficiency (i.e. iron deficiency severe enough to affect erythropoiesis) with pregnancy outcomes or mortality as primary objectives. There is surprisingly little evidence to either support or refute a causal link between iron deficiency and these important adverse health outcomes. As processes like this comparative risk assessment (CRA) bring to light the overall weakness of evidence either supporting or refuting the relationship, new research priorities may emerge.

## 2. Health outcomes considered

In May 2000, a meeting commissioned by WHO, INACG and the Edna McConnell Clark Foundation systematically reviewed the evidence of a causal relationship between iron deficiency or anaemia and six health outcomes: child mortality, maternal mortality, birth outcomes, morbidity, work productivity and child development. These papers were subsequently published in a supplement to the Journal of Nutrition (Beard and Stoltzfus 2001). We began by considering those six outcomes.

Malnutrition (in this case, iron deficiency) may contribute to death and disability through direct sequelae or as a risk factor for death and disability from other causes. The total death and disability attributed to iron deficiency is therefore the sum of its actions as a risk factor and its direct sequelae. The objective of the present paper is to consider iron deficiency as a risk factor for death and disability from other causes.

Of the six outcomes considered above, child mortality, maternal mortality, birth outcomes and morbidity were considered in the framework of iron deficiency as a risk factor. For example, women do not die in childbirth as a direct effect of iron deficiency, but rather die of heart failure due to blood loss, which is made more precipitous by iron deficiency anaemia. Similarly, babies do not die in the perinatal period from iron deficiency, but rather die of other causes, some of which are related to preterm birth, for which maternal iron deficiency is a risk factor. In contrast, decreased work productivity and altered child development (or intelligence) were considered to be direct sequelae of iron deficiency, the assumption being that iron deficiency directly causes decreased oxygen delivery to muscles and the brain.

### 2.1 Outcomes considered

## Child mortality

There is a body of observational evidence linking child mortality to anaemia. However, we could find no published or unpublished studies of reasonably large size that described the relationship between anaemia and cause-specific mortality. Furthermore, nearly all the evidence linking anaemia to overall mortality comes from sub-Saharan Africa or Papua New Guinea, where P. falciparum malaria is a major cause of anaemia,
especially severe anaemia, and malaria is also a major cause of mortality. Thus it seemed unjustified to attribute the observed relationship between anaemia and mortality to iron deficiency in this context, or to generalize to other regions of the world. Brabin et al. (2001a) summarized these data and their interpretation. There are no observational studies linking iron deficiency per se to child mortality, nor any iron supplementation trials with child mortality as outcome. For these reasons, we were unable to estimate the relationship between iron deficiency and cause-specific child mortality. However, it is important to note that the available evidence does not preclude that relationship.

## Maternal mortality

There are a number of observational studies of anaemia and maternal mortality from both Africa and Asia. None of these studies attempted to distinguish iron deficiency anaemia from other causes of anaemia, although several discussed the multiple causes of anaemia within their study population. Where P. falciparum malaria is endemic, it is an important cause of anaemia, especially in first pregnancies (Brabin 1983). However, there is no evidence that malarial infection contributes directly to maternal mortality, even in areas where $P$. falciparum is endemic (Brabin et al. 2001b). Thus in contrast to the situation with child mortality, it is less plausible that malaria would confound the observed relationship between anaemia and maternal mortality (as it likely does with anaemia and child mortality). There are no observational studies linking iron deficiency per se to maternal mortality, nor are there any iron intervention trials with maternal mortality as outcome. Thus we used the available observational data to estimate the relationship between anaemia and maternal mortality.

## Perinatal mortality

There are several observational studies of maternal anaemia and stillbirths, neonatal or perinatal mortality from Africa, Asia, North America, and the United Kingdom of Great Britain and Northern Ireland. None of these studies attempted to distinguish iron deficiency anaemia from other causes linking anaemia. There are no observational studies linking maternal iron deficiency per se to perinatal mortality, and the few published iron intervention trials with perinatal morality as outcome are small or poorly designed. As with maternal mortality, we used the available observational data to estimate the anaemia-perinatal mortality relationship.

## LOW BIRTH WEIGHT

We did not consider low birth weight as a separate outcome in these analyses, but rather assumed that the mortality risk and morbidity burden of anaemia-associated low birth weight was subsumed in our estimate of perinatal mortality as outcome. The relationship between iron status and
low birth weight has been examined in several clinical trials, and it has been the subject of two recent systematic reviews (Mahomed 2000a; Rasmussen 2001). Both concluded that causal evidence from trials is lacking. Rasmussen (2001) noted that there was insufficient evidence from these trials either to support or refute the relationship, because most trials conducted in populations with a significant burden of anaemia have suffered from poor research designs. There is however a substantial body of observational data relating pregnancy anaemia to low birth weight, similar to the observational data relating pregnancy anaemia to maternal and perinatal mortality. Scott Poe and Mary Cogswell (personal communication) have recently completed a meta-analysis of these observational studies. They found that pregnancy anaemia assessed in the first two trimesters of pregnancy was significantly associated with preterm birth (but not intrauterine growth retardation), and that the risk of preterm birth increased with increasing severity of anaemia.

### 2.2 Outcomes considered to be direct sequelae of IRON DEFICIENCY

## Work productivity

There is a substantial body of evidence in animals and humans demonstrating that iron deficiency decreases fitness and aerobic work capacity through mechanisms that include oxygen transport and respiratory efficiency within the muscle (Beard 2001; Haas and Brownlie 2001). This relationship is directly and linearly related to the severity of iron deficiency anaemia. The personal and socioeconomic consequences of this relationship are likely to be real and measurable (Horton and Levin 2001). This consequence of iron deficiency is estimated as a direct sequela of iron deficiency, and is not presented here as a risk factor.

## Intelligence or cognitive capacity

There is a growing body of evidence from animal and human studies that supports a causal relationship between iron deficiency anaemia in early childhood and intelligence in mid-childhood (Beard 2001; Grantham-McGregor and Ani 2001). Although this effect of iron deficiency will be considered as a direct sequela of iron deficiency, we used the observational studies of iron deficiency anaemia in early childhood and measures of intelligence in mid-childhood to obtain a quantitative estimate of this relationship. Different investigators used different measures of cognition, learning or intelligence, making it impossible to summarize all the results. We therefore limited the outcome to global measures of intelligence that were either IQ or on the same scale as IQ (i.e. mean of 100 with standard deviation of 15 points). However, all of these studies measured deviations in intelligence within the clinically normal range (i.e. $\geq 70$ points). In terms of the International Statistical Classification of Diseases and Related Health Problems, ninth revision
(ICD-9) classification, disability in this domain is limited to mental retardation, which these studies do not directly address. We have therefore provided an estimate of the effect of early iron deficiency on mental retardation in mid-childhood, making the controversial assumption that the reported association of iron deficiency anaemia and mean IQ does not affect the variance of IQ. Based on this assumption we can estimate the expected increased risk of $\mathrm{IQ}<70$ (i.e. mild mental retardation) associated with shifts in mean IQ.

## MORBIDITY

The bulk of experimental evidence from iron supplementation trials regards morbidity. There is evidence that sufficient iron is essential for immune function (Beard 2001), and also that excess iron may exacerbate some diseases. The evidence from experimental trials does not suggest that iron supplementation reduces morbidity; in some cases it has been associated with increased morbidity, most notably malaria and respiratory infections in malarious areas (Oppenheimer 2001). However, it is most plausible that this excess risk is associated with therapeutic iron supplementation intended to treat iron deficiency; not that iron deficiency itself is beneficial. Those trials that used low-dose oral supplementation in currently recommended dosages found no adverse effect (INACG 1999). Therefore, we did not estimate a morbidity risk associated with iron deficiency; nor does the available evidence support a risk associated with the correction of iron deficiency by currently recommended public health strategies.

To summarize, we have presented estimates of risk relationships for maternal mortality and perinatal mortality. These were based on anaemia as the indicator of iron deficiency, and only a proportion of that risk is therefore attributed to iron deficiency. We have also presented an estimate of the relationship between iron deficiency anaemia, decreased IQ and mental retardation, although we believe it should be interpreted with extreme caution.

## 3. Nature and definition of the risk factor

Iron is required in all tissues of the body for cellular respiration and many other reduction-oxidation enzyme systems, and has particular functions in muscle, brain and red cells. Critical metabolic functions in these three organs become perturbed at about the same time as animals are depleted of iron (Beard 2001). Anaemia has been used as the hallmark of iron deficiency severe enough to affect tissue function, because red cells are the tissue most amenable to sampling. Although more specific indicators of functional iron deficiency are available, notably erythrocyte protoporphyrin and serum transferrin receptor, there is insufficient data to link those indicators to the health outcomes that fit the construct of this project, namely maternal and perinatal mortality.

It is problematic that all of the available data on maternal and perinatal mortality use anaemia as the indicator, because iron deficiency is not the sole cause of anaemia in most populations. Even within individuals, anaemia may be caused by multiple factors. The available studies do not attempt to separate iron deficiency from anaemia, and there are no good regional estimates of the proportion of anaemia attributable to iron deficiency, although the topic has been discussed and debated (Gillespie and Johnston 1998).

We therefore took the approach of estimating the risk function associated with low haemoglobin, using haemoglobin as a continuous variable. Where studies reported haematocrit instead of haemoglobin, we converted to haemoglobin by dividing haematocrit by 3 .

Because our task was to estimate the burden due to iron deficiency, the counterfactual (i.e. theoretical minimum) distribution should therefore represent the haemoglobin distribution if iron deficiency were eliminated. We assumed that the change in haemoglobin distribution following a supervised period of iron supplementation was a conservative approximation of the virtual elimination of iron deficiency. ${ }^{1}$ By conservative we mean that it is more likely to underestimate the contribution of iron deficiency than to overestimate it, due to problems of noncompliance with supplementation, insufficient dosage, or insufficient duration.

We approached this calculation from two angles. The first approach was to estimate the percentage of anaemia attributable to iron deficiency. Knowing this, we could estimate the shift in the haemoglobin distribution needed to reduce anaemia by that proportion. Beaton recently summarized the per cent reduction in anaemia observed in nine controlled supplementation trials, all conducted in children (Table 3.1). The range of values was wide, $21-85 \%$. Although there is regional diversity in the studies, these data are not sufficient to generate regional estimates. They provide a global average of $51 \%$.

A second and complementary approach was to examine mean changes in haemoglobin attributable to iron supplementation in iron supplementation trials. Sloan et al. (2002) conducted a meta-analysis of haemoglobin response to iron supplementation to pregnant women in randomized controlled trials. Of 70 trials in the literature, 23 met their inclusion criteria, and 15 of those were from developing countries. In studies from developing countries, haemoglobin response was smaller in study samples with higher initial haemoglobin (summarized mean change of $1.13 \mathrm{~g} / \mathrm{dl}$ in those with initial mean haemoglobin $<10.0 \mathrm{~g} / \mathrm{dl}$ compared to $0.85 \mathrm{~g} / \mathrm{dl}$ in those with haemoglobin $11.0-11.9 \mathrm{~g} / \mathrm{dl})$. However, in studies from developed countries, the initial haemoglobin concentrations were uniformly $\geq 11.0 \mathrm{~g} / \mathrm{dl}$, and the effect size was large $(1.17 \mathrm{~g} / \mathrm{dl})$. The response to iron supplementation was strongly related to iron dose, with maximum effects observed in the eight studies that provided a daily dose $\geq 91 \mathrm{mg}$. In these studies the mean haemoglobin response was around

Table 3.1 Estimated proportion of anaemia attributable to iron deficiency

| Site | Age group | Estimated attribution to iron ${ }^{\text {a }}$ |
| :--- | :--- | :---: |
| Bolivia | School children | 84 |
| India (Baroda) | Adolescents | 26 |
| India (Bombay) | Adolescents | 55 |
| India (Delhi) | Adolescents | 41 |
| Indonesia | Adolescents | 63 |
| Mali | Adolescents | 21 |
| Peru | School children | 52 |
| Sri Lanka | Adolescents | 36 |
| Viet Nam | Preschool children | 85 |
| Average |  | 51 |
|  |  |  |
|  | Percentage reduction in anaemia in supplemented group minus per cent age reduction in control |  |
| Source: | Beaton (2002). |  |

$1.8 \mathrm{~g} / \mathrm{dl}$. This maximal response might be considered the best theoretical basis for predicting the effect of eradicating iron deficiency. However, it is likely that the highest doses were also used in studies of more severely anaemic populations. A case can also be made for using 1.17 as the predicted effect size, as this was the average effect seen in women from developed countries, and very similar to that seen in women from developing countries with initially low haemoglobin (1.10-1.13g/dl).

An important question for the present exercise is whether the predicted haemoglobin shift should vary by region. Sloan et al. (2002) did not disaggregate their data by global region, and the studies examined by Beaton are too few to disaggregate (see Table 3.1). If significant regional differences exist in the percentage of anaemia attributed to iron deficiency, the iron-attributable portion might be smaller in Africa, where malaria contributes greatly to the burden of anaemia. Therefore iron would logically claim a smaller portion of total anaemia. Shankar recently summarized data from controlled iron supplementation trials conducted in P. falciparum malaria-endemic populations, including the haemoglobin response attributable to iron (INACG 1999). Eleven studies were included, nine of them from sub-Saharan Africa (Adam 1997; Fleming et al. 1986; Harvey et al. 1989; Lawless et al. 1994; Menendez et al. 1994, 1997; Murray et al. 1978; Oppenheimer et al. 1986; Smith et al. 1989). The studies included younger and older children, and adults, including pregnant women. The change in haemoglobin attributable to iron supplementation in individual studies ranged from 0.3 to $3.6 \mathrm{~g} / \mathrm{dl}$, yielding a weighted average of 1.24 ( $95 \%$ CI 1.16-1.33). Although malaria certainly contributed to anaemia in these populations, it is
remarkable that the haemoglobin response to iron supplementation was similar to that reported by Sloan et al. (2002). Two of the 11 studies in the Shankar analysis were in pregnant women-the subject of analysis by Sloan et al. and the group at risk for the outcomes estimated in this chapter. These two studies from the Shankar analysis both used an iron dose of $60 \mathrm{mg} /$ day, and yield a weighted average haemoglobin response of $0.83 \mathrm{~g} / \mathrm{dl}$. These studies were excluded from the analysis by Sloan et al., but are consistent with the haemoglobin responses at the dosage reported by them: 0.41 in studies of doses $\leq 60 \mathrm{mg} /$ day , and 0.86 in studies of doses $61-90 \mathrm{mg} /$ day.

Thus, the data suggest that iron deficiency is responsible for about $50 \%$ of anaemia, and that, where anaemia is prevalent, elimination of iron deficiency results in a change in mean haemoglobin of about $1.17 \mathrm{~g} / \mathrm{dl}$ or perhaps even higher. The data are lacking in several respects: notably, the studies included in these three meta-analyses did not include non-pregnant adults outside of Africa. From the data at hand, there is no strong basis for altering these values by region. We used both of these lines of evidence to estimate the proportion of the risk associated with anaemia that is attributable to iron deficiency (see section 8).

## 4. Search strategy

### 4.1 Maternal mortality

We based our work on the recent systematic review by Brabin et al. (2001b). Because there are no experimental trials of iron deficiency and maternal mortality, we estimated the risk relationship from observational data. As described by Brabin et al. (2001b), several studies from Nigeria have documented extremely high risks of maternal mortality at haemoglobin concentrations $<5 \mathrm{~g} / \mathrm{dl}$. We decided to limit our description of the haemoglobin-mortality risk relationship to the haemoglobin range of $5-12 \mathrm{~g} / \mathrm{dl}$. Values $<5 \mathrm{~g} / \mathrm{dl}$ are rare on a population basis, being more than 2 standard deviations below the mean in even the most severely anaemic communities, and we did not want those data to influence our risk estimates for the common population ranges. Pregnancy haemoglobin values $>12 \mathrm{~g} / \mathrm{dl}$ were excluded because we judged that variation in this high range is mostly unrelated to iron status, and our ultimate objective was to estimate risk associated with iron deficiency. We thus included those studies that reported mortality rates in at least two haemoglobin groups in the range of $5-12 \mathrm{~g} / \mathrm{dl}$ rates. This excluded three studies: Fullerton and Turner (1962), Johnson and Ojo (1967) and Tasker (1958). We further excluded the study of Chi et al. (1981) from Indonesia, because the haemoglobin categories presented in their Table 5 disagreed with that in the text and we had no basis for determining which was correct. In summary, 10 studies were identified, and six of these were included in the meta-analysis (see Table 3.2 for study descriptions).
Table 3.2 Observational studies of anaemia and maternal mortality

|  |  | Period of data <br> collection | Selection criteria | Time in pregnancy of <br> anaemia assessment | Etiologies of anaemia at population level |
| :--- | :--- | :--- | :--- | :--- | :--- |


| Country (reference) | Method of anaemia assessment | Maternal mortality definition | Level of care | Number of deaths/ total sample | Reasons for exclusion (see text for details) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| India (Konar et al. 1980) | Haemoglobin level at time of admission, method not stated | Death of any woman dying of any cause while pregnant or within 42 days of term of pregnancy | Eden Hospital, Calcutta (referral hospital for urban and rural population) | 637/114698 |  |
| India (Sarin 1995) | Haemoglobin determined by Tallqvist screen for anaemia followed by Sahli acid haematin method for anaemics | Death of pregnant women at delivery according to hospital records (1982-1994) | Referral hospital in area | 339/38565 |  |
| Indonesia (Chi et al. 1981) | Haemoglobin at time of admission | Hospital puerperal mortality rate | 12 teaching hospitals (urban and rural) | 135/36062 | Haemoglobin categories difficult to interpret |
| Malaysia (Llewellyn-Jones 1965) | Haemoglobin, method not stated | Death at childbirth | Maternity Hospital, Kuala Lumpur | 283/73 048 |  |
| Malaysia (Tasker 1958) | Haemoglobin determined by Tallqvist screen for anaemia followed by Sahli acid haematin method for anaemics | Death at childbirth | Institute of Medical Research, Kuala Lumpur | 132/28720 | Less than two groups with haemoglobin midpoint $>5 \mathrm{~g} / \mathrm{dl}$ |

Observational studies of anaemia and maternal mortality (continued)

| Country (reference) | Method of anaemia assessment | Maternal mortality definition | Level of care | Number of deaths/ total sample | Reasons for exclusion (see text for details) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Nigeria (Fullerton and Turner 1962) | Haematocrit | Death at childbirth | University College Hospital, Ibadan | 18/92 | Less than two groups with haemoglobin midpoint $>5 \mathrm{~g} / \mathrm{dl}$ |
| Nigeria (Johnson and Ojo 1967) | Microhaematocrit prior to amniocentesis, bone marrow at delivery | Deaths in anaemic and non-anaemic women 20-32 weeks pregnant | University College Hospital, Ibadan | 9/234 | Less than two groups with haemoglobin midpoint $>5 \mathrm{~g} / \mathrm{dl}$ |
| Nigeria (Harrison 1975) | Haematocrit | Death in pregnancy, labour or puerperium | University College Hospital, Ibadan | 10/401 <br> (excluding group with haematocrit <14\%) |  |
| Nigeria (Harrison 1982) | Haemoglobin from capillary sample, method not stated | Death in pregnancy, labour or puerperium | Hospital, Zaria | 8/258 |  |
| Nigeria (Harrison 1985) | Haematocrit | Deaths during pregnancy and up to 42 days afterward | Hospital access but many home deliveries | 121/1 777 |  |
| - No data. |  |  |  |  |  |
| Note: Shaded rows indicate excluded studies. |  |  |  |  |  |

### 4.2 Perinatal mortality

As with maternal mortality, we estimated the risk relationship from observational studies, using pregnancy or delivery maternal haemoglobin concentration as the risk factor. Published trials of iron supplementation that reported perinatal or neonatal mortality as an outcome were not used as a basis for our risk estimate because the women in the trial were not anaemic (Hemminki and Rimpelä 1991) or because they were small or poorly designed to test the effect of iron (Agarwal et al. 1991; Fleming et al. 1986).

We based our search on the recent systematic reviews of Brabin et al. (2001b), Rasmussen (2001) and Xiong et al. (2000). We added to this one unpublished study by Dreyfuss et al. in which one of us was involved. Xiong et al. described the relationship between pregnancy anaemia and perinatal outcomes in 16936 women in China. This study is published only as an abstract (Xiong et al. 1996); however, the authors provided the data we needed to include here. In summary, 13 studies were identified and 10 were included in the meta-analysis (see Table 3.3 for study descriptions).

### 4.3 Child development

We based our work on the recent systematic review by GranthamMcGregor and Ani (2001). We were interested in estimating the risk of continuous decrement in cognitive function or capacity in children who were iron-deficient anaemic in early childhood. Thus we limited our meta-analysis to those studies that identified iron-deficient anaemic and non-anaemic infants and toddlers and then compared their intelligence at age $2-7$ years. We further limited the meta-analysis to studies that used standardized tests on a scale of 100 with standard deviation 15 (i.e. IQ tests and the Bayley Mental Development Index). Seven different longitudinal studies were described by Grantham-McGregor and Ani (Table 3.4). The study by Hurtado et al. (1999) was excluded because the outcome measure was placement in special education, rather than a measure of intelligence. Similarly, the study by Dommergues et al. (1989) was excluded because the outcome measure (Brunet-Lezine test) did not meet our criterion for a summarizable outcome. The two longitudinal studies by Lozoff et al. followed the same cohort of children; the data from the 1991 publication were used in this analysis. The two studies of Wasserman et al. also followed the same cohort of children. The data from the 1992 publication were used in this analysis because nearly $40 \%$ of the cohort was lost to follow-up by the time of the evaluation of the children at four years of age in the 1994 publication. The five studies that were included in our meta-analysis are described in Table 3.4.
Observational studies of anaemia and perinatal mortality

| Country (reference) | Site | Period of data collection | Selection criteria | Time in pregnancy of anaemia assessment | Etiologies of anaemia |
| :---: | :---: | :---: | :---: | :---: | :---: |
| China (Xiong 1996) | Suzhou | 1989-1990 | Perinatal care monitoring records | At entry into prental care; about 12 <br> gestational weeks, and again at 32 weeks | Not described |
| India (Sarin 1995) | Punjab | 1990-1994 | Population-based survey of pregnant women | Different stages during pregnancies | Iron deficiency anaemia |
| Kenya (Macgregor 1963) | Mombasa | 1957-196\| | Patients at Lady Grigg Maternity Hospital | Within 48 hours of onset of labour | Malaria and iron deficiency are discussed |
| Malaysia <br> (Llewellyn-Jones <br> 1965) | Kuala Lumpur | 1953-1962 | Women treated in maternity hospital in given time period | Not stated | Iron deficiency anaemia: (i) hookworm; (ii) diet. Megaloblastic anaemia: (i) liver damage in malnutrition; (ii) diet-amount of folic acid, haemolytic, normoblastic anaemia |
| Malaysia (Tasker 1958) | Kuala Lumpur | 1952-1958 | Pregnant women with haemoglobin levels $<45 \%$ | At delivery, upon admission | Iron deficiency and some megaloblastic anaemia |
| Nepal (Dreyfuss and West 2001) | Sarlahi | 1994-1996 | Community-based sample of pregnant women enrolled in vitamin A trial | Mid-pregnancy | Iron deficiency, other micronutrient deficiencies, Plasmodium vivax malaria, hookworms |
| Nigeria (Harrison 1975) | Ibadan | 1957-1968 | Pregnant women with haematocrit $\leq 26 \%$ | At delivery, upon admission | Red cell haemolysis due to malaria, dietary deficiency of folates, and haemoglobinopathies |


| Nigeria (Johnson and Ojo 1967) | Ibadan | 1961 | 20-32 weeks pregnant, anaemic (haematocrit $\leq 24 \%$ ) women | 20-32 weeks | Haemolysis; folic acid deficiency |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Nigeria (Harrison 1982) | Zaria | 1976 | Pregnant women who develop anaemia | Mid-pregnancy | Nutritional deficiency of iron and folates, malaria, haemoglobinopathies, blood loss, bacterial infections, and socioeconomic deprivation |
| Nigeria (Harrison et al. 1985) | Zaria | - | Pregnant women in Zaria area | First attendance at hospital to book for antenatal care or seek emergency care | Malaria, iron deficiency, haemoglobinopathies |
| Papua New Guinea (Mola et al. 1999) | Port Moresby | 1987-1992 | Pregnant women booked at antenatal clinics in or around Port Moresby and delivered in Port Moresby General Hospital | Lowest haemoglobin concentration from multiple values, mostly in second half of pregnancy | Malaria, alpha thallasemia, hookworm infection, iron and folate deficiencies |
| United Kingdom (Murphy et al. 1986) | Cardiff, Wales | 1970-1982 | All singleton births to South Glamorgan residents (Cardiff Births Study) | At first booking; 70\% within first 13 weeks, 24\% at 13-19 weeks, 5\% at 20-24 weeks | Social disadvantage |
| USA (Garn et al. 1981) | multicentre | - | National Collaborative Perinatal Project | Lowest haemoglobin concentration in pregnancy from multiple antenatal values | Not described |

Table 3.3 Observational studies of anaemia and perinatal mortality (continued)

| Country (reference) | Method of anaemia assessment | Perinatal mortality definition | Level of care | Number of events/total sample | Reason for exclusion |
| :---: | :---: | :---: | :---: | :---: | :---: |
| China (Xiong 1996) | Not stated | Perinatal mortality | Hospital | 209/16936 |  |
| India (Sarin 1995) | Haemoglobin by Sahli acid haematin method | Perinatal mortality | Referral hospital in area | \| 529/33 160 |  |
| Kenya (Macgregor 1963) | Tallqvist method, confirmed by lab method if $<6 \mathrm{~g} / \mathrm{dl}$ | Stillbirths and neonatal deaths | Maternity hospital | 339/3 950 |  |
| Malaysia (Llewellyn-Jones 1965) | Haemoglobin, method not stated | "Perinatal loss" (premature and mature stillbirth + neonatal) | Maternity Hospital, Kuala Lumpur | $5109 / 73048$ |  |
| Malaysia (Tasker 1958) | Haemoglobin determined by Tallqvist screen for anaemia followed by Sahli acid haematin method for anaemics | Fetal loss (premature and mature) | Institute of Medical Research, Kuala Lumpur | $1676 / 26442$ | Less than two groups with haemoglobin midpoint $>5 \mathrm{~g} / \mathrm{dl}$ |
| Nepal (Dreyfuss and West 2001) | Haemoglobin from venous sample, by Hemocue | Neonatal death, i.e. death in first 28 post-natal days | Rural with little access to obstetric care; nearly all deliveries occurred at home | 59/1 081 |  |
| Nigeria (Harrison 1975) | Haematocrit | Fetal loss | University College Hospital, Ibadan | 17/301 |  |


| Nigeria (Johnson and Ojo 1967) | Microhaematocrit prior to amniocentesis, bone marrow at delivery | Abortions/immature deliveries, stillbirths and neonatal deaths ("total pregnancy wastage") | University College Hospital, Ibadan | 19/145 | Less than two groups with haemoglobin midpoint $>5 \mathrm{~g} / \mathrm{dl}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Nigeria (Harrison 1982) | Haemoglobin from capillary sample, method not stated | Fetal loss | Hospital, Zaria | 36/221 |  |
| Nigeria (Harrison et al. 1985) | Haematocrit | Stillbirth and neonatal deaths | General hospital | \| 834/18|16 <br> (excluding groups with haemoglobin midpoints $<5 \mathrm{~g} / \mathrm{dl}$ ) |  |
| Papua New Guinea (Mola et al. 1999) | Haemoglobin, method not stated | Stillbirths | General hospital | 246/13311 <br> (excluding groups with haemoglobin midpoints $>12 \mathrm{~g} / \mathrm{dl}$ ) |  |
| United Kingdom (Murphy et al. 1986) | Haemoglobin, method not stated | Perinatal mortality | Modern United Kingdom health system | 4 195/36466 (excluding group with haemoglobin midpoint $>12 \mathrm{~g} / \mathrm{dl}$ ) |  |
| USA (Garn et al. 1981) | Haemoglobin or haematocrit | Fetal death | Modern USA health system | \| 196 / >50 000 | Numerators and denominators not tabulated for haemoglobin groups |
| - No data. |  |  |  |  |  |
| Note: Shaded rows in | cate excluded studies. |  |  |  |  |

Table 3.4 Longitudinal observational studies of iron deficiency anaemia and child intelligence ${ }^{a}$

| Country (reference) | Sample size | Period of follow-up | Exclusions | Study design |
| :---: | :---: | :---: | :---: | :---: |
| Chile (de Andraca et al. 1990) | Total $=77$. <br> Formerly anaemic $=4 \mathrm{I}$. <br> Formerly non-anaemic $=29$. <br> All anaemic treated at 12 months | Birth to 5-6 years | BW $<2500 \mathrm{~g}$, chronic ill health, intermediate levels of anaemia | Part of a randomized trial of iron fortification in early infancy, at one year, $25 \%$ of the non-fortified group had anaemia. The anaemic children all received 3 months of iron treatment. Selected children re-examined at 5 to 6 years of age |
| Costa Rica (Lozoff et al. 1991) | 163 of 191 children originally evaluated at 12-24 months. 30 had moderate anaemia $=\mathrm{Hb} \leq 10.0 \mathrm{~g} / \mathrm{dl}$, ferritin $\leq 12 \mathrm{mcg}$, EP $>1.77 \mathrm{mcgmol}$ or transferrin $\leq 10 \%$. 133 comparison group | 12 months to 5 years | BW $<2.5 \mathrm{~kg}$, multiple pregnancy, complicated births, acute or chronic medical problem | Follow-up at 5 years of Lozoff et al. (1987). The IDA group was initially treated for 3 months to to correct their anaemia. Current evaluators blind to original iron status. All now free of anaemia |
| Former Yugoslavia (Wasserman et al. 1992) | Children whose mothers were followed up from pregnancy in two areas of Kosovo. <br> Mitrovica $=$ lead exposed. <br> Pristina = nonlead exposed. <br> 54I agreed to participate. <br> $392(208+184)$ seen at 24 months | Birth to 24 months | Major CNS defects, chromosomal abnormalities, multiple pregnancy | Follow-up two cohorts from birth measuring serum lead, iron status and developmental indices. Related Hb at each age with DQ at 24 months. Anaemic children treated |
| Israel (Palti et al. 1983) | Routine health service screen for Hb at 9 months. <br> Tested at 2 years $=873$. <br> At 3 years $=388$. <br> At 5 years $=239$. <br> Hb only measure of iron status | 9-10 months to 5 years | Not given | Follow-up of all children from 9-10 months to 2,3 and 5 years. All with $\mathrm{Hb}<11 \mathrm{~g} / \mathrm{dl}$ treated with iron at 9 months for 3 months. At 5 years took a random sample of remaining children |
| USA (Cantwell 1974) | 6I full-term neonates from comparable socioeconomic groups: 29 given IM iron in neonatal period 32 infants developed Fe deficiency anaemia between 6-18 months. (Hb $6.1-9.5 \mathrm{~g} \%$ ) without PEM. No details of iron status | Birth to 7 years | Preterm | 29 of 61 infants received iron injections (method of assignment not given) and were not anaemic (Hb II.5-I2.9). 32 infants developed IDA. Examined at 6-7 years by examiners blind to the groups |


| Country (reference) | Outcome measures | Covariates adjusted for | Findings | Remarks |
| :---: | :---: | :---: | :---: | :---: |
| Chile (de Andraca et al. 1990) | Stanford-Binet IQ, Illinois psycholinguistic abilities test, psychoeducational abilities test, Bruininks-Osteretsky test of motor proficiency, VMI, neurological exam | Home, maternal depression and stress. Not clear if used in analysis | Hb at $I$ year $=10.1 \pm 0.7$ vs $13.0 \pm 0.8$. Hb a 15 months $=12.8 \pm 0.7$ vs $13.0 \pm 0.8$. Current Hb level not given. Formerly anaemic children performed significantly worse in IQ $(P=0.02)$, psychoeducational abilities ( $P<0.0 \mathrm{I}$ ), VMI ( $P<0.0 \mathrm{I}$ ), motor proficiency ( $P<0.0 \mathrm{I}$ ), language abilities ( $P<0.01$ ). They were more neurologically immature ( $P<0.01$ ). Their homes were significantly less stimulating and their mothers were more depressed and less affectionate |  |
| Costa Rica (Lozoff et al. 1991) | Current iron status. WISC test, Woodcock Johnson psychoeducational battery, Goodenough-Harris draw-a-man test, Beery developmental test of VMI , Bruininks-Oseretsky test of motor proficiency | Sex, birth weight, mother's IQ, height and education, breastfeeding, absence of father, home | No current difference in Hb and other measures of iron status. After controlling for covariates, previously anaemic group had lower scores on performance IQ, quantitative and visual matching subtests of the Woodcock Johnson battery, the VMI and the Bruininks-Oseretsky test of motor proficiency. In post hoc analyses, children who were non-anaemic but continued to have iron deficiency after treatment also had significantly lower scores | Good covariate control. Verbal skills less affected |
| Former Yugoslavia (Wasserman et al. 1992) | Bayley MDI at 6, 12, 18, 24 months with iron and lead status | Ethnic group, home, birth order, BW, sex, maternal IQ , education and age, lead levels | Hb at 6,12 and 24 months was not significantly associated with MDI at 24 months but Hb at 18 months was significant. Controlling for all covariates, in both | At all ages mothers' education had the most significant effect on DQ |

Table 3.4 Longitudinal observational studies of iron deficiency anaemia and child intelligence ${ }^{\text {a }}$ (continued)

| Country (reference) | Outcome measures | Covariates adjusted for | Findings | Remarks |
| :---: | :---: | :---: | :---: | :---: |
|  |  |  | Mitrovica and Pristina a change in Hb at 18 months of $2.0 \mathrm{~g} / \mathrm{dl}$ was associated with a change of 3.4 MDI points $(~ P=0.02)$. Other indices of iron deficiency not associated with development |  |
| Israel (Palti et al. 1983) | Brunet-Lezine test at 2 years. MILI test (an Israeli intelligence test) at 3 years. Wechsler with Israeli adaptation (WPSSI) at 5 years | Maternal education, father's occupation, BW, sex | When controlling for covariates: Hb at 9 months not significantly associated with DQ at 2 years ( $P=0.105$ ) and at 3 years ( $P=0.07$ ) but at 5 years had a significant effect on IQ ( $P=0.02$ ). At 5 years an increase of $1.0 \mathrm{~g} / \mathrm{dl}$ of Hb associated with 1.75 change in IQ points | A large number of cognitive tests. Previously anaemic not reported alone |
| USA (Cantwell 1974) | Neurological examination and Stanford-Binet IQ | None reported | The formerly anaemic group had a higher incidence of "soft signs" e.g. clumsiness with balancing on one foot, in tandem walking, and repetitive hand and foot movement and were more inattentive and hyperactive than the non-anaemic group. IQ scores averaged 98 in the non-anaemic and 92 in the anaemic. No significance levels levels reported |  |

[^8]
## 5. Methods for COMBINING RISK ESTIMATES FROM INDIVIDUAL STUDIES

### 5.1 Anaemia and maternal mortality

In each study, maternal mortality data were given in aggregate for each of the ranges. All ranges were converted to haemoglobin by dividing haematocrit values by 3 . The midpoint of each range was used as the independent variable. The midpoints were estimated from the information provided in the articles. A logistic regression model was then used to fit the observed data, weighting each haemoglobin midpointmortality point by the total number of women in that range. Within each study, an estimate of the risk ratio associated with a one-unit difference in haemoglobin was calculated. These individual estimates were initially combined in a fixed-effects model, weighting individual estimates by the inverse of their within-study variance, to estimate an overall risk ratio. The heterogeneity statistic indicated that individually observed effect sizes varied significantly around the overall fixed-effects model estimate. Dropping the assumption of a fixed-treatment effect, the individual effect sizes were then assumed to be normally distributed and a random-effects combined estimate was calculated using the method of DerSimonian and Laird (1986).

### 5.2 AnAEMIA and perinatal mortality

Analyses for perinatal mortality were conducted in a similar manner as that for maternal mortality. The nine studies included in the metaanalysis were sufficiently heterogeneous that a random-effects model was used to generate combined estimates (DerSimonian and Laird 1986).

### 5.3 IRON DEFICIENCY ANAEMIA AND CHILD INTELLIGENCE

The beta coefficients from multivariate regression models associated with a $1 \mathrm{~g} / \mathrm{dl}$ change in haemoglobin were obtained directly from the original published paper (Palti et al. 1983; Wasserman et al. 1992) or estimated indirectly from means and $P$-values (Cantwell 1974; de Andraca et al. 1990; Lozoff et al. 1991). Standard deviations were obtained directly from the original paper (Wasserman et al. 1992), estimated from $P$ values and beta coefficients (de Andraca et al. 1990; Lozoff et al. 1991; Palti et al. 1983), or estimated by assuming a significance level of 0.05 (Cantwell et al. 1974). As original data were not available, variability in baseline anaemia levels was not considered; rather, baseline mean haemoglobin levels were compared to follow-up IQ scores with standard deviations to estimate individual study regression coefficients. The estimates were combined in a fixed-effects model, weighting studies according to the reciprocal of their within-study variance. A chi-squared test for heterogeneity found no significant between-study variance; thus a randomeffects model was not necessary.

## 6. Results

### 6.1 Maternal mortality

We computed odd ratios for maternal mortality associated with a $1 \mathrm{~g} / \mathrm{dl}$ increase in pregnancy haemoglobin. Of the six studies included in our meta-analysis, all had individual study ORs $<1.0$, and three of those were statistically significant (Table 3.5 and Figure 3.1). The estimated OR from combining data points from all the studies was 0.75 , with a CI that clearly excluded unity ( $0.62-0.89$ ). The studies were not geographically diverse, coming from only three countries (India, Malaysia and Nigeria). However there was not a systematic difference between the risk estimates from the Nigerian vs the Asian studies. Two Nigerian studies had markedly lower ORs than the other four studies; however these two studies carried little weight in the combined OR.

### 6.2 Perinatal mortality

Ten studies were included in our meta-analysis. The individual ORs for perinatal mortality associated with a $1 \mathrm{~g} / \mathrm{dl}$ increase in haemoglobin ranged from 0.55 to 0.87 (Table 3.6 and Figure 3.2). Nine of the 10 individual study estimates were statistically different from unity. The estimated OR from the 10 studies combined was 0.72 ( $95 \%$ CI $0.65-0.81$ ).

The nine studies included in the meta-analysis were sufficiently heterogeneous that a random-effects model was used to generate combined estimates. We explored three factors that might explain this heterogeneity; these subgroup analyses are presented in Table 3.7. First, we were liberal in accepting various outcome definitions related to perinatal mortality. Only three of the 10 studies used the correct definition, which includes fetal death after 22 (or 28) weeks' gestation and neonatal mortality in the first seven days of life. Use of the correct definition of

Table 3.5 Individual and combined estimates of odds ratio of maternal death

| Country (study) | Point estimate (OR) | $95 \% ~ C l$ |
| :--- | :---: | ---: |
| India (Konar et al. 1980) | 0.61 | $0.57-0.64$ |
| India (Sarin 1995) | 0.84 | $0.81-0.88$ |
| Malaysia (Llewellyn-Jones 1965) | 0.74 | $0.69-0.80$ |
| Nigeria (Harrison 1975) | 0.46 | $0.15-1.42$ |
| Nigeria (Harrison 1982) | 0.38 | $0.14-1.03$ |
| Nigeria (Harrison and Rossiter 1985) | 0.95 | $0.83-1.09$ |
| Combined | 0.75 | $0.62-0.89$ |
| Odds ratio for maternal mortality associated with a I g/dl improvement in haemoglobin concentration, in |  |  |
| the range of 5-I2 g/dl haemoglobin. |  |  |

Figure 3.I Individual and combined estimates of odds ratio of maternal death


Table 3.6 Individual and combined estimates of odds ratio of perinatal death

| Country (study) | Point estimate (OR) ${ }^{\text {a }}$ | 95\% Cl |
| :---: | :---: | :---: |
| China (Xiong et al. 1996) | 0.82 | 0.66-1.03 |
| India (Sarin 1995) | 0.87 | 0.85-0.89 |
| Kenya (Macgregor 1963) | 0.66 | 0.60-0.73 |
| Malaysia (Llewellyn-Jones 1965) | 0.86 | 0.84-0.89 |
| Nepal (Dreyfuss and West 2001) | 0.82 | 0.70-0.97 |
| Nigeria (Harrison 1975) | 0.55 | 0.33-0.92 |
| Nigeria (Harrison 1982) | 0.59 | 0.45-0.78 |
| Nigeria (Harrison et al. 1985) | 0.58 | 0.55-0.61 |
| Papua New Guinea (Mola et al. 1999) | 0.77 | 0.70-0.85 |
| United Kingdom (Murphy et al. 1986) | 0.62 | 0.57-0.68 |
| Combined ${ }^{\text {b }}$ | 0.72 | 0.65-0.81 |

[^9]Figure 3.2 Individual and combined estimates of odds ratio of perinatal death


Table 3.7 Subgroup analyses for perinatal mortality

| Group of studies (n) | Point estimate (OR) | 95\% Cl | \% change from overall estimate |
| :---: | :---: | :---: | :---: |
| All (10) ${ }^{\text {a }}$ | 0.72 | 0.65-0.81 |  |
| True definition of perinatal mortality (3) ${ }^{\text {b }}$ | 0.76 | 0.59-0.98 | +5.5 |
| Outcome includes some components of perinatal mortality (7) ${ }^{\text {c }}$ | 0.70 | 0.58-0.84 | -2.9 |
| P. falciparum endemic (5) ${ }^{\text {d }}$ | 0.65 | 0.56-0.75 | -10.3 |
| P. falciparum not endemic (5) ${ }^{\text {e }}$ | 0.80 | 0.73-0.87 | +10.8 |
| Haemoglobin assessed at delivery (2) ${ }^{\text {f }}$ | 0.66 | 0.59-0.73 | -8.7 |
| Haemoglobin assessed in early-mid pregnancy (8) ${ }^{\text {g }}$ | 0.74 | 0.65-0.83 | +2.5 |

a Includes all 10 studies in Table 3.6.
b China 1996; India 1995; United Kingdom 1986.
c Kenya 1963; Malaysia 1965; Nepal 1998, Nigeria 1975, 1976, 1985; Papua New Guinea 1999.
d Kenya 1963; Nigeria 1975, 1976, 1985; Papua New Guinea 1999.
e China 1996; India 1995; Malaysia 1965; Nepal 1998; United Kingdom 1986.
f Kenya 1963; Nigeria 1975.
${ }^{8}$ China 1996; India 1995; Malaysia 1965; Nepal 1998; Nigeria 1976, I985; Papua New Guinea 1999; United Kingdom 1986.

Table 3.8 Comparison of adjusted and unadjusted odds ratios in two studies of pregnancy anaemia and perinatal mortality

| Hb category | Live births | Neonatal deaths | Death rate (000s) | OR | Adjusted OR |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Study: Nepal $1998^{\mathrm{a}}$ |  |  |  |  |  |
| $\geq 11.0 \mathrm{~g} / \mathrm{dl}$ | 330 | 12 | 36.4 | 1.00 | 1.00 |
| $9.0-10.9$ | 534 | 32 | 60.6 | 1.69 | 1.57 |
| $7.0-8.9$ | 176 | 9 | 51.1 | 1.43 | 1.40 |
| $<7.0$ | 41 | 6 | 146.3 | 4.54 | 4.61 |
| Study: China $1996^{\mathrm{b}}$ |  |  |  |  |  |
| $\geq 10.0 \mathrm{~g} / \mathrm{dl}$ | 15236 | 184 | 12 | 1.00 | 1.00 |
| $7.0-9.9$ | 1332 | 21 | 16 | 1.31 | 1.21 |
| $<7.0$ | 159 | 4 | 25 | 2.08 | 1.81 |

a Adjusted ORs were adjusted for randomized supplement group (i.e. maternal supplementation with vitamin A or beta-carotene during pregnancy; this treatment did not affect neonatal death), primiparity, gestational age at Hb assessment, reported severe illness in late pregnancy, contribution of $\geq 1$ pregnancy to the data set, preterm birth. Source: M. Dreyfuss, personal communication.
b Adjusted for maternal age, gestational age, parity, hypertensive disorders of pregnancy, gestational age at enrolment into prenatal care, hospital and maternal education. Source: X. Xiong, personal communication.
perinatal mortality did not significantly change the effect estimate. Second, P. falciparum malaria contributes to anaemia in pregnancy and may also affect perinatal mortality. Five of the studies were conducted in populations with endemic $P$. falciparum malaria, and these studies had a combined risk estimate that was substantially further from unity than those studies conducted in populations not heavily exposed to this form of malaria. We concluded that the risk relationship with anaemia is greater in $P$. falciparum malaria-endemic regions. Third, two studies assessed haemoglobin at delivery whereas the other eight assessed haemoglobin earlier in gestation. Xiong et al. (2000) have suggested that anaemia early in pregnancy carries a greater risk of adverse perinatal outcomes; however this inference was based on low birth weight as outcome. We did not find this in our data. In fact, the risk relationship was slightly stronger in the two studies that assessed haemoglobin at delivery. However, those two studies were also in P. falciparum malariaendemic populations, which could bias the analysis.

The ORs presented in Table 3.6 and Figure 3.2 are unadjusted for potential confounding factors. Two of the 10 studies, Nepal 1998 and China 1996, could provide us with both unadjusted and adjusted ORs by haemoglobin category. These are displayed in Table 3.8. In the Nepal study, multivariate adjustment for a number of variables had no effect on the estimated ORs. However, in the China study, in which more

Table 3.9 Final odds ratios and confidence intervals used to generate burden of disease estimates

| Outcome | OR estimate | $95 \% \mathrm{Cl}$ |
| :--- | :---: | ---: |
| Maternal mortality | 0.80 | $0.70-0.91$ |
| Perinatal mortality, Africa | 0.72 | $0.65-0.80$ |
| Perinatal mortality, other regions | 0.84 | $0.78-0.90$ |

covariates were measured, multivariate adjustment attenuated the ORs by about $20 \%$.

We have therefore attenuated the ORs for both perinatal and maternal mortality by $20 \%$, as the evidence at hand suggests that the unadjusted estimates may be overestimated to about that degree. The final ORs and confidence intervals used to generate the burden of disease estimates are shown in Table 3.9.

### 6.3 Child intelligence

Five studies were included in our meta-analysis. The studies were geographically diverse, including Europe, Latin America, the Middle East and North America (the United States of America). However, there were no studies included from Africa or Asia. We estimated the expected change in IQ points associated with a $1 \mathrm{~g} / \mathrm{dl}$ change in haemoglobin. The individual study estimates ranged from 1.36 to 2.52 , and all were statistically different from zero (Table 3.10 and Figure 3.3). The combined estimate was 1.73 points, with a $95 \%$ CI of $1.04-2.41$. The quality of the studies was high, and all of them provided estimates that were adjusted for multiple covariates.

The relevant disease outcome for this analysis would be mild mental retardation, defined as $\mathrm{IQ}<70$, or more than 2 standard deviations below the expected population mean of 100 . The increase in risk of mental retardation can be estimated from the expected mean change in IQ if one assumes that the mean change represents a shift in the entire distribution of values with no change in variance of the distribution. These calculations are displayed in Table 3.10. The resultant relative risk associated with a $1 \mathrm{~g} / \mathrm{dl}$ increment in haemoglobin concentration is 0.78 . The relative risk estimates associated with the upper and lower confidence limits of the estimated change in IQ are 0.70 and 0.86 , respectively (Table 3.11).

## 7. DESCRIPTION OF MALNUTRITION TABLES

To describe the distribution of low haemoglobin by region we used the anaemia prevalence data that were published in the 1990 Global Burden

Table 3.10 Individual and combined estimates of the expected difference in IQ points per unit ( $\mathrm{g} / \mathrm{dl}$ ) of haemoglobin

| Study | Point estimate (IQ points) ${ }^{\mathrm{a}}$ | $95 \% \mathrm{Cl}$ |
| :--- | :---: | :---: |
| Wasserman et al. (I992) | 1.68 | $0.32-3.04$ |
| Palti et al. (1983) | 1.75 | $0.29-3.21$ |
| Lozoff et al. (199I) | 2.52 | $0.02-5.02$ |
| Cantwell (1974) | 1.36 | $0.03-2.70$ |
| de Andraca et al. (1990) | 1.96 | $0.35-3.57$ |
| Combined | 1.73 | $1.04-2.41$ |
| Estimated increase in IQ associated with a I g/dl increase in haemoglobin concentration. |  |  |

Figure 3.3 Individual and combined estimates of IQ point difference associated with haemoglobin levels in infancy

of Disease (GBD) books (Murray and Lopez 1996a, 1996b). This database is currently being revised and updated, but the complete new data will not be available within the time frame of this project. We applied the following steps: (i) converted the prevalence data from the GBD 1990 regions to the subregions ${ }^{2}$ used currently; (ii) converted data on prevalence of anaemia into mean haemoglobin values; and (iii) estimated the

Table 3.1I Predicted risks and relative risks of mental retardation, based on combined estimate of difference in mean IQ points in Table 3.10

|  | Point estimate | Lower bound | Upper bound |
| :---: | :---: | :---: | :---: |
| Expected rate of mental retardation in a healthy population ${ }^{\text {a }}$ | 2.3\% |  |  |
| Estimated average decrement in IQ per Ig/dl decrement in haemoglobin ${ }^{\text {b }}$ | 1.73 points | 1.04 points | 2.41 points |
| Predicted rate of mental retardation in population with haemoglobin distribution shifted I g/dl downward due to iron deficiency ${ }^{\text {c }}$ | 2.94\% | 2.68\% | 3.29\% |
| Predicted relative risk of mental retardation associated with I g/dl increment in haemoglobin ${ }^{\text {d }}$ | 0.78 | 0.86 | 0.70 |
| a Assumes mean IQ of 100 , with $\mathrm{SD}=15$. |  |  |  |
| ${ }^{\text {b }}$ From Table 3.10. |  |  |  |
| c Assumes shift in mean given in row 2 with constant $\mathrm{SD}=15$. |  |  |  |
| ${ }^{\text {d }}$ Point estimate in row I divided by percentages in row 3. |  |  |  |

counterfactual haemoglobin distribution, representing the elimination of iron deficiency.

The anaemia database that we used was based on surveys conducted prior to 1990. Global monitoring of anaemia trends suggests that the prevalence of anaemia in the world has not decreased in the past decade (UNICEF 1998). However, the representativeness and reliability of these data (in terms of sample sizes) are less than we would hope for. Table 3.12 summarizes the numbers of country surveys included in the available data set, and the conversion from previous regions to the subregions used in the present analysis. Data for EUR-B and EUR-C are especially scarce. In the case of SEAR-D, the previous data source included only one country, India, but several surveys contributed to the country estimate.

Nationally representative anaemia data from the United States are available that are more recent than the data in the 1990 WHO database, and demonstrate lower anaemia prevalences (Looker et al. 1997). So as not to overestimate the burden of anaemia in economically developed regions, we used these data from the United States for three subregions, AMR-A, EUR-A and WPR-A. The haemoglobin cut-offs used to define anaemia globally have not changed since the 1990 database was created (Table 3.13). The resulting anaemia prevalence estimates are given in Table 3.14.

The second step was to convert the prevalence of anaemia to a mean haemoglobin value. We assumed all haemoglobin distributions to be approximately normal (Yip et al. 1996). We then needed to make

Table 3.12 Conversion of anaemia prevalence data from GBD 1990 regions to subregions

| Subregion | Previous (GBD 1990) region <br> used as data source | Number of countries contributing <br> surveys to regional estimate |
| :--- | :--- | :---: |
| AFR-D | Sub-Saharan Africa | 16 |
| AFR-E | Sub-Saharan Africa | 16 |
| AMR-A | a |  |
| AMR-B | Latin America and Caribbean | 21 |
| AMR-D | Middle Eastern Crescent | 21 |
| EMR-B | Middle Eastern Crescent | 6 |
| EMR-D | a | 6 |
| EUR-A | Former Soviet Economy | 1 |
| EUR-B | Former Soviet Economy | 1 |
| EUR-C | Other Asia and islands | 8 |
| SEAR-B | India | 1 |
| SEAR-D | a |  |
| WPR-A | Other Asia and islands | 8 |
| WPR-B | Lerica and Caribbean |  |

a For these subregions, recent nationally representative data from the USA were used (Looker et al. 1997).

Table 3.13 Haemoglobin cut-offs to define anaemia in populations living at sea level

| Population group | Haemoglobin cut-off $(\mathrm{g} / \mathrm{l})$ |
| :--- | :---: |
| Children 0-4 years | 110 |
| Children 6-14 years | 120 |
| Non-pregnant women | 120 |
| Pregnant women | 110 |
| Men | 130 |

Source: Stoltzfus and Dreyfuss (1998).
assumptions about the standard deviation of the distributions. The standard deviation of haemoglobin in a population depends on at least two factors. The first factor is the proportion of individuals who are at their homeostatic haemoglobin concentration (i.e. non-anaemics). There is a certain amount of variability in haemoglobin that is set by individual characteristics, including genetics. This would be represented by the standard deviation in a population with no anaemia. Added to this "inherent" variability is the variability associated with non-physiologic states,

Table 3.14 Estimated anaemia prevalence by subregion, sex and age

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 60.0 | 52.0 | 28.0 | 28.0 | 28.0 | 47.0 | 47.0 | 47.0 |
|  | Female | 60.0 | 52.0 | 41.0 | 41.0 | 41.0 | 47.0 | 47.0 | 47.0 |
| AFR-E | Male | 60.0 | 52.0 | 28.0 | 28.0 | 28.0 | 47.0 | 47.0 | 47.0 |
|  | Female | 60.0 | 52.0 | 41.0 | 41.0 | 41.0 | 47.0 | 47.0 | 47.0 |
| AMR-A | Male | 6.5 | 5.0 | 5.0 | 5.0 | 5.0 | 6.0 | 7.0 | 7.0 |
|  | Female | 6.5 | 5.0 | 8.0 | 8.0 | 8.0 | 7.0 | 7.0 | 7.0 |
| AMR-B | Male | 46.0 | 24.0 | 11.0 | 11.0 | 11.0 | 18.0 | 18.0 | 18.0 |
|  | Female | 46.0 | 24.0 | 23.0 | 23.0 | 23.0 | 18.0 | 18.0 | 18.0 |
| AMR-D | Male | 46.0 | 24.0 | 11.0 | 11.0 | 11.0 | 18.0 | 18.0 | 18.0 |
|  | Female | 46.0 | 24.0 | 23.0 | 23.0 | 23.0 | 18.0 | 18.0 | 18.0 |
| EMR-B | Male | 63.0 | 39.0 | 17.0 | 17.0 | 17.0 | 25.0 | 25.0 | 25.0 |
|  | Female | 63.0 | 39.0 | 44.0 | 44.0 | 44.0 | 25.0 | 25.0 | 25.0 |
| EMR-D | Male | 63.0 | 39.0 | 17.0 | 17.0 | 17.0 | 25.0 | 25.0 | 25.0 |
|  | Female | 63.0 | 39.0 | 44.0 | 44.0 | 44.0 | 25.0 | 25.0 | 25.0 |
| EUR-A | Male | 6.5 | 5.0 | 5.0 | 5.0 | 5.0 | 6.0 | 7.0 | 7.0 |
|  | Female | 6.5 | 5.0 | 8.0 | 8.0 | 8.0 | 7.0 | 7.0 | 7.0 |
| EUR-B | Male | 22.0 | 20.0 | 5.0 | 5.0 | 5.0 | 12.0 | 12.0 | 12.0 |
|  | Female | 22.0 | 20.0 | 10.0 | 10.0 | 10.0 | 12.0 | 12.0 | 12.0 |
| EUR-C | Male | 22.0 | 20.0 | 5.0 | 5.0 | 5.0 | 12.0 | 12.0 | 12.0 |
|  | Female | 22.0 | 20.0 | 10.0 | 10.0 | 10.0 | 12.0 | 12.0 | 12.0 |
| SEAR-B | Male | 49.0 | 33.0 | 32.0 | 32.0 | 32.0 | 48.0 | 48.0 | 48.0 |
|  | Female | 49.0 | 33.0 | 49.0 | 49.0 | 49.0 | 48.0 | 48.0 | 48.0 |
| SEAR-D | Male | 66.0 | 65.0 | 36.0 | 36.0 | 36.0 | 65.0 | 65.0 | 65.0 |
|  | Female | 66.0 | 65.0 | 60.0 | 60.0 | 60.0 | 65.0 | 65.0 | 65.0 |
| WPR-A | Male | 6.5 | 5.0 | 5.0 | 5.0 | 5.0 | 6.0 | 7.0 | 7.0 |
|  | Female | 6.5 | 5.0 | 8.0 | 8.0 | 8.0 | 7.0 | 7.0 | 7.0 |
| WPR-B | Male | 49.0 | 33.0 | 32.0 | 32.0 | 32.0 | 48.0 | 48.0 | 48.0 |
|  | Female | 49.0 | 33.0 | 49.0 | 49.0 | 49.0 | 48.0 | 48.0 | 48.0 |

including iron deficiency. Thus it is logical to expect the standard deviation of haemoglobin to be higher in populations with more anaemia compared to those with infrequent anaemia. The second factor is the precision of the haemoglobin assay, with greater precision yielding smaller observed standard deviations.

The variation in haemoglobin standard deviations is illustrated in Table 3.15, which relates the prevalence of anaemia to the standard deviation, using data from a draft version of the updated (but still incomplete) anaemia database for the WHO African and Eastern Mediterranean Regions (B. de Benoist, personal communication) and data from the most recent national health and nutrition examination survey in the United States, after excluding iron-deficient individuals. The standard deviations from the data from the United States are significantly smaller,

Table 3.15 Standard deviations of haemoglobin values

|  | Africa and EMRO surveys ${ }^{\mathrm{a}}$ | USA NHANES III data ${ }^{\text {b }}$ |
| :--- | :--- | :---: |
| Prevalence of anaemia | $>50 \%$ | $<5 \%$ |
| Standard deviation (g/dl) |  |  |
| Children 0-4 years | 1.45 | 0.75 |
| Children 5-14 years | 1.58 | 0.85 |
| Non-pregnant women | 1.66 | 0.91 |
| Men | 1.60 | 0.97 |
| EMRO | WHO Eastern Mediterranean Region. |  |
| a $\quad$ B. de Benoist, personal communication. Values are weighted averages from available country |  |  |
| surveys. |  |  |
| brom Looker et al. (1997), Table 2. |  |  |

Table 3.16 Standard deviations used to convert anaemia prevalence to mean haemoglobin

| Anaemia prevalence | Standard deviation of haemoglobin |
| :--- | :---: |
| $<15 \%$ | $1.0 \mathrm{~g} / \mathrm{dl}$ |
| $15-30 \%$ | $1.2 \mathrm{~g} / \mathrm{dl}$ |
| $>30 \%$ | $1.5 \mathrm{~g} / \mathrm{dl}$ |

probably due to both of the factors discussed above. Namely, the proportion of anaemic individuals in this United States sample was very low, and the haemoglobin assessment method (venous blood collection, laboratory-based assay with rigorous quality control) was more precise than most field-based studies in developing countries.

Based on these considerations, we assumed certain standard deviations based on the anaemia prevalence of each population group. These are given in Table 3.16.

Knowing the proportion below a certain haemoglobin cut-off (i.e. prevalence of anaemia in a population) and the standard deviation of the normal distribution, we estimated the mean haemoglobin associated with the current prevalence of anaemia (Snedecor et al. 1980). These values are given in Table 3.17.

We then estimated the theoretical-minimum-risk distribution, representing the haemoglobin distribution if iron deficiency were eliminated. Assuming that $50 \%$ of anaemia in the world is attributable to iron deficiency (see section 3), we divided the current anaemia prevalences by 2. Because anaemia cut-offs are defined as the 5th percentile of a normative reference distribution (i.e. the distribution of individuals known to

Table 3.17 Estimated mean haemoglobin values (g/dl) by subregion, sex and age in 2000

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 10.62 | 11.67 | 13.70 | 13.70 | 13.70 | 13.11 | 13.11 | 13.11 |
|  | Female | 10.62 | 11.67 | 12.34 | 12.34 | 12.34 | 12.11 | 12.11 | 12.11 |
| AFR-E | Male | 10.62 | 11.67 | 13.70 | 13.70 | 13.70 | 13.11 | 13.11 | 13.11 |
|  | Female | 10.62 | 11.67 | 12.34 | 12.34 | 12.34 | 12.11 | 12.11 | 12.11 |
| AMR-A | Male | 12.51 | 13.39 | 14.64 | 14.64 | 14.64 | 14.55 | 14.55 | 14.55 |
|  | Female | 12.51 | 13.39 | 13.41 | 13.41 | 13.41 | 13.48 | 13.48 | 13.48 |
| AMR-B | Male | 11.15 | 12.60 | 14.47 | 14.47 | 14.47 | 14.10 | 14.10 | 14.10 |
|  | Female | 11.15 | 12.60 | 12.89 | 12.89 | 12.89 | 13.10 | 13.10 | 13.10 |
| AMR-D | Male | 11.15 | 12.60 | 14.47 | 14.47 | 14.47 | 14.10 | 14.10 | 14.10 |
|  | Female | 11.15 | 12.60 | 12.89 | 12.89 | 12.89 | 13.10 | 13.10 | 13.10 |
| EMR-B | Male | 10.50 | 12.17 | 14.14 | 14.14 | 14.14 | 13.81 | 13.81 | 13.81 |
|  | Female | 10.50 | 12.17 | 12.23 | 12.23 | 12.23 | 12.81 | 12.81 | 12.81 |
| EMR-D | Male | 10.50 | 12.17 | 14.14 | 14.14 | 14.14 | 13.81 | 13.81 | 13.81 |
|  | Female | 10.50 | 12.17 | 12.23 | 12.23 | 12.23 | 12.81 | 12.81 | 12.81 |
| EUR-A | Male | 12.51 | 13.39 | 14.64 | 14.64 | 14.64 | 14.55 | 14.55 | 14.55 |
|  | Female | 12.51 | 13.39 | 13.41 | 13.41 | 13.41 | 13.48 | 13.48 | 13.48 |
| EUR-B | Male | 11.93 | 12.76 | 14.64 | 14.64 | 14.64 | 14.41 | 14.41 | 14.41 |
|  | Female | 11.93 | 12.76 | 13.28 | 13.28 | 13.28 | 13.41 | 13.41 | 13.41 |
| EUR-C | Male | 11.93 | 12.76 | 14.64 | 14.64 | 14.64 | 14.41 | 14.41 | 14.41 |
|  | Female | 11.93 | 12.76 | 13.28 | 13.28 | 13.28 | 13.41 | 13.41 | 13.41 |
| SEAR-B | Male | 11.04 | 12.41 | 13.70 | 13.70 | 13.70 | 13.08 | 13.08 | 13.08 |
|  | Female | 11.04 | 12.41 | 12.04 | 12.04 | 12.04 | 12.08 | 12.08 | 12.08 |
| SEAR-D | Male | 10.38 | 11.17 | 13.54 | 13.54 | 13.54 | 12.42 | 12.42 | 12.42 |
|  | Female | 10.38 | 11.17 | 11.62 | 11.62 | 11.62 | 11.42 | 11.42 | 11.42 |
| WPR-A | Male | 12.51 | 13.39 | 14.64 | 14.64 | 14.64 | 14.55 | 14.55 | 14.55 |
|  | Female | 12.51 | 13.39 | 13.41 | 13.41 | 13.41 | 13.48 | 13.48 | 13.48 |
| WPR-B | Male | 11.04 | 12.41 | 13.70 | 13.70 | 13.70 | 13.08 | 13.08 | 13.08 |
|  | Female | 11.04 | 12.41 | 12.04 | 12.04 | 12.04 | 12.08 | 12.08 | 12.08 |

be free of disease), we set the minimum prevalence of anaemia in all world subregions to be $5.0 \%$. We assumed a normal distribution, and applied the standard deviations in Table 3.15. This yields values in Table 3.18.

To check the plausibility of this theoretical minimum distribution, we examined the shift in population mean haemoglobin concentration from current reality to the theoretical minimum, representing the eradication of iron deficiency. These shifts are summarized in Table 3.19.

The predicted shifts in mean haemoglobin ranged from 0.0 , in young adult men in affluent subregions, to $1.28 \mathrm{~g} / \mathrm{dl}$ in children in SEAR-D. The predicted haemoglobin shift for young children in AFR was 1.17 , which is consistent with the evidence from iron supplementation trials in the

Table 3.18 Estimated mean haemoglobin (g/dl) values if iron deficiency were eliminated (theoretical minimum distribution), by subregion, sex and age

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 11.79 | 12.52 | 14.08 | 14.08 | 14.08 | 13.87 | 13.87 | 13.87 |
|  | Female | 11.79 | 12.52 | 12.99 | 12.99 | 12.99 | 12.87 | 12.87 | 12.87 |
| AFR-E | Male | 11.79 | 12.52 | 14.08 | 14.08 | 14.08 | 13.87 | 13.87 | 13.87 |
|  | Female | 11.79 | 12.52 | 12.99 | 12.99 | 12.99 | 12.87 | 12.87 | 12.87 |
| AMR-A | Male | 12.58 | 13.39 | 14.64 | 14.64 | 14.64 | 14.60 | 14.55 | 14.55 |
|  | Female | 12.58 | 13.39 | 13.51 | 13.5 \| | 13.51 | 13.55 | 13.55 | 13.55 |
| AMR-B | Male | 11.89 | 12.92 | 14.60 | 14.60 | 14.60 | 14.34 | 14.34 | 14.34 |
|  | Female | 11.89 | 12.92 | 13.20 | 13.20 | 13.20 | 13.34 | 13.34 | 13.34 |
| AMR-D | Male | 11.89 | 12.92 | 14.60 | 14.60 | 14.60 | 14.34 | 14.34 | 14.34 |
|  | Female | 11.89 | 12.92 | 13.20 | 13.20 | 13.20 | 13.34 | 13.34 | 13.34 |
| EMR-B | Male | 11.72 | 12.78 | 14.37 | 14.37 | 14.37 | 14.15 | 14.15 | 14.15 |
|  | Female | 11.72 | 12.78 | 12.93 | 12.93 | 12.93 | 13.15 | 13.15 | 13.15 |
| EMR-D | Male | 11.72 | 12.78 | 14.37 | 14.37 | 14.37 | 14.15 | 14.15 | 14.15 |
|  | Female | 11.72 | 12.78 | 12.93 | 12.93 | 12.93 | 13.15 | 13.15 | 13.15 |
| EUR-A | Male | 12.58 | 13.39 | 14.64 | 14.64 | 14.64 | 14.60 | 14.55 | 14.55 |
|  | Female | 12.58 | 13.39 | 13.51 | 13.51 | 13.51 | 13.55 | 13.55 | 13.55 |
| EUR-B | Male | 12.23 | 13.03 | 14.96 | 14.96 | 14.96 | 14.55 | 14.55 | 14.55 |
|  | Female | 12.23 | 13.03 | 13.64 | 13.64 | 13.64 | 13.55 | 13.55 | 13.55 |
| EUR-C | Male | 12.23 | 13.03 | 14.96 | 14.96 | 14.96 | 14.55 | 14.55 | 14.55 |
|  | Female | 12.23 | 13.03 | 13.64 | 13.64 | 13.64 | 13.55 | 13.55 | 13.55 |
| SEAR-B | Male | 11.83 | 12.92 | 14.19 | 14.19 | 14.19 | 13.85 | 13.85 | 13.85 |
|  | Female | 11.83 | 12.92 | 12.83 | 12.83 | 12.83 | 12.85 | 12.85 | 12.85 |
| SEAR-D | Male | 11.66 | 12.43 | 14.10 | 14.10 | 14.10 | 13.68 | 13.68 | 13.68 |
|  | Female | 11.66 | 12.43 | 12.79 | 12.79 | 12.79 | 12.68 | 12.68 | 12.68 |
| WPR-A | Male | 12.58 | 13.39 | 14.64 | 14.64 | 14.64 | 14.60 | 14.55 | 14.55 |
|  | Female | 12.58 | 13.39 | 13.51 | 13.51 | 13.51 | 13.55 | 13.55 | 13.55 |
| WPR-B | Male | 11.83 | 12.92 | 14.19 | 14.19 | 14.19 | 13.85 | 13.85 | 13.85 |
|  | Female | 11.83 | 12.92 | 12.83 | 12.83 | 12.83 | 12.85 | 12.85 | 12.85 |

P. falciparum-endemic populations (1.24, $95 \%$ CI 1.16-1.33, see section 3 ). The predicted haemoglobin shifts for women of reproductive age range from 0.11 to $1.17 \mathrm{~g} / \mathrm{dl}$. These estimates were somewhat lower than the average haemoglobin response of pregnant women to iron supplementation ( 0.85 to $1.17 \mathrm{~g} / \mathrm{dl}$ ), and were substantially lower than the haemoglobin responses in women provided iron doses of $\geq 90 \mathrm{mg} /$ day (i.e. $1.8 \mathrm{~g} / \mathrm{dl}$ ) estimated by Sloan et al. (2002).

We are thus faced with some uncertainty about how to estimate the haemoglobin shift that would occur if iron deficiency were eliminated. Basing our estimates on a $50 \%$ reduction in anaemia yields a smaller shift than we would obtain if we assumed that the responses of pregnant women to maximal daily iron doses in the trials summarized by Sloan

Table 3.19 Shifts in mean haemoglobin (g/dl) from 2000 estimates (Table 3.17) if iron deficiency were eliminated (Table 3.18)

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | Row means |
| AFR-D | Male | 1.17 | 0.85 | 0.38 | 0.38 | 0.38 | 0.75 | 0.75 | 0.75 | 0.68 |
|  | Female | 1.17 | 0.85 | 0.65 | 0.65 | 0.65 | 0.75 | 0.75 | 0.75 | 0.78 |
| AFR-E | Male | 1.17 | 0.85 | 0.38 | 0.38 | 0.38 | 0.75 | 0.75 | 0.75 | 0.68 |
|  | Female | 1.17 | 0.85 | 0.65 | 0.65 | 0.65 | 0.75 | 0.75 | 0.75 | 0.78 |
| AMR-A | Male | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.08 | 0.08 | 0.03 |
|  | Female | 0.06 | 0.00 | 0.11 | 0.11 | 0.11 | 0.08 | 0.08 | 0.08 | 0.08 |
| AMR-B | Male | 0.74 | 0.33 | 0.13 | 0.13 | 0.13 | 0.24 | 0.24 | 0.24 | 0.27 |
|  | Female | 0.74 | 0.33 | 0.31 | 0.31 | 0.31 | 0.24 | 0.24 | 0.24 | 0.34 |
| AMR-D | Male | 0.74 | 0.33 | 0.13 | 0.13 | 0.13 | 0.24 | 0.24 | 0.24 | 0.27 |
|  | Female | 0.74 | 0.33 | 0.31 | 0.31 | 0.31 | 0.24 | 0.24 | 0.24 | 0.34 |
| EMR-B | Male | 1.22 | 0.61 | 0.23 | 0.23 | 0.23 | 0.34 | 0.34 | 0.34 | 0.44 |
|  | Female | 1.22 | 0.61 | 0.70 | 0.70 | 0.70 | 0.34 | 0.34 | 0.34 | 0.62 |
| EMR-D | Male | 1.22 | 0.61 | 0.23 | 0.23 | 0.23 | 0.34 | 0.34 | 0.34 | 0.44 |
|  | Female | 1.22 | 0.61 | 0.70 | 0.70 | 0.70 | 0.34 | 0.34 | 0.34 | 0.62 |
| EUR-A | Male | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.08 | 0.08 | 0.03 |
|  | Female | 0.06 | 0.00 | 0.11 | 0.11 | 0.11 | 0.08 | 0.08 | 0.08 | 0.08 |
| EUR-B | Male | 0.30 | 0.27 | 0.32 | 0.32 | 0.32 | 0.14 | 0.14 | 0.14 | 0.24 |
|  | Female | 0.30 | 0.27 | 0.36 | 0.36 | 0.36 | 0.14 | 0.14 | 0.14 | 0.26 |
| EUR-C | Male | 0.30 | 0.27 | 0.32 | 0.32 | 0.32 | 0.14 | 0.14 | 0.14 | 0.24 |
|  | Female | 0.30 | 0.27 | 0.36 | 0.36 | 0.36 | 0.14 | 0.14 | 0.14 | 0.26 |
| SEAR-B | Male | 0.79 | 0.51 | 0.49 | 0.49 | 0.49 | 0.77 | 0.77 | 0.77 | 0.64 |
|  | Female | 0.79 | 0.51 | 0.79 | 0.79 | 0.79 | 0.77 | 0.77 | 0.77 | 0.75 |
| SEAR-D | Male | 1.28 | 1.26 | 0.56 | 0.56 | 0.56 | 1.26 | 1.26 | 1.26 | 1.00 |
|  | Female | 1.28 | 1.26 | 1.17 | 1.17 | 1.17 | 1.26 | 1.26 | 1.26 | 1.23 |
| WPR-A | Male | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.08 | 0.08 | 0.03 |
|  | Female | 0.06 | 0.00 | 0.11 | 0.11 | 0.11 | 0.08 | 0.08 | 0.08 | 0.08 |
| WPR-B | Male | 0.79 | 0.51 | 0.49 | 0.49 | 0.49 | 0.77 | 0.77 | 0.77 | 0.64 |
|  | Female | 0.79 | 0.51 | 0.79 | 0.79 | 0.79 | 0.77 | 0.77 | 0.77 | 0.75 |
| Column means |  | 0.71 | 0.46 | 0.38 | 0.38 | 0.38 | 0.43 | 0.43 | 0.43 | 0.45 |

et al. (2002) represented the true effect. We used the $50 \%$ reduction in anaemia for two reasons. First, it yields more conservative estimates for the overall burden of disease attributable to iron deficiency. Second, we expect that the responses seen in iron supplementation trials are not representative of all women in the region in which they were conducted. Researchers frequently select study populations that are unusual in their potential to respond to an intervention; in this case populations that are more anaemic than the average population. Therefore it is reasonable to expect that data from randomized trials would overestimate global or regional average effects.

Figure 3.4 Relationship between mean haemoglobin values in pregnant and non-pregnant women within populations


Note: $\quad n=43$ surveys, $r=0.84, P<0.00$ I.
Source: Data from WHO (B. de Benoist, personal communication).

A final consideration is that these global tables report values for all women in the reproductive age ranges, whereas the risks of perinatal and maternal mortality apply uniquely to pregnancy anaemia. Population surveys that include both pregnant and non-pregnant women demonstrate a strong linear correlation between the haemoglobin values in these two population subgroups, as shown in Figure 3.4. Across the range of haemoglobin for women in Table 3.17, the difference between haemoglobin values in non-pregnant and pregnant states is nearly constant (range: $1.39 \mathrm{~g} / \mathrm{dl}$ to $1.34 \mathrm{~g} / \mathrm{dl}$ ). The haemoglobin cut-off used to define anaemia is also lower in pregnancy. The WHO cut-off, for simplicity, is $1.0 \mathrm{~g} / \mathrm{dl}$ lower in pregnancy. However, the physiologic haemodilution of pregnancy reduces the haemoglobin concentration by $1.5 \mathrm{~g} / \mathrm{dl}$ in midpregnancy (Institute of Medicine 1990), when many pregnancy values are obtained in field surveys. Thus we can expect the prevalence of true anaemia to be approximately the same in pregnant and non-pregnant women. Because pregnant women are more iron-deficient than non-pregnant women, there is reason to believe that the iron-attributable fraction of anaemia would be higher in pregnancy than in other states. However, lacking firm data on this relationship, we have made the conservative assumption that the anaemia prevalence and predicted haemoglobin shift associated with the elimination of iron deficiency are the same for pregnant women as non-pregnant women.

## 8. Burden of disease estimates

The estimated deaths and DALYs attributable to iron deficiency are shown in Tables 3.20 and 3.21. Those attributed to iron deficiency as a risk factor for perinatal and maternal conditions are shown separately from those attributed directly to iron deficiency anaemia. For perinatal and maternal conditions, only mortality (as opposed to morbidity) from maternal conditions was attributed to iron deficiency. In the absence of any data on anaemia and morbidity related to childbirth or the puerperium, we assumed that such a relationship did not exist. As discussed earlier, the effects of iron deficiency anaemia on work productivity and child development and cognition are included in the direct attributions to iron deficiency anaemia (i.e. in the iron deficiency anaemia columns). These direct estimates are updated from the 1990 GBD (Murray et al. 1996a, 1996b), using 2000 demographic statistics, but otherwise the same methods.

The total burden of death and disability attributable to iron deficiency is higher than the 1990 estimate. This is due to differences in the risk estimates, not the prevalence estimates, which remain unchanged. The inclusion of iron deficiency anaemia as a risk factor for perinatal mortality is new, and contributes substantially to estimated deaths and DALYs lost. Furthermore, in the 1990 estimate, deaths from iron defi-

Table 3.20 Deaths from perinatal conditions, maternal conditions and iron deficiency anaemia, attributable to iron deficiency, by subregion

| Subregion | Perinatal causes <br> $(000 \mathrm{~s})$ | Maternal causes <br> $(000 \mathrm{~s})$ | Iron deficiency <br> anaemia (000s) | Total (000s) |
| :--- | :---: | :---: | :---: | :---: |
| AFR-D | 97 | 20 | 8 | 125 |
| AFR-E | 103 | 29 | 13 | 145 |
| AMR-A | 1 | 0 | 3 | 4 |
| AMR-B | 17 | 2 | 7 | 26 |
| AMR-D | 3 | 1 | 3 | 7 |
| EMR-B | 4 | 1 | 1 | 7 |
| EMR-D | 56 | 14 | 9 | 6 |
| EUR-A | 1 | 0 | 3 | 79 |
| EUR-B | 4 | 0 | 1 | 4 |
| EUR-C | 2 | 5 | 13 | 6 |
| SEAR-B | 16 | 38 | 66 | 34 |
| SEAR-D | 222 | 0 | 0 | 326 |
| WPR-A | 0 | 5 | 5 | 0 |
| WPR-B | 63 | 115 | 135 | 73 |
| World | 591 |  |  | 841 |

Table 3.2 I DALYs from perinatal conditions, maternal conditions and iron deficiency anaemia, attributable to iron deficiency, by subregion

| Subregion | Perinatal causes <br> $(000 \mathrm{~s})$ | Maternal causes <br> $(000 \mathrm{~s})$ | Iron deficiency <br> anaemia (000s) | Total (000s) |
| :--- | ---: | ---: | ---: | ---: |
| AFR-D | 3237 | 614 | 934 | 4785 |
| AFR-E | 3442 | 881 | 1033 | 5356 |
| AMR-A | 47 | 2 | 430 | 479 |
| AMR-B | 561 | 59 | 291 | 911 |
| AMR-D | 102 | 36 | 201 | 339 |
| EMR-B | 149 | 22 | 344 | 515 |
| EMR-D | 1882 | 418 | 895 | 3195 |
| EUR-A | 29 | 1 | 269 | 299 |
| EUR-B | 145 | 6 | 286 | 437 |
| EUR-C | 61 | 148 | 205 | 675 |
| SEAR-B | 545 | 1114 | 835 | 1528 |
| SEAR-D | 7428 | 0 | 3955 | 12497 |
| WPR-A | 6 | 145 | 105 | 111 |
| WPR-B | 2101 | 3448 | 2092 | 4338 |
| World | 19736 |  |  |  |

ciency anaemia were included only as "direct" deaths due to severe and very severe anaemia. The increased maternal mortality from iron deficiency anaemia as an underlying risk factor is newly included in these estimates.

It is apparent that while the death and especially the disability directly attributable to iron deficiency anaemia is large, death and disability attributable to iron deficiency anaemia acting as a risk factor for perinatal and maternal causes is much larger. Because of the great numbers of perinatal deaths globally, perinatal causes account for $70 \%$ of the total deaths and $56 \%$ of the total DALYs attributable to iron deficiency anaemia. Numbers of maternal deaths, while very high, are much lower than perinatal deaths. Therefore the absolute contribution of maternal causes to the totals is smaller, namely $10 \%$ of DALYs and $14 \%$ of deaths. The relative contribution of maternal and perinatal causes to the total DALYs lost is largest where death rates are high (e.g. 80\% in SEAR-D compared to only $25 \%$ in EUR-A). Indeed, as expected, the contribution of iron deficiency anaemia to maternal deaths is zero in AMR-A, EUR-A and WPR-A, where maternal mortality rates are very low.

When examined by region, the global burden of iron deficiency anaemia and its consequences are most heavily borne by those in SouthEast Asia and Africa. For maternal causes, $43 \%$ and $37 \%$ of the mater-
nal deaths (and DALYs) attributed to iron deficiency anaemia occur in Africa (AFR-D and AFR-E combined) and South-East Asia (SEAR-B and SEAR-D combined), respectively. For perinatal causes, the relative burden in the two regions is reversed, but they still suffer the greatest toll, with $34 \%$ of the iron deficiency-related perinatal deaths (and DALYs lost) in Africa and $40 \%$ in South-East Asia. WPR-B and EMR-D also bear a large burden of iron deficiency-related perinatal deaths, with approximately $10 \%$ of the global total in each of these two subregions.

## 9. Expected changes in the prevalence of IRON DEFICIENCY ANAEMIA

Based on the trends observed over the past 20 years there is no reason to expect anaemia or iron deficiency anaemia prevalence to decrease in the coming decade. The United Nations Subcommittee on Nutrition (ACC/SCN) reviewed trends in data from the 1970s and 1980s (ACC/SCN 1992). Iron density in the diet was decreasing during this period for every global region except the Near East and North Africa. Furthermore, trends in anaemia from 1977 to 1987 were increasing in the two regions where the problem is most severe: south Asia, and subSaharan Africa. In 1990, the World Summit for Children set goals for reducing malnutrition, sickness and death in children that included a goal to reduce iron deficiency anaemia by one-third by the year 2000. A progress report in 1995 (UNICEF 1995) stated:

Very few countries have so far taken nation-scale action to eliminate iron deficiency anaemia. . . . No mid-decade target was established for progress against anaemia; but the goal . . is unlikely to be met without a significant acceleration of effort over the next six years.

This needed acceleration of effort did not take place. A more recent progress report (WHO 2000) concluded the following:

> Unfortunately, there has been little appreciable change over the last two decades in the high worldwide prevalence of iron deficiency anaemia. Few active programmes in both developed and developing countries have succeeded in reducing iron deficiency and anaemia. Important factors contributing to the lack of progress include failure to recognize the causes of iron deficiency and anaemia, lack of political commitment to control it, inadequate planning of control programmes, insufficient mobilization and training of health staff, and insufficient community involvement in solving the problem.

Thus the evidence strongly suggests that under a "business-as-usual" scenario, the prevalence of iron deficiency anaemia will not decrease over the next decade. We hope that the new estimates of the burden of disease attributed to iron deficiency will result in a significant deviation from business-as-usual.

## 10. Discussion

Our analysis of the relationship between pregnancy anaemia and maternal mortality differed from that of the recent analysis by Brabin et al. (2001b), although we drew upon the same published studies. The salient difference is that we began by estimating the mortality-haemoglobin relationship within each study (expressed as an OR per g/dl increment in haemoglobin), and then used meta-analytic techniques to obtain a weighted average of those ORs. Restricting the analysis to observations between 5 and $12 \mathrm{~g} / \mathrm{dl}$ haemoglobin, the risk increased with decreasing haemoglobin within each study.

As noted previously, there are no trials of iron supplementation with maternal mortality as outcome. Furthermore, because of cost and ethical considerations, we will likely have to continue to rely on observational data to refine these estimates. However, better observational data are badly needed. The currently available data are generally quite old, are predominantly from Asia (India and Malaysia) and are not controlled for many potentially confounding or modifying factors.

This analysis differs from most previous statements about anaemia and maternal mortality by positing a continuous relationship between haemoglobin concentration and mortality risk. INACG issued the following statement about severe anaemia and death in childbirth (1989):

> At $6.0 \mathrm{~g} / \mathrm{dl}$, evidence of circulatory decompensation becomes apparent.
> Women experience breathlessness and increased cardiac output at rest.
> At this stage, added stress from labor ... can result in maternal death.
> Without effective treatment, maternal death from anemic heart failure $\ldots$ is likely with a haemoglobin concentration of $4.0 \mathrm{~g} / \mathrm{dl}$. Even a blood loss of 100 ml can cause circulatory shock and death.

A more recent summary statement (Stoltzfus 2001), based on the systematic review of Brabin et al. (2001b) also concluded that "A significant body of causal evidence exists for . . . severe anaemia and maternal mortality" but that "causal evidence is lacking or contradictory for ... mild-moderate anaemia and maternal mortality" (Table 1). However, considering that death from cardiovascular causes is a function of blood volume, blood loss, cardiac fitness and haemoglobin concentration, it seems plausible that the relationship between haemoglobin concentration and maternal death would be continuous in nature, although not necessarily linear. Indeed the relationship as we have modelled it in this analysis is log-linear, with risk increasing exponentially with decreasing haemoglobin concentration.

We are not aware of another systematic analysis of the observational data linking pregnancy anaemia to perinatal mortality. The risk estimates from geographically diverse studies were remarkably consistent, and we believe provided the best evidence available for our purpose. Rasmussen (2001) identified four controlled iron supplementation trials that reported perinatal deaths as an outcome. All four studies had major
design concerns (low rates of follow-up or lack of anaemic individuals enrolled in the trial), and all were of insufficient sample size to draw clear conclusions. Rasmussen did not draw any firm conclusion from these data.

One relatively large trial was recently completed in rural Nepal, where the incidence of pregnancy anaemia and perinatal mortality are both high. This trial has only been reported in abstract form (Christian et al. 2002) and also was not sufficiently large to draw clear conclusions about perinatal mortality. However, it represents the strongest randomized trial evidence to date, because it was placebo-controlled and had high followup rates. In this trial, all women were provided anthelminthic treatment and vitamin A supplements. Women supplemented with 60 mg iron and $400 \mu \mathrm{~g}$ folic acid had $20 \%$ ( $95 \%$ CI $0.55-1.17$ ) lower incidence of perinatal mortality than the placebo group (P. Christian, personal communication). This point estimate is larger than the per cent reduction we estimated in this analysis (i.e. $16 \%, 95 \%$ CI $10-22 \%$ ). However, the trial also included several additional treatment arms, including one group that received folic acid without iron. The folic acid-supplemented group also had lower perinatal mortality rates than the placebo group, a reduction of $11 \%$ ( $95 \%$ CI $0.63-1.26$ ). Micronutrient supplementation began after the first trimester of pregnancy, and therefore these reductions cannot be attributed to the demonstrated effects of folic acid in preventing neural tube defects. Although neither of these reductions is statistically significant, they are important public health findings. If we accept these point estimates as the best available estimates from trial data, we conclude that our perinatal risk estimates are in accord with the benefits seen with effective iron-folic acid supplementation, but that a large part of the benefit may in fact be attributable to the folic acid.

The biological mechanisms linking iron deficiency anaemia (or anaemia from any cause) to perinatal mortality remain to be elucidated. One possible pathway is through preterm delivery. Recent reviews of clinical trial evidence have found the evidence inconclusive in support of a role for iron supplementation in preventing preterm birth or low birth weight (Mahomed 2000b; Rasmussen 2001). However, the evidence for or against is remarkably weak, and does not rule out a causal relationship (Stoltzfus 2001). Scott Poe and Mary Cogswell (personal communication) have recently completed a meta-analysis of the observational evidence relating pregnancy anaemia to low birth weight, intrauterine growth retardation and preterm birth. They found that pregnancy anaemia assessed in the first two trimesters of pregnancy was more strongly associated with preterm birth than intrauterine growth retardation or low birth weight, and that the risk of preterm birth increased with increasing severity of anaemia. Thus, while controlled trial evidence is lacking, the observational evidence suggests that preterm birth is one plausible mechanism for the anaemia-related risk of perinatal mortality.

If we accept preterm birth as a plausible causal pathway, the question remains as to how iron deficiency anaemia causes preterm birth. Allen suggested three possible mechanisms (Allen 2001). Iron deficiency anaemia might activate a hormonal stress response, might increase oxidative stress or might increase the risk of maternal infections. Further research is needed to elucidate these mechanisms or to suggest alternative ones.

While preterm birth may be considered the leading hypothesis to explain the link between maternal anaemia and perinatal mortality, it is not the only hypothesis that should be pursued. If these above proposed mechanisms exist, it is plausible that they cause adverse effects on the fetus that go beyond preterm birth, for example by impairing the neonatal immune system, endocrine function, temperature regulation, or other systems critical to a successful transition from intra to extrauterine life. Additionally, the adverse effects of maternal anaemia on the mother's function and well-being may also increase risks to her neonate through her decreased capacity to actively care for and breastfeed the infant (Henly et al. 1995). In the extreme case, maternal death associated with anaemia would increase the neonate's risk of death.

We included here an estimate of the continuous relationship between iron deficiency anaemia in early childhood and later intelligence, even though this relationship is included in the direct disability score attributed to iron deficiency anaemia. The possible biological mechanisms underlying this relationship are the subject of a large body of ongoing research in animals and children (e.g. see Beard 2001). To date, two published placebo-controlled randomized trials have measured the effect of longer-term (i.e. $>2$ months) iron supplementation on cognitive development of young children in samples that included anaemic children (Idjradinata and Pollitt 1993; Stoltzfus et al. 2001). Both trials found significant benefits in the iron-supplemented group. In Indonesia (Idjradinata and Pollitt 1993), the positive effect was measured on the Bayley Scales of Infant Development, while in Zanzibar (Stoltzfus et al. 2001), parental reports of motor and language milestones were used. These trials support a causal link between iron deficiency anaemia and child development that is at least partly preventable with early treatment.

The major difficulty has been quantifying this relationship in meaningful epidemiological and socioeconomic terms, especially in the sociocultural contexts in which iron deficiency anaemia is most prevalent. Our estimated OR is an attempt to put the relationship in epidemiological terms. This estimate relies on the assumption that a mean shift in IQ can be converted into increased risk of mild mental retardation, assuming that the entire IQ distribution was shifted equally. To our knowledge there has been only one published report of the relationship between early childhood anaemia and mild mental retardation as a dichotomous outcome (as opposed to intelligence or developmental scores as continuous outcomes). Hurtado et al. (1999) assessed the association between
haemoglobin concentration of children in the United States enrolled in the Special Supplemental Program for Women, Infants, and Children, a programme of the U.S. Government that provides food supplements to low-income pregnant women and their young children (Hurtado et al. 1999). After adjusting for several important covariates, the OR for mild or moderate mental retardation at school age remained significant in their model, which treated haemoglobin concentration as a continuous risk factor (as did ours). This finding supports the plausibility of our assumption that the association between mean cognitive scores and anaemia is associated with increased risks of mild mental retardation. However, it is only a single observational study, and we believe our estimate should be interpreted with extreme caution.

In summary, the available evidence suggests that iron deficiency anaemia contributes substantially to death and disability in the world. The great majority of this disease burden is in Africa and Asia and derives from anaemia in pregnancy and early childhood. This evidence is based on critical assumptions, most importantly, that the observed prospective relationships are causal in nature, and that the relationships analysed using anaemia as the risk factor pertain equally to iron deficiency anaemia as one particular form of anaemia.

The high global prevalence of anaemia and its potentially associated disease burden, as reflected in these estimates, constitute an urgent agenda for both research and action. First, we must clarify the assumptions above, the first and foremost of these being causality, and refine these estimates with stronger evidence. Because these estimates are uncertain in many respects, their most important use may be to motivate public health scientists to provide definitive causal evidence. Second, we must establish ways to effectively reduce iron deficiency anaemia to prevent these apparent consequences. We hope these new estimates of the burden of disease due to iron deficiency will motivate these actions.

## Acknowledgements

This work was supported by the Family Health and Child Survival (FHACS) Cooperative Agreement, between the United States Agency for International Development (USAID), Office of Health and Nutrition, and the Johns Hopkins Bloomberg School of Public Health, Department of International Health and by the World Health Organization Global Programme on Evidence for Health Policy. The authors are grateful to Steve Goodman for his advice on statistical issues, and to Ines Egli and Bruno de Benoist for data used in estimation of the global anaemia prevalence. We also thank Michele Dreyfuss, Parul Christian, Xu Xiong, Mary Cogswell and Scott Poe for sharing unpublished data or analyses for inclusion in this chapter. Finally, we thank Steve Fishman, Jeffrey McGuckin and Justine Kavle for their help in facilitating communication and compiling data for this project.

## Notes

1 We recognize that iron supplementation is not the only means of shifting the haemoglobin distribution, but randomized trials of iron supplementation provide the best estimate of the distribution if dietary iron deficiency were nearly eliminated.
2 See preface for an explanation of this term.

## References

ACC/SCN (1992) Second report on the world nutrition situation. Vol. 1: Global and regional results. The United Nations Subcommittee on Nutrition, ACC/SCN, Geneva.
Adam Z (1997) Iron supplementation and malaria: a randomized, placebocontrolled field trial in rural Ethiopia [Dissertation]. London School of Hygiene and Tropical Medicine, London.
Agarwal KN, Agarwal DK, Mishra KP (1991) Impact of anaemia prophylaxis in pregnancy on maternal haemoglobin, serum ferritin and birth weight. Indian Journal of Medical Research, 94:277-280.
Allen LH (2001) Biological mechanisms that might underlie iron's effects on fetal growth and preterm birth. Journal of Nutrition, 131:S581-589.
Beard JL (2001) Iron biology in immune function, muscle metabolism and neuronal functioning. Journal of Nutrition, 131:S568-580.

Beard JL, Stoltzfus R, eds. (2001) Iron-deficiency anaemia: reexamining the nature and magnitude of the public health problem. Proceedings of the WHOINACG Conference, Belmont, MD, USA, 2000. Journal of Nutrition, 131:2SII.

Beaton GH (2002) Functional outcomes of iron deficiency and iron deficiency in pregnancy and beyond. (Proceedings of the INACG Symposium, February 15-26, 2001 in Hanoi, Viet Nam) Hanoi.
Brabin BJ, Hakimi M, Pelletier D (2001b) An analysis of anaemia and preg-nancy-related maternal mortality. Journal of Nutrition, 131:S604-614.
Brabin BJ, Premji Z, Verhoeff F (2001a) An analysis of anaemia and child mortality. Journal of Nutrition, 131:S636-645.
Brabin BJ (1983) An analysis of malaria in pregnancy in Africa. Bulletin of World Health Organization, 61:1005-1016.

Cantwell RJ (1974) The long term neurological sequelae of anaemia in infancy. Pediatric Research, 342:68.
Chi I-C, Agoestina T, Harbin J (1981) Maternal mortality at twelve teaching hospitals in Indonesian epidemiologic analysis. International Journal of Gynaecology and Obstetrics, 19:259-266.
Christian P, West KP Jr, Khatry SK, LeClerq SC, Kimbrough-Pradhan E, Katz J (2002) The effect of maternal micronutrient supplementation on fetal loss and infant mortality in rural Nepal: a randomized trial. FASEB Journal, 16:A748.
de Andraca I, Walter T, Castillo M, Pino P, Rivera P, Cobo C (1990) Iron deficiency anaemia and its effects upon psychological development at pre-school age: a longitudinal study. Nestlé Foundation, Lausanne.

DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Controlled Clinical Trials, 7:177-188.
Dommergues MP, Archambeaud B, Ducot Y et al. (1989) Carence en fer et tests de developpement psychomoteur: etude longitudinale entre l'age de 10 mois et l'age de 4 ans. [Iron deficiency and psychomotor development tests: longitudinal study between 10 months and 4 years of age.] Archives Francaises de Pediatrie, 46:487-490.

Dreyfuss ML, Shrestha JB, Khatry SK et al. (1996) Relationship between iron status and helminth infection among pregnant women in Nepal. FASEB Journal, 10:A730.

Fleming AF, Ghatoura GB, Harrison KA, Briggs ND, Dunn DT (1986) The prevention of anaemia in pregnancy in primigravidae in the guinea savanna of Nigeria. Annals of Tropical Medicine and Parasitology, 80:211-233.
Fullerton WT, Turner AG (1962) Exchange transfusion in treatment of severe anaemia in pregnancy. The Lancet, 1:75-78.
Garn SM, Ridella SA, Petzold AS, Falkner F (1981) Maternal hematologic levels and pregnancy outcomes. Seminars in Perinatology, 5:155-162.
Gillespie S, Johnston JL (1998) Expert consultation on anaemia determinants and interventions. Micronutrient Initiative, Ottawa, ON.
Grantham-McGregor S, Ani C (2001) A review of studies on the effect of iron deficiency on cognitive development in children. Journal of Nutrition, 131:S649-666.

Haas JD, Brownlie T (2001) Iron deficiency and reduced work capacity: a critical review of the research to determine a causal relationship. Journal of Nutrition, 131:S676-690.

Harrison KA, Lister UG, Rossiter CE, Chong H (1985) Perinatal mortality. British Journal of Obstetrics and Gynaecology, 5:86-99.
Harrison KA, Rossiter CE (1985) Maternal mortality. British Journal of Obstetrics and Gynaecology, 5:100-115.
Harrison KA (1975) Maternal mortality in anaemia in pregnancy. West Africa Medical Journal, 27-31.
Harrison KA (1982) Anaemia, malaria, and sickle cell disease. Clinics in Obstetrics and Gynaecology, 9:445-477.

Harvey PWJ, Heywood PF, Nesheim MC et al. (1989) The effect of iron therapy on malarial infection in Papua New Guinean schoolchildren. American Journal of Tropical Medicine and Hygiene, 40:12-18.

Hemminki E, Rimpelä U (1991) A randomized comparison of routine versus selective iron supplementation during pregnancy. Journal of the American College of Nutrition, 10:3-10.

Henly SJ, Anderson CM, Avery MD, Hills-Bonczyk SG, Potter S, Duckett LJ (1995) Anaemia and insufficient milk in first-time mothers. Birth, 22:86-92.

Horton S, Levin C (2001) Commentary on "Evidence that iron deficiency anaemia causes reduced work capacity". Journal of Nutrition, 131: S691-696.

Hurtado EK, Claussen AH, Scott KG (1999) Early childhood anaemia and mild or moderate mental retardation. American Journal of Clinical Nutrition, 69:115-119.

Idjradinata P, Pollitt E (1993) Reversal of developmental delays in iron-deficient anaemic infants treated with iron. The Lancet, 341:1-4.
INACG (1989) Guidelines for the control of maternal nutritional anaemia. International Nutritional Anemia Consultative Group. International Life Sciences Institute (ILSI), Washington, DC.

INACG (1999) Safety of iron supplementation programs in malaria-endemic regions. ILSI, Washington, DC.

Institute of Medicine (1990) Nutrition during pregnancy. National Academy Press, Washington, DC.

Johnson JWC, Ojo OA (1967) Amniotic fluid oxygen tensions in severe maternal anaemia. American Journal of Obstetrics and Gynecology, 97:499-506.

Konar M, Sikdar K, Basak S, Lahiri D (1980) Maternal Mortality. Journal of the Indian Medical Association, 75:45-51.

Lawless JW, Latham MC, Stephenson LS, Kinoti SN, Pertet AM (1994) Iron supplementation improves appetite and growth in anaemic Kenyan primary school children. Journal of Nutrition, 124:645-654.
Llewellyn-Jones D (1965) Severe anaemia in pregnancy. Australian and New Zealand Journal of Obstetrics and Gynaecology, 5:191-197.
Looker AC, Dallman PR, Carroll MD, Gunter EW, Johnson CL (1997) Prevalence of iron deficiency in the United States. Journal of the American Medical Association, 277:973-976.

Lozoff B, Jimenez E, Wolf AW (1991) Long-term developmental outcome of infants with iron deficiency. New England Journal of Medicine, 325:687-694.

Macgregor MW (1963) Maternal anaemia as a factor in prematurity and perinatal mortality. Scottish Medical Journal, 8:134-140.

Mahomed K (2000a) Iron supplementation in pregnancy. Cochrane Database System Reviews, (2):CD000117.
Mahomed K (2000b) Iron and folate supplementation in pregnancy. Cochrane Database System Reviews, (2):CD001135.
Menendez C, Kahigwa E, Hirt R et al. (1997) Randomised placebo-controlled trial of iron supplementation and malaria chemoprophylaxis for prevention of severe anaemia and malaria in Tanzanian infants. The Lancet, 350: 844-850.

Menendez C, Todd J, Alonso PL et al. (1994) The effects of iron supplementation during pregnancy, given by traditional birth attendants, on the prevalence of anaemia and malaria. Transactions of the Royal Society of Tropical Medicine and Hygiene, 88:590-593.

Mola G, Permezel M, Amoa AB, Klufio CA (1999) Anaemia and perinatal outcome in Port Moresby. Australian and New Zealand Journal of Obstetrics and Gynaecology, 39:31-34.
Murphy JF, O’Riordan J, Newcombe RG, Coles EC, Pearson JF (1986) Relation of haemoglobin levels in first and second trimesters to outcome of pregnancy. The Lancet, 1:992-994.
Murray CJ, Lopez AD, eds. (1996a) The global burden of disease a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, Vol 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Murray CJ, Lopez AD, eds. (1996b) Global health statistics: a compendium of incidence, prevalence and mortality estimates for over 200 conditions. Global Burden of Disease and Injury, Vol 2. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Murray MJ, Murray AB, Murray NJ, Murray MB (1978) The effect of iron status of Nigerian mothers on that of their infants at birth and 6 months, and on the concentration of Fe in breast milk. British Journal of Nutrition, 39:627-630.
Oppenheimer SJ (2001) Iron and its relation to immunity and infectious disease. Journal of Nutrition, 131:S616-635.
Oppenheimer SJ, Macfarlane SB, Moody JB, Bunari O, Hendrickse RG (1986) Effect of iron prophylaxis on morbidity due to infectious disease: report on clinical studies in Papua New Guinea. Transactions of the Royal Society of Tropical Medicine and Hygiene, 80:596-602.

Palti H, Pevsner B, Adler B (1983) Does anaemia in infancy affect achievement on developmental and intelligence tests? Human Biology, 55:183-194.
Rasmussen KM (2001) Is there a causal relationship between iron deficiency or iron-deficiency anaemia and weight at birth, length of gestation and perinatal mortality? Journal of Nutrition, 131:S590-603.
Sarin AR (1995) Severe anaemia of pregnancy, recent experience. International Journal of Gynaecology and Obstetrics, 50:S45-49.
Sloan NL, Jordan E, Winikoff B (2002) Effects of iron supplementation on maternal hematologic status in pregnancy. American Journal of Public Health, 92:288-293.

Smith AW, Hendrickse RG, Harrison C, Hayes RJ, Greenwood BM (1989) The effects on malaria of treatment of iron-deficiency anaemia with oral iron in Gambian children. Annals of Tropical Paediatrics, 9:17-23.

Snedecor GW, Cochran WG (1980) Statistical Methods. Iowa State University Press, Ames.

Stoltzfus RJ (2001) Iron-deficiency anaemia: reexamining the nature and magnitude of the public health problem. Summary: implications for research and programs. Journal of Nutrition, 131:S697-700.
Stoltzfus RJ, Dreyfuss ML (1998) Guidelines for the use of iron supplements to prevent and treat iron deficiency anaemia. ILSI Press, Washington, DC.

Stoltzfus RJ, Kvalsvig JD, Chwaya HM et al. (2001) Effects of iron supplementation and anthelmintic treatment on motor and language development of preschool children in Zanzibar: double blind, placebo controlled study. British Medical Journal, 323:1389-1393.
Tasker PWG (1958) Anaemia in pregnancy. A five year appraisal. Medical Journal of Malayasia, 8:3-8.
UNICEF (1995) The state of the world's children 1995. Oxford University Press, Oxford.
UNICEF (1998) The state of the world's children 1998. Oxford University Press, Oxford.
Wasserman G, Graziano JH, Factor-Litvak P et al. (1992) Independent effects of lead exposure and iron deficiency anaemia on developmental outcome at age 2 years. Journal of Pediatrics, 121:695-703.
WHO (2000) Nutrition for health and development. A global agenda for combating malnutrition. A progress report. (WHO/NHD/00.6). World Health Organization, Geneva.

Xiong X (1996) Anaemia during pregnancy and birth outcomes: new data from China *̛ a meta-analysis [Dissertation]. Université Libre de Bruxelles Health Science Library, Brussels.
Xiong X, Buekens P, Alexander S, Demianczuk N, Wollast E (2000) Anaemia during pregnancy and birth outcome: a meta-analysis. American Journal of Perinatology, 17:137-46.
Xiong X, Buekens P, Alexander S, Wollast E (1996) The relationship between anaemia during pregnancy and birth outcomes. Archives of Public Health, 53:S136.

Yip R, Stoltzfus RJ, Simmons WK (1996) Assessment of the prevalence and the nature of iron deficiency for populations: the utility of comparing haemoglobin distributions. In: Iron nutrition in health and disease. Hallberg L, ed. John Libby \& Co., London.

## Chapter 4

# Vitamin A Deficiency 

Amy L. Rice, Keith P. West Jr. and Robert E. Black

## Summary

Vitamin A deficiency is a common form of micronutrient malnutrition affecting $21.1 \%$ of preschool-age children and $5.6 \%$ of pregnant women worldwide.

The published literature linking vitamin A interventions to causespecific child mortality due to measles, diarrhoea, malaria and other infectious diseases and to all-cause maternal mortality was comprehensively reviewed. Randomized controlled trial data of vitamin A interventions and survival were used to estimate the risk of mortality associated with vitamin A deficiency. The published relative risks were adjusted for the estimated prevalence of deficiency at study baseline. Summary relative risks were calculated from meta-analyses (for measles, diarrhoea, and other infectious disease causes of child mortality) or from single studies (malaria mortality among children and all-cause maternal mortality).

The estimated relative risks associated with vitamin A deficiency in children were 1.86 ( $95 \%$ CI $1.32-2.59$ ) for measles mortality, 2.15 ( $95 \%$ CI 1.83-2.58) for diarrhoea mortality, 1.78 ( $95 \%$ CI 1.43-2.19) for malaria mortality, 1.13 ( $95 \%$ CI 1.01-1.32) for other infectious disease mortality and $4.51(95 \%$ CI $2.91-6.94)$ for all-cause maternal mortality.

The available evidence suggests that nearly 800000 deaths worldwide can be attributed to vitamin A deficiency among women and children. Approximately $20-24 \%$ of child mortality from measles, diarrhoea and malaria and $20 \%$ of all-cause maternal mortality can be attributed to this preventable condition. Africa and South-East Asia have the highest burden of disease.

## 1. Introduction

The objective of this chapter is to summarize the available evidence that can be used to quantitatively estimate the risk of adverse health outcomes associated with vitamin A deficiency and to calculate the associated burden of disease for different regions of the world. Vitamin A is an essential nutrient required for maintaining immune function, eye health, vision, growth and survival in human beings. Over the years, numerous studies have been conducted to identify the biological functions of vitamin A , the health consequences associated with deficiency, and the mechanisms that explain these relationships. Causal relationships have been clearly demonstrated in some instances and comprehensive reviews on the subject have been published (Sommer and West 1996).

According to World Bank estimates, vitamin A supplementation for preschool-age children is one of the most cost-effective child survival interventions (World Bank 1993). National level public health programmes to prevent and treat vitamin A deficiency are currently being implemented in countries in Asia, Africa and elsewhere. International donors and agencies including the Canadian International Development Agency (CIDA), United Nations Children's Fund (UNICEF), the United States Agency for International Development (USAID), the World Health Organization (WHO) and others have actively supported both national and global level initiatives to raise awareness about the problem of vitamin A deficiency and to promote efforts to implement effective and affordable solutions (Mason et al. 2001). Reducing the prevalence of vitamin A deficiency will lessen disease burden by improving immune function, lowering mortality rates and preventing blindness, especially among children. This chapter will contribute to ongoing efforts to assess the global problem of vitamin A deficiency by using existing data to estimate global prevalence rates, to identify and quantify the adverse health consequences associated with deficiency, and to estimate the future health benefits that could be gained by implementing even more effective control programmes.

## 2. Definition of malnutrition and vitamin A deficiency

Malnutrition is a complex phenomenon. Broadly defined, malnutrition refers to the condition of inappropriate nutrition. In the past, discussions of malnutrition in the context of health issues in low-income countries often used this term to refer to the condition of "undernutrition" associated with what was presumed to be protein-energy malnutrition and operationally defined as a deficit in anthropometric status or by the presence of clinical signs such as oedema or altered hair colour. In more recent years, various vitamin and mineral deficiencies, including vitamin A, iron, iodine and zinc have been recognized as discrete types of mal-
nutrition that adversely affect human health and contribute to disease and mortality. Some of these nutrients affect closely related biological systems; for example both vitamin A and zinc play important roles in maintaining different aspects of immune function (Shankar 2001) and both vitamin A and iron affect haemoglobin metabolism (Semba and Bloem 2002). Ecological-level studies have demonstrated that the prevalence of these micronutrient deficiencies are high in many of the same countries, thus many individuals may suffer from multiple micronutrient deficiencies at the same time. However, relatively few data are currently available for quantifying either the joint distribution of multiple deficiencies or the impact that multiple micronutrient deficiencies have on specific health outcomes. Therefore, the comparative risk assessment (CRA) project will estimate the separate contribution of these risk factors to the global burden of disease. Individual reviews are available for the risk factors of iron deficiency (chapter 3), zinc deficiency (chapter 5) and underweight status (chapter 2) in addition to vitamin A deficiency, which is the subject of the current chapter.

Vitamin A is an essential nutrient required for maintaining immune function, eye health, vision, growth and survival in human beings (National Research Council 1989). Although animal studies that identified vitamin A as a necessary factor for rat growth were conducted in the early 1900 s and the chemical structure of the vitamin was elucidated over 20 years later, reports describing the link between xerophthalmia (signs in the eye of disease due to a severe lack of the vitamin) and successful treatment with animal liver (a rich source of the vitamin) date back to the medical writings of ancient Egypt (Olson 1996).

At present, vitamin A deficiency remains a widespread public health problem, especially in countries of South Asia and Africa. Globally, preschool-age children and women of reproductive age are the two population groups most commonly recognized to be at risk of this nutritional deficiency and its adverse health consequences. A combination of chronically lower than required dietary intakes of vitamin A-rich foods (eggs, milk, liver, deep orange fruits and dark green leafy vegetables, etc.) combined with malabsorption and increased vitamin A excretion rates associated with some common illnesses places many women and children at risk of developing vitamin A deficiency (Christian et al. 1998b; IVACG 1997; Sommer and West 1996; Stephensen 2001).

No single indicator can be reliably used to assess the full spectrum of vitamin A deficiency. Different aspects of vitamin A status are assessed using clinical indicators, biochemical indicators, functional indicators and histological indicators (WHO 1996). In humans, vitamin A is stored almost exclusively ( $>90 \%$ ) in the liver and some investigators propose liver and/or total body stores as a primary indicator of vitamin A status. Although recent isotope dilution techniques to indirectly measure liver vitamin A stores have yielded promising results (Haskell et al. 1999;

Ribaya-Mercado et al. 1999) these techniques have not yet been used in large-scale population-based surveys.

Severe vitamin A deficiency can be identified by the presence of the classical eye signs of xerophthalmia in individuals. However, because severe vitamin A deficiency is relatively rare in most populations, a large number of individuals must be surveyed in order to generate a reliable prevalence estimate. Depending on the severity of vitamin A deficiency in a population, the sample size requirement for a xerophthalmia survey may be nearly ten times higher than what would be required to generate a reliable prevalence estimate for other indicators of vitamin A status, such as low serum retinol concentrations, which may occur more frequently in the same population (Sommer and Davidson 2002; WHO 1996).

Milder vitamin A deficiency is far more common, but the assessment of vitamin A deficiency that does not result in relatively easily observable eye signs is also more problematic. One way to identify milder forms of vitamin A deficiency is to collect blood samples and measure the concentration of circulating serum retinol in an individual. Values $<0.70 \mu \mathrm{~mol} / 1$ have traditionally been considered indicative of deficiency in children, based on empirical data from population-based studies that did not exclude individuals based on measurements of acute phase proteins. In adults, appropriate cut-off levels are less firmly established, but values $<0.70 \mu \mathrm{~mol} / 1$ and $<1.05 \mu \mathrm{~mol} / \mathrm{l}$ have been used for different purposes. Because serum retinol concentrations are transiently depressed during the acute phase response to certain infections, some investigators have questioned the validity of using this indicator to assess the vitamin A status of individuals (Stephensen 2001). However, determining the prevalence of serum retinol concentrations below a defined cut-off point remains one of the most commonly used and widely accepted approaches for assessing the vitamin A status of entire populations (Sommer and Davidson 2002).

At the population level, the prevalence of vitamin A deficiency can be determined based on the prevalence of either: (i) night blindness, usually obtained by verbal recall; (ii) other eye signs of xerophthalmia (Bitot's spots or corneal lesions); or (iii) biochemical indicator values (serum retinol, breast milk retinol, relative dose-response test, modified relative dose-response test, or serum 30-day response), or histological indicator values (conjunctival impression cytology [CIC] that fall below a defined cut-off point (WHO 1996). Until recently the majority of nationally representative, large-scale surveys related to vitamin A deficiency were conducted primarily among preschool-age children. However, in the past few years some large-scale surveys, including recent demographic and health surveys, have also attempted to estimate the prevalence of night blindness among pregnant women. More limited survey findings are available for serum and breast milk retinol concentrations among women. Surveys that included data on any or all of the indicators listed above were used
as a basis for estimating the global prevalence of vitamin A deficiency as a risk factor for the CRA project. The categorical definitions chosen to represent "vitamin A deficiency" among children (aged 0-4 years) and women (aged 15-44 years) for the CRA project are described in the following sections. A description of the data, indicators, and the process used to estimate the current prevalence of vitamin A deficiency among preschool-age children and pregnant women is presented in section 3.5.

### 2.1 Definitions for children (0-4 Years)

Globally, the most reliable population-based survey data provide estimates of vitamin A deficiency among children aged <5 years, primarily because this is the most well-established high-risk age group for this nutritional risk factor. In the CRA project prevalence estimates have been developed only for the $0-4$-year age group, although there is some evidence that slightly older children also suffer from vitamin A deficiency and its adverse health consequences.

In order to calculate the attributable fraction of an adverse health outcome that is due to a risk factor, compatible definitions must be used when estimating the relative risk of the adverse outcome associated with the risk factor and the prevalence of the risk factor itself in a population. This stringent requirement for a compatible definition greatly influenced the data that were suitable for use in the CRA project. The majority of large-scale vitamin A intervention trials involving preschoolage children have been mortality studies that assumed (but did not confirm for all participants) that the children were mildly deficient (i.e. had low serum retinol concentrations). In general, very few participants in those studies exhibited eye signs of xerophthalmia, which is consistent with the expected epidemiological pattern of vitamin A deficiency. Vitamin A receipt was associated with a lower relative risk of adverse outcomes in those trials, even among children who had no eye signs of deficiency.

After considering the availability of intervention trial data and global prevalence data for vitamin A deficiency among preschool-age children, a definition of vitamin A deficiency related to low serum retinol concentrations among children in the $0-4$-year age range emerged as the most appropriate choice for use in the CRA project:

- Vitamin A deficient: serum retinol concentration $<0.70 \mu \mathrm{~mol} / 1$.
- Vitamin A sufficient: serum retinol concentration $\geq 0.70 \mu \mathrm{~mol} / \mathrm{l}$.


### 2.2 Global prevalence estimates for children

The 1995 WHO report Global prevalence of vitamin A deficiency included prevalence estimates of vitamin A deficiency among preschoolage children for two classes of indicators: (i) clinical eye signs of disease (xerophthalmia); and (ii) low serum retinol concentrations. However, data were reported only for the individual countries that met the defini-
tion of a significant public health problem, which was defined as a population prevalence of low serum retinol ( $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ ) $\geq 10 \%$, of night blindness (XN) $>1 \%$, of Bitot's spots (X1B) $>0.5 \%$, of corneal xerosis and/or ulceration (X2, X3A, X3B) $>0.01 \%$, or of xerophthalmiarelated corneal scars (XS) $>0.05 \%$ (WHO 1995). Estimates for countries that did not meet those definitions were not incorporated into the global prevalence estimate. A global estimate for the total number of children at risk of vitamin A deficiency ( $\sim 254$ million) was generated by adding the estimated number of children with low serum retinol concentrations ( $\sim 251$ million) and the estimated number with clinical eye signs of vitamin A deficiency ( $\sim 3$ million).

In 1998, the Micronutrient Initiative, UNICEF and Tulane University published a joint report reviewing the recent progress of vitamin A intervention programme activities (Micronutrient Initiative/UNICEF/Tulane University 1998). To generate prevalence estimates for vitamin A deficiency among children, this group used a different approach and developed a modelling process that attempted to take into account the effects of time trends and vitamin A intervention programmes. Data from a subset of countries that had conducted prevalence surveys since the mid-1980s were used to estimate the regional and global prevalence of vitamin A deficiency among children. The number of children affected by low serum retinol concentrations was estimated to range from 75 to 140 million, while the number of children with clinical eye signs of deficiency was estimated as 3 million. The sizeable discrepancy in the estimated prevalence of vitamin A deficiency from these two sources, 254 million and 78 to 143 million children, respectively, was due in part to differences in the data and methodology used, but also to a calculation error in the WHO report (West 2002).

An updated estimate for the global prevalence of vitamin A deficiency in children appears in the 2002 publication Extent of vitamin A deficiency among preschool children and women of reproductive age (West 2002), hereafter referred to as the 2002 West report. A brief description of the methodology used in developing the estimate is described in this chapter in section 3.5. Those prevalence data served as the basis for calculating the global burden of disease attributable to vitamin A deficiency among children aged 0-4 years for the CRA project.

### 2.3 Defintitions for pregnant women (15-44 Years)

In recent years, women of reproductive age have increasingly been recognized as an important group at risk of vitamin A deficiency and the adverse health outcomes associated with this condition (West 2002). However, when compared to preschool-age children, far less information is available to quantitatively estimate the burden of disease among women. Current areas of active research include assessing the magnitude of the problem, investigating the causes of deficiency in women, describing the range of associated adverse health outcomes, and identifying
appropriate interventions for preventing and treating vitamin A deficiency.

Standardized indicators and definitions of vitamin A deficiency among women are only beginning to be developed. To date, relatively few largescale surveys have been conducted to estimate the prevalence of vitamin A deficiency in women-primarily in Asia and Africa. Although many surveys used the presence of night blindness as an indicator of poor vitamin A status among women, some survey data related to low serum retinol and breast milk vitamin A concentrations are also available. However, very few studies have been conducted to date that quantitatively relate the risk of vitamin A deficiency (defined by any indicator) to adverse health outcomes in women. Despite the inherent limitations in the current data related to vitamin A deficiency in women, the evidence was considered strong enough to generate initial estimates of the related global burden of disease. Further work in this area of research will certainly lead to a refinement of these estimates.

The requirement to have a compatible definition of vitamin A deficiency for estimating the relative risk of adverse health outcomes and the global prevalence of vitamin A deficiency among women determined which data were suitable for use in the CRA project. After considering the availability of intervention trial data and global prevalence data for vitamin A deficiency in women, a definition related to low serum retinol concentrations among pregnant women in the 15-44-year age range emerged as the most appropriate choice for use in the CRA project.

Estimates for the prevalence of vitamin A deficiency have been generated only for pregnant women in the $15-44$-year age range primarily because the strongest information is available for this particular group of women. Future projects that quantify the global burden of disease may include non-pregnant and older women as well, if stronger data have become available for quantitatively estimating the prevalence of deficiency and adverse health outcomes in these groups. Vitamin A deficiency in pregnant women aged 15-44 years was operationally defined as:

- Vitamin A deficient: Serum retinol concentration $<0.70 \mu \mathrm{~mol} /$.
- Vitamin A sufficient: Serum retinol concentration $\geq 0.70 \mu \mathrm{~mol} / \mathrm{l}$.


### 2.4 Global PREvalence estimates For pregnant women

The first comprehensive estimate for the global prevalence of vitamin A deficiency among women of reproductive age appeared in the 2002 West report (West 2002). A brief description of the methodology used in developing the estimate has been described in this chapter in section 3.5. Those prevalence data will serve as the basis for calculating the global burden of disease attributable to vitamin A deficiency among pregnant women aged 15-44 years for the CRA project.

## 3. Adverse health outcomes considered FOR REVIEW

Within the framework of the larger Global Burden of Disease (GBD) project, health states (in this case vitamin A deficiency) can contribute directly or indirectly to death and disability. The total amount of death and disability attributed to vitamin A deficiency is therefore a sum of its direct and indirect effects. The direct contribution is measured by estimating the burden associated with the sequelae assigned to vitamin A deficiency as defined by the corresponding chapter in the International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10), whereas the indirect contribution is measured by considering vitamin A deficiency as a risk factor for other causes of death and disability.

Adverse health outcomes generally associated with micronutrient and mineral deficiencies were initially considered for review in relation to vitamin A deficiency. These included blindness; impaired cognitive function; impaired physical work capacity; morbidity (incidence and/or severity) due to diarrhoea, measles, acute respiratory infections, malaria and other infectious diseases; cause-specific mortality related to these diseases; and all-cause mortality. Outcomes potentially associated with vitamin A deficiency in pregnant women included fetal loss, low birth weight, preterm birth, all-cause infant mortality, maternal morbidity and maternal mortality.

### 3.1 Selection criteria for adverse health outcomes

The adverse health outcomes selected for review and inclusion in the CRA project were those that fulfilled the following criteria: (i) a health outcome included in the 2000 GBD study; (ii) a health outcome where data exist to support a probable causal relationship with vitamin A deficiency; and (iii) a health outcome where data were available to quantitatively estimate the relationship with vitamin A deficiency.

Blindness and an increased risk of all-cause mortality are two of the most well-established adverse outcomes associated with vitamin A deficiency in preschool-age children (Sommer 1982; Sommer and West 1996). According to recent estimates, vitamin A deficiency that leads to corneal scarring remains one of the most common preventable causes of childhood blindness in developing countries (Gilbert and Foster 2001). In the 1980s, vitamin A supplementation was shown to significantly reduce all-cause child mortality in a series of eight large-scale trials conducted in Asia and Africa. Meta-analyses of those trial data suggest a $23 \%$ to $34 \%$ reduction in all-cause mortality among children 6 months to 5 years of age (Beaton et al. 1993; Fawzi et al. 1993; Glasziou and Mackerras 1993; Tonascia 1993). Studies that provided vitamin A supplements to newborn infants immediately after birth have also demonstrated a reduction in mortality (Humphrey et al. 1996; Tielsch et al.
2001), although other trials involving vitamin A supplementation of young infants post-neonatally through five months of age have not shown a survival benefit (Anonymous 1998b; Daulaire et al. 1992; West et al. 1995).

Blindness was excluded as a health outcome from the risk factor analysis because in the GBD study, vitamin A-related blindness is considered to be a direct functional outcome of the deficiency, and the disability associated with blindness was measured in this manner (Murray et al. 1996a, 1996b). In the case of child mortality, the contribution of vitamin A deficiency to the global burden of disease was measured primarily through its contribution as a risk factor for several types of cause-specific mortality, rather than as a risk factor for all-cause mortality. However, a small number of child deaths have also been directly attributed to vitamin A deficiency in the GBD database, because vitamin A deficiency itself appears as the underlying cause of death in some vital registration data sets. For this specific cause of death, by definition, the total number of deaths in a subregion ${ }^{1}$ was directly assigned the value obtained from the relevant child mortality statistics for that subregion (Murray et al. 2001). Thus, for the CRA project the total number of child deaths attributable to vitamin A deficiency is the sum of those that are directly and indirectly attributed to the deficiency.

Five outcomes potentially associated with maternal vitamin A deficiency during pregnancy were also excluded from the risk factor analysis because the outcomes were not assessed in the 2000 GBD study or the data were insufficient to quantitatively assess the risk of their occurrence. These were fetal loss, low birth weight, preterm birth, all-cause infant mortality and general maternal morbidity (Christian et al. 1998b, 2000, 2001; Katz et al. 2000; Semba et al. 1998). Impaired cognitive function and impaired physical work capacity were excluded from the analysis because there is little evidence to suggest a biologically plausible association with vitamin A deficiency, except for the condition of maternal night blindness, which appears to limit the time for performing household chores to daylight hours (Christian et al. 1998a). Although experimental animal data and observational human data suggest a biologically plausible role for vitamin A deficiency predisposing children to acute respiratory infections (Bloem et al. 1990; Milton et al. 1987; Sommer and West 1996; Sommer et al. 1987), morbidity and causespecific mortality related to acute respiratory infections were excluded from the analysis because the data from controlled intervention trials in humans have not, to date, consistently revealed measurable effects of vitamin A on incidence, duration or severity of acute respiratory infections (Anonymous 1995).

After excluding those outcomes from the initial list of outcomes under consideration, the following remained for a more detailed review: (i) morbidity and mortality associated with measles in children; (ii) morbidity and mortality associated with diarrhoea in children; (iii) morbid-
ity and mortality related to malaria in children; (iv) selected other infectious disease causes of death in children (other than measles, diarrhoea, malaria or acute respiratory infections); and (v) all-cause maternal mortality. All-cause maternal mortality was a compilation of three separate conditions: maternal sepsis; maternal haemorrhage; and obstructed labour. These last three outcomes related to maternal mortality were selected from the limited number of maternal health conditions that are coded separately in the ICD-10 coding scheme.

The health outcomes selected for children aged $<5$ years represent the most common preventable causes of death among this age group (Murray et al. 2001). Although some studies have also explored the link between vitamin A and other specific infectious diseases coded as individual cause of death categories in the GBD study (for example helminthic infections and tuberculosis), the strength of the current evidence was considered insufficient to demonstrate a causal link and quantitatively estimate the associated risk with these specific outcomes. HIV/AIDS was also excluded as a quantifiable health outcome for the same reason. The findings relevant to the outcomes that were chosen for inclusion are presented below.

### 3.2 Methods for identifying relevant studies and REVIEW MATERIALS

The following sources were initially consulted to identify relevant materials for this chapter: Medline database; published books about vitamin A, international health, and nutrition; International Vitamin A Consultative Group (IVACG) statements; meeting reports; abstracts and conference proceedings; and other non-peer reviewed literature sources related to vitamin A programme implementation and cost-effectiveness analyses. The Medline database was searched for literature published between 1966 and 2001 in English or with an English language abstract. Combinations of the following keywords were used: vitamin A, vitamin A deficiency, blindness, mortality, acute respiratory infection, pneumonia, diarrhoea, measles, malaria, stillbirth, fetal loss, miscarriage, low birth weight, women.

Abstracts of articles concerning the relationship between vitamin A deficiency in humans, intervention trials, and the health outcomes of interest were reviewed and copies of relevant articles were obtained. Additional publications and reference materials were identified from the citation lists in those sources and through discussions with investigators working in the field.

### 3.3 Inclusion criteria for individual studies

The individual studies and reports presented in this chapter were restricted to the results of controlled intervention trials because these findings provide strong evidence for a causal relationship and the data
can be used to quantify the risk associated with either documented or suspected vitamin A deficiency (Rothman and Greenland 1998). For the outcomes where published meta-analyses or international consensus statements from IVACG exist, the results from those sources have been included in this chapter rather than a detailed presentation of data from the individual trials.

### 3.4 Description of excluded studies

Numerous observational cohort studies, case-control studies, and caseseries investigations have been conducted over the years to explore the relationships between vitamin A and morbidity or mortality from specific diseases in children. However, the results of those studies are not presented or discussed in detail in this chapter because such designs provide weaker evidence for a causal relationship as compared to randomized controlled intervention trials. Comprehensive reviews of the numerous cohort studies, case-control studies, and case-series reports related to various adverse health outcomes can be found elsewhere (Bauernfiend 1986; Sommer and West 1996). The findings from intervention studies related to vitamin A deficiency and child morbidity have also been summarized elsewhere (Nalubola and Nestel 1999).

### 3.5 Estimating risk factor levels

The prevalence of vitamin A deficiency (defined as serum retinol concentrations $<0.70 \mu \mathrm{~mol} / 1$ ) among children aged $0-4$ years and pregnant women aged 15-44 years was estimated for each of the 14 subregions. This process involved several steps.

First, country-specific prevalence rates were estimated for each one of the 191 WHO Member States. Updated country-specific prevalence rates were obtained from the recent review, 2002 West report (West 2002), which includes the estimated prevalence and number of deficient children and pregnant women in countries where vitamin A deficiency is either documented or presumed to exist, based on non-population-based vitamin A status data or other indirect indicators. Since separate prevalence estimates were not reported for boys and girls, the same prevalence rate was applied to both of these groups in the CRA project analyses. Next, the prevalence of vitamin A deficiency for the countries not included in that review was assumed to be zero. A suitable database was created for analysis and the findings were summarized across the 14 subregions.

A detailed description of the methods used to compile the data for the 2002 West report appears in that document (West 2002), but a brief summary is presented here. In addition, a file that contains a complete listing of the contributing studies and technical notes associated with the 191 countries can be found at http://www.jhsph.edu/chn/ GlobalVAD.html.

## Data sources

A wide variety of data sources related to vitamin A deficiency were reviewed in order to obtain the most current information possible. These included: (i) the 1995 comprehensive survey report compiled by the WHO Micronutrient Deficiency Information System (MDIS) (WHO 1995); (ii) a 2001 update from the MDIS group at WHO that included national survey data published after 1995; (iii) the 1998 report published by the Micronutrient Initiative, UNICEF and Tulane University (Micronutrient Initiative/UNICEF/Tulane University 1998), which incorporated both the 1995 MDIS data and more recent country updates from 107 countries with UNICEF offices and programmes; (iv) published surveys and field studies that reported vitamin A status indicators in women or children; and (v) unpublished reports, meeting presentations and personal communication about recent field surveys and studies not included in other data sources.

## Indicators

The 2002 West report presents prevalence data for xerophthalmia rates and serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ among children and for night blindness rates and serum retinol concentrations $<0.70 \mu \mathrm{~mol} / /$ and $<1.05 \mu \mathrm{~mol} / \mathrm{l}$ among populations of pregnant women. Since the CRA project only utilized serum retinol data, the methodology used to generate xerophthalmia estimates is not discussed further in this chapter.

In preschool-age children the prevalence of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / 1$ was directly estimated from survey data whenever possible. When such data were unavailable, the prevalence was assigned a value equivalent to the population prevalence of abnormal CIC results. Survey data referring to children in a narrower age than 0-4 years or surveys that included data that extended slightly beyond the fifth year of life were used as the prevalence estimate for children aged $0-4$ years.

For pregnant women the prevalence of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / 1$ was directly estimated from appropriate survey data whenever possible. When such data were lacking, the prevalence was assigned a value equivalent to the prevalence among non-pregnant or lactating women in the early postpartum period or a value equivalent to the prevalence of breast milk retinol concentrations $<1.05 \mu \mathrm{~mol} / \mathrm{l}$. When serological or breast milk data were reported as a mean and SD, rather than as a prevalence rate, the prevalence was derived by assuming the data were normally distributed and calculating the standard normal deviate ( $z$ score) and the probability associated with the area under the left tail of the normal curve.

## Data extrapolation

Separate algorithms were developed for the different subregions of the world to estimate the country-specific prevalence estimates for serum retinol concentrations $<0.70 \mu \mathrm{~mol} / 1$ among children and pregnant
women (West 2002). In brief, nationally representative survey data, as deduced from individual reports or stated as such from aggregate WHO or other agency reports, were reported whenever possible. In the absence of national level data, results from sub-national or smaller surveys were used and adjustment factors were applied following the precedent of previous analysts (Micronutrient Initiative/UNICEF/Tulane University 1998; WHO 1995), although the subjective weight may have changed from the previous reports owing to new or reinterpreted results for a country. For countries where no data were available, estimates were generated by extrapolation in situations where cultural, dietary, demographic, health and development patterns as well as existing rates of adult and child mortality suggested that vitamin A deficiency is likely to exist. Prevalence rates were extrapolated either by assigning a value from a nearby country with comparable characteristics (primarily in the WHO Region of South-East Asia) or by assigning a median value from neighbouring country national surveys. Prevalence estimates among preschool-age children were also adjusted downwards in countries where the survey data preceded coverage reports from recent vitamin A supplementation programmes that reported coverage rates $>75 \%$. The prevalence data for women were not adjusted to account for any potential programmatic impact, because although postpartum maternal vitamin A supplementation programmes are slowly emerging, programmes to prevent vitamin A deficiency during pregnancy are virtually non-existent in the developing world at the present time.

The 2002 West report did not generate prevalence estimates for countries where there was no plausible evidence to suggest the presence of vitamin A deficiency. Therefore, those subregional prevalence estimates do not reflect the contribution of other countries in the world that were assumed to have a $0 \%$ prevalence rate of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / /$. Over half of the countries assigned a $0 \%$ prevalence rate are located in the subregions classified as EUR-A, EUR-B, EUR-C or AMR-A, where child mortality rates are low (Murray et al. 2001).

The distribution of the 191 countries included in the CRA project is shown in Table 4.1 by the type of survey data used for estimating serum retinol prevalence rates $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ in children. To calculate subregional prevalence rates of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$, the number of affected individuals was totalled for the countries in each subregion and then divided by the subregional population base. For children $<5$ years of age, the population base was obtained from the 2001 State of the world's children report (UNICEF 2001), which reported data for the year 1999. For pregnant women the annual number of live births was chosen to represent the population global base for pregnant women, and data were obtained from the same source (UNICEF 2001). The countries that were assigned $0 \%$ prevalence accounted for only approximately $17 \%$ of the base population of children (Table 4.2) and pregnant women (Table 4.3).
Table 4.I
concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ was estimated by subregion and type of data used to generate subregional prevalence estimates

| Subregion | Population group |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Children (0-4 years) |  |  |  |  | Pregnant women (15-44 years) |  |  |  |  |
|  | Total | National level data | Sub-national level data | Imputed prevalence estimates >0 | No available survey data, assigned 0\% prevalence estimate | Total | National level data | Sub-national level data | Imputed prevalence estimates >0 | No available survey data, assigned $0 \%$ prevalence estimate |
| AFR-D | 26 | 5 | 7 | 14 | 0 | 26 | 1 | 5 | 20 | 0 |
| AFR-E | 20 | 10 | 7 | 3 | 0 | 20 | 3 | 7 | 10 | 0 |
| AMR-A | 3 | 0 | 0 | 0 | 3 | 3 | 0 | 0 | 0 | 3 |
| AMR-B | 26 | 9 | 2 | 0 | 15 | 26 | 0 | 2 | 9 | 15 |
| AMR-D | 6 | 4 | 1 | 1 | 0 | 6 | 0 | 1 | 5 | 0 |
| EMR-B | 13 | 1 | 1 | 0 | 11 | 13 | 0 | 0 | 0 | 13 |
| EMR-D | 9 | 2 | 3 | 4 | 0 | 9 | 1 | 3 | 5 | 0 |
| EUR-A | 26 | 0 | 0 | 0 | 26 | 26 | 0 | 0 | 0 | 26 |
| EUR-B | 16 | 1 | 0 | 0 | 15 | 16 | 0 | 0 | 0 | 16 |
| EUR-C | 9 | 0 | 0 | 0 | 9 | 9 | 0 | 0 | 0 | 9 |
| SEAR-B | 3 | 2 | 1 | 0 | 0 |  | 0 | 3 | 0 | 0 |
| SEAR-D | 7 | 3 | 0 | 3 | 1 | 7 | 2 | 2 | 3 | 0 |
| WPR-A | 5 | 0 | 0 | 0 | 5 | 5 | 0 | 0 | 0 | 5 |
| WPR-B | 22 | 5 | 2 | 9 | 6 | 22 | 2 | 2 | 12 | 6 |
| World | 191 | 42 | 24 | 34 | 91 | 191 | 9 | 25 | 64 | 93 |

Table 4.2

|  |  | Countries with assigne | evalence | mates ${ }^{\text {a }}$ | Countri assigned estimates set | with no revalence revalence 0\%) |  | Subregional and | obal total |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Subregion | Countries | Children with serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ (000s) | Children <br> $<5$ years (000s) | Prevalence of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ (\%) | Countries | Children $<5$ years (000s) | Countries | Children with serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}(000 \mathrm{~s})$ | Children $<5$ years (000s) | Prevalence of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ (\%) |
| AFR-D | 26 | 13552 | 47653 | 28.4 | 0 | NA | 26 | 13552 | 47653 | 28.4 |
| AFR-E | 20 | 19853 | 56281 | 35.3 | 0 | NA | 20 | 19853 | 56281 | 35.3 |
| AMR-A | 0 | NA | NA | NA | 3 | 21886 | 3 | 0 | 21886 | 0.0 |
| AMR-B | 11 | 7010 | 38256 | 18.3 | 15 | 6565 | 26 | 7010 | 44821 | 15.6 |
| AMR-D | 6 | 1209 | 9319 | 13.0 | 0 | NA | 6 | 1209 | 9319 | 13.0 |
| EMR-B | 2 | 449 | 7412 | 6.1 | 11 | 9022 | 13 | 449 | 16434 | 2.7 |
| EMR-D | 9 | 12215 | 52406 | 23.3 | 0 | NA | 9 | 12215 | 52406 | 23.3 |
| EUR-A | 0 | NA | NA | NA | 26 | 21852 | 26 | 0 | 21852 | 0.0 |
| EUR-B | 1 | 45 | 152 | 29.5 | 15 | 17935 | 16 | 45 | 18087 | 0.2 |
| EUR-C | 0 | NA | NA | NA | 9 | 12565 | 9 | 0 | 12565 | 0.0 |
| SEAR-B | 3 | 13538 | 28434 | 47.6 | 0 | NA | 3 | 13538 | 28434 | 47.6 |
| SEAR-D | 7 | 42274 | 140575 | 30.1 | 0 | NA | 7 | 42274 | 140575 | 30.1 |
| WPR-A | 0 | NA | NA | NA | 5 | 8019 | 5 | 0 | 8019 | 0.0 |
| WPR-B | 16 | 17128 | 122006 | 14.0 | 6 | 3792 | 22 | 17128 | 125798 | 13.6 |
| World | 101 | 127273 | 502494 | 25.3 | 90 | 101636 | 191 | 127273 | 604130 | 21.1 |
| NA |  |  |  |  |  |  |  |  |  |  |
| Country-specific data used to estimate subregional prevalence rates were based on the review article "Global prevalence of vitamin A defic of reproductive age" (West 2002). Technical notes and data for individual countries are located at www.jhsph.edu/chn/GlobalVAD.html. In the with serum retinol concentrations $<0.70 \mu \mathrm{~mol} / /$ in a country was either estimated based on existing survey data or imputed for countries wher based on cultural, dietary, demographic, health and development patterns as well as existing rates of child mortality. In order to calculate subr the CRA project, the prevalence estimate was set to $0 \%$ for the remaining 90 countries where no data on preschool child vitamin A deficien childhood vitamin A deficiency was considered unlikely to exist. These are all classified as "A" or "B" according to the WHO Comparative Rist <br> Region of the Americas $(n=18)$ Antigua and Barbuda, Argentina, Bahamas, Barbados, Canada, Chile, Cuba, Grenada, Guyana, Jamaica, P Saint Vincent and the Grenadines, Suriname, Trinidad and Tobago, United States of America and Uruguay; Eastern Mediterranean Region Lebanon, Libyan Arab Jamahiriya, Qatar, Saudi Arabia, Syria, Tunisia and United Arab Emirates; European Region ( $n=50$ ) All countries we Republic of Macedonia; Western Pacific Region $(n=1 I)$ Australia, Brunei Darussalam, Fiji, Japan, Mongolia, Nauru, New Zealand, Niue, |  |  |  |  |  |  |  |  |  |  |

### 3.6 Children (0-4 Years)

The subregional and global prevalence rates of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ among children aged $0-4$ years are shown in Table 4.2. The prevalence estimates from the smaller number of countries that contributed to the 2002 West report are compared to the global estimates generated for the CRA project. The results indicate that globally, approximately $21 \%$ of all children have serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$. The highest prevalence rates and the largest number of affected children live in the South-East Asian and African Regions. The estimated number of affected children is similar to what was reported by the Micronutrient Initiative, UNICEF and Tulane University group (Micronutrient Initiative/UNICEF/Tulane University 1998).

### 3.7 Pregnant women (15-44 Years)

The subregional and global prevalence rates of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ among pregnant women aged $15-44$ years are shown in Table 4.3. The prevalence estimates from the smaller number of countries that contributed to the 2002 West report are compared to the global estimates generated for the CRA project. The results indicate that globally, approximately $5.6 \%$ of all pregnant women have serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$. Although the prevalence rates among pregnant women are approximately one-fourth those observed in children aged $<5$ years, a similar risk distribution emerged-with the highest prevalence rates and number of affected women being located in the South-East Asian and African Regions. Although the global prevalence is low, vitamin A deficiency may be an important contributing factor to adverse health outcomes for pregnant women in selected areas of the world.

### 3.8 Other groups

There is very little, if any, information available about the global prevalence of vitamin A deficiency and the associated risk in other population groups. Therefore, the prevalence of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ among children aged $\geq 5$ years, among non-pregnant women aged 15-44 years, among men aged 15-44 years, and among adults aged $\geq 45$ years was not estimated.

## 4. Assessing CaUsAlity And QuANTIFYING RISK FACTOR-DISEASE RELATIONSHIPS

There is often a discrepancy between the problems that motivate a study and the data available for addressing the issue. This applies to the present work, especially when identifying data to be used in estimating the relative risk of adverse health outcomes associated with vitamin A deficiency. Ideally, the results of multiple studies would be available for each outcome of interest from different countries located in different regions
Table 4.3

 reproductive age" (West 2002). Technical notes and data for individual countries are located at www.jhsph.edu/chn/GlobalVAD.html. In the review article the number of pregnant women aged $15-45$ years with serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{I}$ in a country was either estimated based on existing survey data or imputed for countries where maternal opment patterns as w ( Comparative Risk Assessment Index (WHO 2001): Region of the Americas ( $n=18$ ) Antigua and Barbuda, Argentina, Bahamas, Barbados, Canada, Chile, Cuba, Grenada, Guyana, amaica, Paraguay, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Suriname, Trinidad and Tobago, United States and Uruguay; Eastern Mediterranean Region $(n=13)$ Bahrain, Cyprus, Iran, Jordan, Kuwait, Lebanon, Libyan Arab Jamahiriya, Oman, Qatar, Saudi Arabia, Syria, Tunisia and United Arab Emirates; European Region ( $n=5$ I) All countries were excluded; Western Pacific Region ( $n=11$ ) Australia, Brunei Darussalam, Fiji, Japan, Mongolia, Nauru, New Zealand, Niue, Republic of Korea, Samoa and Singapore. NA
of the world. This would allow the subregional estimates to be based on empirical data for each location. However, this type of data is simply not available. In the absence of such information the best data possible should be compiled and any important limitations recognized. There are several important limitations of the data and methodology used to estimate risk estimates for this project that deserve comment.

First, no definitive criteria exist for determining with certainty whether or not a particular risk factor is causally related to an adverse health outcome. However, many investigators have adopted the general principles that were originally proposed by Hill for use as guidelines when evaluating potential causal relationships (Hill 1965). A review of vitamin A studies conducted using different designs reveals that many of these general principles hold true when the entire body of evidence is considered together. However, experimental evidence-in this case the demonstration that a vitamin A-related intervention prevents an adverse health outcome-provides some of the strongest evidence for a causal relationship. Therefore, the present review has been restricted to the results of randomized placebo-controlled vitamin A intervention trials conducted in areas with either documented or suspected vitamin A deficiency. The inference commonly made from such a study is that if the vitamin A intervention prevents the occurrence of an adverse health outcome, then vitamin A deficiency is causally associated with it.

The design used for the prospective trials of vitamin A and child mortality was to assign the participants to either a vitamin A or control group, to implement the intervention, and to then follow the participants over time. Aside from observing the presence of clinical eye signs of deficiency, the vitamin A status (serum retinol concentration) of each and every participant was not assessed. The baseline vitamin A status of the populations under study was either inferred from prior survey data gathered in the same subregion or from a subset of the study participants.

Second, the risk estimates reported in the original publications of vitamin A and child mortality represent the protective effect of the vitamin A interventions against adverse health outcomes, rather than the relative risk of an adverse outcome associated with vitamin A deficiency per se. The risk of an adverse outcome was estimated from the original trial results as the inverse of the protective relative risk (= 1/protective relative risk). The following assumptions were made when using the intervention trial data in this manner: (i) that all children participating in the trials were vitamin A deficient at the beginning of the intervention period; and (ii) that the deficiency was corrected in all children assigned to the intervention group, while those in the placebo group remained deficient. Neither of these assumptions is likely to have been met in all of the trials. Thus the unadjusted risk estimates from the trials may underestimate the true relationship between vitamin A deficiency and an adverse health outcome.

Finally, the data from the original intervention trials were analysed on an intention-to-treat basis, rather than on the basis of achieved compliance. In reality both compliance and the biological efficacy of the particular intervention under study would influence the measured relationship between the vitamin A intervention and an adverse health outcome. Secondary analyses from one of the child mortality studies that took actual compliance into account estimated a far greater reduction in all-cause mortality than was observed in the original intention-to-treat analysis (Sommer and Zeger 1991).

On the other hand, the summary risk estimates from the metaanalyses of the individual intervention trials conducted in a variety of countries provide a certain level of built-in control for unmeasured factors that may have differed across sites. The United Nations Administrative Committee on Coordination/Subcommittee on Nutrition (ACC/SCN) meta-analysis of the child mortality trials examined the effect of age and sex on observed mortality and concluded there were no significant differences of vitamin A supplementation on all-cause mortality between males and females or by age category for children between 6 months and 5 years of age (Beaton et al. 1993). In addition, there was no detectable relationship between the effects of the vitamin A intervention and anthropometric status on child mortality.

For the purpose of the CRA project, relative risk estimates for child and maternal health outcomes used vitamin A intervention trial data as the starting point. However, the relative risk estimates were adjusted to take into account the fact that many, but not all, of the study participants had low serum retinol concentrations at the beginning of the intervention trials. The following section describes how the adjustment process was conducted. The same process was applied to data for both the child and maternal health outcomes.

The adjusted relative risks were calculated using a four-step process.

1. A quantitative estimate of the protective effect that a vitamin $A$ intervention had in preventing an adverse health outcome was found in the published literature.
2. The prevalence of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / 1$ was estimated among the study population at baseline.
3. An adjusted relative risk was calculated by constructing a hypothetical population of 100000 individuals and dividing them into two strata using a serum retinol concentration cut-off of $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ and the prevalence estimate obtained in the second step. The relative risk of an adverse outcome was then calculated for both strata separately by setting the background incidence rate of the adverse outcome to be equivalent among the following groups: (i) the vitamin A intervention group in the entire study population; (ii) the vitamin A intervention and control groups in the strata with serum retinol con-
centrations $\geq 0.70 \mu \mathrm{~mol} / 1$; and (iii) the vitamin A intervention group in the stratum with serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$. The relative risk for children in the stratum with serum retinol concentrations $\geq 0.70 \mu \mathrm{~mol} / \mathrm{l}$ represents the effect of the vitamin A intervention among children who were not deficient before the trial began. In this stratum the relative risk is 1.0 because those children were not expected to benefit (in terms of reducing all-cause mortality) from the intervention. The relative risk for children in the other stratum is lower than the overall trial estimate (representing a greater protective effect) because those children were deficient when the trial began and all of the observed benefit (in terms of reducing all-cause mortality) associated with the vitamin A intervention was presumably observed among this subgroup of children.
4. The final step in the adjustment process was to calculate the relative risk of all-cause mortality associated with vitamin A deficiency by calculating the inverse of the adjusted protective effect (=1/protective relative risk). See Table 4.4 for an example calculation based on a $23 \%$ reduction in child deaths (protective relative risk of 0.77 ) associated with a vitamin A intervention and a $41 \%$ baseline prevalence rate of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ in the study population. In this example, the originally reported protective effect associated with the vitamin A intervention is a relative risk of 0.77 ; the adjusted protective relative risk associated with the vitamin A intervention is 0.58 (equivalent to a $42 \%$ reduction in child deaths); and the adjusted relative risk of child death associated with vitamin A deficiency is 1.72 .

This adjustment process requires estimates for the baseline serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ for the populations contributing relative risk data to each adverse heath outcome. The baseline prevalence estimates are shown in Table 4.5. For some, but not all, adverse health outcomes, baseline prevalence data were directly available from the published reports that contributed relative risk estimates to the adjustment process. In other cases, the prevalence rates were extrapolated accordingly. The process used to derive prevalence rates for each included adverse health outcome is described below.

In order to derive estimates for the measles, diarrhoea, malaria and other infectious disease causes of death and disability in children attributable to vitamin A deficiency, it was necessary to estimate a single underlying prevalence of deficiency that existed in the southern Asian and African populations in which eight large community-based, vitamin A child-mortality intervention trials were conducted. Knowing a single, underlying prevalence of deficiency across these diverse trial populations reveals the background burden of vitamin A deficiency that is understood to account for the overall estimated reduction of $23 \%$ in
Table 4.4
Example of how to calculate the adjusted relative risk of child death associated with vitamin A deficiency assuming a published relative risk of 0.77 for child survival associated with the receipt of vitamin $A$ in a controlled intervention trial and a $41 \%$ baseline prevalence rate of 'vitamin A deficiency' among the children (defined as serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ )

| Study population | Died | Survived | Total | Incidence of death | Protective relative risk | \% reduction in deaths due to vitamin A intervention |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Entire study population |  |  |  |  |  |  |
| Vitamin A group | 385 | 49615 | 50000 | 0.0077 | 0.77 | 23 |
| Control group | 500 | 49500 | 50000 | 0.0100 |  |  |
| Total | 885 | 99115 | 100000 |  |  |  |
| Children with serum retinol concentrations $\geq 0.70 \mu \mathrm{~mol} / \mathrm{l}$ |  |  |  |  |  |  |
| Vitamin A group | 227 | 29273 | 29500 | 0.0077 | 1.00 | 0 |
| Control group | 227 | 29273 | 29500 | 0.0077 |  |  |
| Total | 454 | 58456 | 59000 |  |  |  |
| Children with serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ |  |  |  |  |  |  |
| Vitamin A group | 158 | 20342 | 20500 | 0.0077 | $0.58{ }^{\text {a }}$ | 42 |
| Control group | 273 | 20272 | 20500 | 0.0133 |  |  |
| Total | 431 | 40569 | 41000 |  |  |  |

a In the example above the relative risk of child death associated with vitamin A deficiency was calculated by using an overall observed relative risk of child survival associated with receipt of vitamin $A(0.77)$ and estimating the relative risk of child survival among the $59 \%$ of children who had serum retinol concentrations $\geq 0.70 \mu \mathrm{~mol} / \mathrm{I}$ ( I .00 ) and the $41 \%$ of children with concentrations below that cut-off ( 0.58 ). The relative risk of child death associated with vitamin $A$ deficiency was then calculated as 1.72 , which equals the inverse of the protective effect among "vitamin A deficient" children who received vitamin A: 1.72 [I.00/0.58].

| Table 4.5 | Estimated baseline prevalence of serum retinol <br> concentrations $<0.70 \mu \mathrm{~mol} / / \mathrm{in}$ the intervention studies used <br> to estimate the relative risk of cause-specific mortality <br> associated with vitamin A deficiency ${ }^{\mathrm{a}}$ |
| :---: | :--- |
| Cause of death | Estimated baseline <br> prevalence of serum retinol <br> concentrations $<0.70 \mu / l$ in <br> the intervention trials $(\%)$ |


| Children |  |  |
| :---: | :---: | :---: |
| Diarrhoea | 41 | See Table 4.6 |
| Measles | 41 | See Table 4.6 |
| Malaria | 55 | See Table I in Shankar et al. (1999) |
| Other infectious causes | 41 | See Table 4.6 |
| Women |  |  |
| Maternal conditions | 19 | See Table 2 in West et al. (1999) |
| a The cause of death outcomes were defined using the following GBD codes: measles (UOI5); diarrhoea (UOIO); malaria (U020); other infectious causes (U037); all-cause maternal mortality corresponds to the GBD code for maternal conditions (U042). |  |  |

Children
Diarrhoea
easles

Other infectious causes
Women
Maternal conditions

See Table 4.6
See Table 4.6
See Table I in Shankar et al. (1999)
See Table 4.6

See Table 2 in West et al. (1999) (UO10); malaria (U020); oher infectious cause (U037); all GBD code for maternal conditions (U042).
preschool-age child mortality achieved with vitamin A interventions (Beaton et al. 1993).

An overall background prevalence of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ was derived by estimating the size and prevalence of vitamin A deficiency for each national or regional population judged to be represented by each intervention trial (Table 4.6). The prevalence rate for each country as available, was applied to the estimated number of preschool-age children in order to estimate the number of vitamin A deficient preschool-age children in each population represented by a trial. These steps resulted in a subjective re-weighting of the sizes of the populations at-risk in each area that were independent of sample sizes for each trial. Finally, numbers of deficient children in each population represented by a trial were summed and divided by the sum of the population estimates of children aged $<5$ years, resulting in a prevalence that is, roughly, weighted by the sizes of populations at risk that were represented by the trials. This exercise produced prevalence rates of serum retinol $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ that ranged from $28 \%$ in the Sudan, to $72 \%$ as measured during a vitamin A child mortality trial in Ghana (Ghana VAST Study Team 1993). An estimated 19 million children, or $41 \%$, among the estimated 46.5 million children living in areas at risk of vitamin A deficiency, representing the underlying, local populations of interest, were considered to be vitamin A deficient. In the absence of additional data, this underlying prevalence was further judged to represent the underlying pool of deficient children for whom death from severe episodes of diarrhoea, measles and other non-malarial infectious disease illnesses could be averted each year with vitamin A.

| Table 4.6 | Baseline prevalence estimates for serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{I}$ among populations of children aged $<5$ years from eight vitamin $A$ intervention trials, by country |  |  |
| :---: | :---: | :---: | :---: |
| Host country | $\begin{gathered} \text { Population } \\ <5 \text { years of agea } \\ (000 \mathrm{~s}) \end{gathered}$ | Prevalence of serum retinol concentations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ | Number of vitamin A-deficient children (000s) |
| Ghana, Kitampo | $1063{ }^{\text {b }}$ | $0.72{ }^{\text {c }}$ | 765 |
| India, Andra Pradhesh | $8709^{\text {d }}$ | $0.31{ }^{\text {e }}$ | 2700 |
| India, Tamil Nadu | $7143^{\text {d }}$ | $0.37{ }^{\text {f }}$ | 2643 |
| Indonesia | $22006^{8}$ | $0.48{ }^{\text {h }}$ | 10563 |
| Nepal | $3485{ }^{\text {i }}$ | $0.35{ }^{\text {i }}$ | 1220 |
| Sudan | $4162^{\text {k }}$ | 0.28 | 1165 |
| Total/Overall | 46567 | $0.41{ }^{\text {m }}$ | 19055 |

a Based on Table I of the State of the world's children's report (UNICEF 200I), unless otherwise noted.
b Given that the vitamin A trial in Ghana (Ghana VAST Study Team 1993) was carried out in the central part of the country, which is considered to be at higher risk than the southern, palm-oil consuming areas of the country, only one-third of the Ghanaian child population $<5$ years of age ( 3189000 , UNICEF 2001) was considered to be represented by children in the trial.
c D Ross et al., personal communication, 1995, reported in Sommer and West (1996).
${ }^{\text {d }}$ Based on government of India census data for 2001 indicating $11.5 \%$ of the country's rural population was $<5$ years, applied to statewide census estimates for both Tamil Nadu and Andra Pradesh; statewide populations were assumed to represent the at-risk population for each trial.
e In the absence of serum retinol data from the trial or from representative population surveys of Andra Pradesh, a prevalence of $31 \%$ (West 2002) was applied to the state population.
f Rahmathullah et al. (I990).
$g$ Indonesia is represented by its entire population given that the two mortality trials carried out in Aceh (Sommer et al. 1986) and West Java (Muhilal et al. 1988) were conducted in the north and central parts of the country, respectively.
h Prevalence based on the West Java trial (Muhilal et al. 1988) was assumed to represent Aceh and the rest of Indonesia in the early-mid-1980s, amidst evidence of higher subsequent prevalence rates (Kjolhede et al. 1995; West 2002).
i Because two population-based trials (Daulaire et al. 1992; West et al. 1991) were conducted in diverse and different parts of the country, the entire population of Nepalese preschool children (UNICEF 2001) was considered to be the underlying population at risk.
j In the absence of biochemical data from either child mortality trial in Sarlahi (West et al. 1991) or Jumla (Daulaire et al. 1992), a prevalence of $35 \%$ obtained from the 1998 National Micronutrient Survey (Anonymous 1998a) was taken to represent the prevalence during both trials and for the country.
k In the absence of risk differentials across different population groups of the Sudan, the entire population was assumed to be represented by children in the trial (Herrera et al. 1992).
I Median prevalence (28\%) of distribution of 33 national prevalence estimates obtained for African and Eastern Mediterranean Regions was assumed to represent the status of Sudanese preschool children.
m Calculated by dividing 19055 by 46567.

A prevalence of $55 \%$ was assigned for populations of preschool-age children whose risk of death and disability due to Plasmodium falciparum malaria could be averted by vitamin A supplementation. This prevalence estimate was based on the data reported in Table 1 from a
single, community-based randomized trial in Papua New Guinea that measured effects of vitamin A supplementation on P. falciparum malaria clinic attack rates (Shankar et al. 1999). The assigned prevalence rate was calculated based on the published mean and standard deviation serum retinol concentration for the population at baseline and applying an assumption of normally distributed data.

A prevalence of $19 \%$ was assigned to represent populations of pregnant women living in areas of the world where risk of mortality may be reduced by approximately $40 \%$ with improved, regular, supplemental intakes of vitamin A, based on the data reported in Table 2 from a single, large randomized community recently conducted in rural Nepal (West et al. 1999). The same overall adjustment process was used to calculate adjusted relative risks of adverse health outcomes for each of the included child and maternal health outcomes. For the child health outcomes, the same adjusted relative risk was reported for children aged $0-4$ years (boys and girls) for all regions of the world.

## 5. RISK FACTOR-DISEASE OUTCOME RELATIONSHIPS

The following section describes the individual studies that were used to evaluate and quantify the relative risk of an adverse health outcome associated with vitamin A deficiency. The intervention trials and results are reported separately for children aged $0-4$ years and pregnant women aged 15-44 years. Insufficient data were available to quantitatively evaluate the risks of vitamin A deficiency in other population groups. Table 4.7 describes the individual studies that were used to estimate the relative risk of adverse health outcomes for the CRA project with respect to vitamin A. The studies are listed in the same order that the outcomes are discussed in the chapter. Mortality associated with measles and diarrhoeal disease in children appears first, followed by malaria incidence and malaria mortality in children, mortality associated with other infectious causes of death in children, and finally all-cause maternal mortality among pregnant women.

Unless otherwise noted, the risk estimates cited for individual studies are the findings that were originally reported in the literature and represent the protective effect of the vitamin A intervention against adverse health outcomes. An adjusted relative risk was calculated from these data to represent the risk of adverse health outcomes associated with vitamin A deficiency following the procedure described in section 4.

### 5.1 RISK estimates for preschool children (0-4 years)

Preschool-age children have traditionally been considered to be a highrisk group for vitamin A deficiency and its consequences. Most of the large-scale controlled intervention trials that have been conducted were designed to explore the relationships between vitamin A supplementation, morbidity and mortality among children aged $<6$ years.
Table 4.7

| Outcome (Reference) | Study design | Location | Study population |
| :---: | :---: | :---: | :---: |
| Children |  |  |  |
| Measles mortality (Beaton et al. 1993) | Meta-analysis of randomized, placebo-controlled vitamin A supplementation trials | 4 trials (Ghana; India; Sarlahi, Nepal; Jumla, Nepal) | Children 6-60 months |
| Original studies included in the meta-analysis: <br> (Ghana VAST Study Team 1993) <br> (Rahmathullah et al. 1990) <br> (West et al. 1991) <br> (Daulaire et al. 1992) | Large dose vitamin A every 4 months RDA of vitamin A in weekly doses Large dose vitamin A every 4 months Large dose vitamin A once with a 5 month follow-up | Ghana <br> India <br> Sarlahi, Nepal Jumla, Nepal | Children 6-90 months Children 6-60 months Children 6-72 months Children I-59 months |
| Diarrhoea mortality (Beaton et al. 1993) | Meta-analysis of randomized, placebo-controlled vitamin A supplementation trials | 5 trials (Ghana; India; Sarlahi, Nepal; Jumla, Nepal; Sudan | Children 6-60 months |
| Original studies included in the meta-analysis: <br> (Ghana VAST Study Team 1993) <br> (Rahmathullah et al. 1990) <br> (West et al. 1991) <br> (Daulaire et al. 1992) <br> (Herrera et al. 1992) | Large dose vitamin A every 4 months <br> RDA of vitamin A in weekly doses <br> Large dose vitamin A every 4 months <br> Large dose vitamin A once with a <br> 5 month follow-up <br> Large dose vitamin A every 4 months | Ghana India Sarlahi, Nepal Jumla, Nepal Sudan | Children 6-90 months Children 6-60 months Children 6-72 months Children I-59 months <br> Children 9-72 months |

Table 4.7 Studies used to estimate the relative risk of adverse health outcomes for the CRA project, by outcome ${ }^{\text {a }}$ (continued)


In the early 1990s the ACC/SCN commissioned a review of the existing scientific evidence to assess the effectiveness of vitamin A supplementation on child mortality and morbidity (Beaton et al. 1993). The findings from ten different community-based mortality trials conducted in Africa (Ghana and the Sudan), the Americas (Haiti) and Asia (India, Indonesia and Nepal), and 20 different morbidity trials conducted around the world were considered for inclusion in a formal metaanalysis. Sufficient data were available from five of the mortality trials (Daulaire et al. 1992; Fawzi et al. 1993; Ghana VAST Study Team 1993; Rahmathullah et al. 1990; West et al. 1991) to assess the relationship between vitamin A interventions and cause-specific mortality from measles, diarrhoea and respiratory illness. Separate relative risks were reported in the meta-analysis of these studies for all-cause mortality, mortality attributed to these three specific causes and mortality attributed to all other causes combined.

All five of those mortality trials were conducted in populations where vitamin A deficiency (both clinical and subclinical) was prevalent (assessed either as high rates of xerophthalmia or low serum retinol concentrations). In four of the trials (one in Ghana, two in Nepal and one in the Sudan), the intervention consisted of vitamin A supplementation in age appropriate doses ( 200000 IU for children $\geq 12$ months of age, 100000 IU for children 6 to 11 months of age and, for one trial in Nepal, 100000 IU for infants 1 to 11 months of age and 50000 IU for infants $<1$ month of age) every 4-6 months. In India, the intervention was a weekly supplement that contained a weekly RDA of vitamin A. In the Ghana, India and Nepal trials, a statistically significant reduction in all-cause mortality was observed among children in the vitamin A intervention group. No statistically significant effect on all-cause mortality was observed in the Sudan.

## MEASLES

Measles (and thus measles-related mortality) have nearly been eliminated in the Americas due to the implementation of successful measles control programmes over the past ten years (Hersh et al. 2000). At present, the highest rates of measles mortality occur among children aged $<1$ year in developing countries. Although unvaccinated children aged $>10$ years and adults are also susceptible to measles and measles mortality, estimates for the burden of disease associated with vitamin A deficiency were restricted to children aged 0-4 years for the CRA analysis.

## Incidence and morbidity

Measles is a highly infectious disease caused by the measles virus (Reingold and Phares 2001). Although vitamin A is involved in different aspects of the immune defence system, there is little evidence to suggest that vitamin A status affects the incidence of measles. Therefore,
the relative risk of experiencing a measles episode associated with vitamin A deficiency was not calculated for the CRA analysis.

## Mortality

Vitamin A deficiency appears to increase the risk of measles-related mortality. The ACC/SCN meta-analysis of the four prospective communitybased trials with information on measles (Ghana, India and two in Nepal) suggests that vitamin A supplementation was associated with a $26 \%$ overall reduction in measles mortality ( $\mathrm{RR}=74,95 \% \mathrm{CI}$ 0.53-1.04) (Beaton et al. 1993). In the individual trials, the reduction in measles mortality ranged from $18 \%$ in Ghana to $75 \%$ in Sarlahi, Nepal (Sommer and West 1996). A more recent study from Ghana found no difference in the acute measles case-fatality rate between vitamin A supplemented and placebo groups (Dollimore et al. 1997). Vitamin A supplementation, however, has been found to reduce measles severity. In some (Coutsoudis et al. 1991; Ellison 1932; Hussey and Klein 1990) but not all (Rosales et al. 1996) hospital-based treatment trials, the surviving children who had received vitamin A experienced complications less frequently and recovered faster than their counterparts who received placebo treatment. These findings provide supporting evidence for a causal relationship between vitamin A deficiency and measles mortality.

The relative risk of measles (GBD code U015) mortality among children aged 0-4 years with vitamin A deficiency was estimated for the CRA project from the protective effect of vitamin A interventions (RR $=0.74,95 \%$ CI $0.53-1.04$ ) observed in the ACC/SCN meta-analysis of community-based trials and a $41 \%$ estimated baseline prevalence rate of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / 1$ among those study populations. The adjusted relative risk estimate derived from those data is $R R=1.86$ ( $95 \%$ CI 1.32-2.59). The calculations are shown in Table 4.8.

## Diarrhoea

Diarrhoea and diarrhoea-related mortality are prevalent among preschool-age children, particularly in low income countries. The global burden of disease related to diarrhoea mortality is widely distributed around the world.

## Incidence and morbidity

Diarrhoeal disease can be caused by a variety of bacterial, viral and parasitic agents (Reingold and Phares 2001). Although vitamin A is involved in different aspects of the immune defence system, including the maintenance of the epithelial cell border in the intestinal tract, there is little evidence from intervention trials to suggest that vitamin A status affects the incidence of diarrhoea. Therefore, the relative risk of experiencing a diarrhoea episode associated with vitamin A deficiency was not calculated. However, there is considerable evidence to link vitamin A status to the severity of diarrhoea episodes (Sommer and West 1996). Children
Unadjusted and adjusted relative risks of adverse health outcomes associated with vitamin A deficiency ${ }^{\text {a }}$

| Outcome | Source of relative risk data | Original trial results ( $R R$ of protective effect due to vitamin A intervention) |  | RR associated with vitamin A deficiency ${ }^{b}$ |  | RR associated with vitamin A <br> deficiency (serum retinol concentrations $<0.70 \mu \mathrm{~mol} / /)^{c}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | $R \mathrm{R}=[1.00 /$ published | $95 \% \mathrm{Cl}=$ | $R R=[1.00 /$ published | 95\% Cl |
|  |  | RR | 95\% CI | protective RR] | [1.001published CI] | protective RR] | [1.00/published CI] |
| Children (0-4 years) |  |  |  |  |  |  |  |
| Diarrhoea mortality | Table 5.10 in Beaton et al. (1993) | 0.68 | 0.57-0.80 | 1.47 | 1.25-1.75 | 2.15 | 1.83-2.58 |
| Measles mortality | Table 5.10 in Beaton et al. (1993) | 0.74 | 0.53-1.04 | 1.35 | 0.96-1.89 | 1.86 | 1.32-2.59 |
| Malaria incidence | Table 3 in Shankar et al. (1999) | 0.70 | 0.57-0.87 | 1.43 | 1.15-1.75 | 1.78 | 1.43-2.19 |
| Malaria mortality | Table 3 in Shankar et al. (1999) using incidence of malaria episodes as a proxy for malaria mortality | 0.70 | 0.57-0.87 | 1.43 | 1.15-1.75 | 1.78 | 1.43-2.19 |
| Selected other infectious disease causes of mortality | Table 5.10 in Beaton et al. (1993) | 0.95 | 0.81-1.06 | 1.05 | 0.94-1.23 | 1.13 | 1.01-1.32 |
| Pregnant women (15-44 years) |  |  |  |  |  |  |  |
| All-cause maternal mortality | Text statement about the relative risk for the maternal mortality ratio among vitamin A + beta-carotene vs placebo recipients in West et al. (1999) | 0.60 | 0.39-0.93 | 1.67 | 1.08-2.56 | 4.51 | 2.91-6.94 |

a The outcomes were defined using the following GBD codes: measles (UO15); diarrhoea (UOIO); malaria (U020); other infectious causes (U037); all-cause maternal mortality corresponds to the GBD code for maternal conditions (U042).
b Unadjusted for baseline prevalence of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$.
c Adjusted for baseline prevalence of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ in trial populations.
with poor vitamin A status tend to suffer from more severe and more prolonged episodes of diarrhoea. These findings provide supporting evidence for a causal relationship between vitamin A deficiency and diarrhoea mortality.

## Mortality

Vitamin A deficiency does appear to increase the risk of diarrhoea mortality. The ACC/SCN meta-analysis of the five community-based trials with information on diarrhoea (Ghana, India, Nepal [two trials] and the Sudan) suggest that vitamin A supplementation was associated with a $32 \%$ reduction in diarrhoea mortality ( $\mathrm{RR}=0.68,95 \%$ CI $0.57-0.80$ ) (Beaton et al. 1993). In the individual trials that showed a reduction (all but the Sudan), the reduction in diarrhoea mortality ranged from $34 \%$ in Ghana to $52 \%$ in India (Sommer and West 1996).

The relative risk of diarrhoea (GBD code U010) mortality among children aged $0-4$ years with vitamin A deficiency was estimated from the protective effect of vitamin A interventions ( $\mathrm{RR}=0.68,95 \% \mathrm{CI}$ $0.57-0.80$ ) observed in the ACC/SCN meta-analysis of community-based trials and a $41 \%$ estimated baseline prevalence rate of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / 1 \mathrm{l}$ among those study populations. The adjusted relative risk estimate derived from those data is $\mathrm{RR}=2.15$ ( $95 \%$ CI 1.83-2.56). The calculations are shown in Table 4.8.

## Malaria

Although malaria is endemic in several areas of the world, countries in Africa have particularly high prevalence rates and children aged $<5$ years experience much of the associated burden of disease.

## Incidence and morbidity

Malaria is caused by several species of the protozoa Plasmodium, which are transmitted by mosquitoes from one person to another (Reingold and Phares 2001). In humans, P. falciparum is the species responsible for the majority of malaria-related morbidity and mortality. Both animal and human studies have implicated specific immune mechanisms for how vitamin A could influence the incidence and severity of malarial disease (Davis et al. 1998; Krishnan et al. 1976; Serghides and Kain 2002; Stoltzfus et al. 1989). In addition, cross-sectional studies have also documented inverse associations between plasma retinol levels and P. falciparum parasitaemia in humans (Filteau et al. 1993; Friis et al. 1997; Galan et al. 1990; Samba et al. 1992; Tabone et al. 1992; Thurnham and Singkamani 1991), and one study has reported that low baseline vitamin A status was associated with an increased risk of parasitaemia (Sturchler et al. 1987).

Findings from more recent randomized controlled intervention trials suggest that vitamin A supplementation appears to reduce both the incidence and severity of malaria illness. In Ghana, cohorts of children
received high-dose vitamin A or a placebo every four months as part of either the Child Health (morbidity) or Child Survival (mortality) Study. Blood samples were collected on a monthly basis from different random subsamples of children in a series of cross-sectional surveys conducted over a one-year period. Based on the cross-sectional data, the authors reported no statistically significant differences in the incidence of fever, malaria parasitaemia, parasite density, or probable malaria illness between the two groups and concluded that vitamin A supplementation had no impact on malaria (Binka et al. 1995). However, given the relatively small sample sizes, these studies had limited statistical power and could only detect a statistically significant protective effect of vitamin A supplementation against probable malarial illness that exceeded $70 \%$ (Child Survival Study) or 95\% (Child Health Study) (Shankar 1995).

A more definitive study of vitamin A supplementation and malaria was recently conducted in Papua New Guinea. In that study children aged 6-60 months received either high-dose vitamin A or a placebo every three months and were followed over a 13 -month period. Malaria morbidity was assessed weekly using community-based case detection and surveillance of the patients who reported to the local health centre. Blood smears were prepared for children who had a fever $>37.5^{\circ} \mathrm{C}$ at the time of the weekly home visit or for any clinic visit and for all children who participated in the three cross-sectional surveys conducted at baseline, at the midpoint, and at the end of the study. Slide-confirmed cases of malaria were followed prospectively and observed for adverse health outcomes. Children in the vitamin A group had a 30\% lower frequency of $P$. falciparum febrile illnesses $(R R=0.70,95 \%$ CI $0.57-0.87$ ) and younger children (12-36 months) apparently benefited most from supplementation. The younger children in the vitamin A group also had less severe malaria morbidity, measured as fewer febrile episodes, fewer enlarged spleens and a lower parasite density (Shankar et al. 1999).

The study from Papua New Guinea used a stronger design, evaluated a greater number of disease variables, and had more statistical power to investigate a relationship between vitamin A and malaria morbidity as compared to the Ghana study. Therefore, the findings from Papua New Guinea were used to assess the relative risk of malaria incidence associated with vitamin A deficiency for the CRA project.

## Mortality

To date, no randomized trials have demonstrated a significant reduction in malaria mortality associated with the vitamin A supplementation. However, given the recently observed relationship between vitamin A supplementation and malaria severity, an equivalent relative risk was assigned for malaria mortality for the purpose of the CRA project.

The relative risk of malaria incidence (GBD code U020) among children aged $0-4$ years with vitamin A deficiency was estimated for the CRA project from the protective effect of the vitamin A intervention
( $\mathrm{RR}=0.70,95 \%$ CI $0.57-0.87$ ) observed in the Papua New Guinea trial and a $55 \%$ estimated baseline prevalence rate of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ in that study population. The adjusted relative risk estimate derived from those data is $\mathrm{RR}=1.78$ ( $95 \%$ CI 1.43-2.19). The calculations are shown in Table 4.8.

The relative risk of malaria mortality (GBD code U020) among children aged $0-4$ years with vitamin A deficiency was estimated by extrapolation for the CRA project from the results of the vitamin A intervention on malaria incidence observed in the community-based trial in Papua New Guinea (RR $=0.70,95 \%$ CI $0.57-0.87$ ) and a $55 \%$ estimated baseline prevalence rate of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / / 1$ in that study population. The adjusted relative risk estimate derived from those data is $R R=1.78$ ( $95 \%$ CI 1.43-2.19). The calculations are shown in Table 4.8.

## Other infectious diseases in children

Infectious diseases are a common cause of death among preschool-age children worldwide. The global burden of disease related to the GBD category of "other infectious diseases" among children is widely distributed around the world.

Although acute respiratory infections, diarrhoeal diseases, measles and malaria represent the most common causes of childhood illness and death, many other bacterial, viral and parasitic agents cause disease among children as well. Since vitamin A is involved in a wide variety of different biological processes, vitamin A deficiency may also increase the risk of death due to other less common causes of childhood disease, although studies to date have not specifically quantified those relationships. The GBD study included a group of "other infectious diseases" (GBD code U037) in children that captured the contribution of many low-incidence causes of death under one category. Causes of infectious diseases not explicitly listed in the other GBD categories for children are included under this code (Murray et al. 2001).

## Mortality

Data from intervention trials among children suggest that vitamin A deficiency increases the risk of mortality from causes other than diarrhoea, measles and malaria. The ACC/SCN meta-analysis of the five community-based trials (Ghana, India, Nepal [two trials] and the Sudan) with information on cause-specific mortality found that vitamin A supplementation was associated with a $5 \%$ reduction in mortality from unspecified causes of death ( $\mathrm{RR}=0.95,95 \% \mathrm{CI} 0.81-1.06$ ) (Beaton et al. 1993). This risk estimate was used for the purposes of estimation. The relative risk of mortality due to other infectious diseases (GBD code U037) among children aged 0-4 years with vitamin A deficiency was estimated for the CRA project from the protective effect of vitamin A interventions $(\mathrm{RR}=0.95,95 \% \mathrm{CI} 0.81-1.06)$ observed in the ACC/SCN
meta-analysis of community-based trials and a $41 \%$ estimated baseline prevalence rate of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ among those study populations. The adjusted relative risk estimate derived from those data is $\mathrm{RR}=1.13$ ( $95 \%$ CI 1.01-1.32). The calculations are shown in Table 4.8.

### 5.2 RISK ESTIMATES FOR PREGNANT WOMEN (15-44 YEARS)

Over half a million women around the world die each year from causes associated with pregnancy and childbirth. Most of these maternal deaths occur in the Regions of Africa and South-East Asia.

Although data from either observational or experimental studies relating vitamin A deficiency among women of reproductive age to specific causes of mortality are limited, biologically plausible mechanisms for such relationships do exist (Faisel and Pittrof 2000). A relative risk of maternal mortality was estimated for vitamin A deficiency among pregnant women aged 15-44 years, based on the best data currently available. The risk estimate for these outcomes may be modified in the future when more data that specifically address this issue are available.

To date, the only large-scale randomized controlled trial of vitamin A supplementation among women of reproductive age has been conducted in an area of southern Nepal where high rates of maternal night blindness have been reported (West et al. 1999). More than 44000 women of reproductive age were randomized to continually receive either vitamin $\mathrm{A}(7000 \mu \mathrm{~g}$ retinol equivalents), an equivalent amount of betacarotene, or a placebo capsule on a weekly basis over an approximate three and a half year period; that is, before and during pregnancy, throughout the postpartum period and through any subsequent pregnancy until close-out of the trial. Over 22000 identified pregnancies were followed. Pregnancy-related mortality (defined as a death during pregnancy or prior to 12 weeks postpartum) was decreased by $40 \%$ $(R R=0.60,95 \%$ CI $0.37-0.97)$ in the vitamin A group and by $49 \%$ ( $\mathrm{RR}=0.51,95 \%$ CI 0.30-0.86) in the beta-carotene group (West et al. 1999). These findings provide support for a causal relationship between vitamin A deficiency and pregnancy-related mortality.

Data from the study in Nepal were used to estimate the relative risk of maternal mortality associated with vitamin A deficiency. The study presents several sets of mortality results for women stratified by time and cause of death during pregnancy and the postpartum period. The mortality results used for CRA analysis were those that excluded all deaths $>6$ weeks postpartum and all deaths attributed to reported injury which are the results that most closely conform to the ICD-10 definition of a maternal death (WHO 1992). The combined results for women in the vitamin A and beta-carotene intervention groups were used to represent non-vitamin A deficient women.

The relative risk of maternal mortality among pregnant women with vitamin A deficiency-due to a wide variety of maternal conditions (GBD
code U042)—was estimated for the CRA project from the combined protective effect of the vitamin A and beta-carotene interventions on the maternal mortality ratio ( $\mathrm{RR}=0.60,95 \%$ CI $0.39-0.93$ ) and a $19 \%$ estimated baseline prevalence rate of serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ observed among women in the Nepal trial (West et al. 1999). In the absence of other studies linking vitamin A deficiency to cause-specific deaths among women, the risk of maternal death (from a wide variety of maternal conditions) was extrapolated from the Nepal study. A maternal mortality ratio of 378 deaths per 100000 live births among women in the vitamin A and beta-carotene groups combined (West et al. 1999) was used as the starting point for the adjusted relative risk calculations. The adjusted relative risk estimate derived from those data is $\mathrm{RR}=4.51$ ( $95 \%$ CI 2.91-6.94). The calculations are shown in Table 4.8.

### 5.3 Risk estimates for all other groups

Limited information is available about the global prevalence of vitamin A deficiency or its associated adverse health effects in children aged $\geq 5$ years, in men aged 15-44 years, or in men or women aged $\geq 45$ years. Therefore, quantitative estimates for the prevalence of vitamin A deficiency and the associated risk of adverse health outcomes were not developed for these groups.

## 6. BURDEN OF DISEASE ASSOCIATED WITH VITAMIN A Deficiency

Summary estimates for three burden of disease measurements are shown in this chapter. The attributable fraction of cause-specific mortality associated with vitamin A deficiency among children aged $0-4$ years and pregnant women aged 15-44 years are shown by subregion in Table 4.9. The estimated number of cause-specific deaths and disability-adjusted life years (DALYs) attributed to vitamin A deficiency is shown in Table 4.9 and Table 4.10, respectively.

The attributable fraction results show that worldwide, vitamin A deficiency among children is associated with an estimated $20 \%$ of measlesrelated mortality, $24 \%$ of diarrhoea mortality, $20 \%$ of malaria incidence and mortality, and $3 \%$ of mortality associated with other infectious causes of disease. By definition, the attributable fraction of disease associated with vitamin A deficiency itself is $100 \%$. The attributable fractions vary widely by cause and across the different subregions. In general, the attributable fractions (for disease conditions other than vitamin A deficiency) are highest in the African and South-East Asian Regions, where vitamin A deficiency is most prevalent and child mortality rates for the causes of disease assessed in this chapter are high. Worldwide, vitamin A deficiency (serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ ) among pregnant women is associated with over $20 \%$ of maternal mortality.
Table 4.9 Attributable fraction of cause-specific mortality (\%) due to vitamin A deficiency for measles, diarrhoea, malaria and

| Subregion | Diseases among children ( $0-4$ years) ${ }^{\text {a }}$ |  |  |  | Conditions leading to maternal mortality among women ( $15-44$ years) ${ }^{2}$ <br> Maternal causes | Direct sequalae of vitamin A deficiency among all age groups and both sexes |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Measles | Diarrhoea | Malaria | Other infectious causes |  |  |
| AFR-D | 20 | 25 | 18 | 4 | 24 | 100 |
| AFR-E | 23 | 29 | 22 | 4 | 28 | 100 |
| AMR-A | b | b | b | b | b | b |
| AMR-B | b | 15 | 11 | 2 | 10 | b |
| AMR-D | b | 13 | 9 | 2 | 14 | b |
| EMR-B | 2 | 3 | 2 | b | b | b |
| EMR-D | 17 | 21 | 15 | 3 | 22 | 100 |
| EUR-A | b | b | b | b | b | b |
| EUR-B | b | b | b | b | b | b |
| EUR-C | b | b | b | b | b | b |
| SEAR-B | 29 | 35 | 27 | 6 | 24 | b |
| SEAR-D | 21 | 26 | 19 | 4 | 17 | 100 |
| WPR-A | b | b | b | b | b | b |
| WPR-B | 10 | 14 | 10 | 2 | 14 | b |
| World | 20 | 24 | 20 | 3 | 22, $21{ }^{\text {c }}$ | 100 |

a The outcomes were defined using the following GBD codes: measles ( U 015 ); diarrhoea ( U 010 ); malaria ( U 020 ); other infectious causes (U037); all-cause maternal mortality corresponds to the GBD code for maternal conditions (U042). vitamin A deficiency, by subregion other infectious diseases among children, for conditions leading to maternal mortality, and for direct sequelae of vitan A din A value for the attributable fraction is not includ
A value for the attributable fraction is not included either because the prevalence of vitamin A deficiency in the subregion was zero or the cause-specific mortality was zero or nearly
zero in the associated subregion-age-sex groups. zero in the associated subregion-age-sex groups.
The global attributable fractions are $22 \%$ (for women aged $15-29$ years) and $21 \%$ (for women aged $30-44$ years). They are slightly different because the total number of deaths among women in these age groups differ.
Table 4.I0 Deaths (000s) from measles, malaria, diarrhoea, other infectious diseases among children, from maternal conditions
attributable to vitamin A deficiency, and those directly attributed to vitamin A deficiency in vital registration systems,
The outcomes were defined using the following GBD codes: measles (UOI5); diarrhoea (UOIO); malaria (U020); other infectious causes (U037); all-cause maternal mortality corresponds to the GBD code for maternal conditions (U042).
b Among both sexes and all age groups.
Data for the measles, diarrhoea, malaria, other infectious causes, all-cause maternal mortality directly attributed to vitamin A deficiency, and the totals columns were obtained directly from annex results tables. Subregional subtotals for children aged $0-4$ years were calculated by adding across disease categories and then totalling those results. Regional subtotals for maternal causes of death were calculated by adding across the 15-29 and 30-44 year-old age categories and then totalling those results. Any discrepancies in the marginal totals in this table are attributed to differences in rounding carried over from the individual columns of the background data tables.

Again, the highest attributable fractions are found in the African and South-East Asian Regions, where vitamin A deficiency and maternal mortality rates are highest.

Worldwide the majority of deaths (see Table 4.10) and DALYs (see Table 4.11) associated with vitamin A deficiency occur in Africa and South-East Asia. However, the relative distribution of deaths and DALYs varies somewhat across subregions for the different causes of disease burden. A higher proportion of the burden of disease from measles, malaria and all-cause maternal mortality occurs in Africa, while SouthEast Asia has a higher proportion of the disease burden from diarrhoea.

The total number of maternal and child deaths and disease-related morbidity that could potentially be averted if vitamin A deficiency were eliminated depends on several factors: (i) the attributable fraction (determined by the prevalence of vitamin A deficiency and the relative risk of an adverse health outcome); (ii) the global distribution of the adverse health outcomes; and (iii) the estimated number of affected individuals in each subregion. Subregions with a higher attributable fraction but a smaller number of individuals affected by the adverse health outcomes (due to either a small population base or a low prevalence of the condition) may contribute less in absolute terms to the overall global burden of disease when compared to subregions or specific causes of death with lower attributable fractions but a larger total number of affected individuals. Conversely, subregions with similar attributable fractions but different causes of death and/or population bases will contribute different numbers of affected individuals to a global summary.

For example, the AFR-D and SEAR-D subregions have similar attributable fractions of disease for measles ( $\sim 20 \%$ ), but the number of deaths attributed to vitamin A deficiency is nearly twice as high in AFRD (41000) when compared to SEAR-D (23000) although the population base of children in the AFR-D subregion is only a third (13.5 million) of the population base of children ( 42.3 million) in the SEARD subregion. For diarrhoea, AFR-D and SEAR-D also have similar attributable fractions of disease ( $\sim 25 \%$ ), but in this case the number of deaths attributed to vitamin A deficiency is more than twice as high in SEAR-D (107000) when compared to AFR-D (46000). The relative contribution of these different factors (attributable fractions, global distribution of disease and population base) should be kept in mind when comparing the burden of disease attributed to vitamin A deficiency across subregions and specific causes of death.

## 7. Discussion

This chapter generated quantitative estimates of the global burden of disease due to vitamin A deficiency by applying the analytical framework of the CRA project to the best sources of currently available data. The results show, despite decades of achievement in paediatric detection and
Table 4.I I DALYs (000s) associated with measles, malaria, diarrhoea, and other infectious diseases among children, with
maternal conditions attributable to vitamin A deficiency and with direct sequelae of vitamin A deficiency among all age groups, by subregion

| Subregion | Children (0-4 years) ${ }^{\text {a }}$ |  |  |  |  | Pregnant women (15-44 years) ${ }^{\text {a }}$ <br> Maternal causes of morbidity and mortality | Both sexes (all age groups) Sequelae directly associated with vitamin A deficiency | Total |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Measles | Diarrhoea | Malaria | Other infectious causes | Total |  |  |  |
| AFR-D | 1432 | 1554 | 2914 | 95 | 5995 | 662 | 378 | 7034 |
| AFR-E | 1176 | 3131 | 3345 | 189 | 7841 | 1095 | 438 | 9375 |
| AMR-A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AMR-B | 0 | 132 | 2 | 10 | 144 | 38 | 0 | 182 |
| AMR-D | 0 | 84 | 0 | 3 | 87 | 33 | 0 | 121 |
| EMR-B | 0 | 15 | 0 | 2 | 17 | 0 | 0 | 17 |
| EMR-D | 389 | 1746 | 253 | 88 | 2476 | 401 | 40 | 2917 |
| EUR-A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EUR-B | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 2 |
| EUR-C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SEAR-B | 233 | 314 | 36 | 15 | 598 | 152 | 4 | 754 |
| SEAR-D | 790 | 3603 | 368 | 191 | 4952 | 637 | 100 | 5690 |
| WPR-A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| WPR-B | 81 | 319 | 11 | 37 | 448 | 89 | 11 | 546 |
| World | 4101 | 10898 | 6931 | 630 | 22560 | 3106 | 972 | $26638{ }^{\text {b }}$ |

The outcomes were defined using the following GBD codes: measles (UOI5); diarrhoea (UOIO); malaria (U020); other infectious causes (U037); all-cause maternal mortality corresponds
to the GBD code for maternal conditions (U042).
$\quad\left(15-44\right.$ years) ${ }^{\text {a }}$
$\begin{gathered}\text { Maternal causes of } \\ \text { morbidity and mortality }\end{gathered}$
$3106 \quad 972 \quad 26638^{\circ}$
Data for the measles, diarrhoea, malaria, other infectious causes, maternal causes of morbidity and mortality, and sequelae directly attributed to vitamin A deficiency, and the totals columns were obtained directly from the annex results tables. Subregional subtotals for children aged $0-4$ years were calculated by adding across disease categories and then totalling those results. Subregional subtotals for maternal conditions were calculated by adding across the 15-29 and 30-44 year-old age categories and then totalling those results. Any discrepancies in the marginal totals in this table are attributed to differences in rounding carried over from the individual columns of the background data tables.
control, that vitamin A deficiency remains a major public health problem among preschool-age children throughout the world. Globally, $21 \%$, or over 127 million children aged $<5$ years are vitamin A deficient. Twenty per cent to $24 \%$ of early childhood deaths due to measles, diarrhoea and malaria are attributable to vitamin A deficiency, plus an additional 3\% of deaths due to other infectious causes, accounting for 647000 deaths of preschool-age children each year. Vitamin A deficiency appears to especially put children at risk of mortality due to diarrhoea, which accounts for $50 \%$ of all childhood deaths attributable to vitamin A deficiency.

New to the global view of undernutrition is a previously unrecognized, substantial burden of maternal vitamin A deficiency. Not surprisingly the burden of disease among pregnant women appears to parallel the geographic distribution seen in young children, with Asia and Africa bearing a disproportionate burden. This first attempt to define the problem suggests, conservatively, that $5.6 \%$ of all pregnant women, or approximately 7.3 million, are vitamin A deficient in a given year. Vitamin Adeficient women are at a 4.5 -fold higher risk of pregnancy-related mortality than non-deficient women. The estimates generated for this project suggest that more than $20 \%$ of all maternal deaths in the world may be attributable to vitamin A deficiency. The method used to generate this particular estimate (applying the relative risk of death associated with vitamin A deficiency to all causes of maternal death) may somewhat overestimate the burden of disease related to vitamin A deficiency. It also does not address the prevalence, burden, or potential health consequences of vitamin A deficiency among non-pregnant women of reproductive age, which remain largely unknown.

In addition to estimating the number of preventable deaths, the results from this project suggest that approximately 27 million DALYs are associated with vitamin A deficiency. Given that DALYs provide a composite measure of disease burden associated with both fatal and non-fatal health conditions and that childhood mortality related to infectious diseases is the primary health outcome known to be associated with vitamin A deficiency, it is not surprising that the vast majority of the estimated DALYs ( $\sim 23$ million) are related to infectious disease causes of death in children.

In summary, vitamin A deficiency affects vulnerable populations throughout critical stages of life, as revealed by these estimates of disease burden in young children and pregnant women. Successful efforts to control and reduce vitamin A deficiency have the potential to improve the health and well-being of women and children around the world and to reduce the global burden of disease associated with this nutritional risk factor.

## 8. Expected changes in the prevalence of vitamin A deficiency

Based on the trends observed over the past 20 years, and more specifically on changes over the past $5-10$ years, there is reason for optimism and the expectation that the global prevalence of vitamin A deficiency (defined as serum retinol concentrations $<0.70 \mu \mathrm{~mol} / \mathrm{l}$ ) will decrease in the years leading up to 2030.

Numerous factors contribute to the current situation. In recent years a favourable global policy environment has been created and global partnerships have emerged to help guide activities aimed at the control and prevention of vitamin A deficiency. Various groups have contributed to the positive policy environment including IVACG, international organizations, bilateral agencies and individual country governments around the world.

Although these groups initially focused their attention almost exclusively on the more obvious problem of vitamin A deficiency among children, the situation is now changing. Policy-makers and programme managers are increasingly expanding their efforts in recognition of the fact that vitamin A deficiency is also prevalent among women of reproductive age and has potentially severe health consequences for them as well.

Many countries have already initiated national supplementation programmes for children (most commonly using community-based health services as a routine delivery channel or special vitamin A distribution initiatives combined with national immunization days or child health days or weeks) in an effort to prevent severe vitamin A deficiency among this age group. Although child-based supplementation programmes do not necessarily directly address the problem of vitamin A deficiency among women, other more general approaches may, if the accompanying health messages are modified to specifically encourage women as participants and programme beneficiaries. These include widespread food fortification initiatives, efforts to improve the availability of vitamin A rich foods through agricultural programmes, nutrition education programmes, etc.

Projections for the global prevalence of vitamin A deficiency for the 30 -year time period from 2000 to 2030 were developed for the CRA project based on the assumption that the policy and programming environment will remain positive and that prevention and control activities will not only keep pace, but will exceed future population growth.

The global prevalence of $21.1 \%$ for serum retinol concentrations $<0.70 \mu \mathrm{~mol} / 1$ (in preschool-age children) in the year 2000 was used as the baseline prevalence rate. Reductions of $10 \%, 20 \%$ and $30 \%$ were projected over the next three decades, resulting in estimated global prevalence rates of $19.0 \%, 16.9 \%$ and $14.8 \%$, in the years 2010,2020 and 2030, respectively. Although specific programmes to address vitamin A
deficiency in women have started much more recently, the same percentage reductions $(10 \%, 20 \%$ and $30 \%)$ in prevalence rates were assumed for the purposes of the CRA project. Starting with an assumed baseline rate of $5.6 \%$ in 2000 , prevalence rates of $5.0 \%, 4.5 \%$ and $3.9 \%$ were projected for 2010, 2020 and 2030, respectively.

## Acknowledgements

Supported by the Family Health and Child Survival (FHACS) Cooperative Agreement, between the United States Agency for International Development (USAID), Office of Health and Nutrition, and The Johns Hopkins Bloomberg School of Public Health, Department of International Health (JHU), and by the World Health Organization Global Programme on Evidence for Health Policy. The authors thank Maria Andersson and Bruno de Benoist for data used to update the vitamin A deficiency prevalence estimates, and Anuraj H. Shankar for helpful comments on malaria issues. The authors are grateful for the assistance of Steve Fishman, Jeffrey McGuckin and Jonathan Sugimoto in compiling data for the project, and for the valuable insight and advice from other colleagues in the field of international health.

## Note

1 See preface for an explanation of this term.

## References

Anonymous (1995) Potential interventions for the prevention of childhood pneumonia in developing countries: a meta-analysis of data from field trials to assess the impact of vitamin A supplementation on pneumonia morbidity and mortality. The vitamin A and pneumonia working group. Bulletin of the World Health Organization, 73:609-619.
Anonymous (1998a) Nepal micronutrient status survey. Ministry of Health, Child Health Division, HMG/N, New ERA, Micronutrient Initiative, UNICEF Nepal. World Health Organization, Kathmandu.
Anonymous (1998b) Randomised trial to assess benefits and safety of vitamin A supplementation linked to immunisation in early infancy. WHO/CHD immunisation-linked vitamin A supplementation study group. The Lancet, 352(9136):1257-1263. Erratum in: The Lancet, 1999; 353:154.

Bauernfiend C (1986) Vitamin A deficiency and its control. Academic Press, Inc., Gainesville, FL.

Beaton GH, Martorell R, Aronson KJ et al. (1993) Effectiveness of vitamin A supplementation in the control of young child morbidity and mortality in developing countries. ACC/SCN State-of-the-art Series. (Nutrition Policy Discussion Paper No. 13.) United Nations Administrative Committee on Coordination/Sub-Committee on Nutrition, Geneva.

Binka FN, Ross DA, Morris SS et al. (1995) Vitamin A supplementation and childhood malaria in northern Ghana. American Journal of Clinical Nutrition, 61:853-859.

Bloem MW, Wedel M, Egger RJ et al. (1990) Mild vitamin A deficiency and risk of respiratory tract diseases and diarrhea in preschool and school children in northeastern Thailand. American Journal of Epidemiology, 131:332-339.
Christian P, Thorne-Lyman AL, West KP Jr. et al. (1998a) Working after the sun goes down: exploring how night blindness impairs women's work activities in rural Nepal. European Journal of Clinical Nutrition, 52:519-524.

Christian P, West KP Jr., Khatry SK et al. (1998b) Night blindness of pregnancy in rural Nepal-nutritional and health risks. International Journal of Epidemiology, 27:231-237.

Christian P, West KP Jr., Khatry SK et al. (2000) Vitamin A or beta-carotene supplementation reduces symptoms of illness in pregnant and lactating Nepali women. Journal of Nutrition, 130:2675-2682.
Christian P, West KP Jr., Khatry SK et al. (2001) Maternal night blindness increases risk of mortality in the first 6 months of life among infants in Nepal. Journal of Nutrition, 131:1510-1512.
Coutsoudis A, Broughton M, Coovadia HM (1991) Vitamin A supplementation reduces measles morbidity in young African children: a randomized, placebocontrolled, double-blind trial. American Journal of Clinical Nutrition, 54: 890-895.

Daulaire NM, Starbuck ES, Houston RM, Church MS, Stukel TA, Pandey MR (1992) Childhood mortality after a high dose of vitamin A in a high risk population. British Medical Journal, 304:207-210.

Davis TM, Skinner-Adams TS, Beilby J (1998) In vitro growth inhibition of Plasmodium falciparum by retinol at concentrations present in normal human serum. Acta Tropica, 69:111-119.
Dollimore N, Cutts F, Binka FN, Ross DA, Morris SS, Smith PG (1997) Measles incidence, case fatality, and delayed mortality in children with or without vitamin A supplementation in rural Ghana. American Journal of Epidemiology, 146:646-654.
Ellison JB (1932) Intensive vitamin therapy in measles. British Medical Journal, 2:708-711.

Faisel H, Pittrof R (2000) Vitamin A and causes of maternal mortality: association and biological plausibility. Public Health Nutrition, 3:321-327.
Fawzi WW, Chalmers TC, Herrera MG, Mosteller F (1993) Vitamin A supplementation and child mortality. A meta-analysis. Journal of the American Medical Association, 269:898-903.
Filteau SM, Morris SS, Abbott RA et al. (1993) Influence of morbidity on serum retinol of children in a community-based study in northern Ghana. American Journal of Clinical Nutrition, 58:192-197.
Friis H, Mwaniki D, Omondi B et al. (1997) Serum retinol concentrations and Schistosoma mansoni, intestinal helminths, and malarial parasitemia: a cross-
sectional study in Kenyan preschool and primary school children. American Journal of Clinical Nutrition, 66:665-671.

Galan P, Samba C, Luzeau R, Amedee-Manesme O (1990) Vitamin A deficiency in pre-school age Congolese children during malarial attacks. Part 2: Impact of parasitic disease on vitamin A status. International Journal for Vitamin and Nutrition Research, 60:224-228.
Ghana VAST Study Team (1993) Vitamin A supplementation in northern Ghana: effects on clinic attendances, hospital admissions, and child mortality. The Lancet, 342:7-12. Erratum in: The Lancet, 1993;342:250.
Gilbert C, Foster A (2001) Childhood blindness in the context of VISION 2020-The right to sight. Bulletin of the World Health Organization, 79: 227-232.
Glasziou PP, Mackerras DEM (1993) Vitamin A supplementation in infectious diseases: a meta-analysis. British Medical Journal, 306:366-370.

Haskell MJ, Mazumder RN, Peerson JM et al. (1999) Use of the deuterated-retinol-dilution technique to assess total-body vitamin A stores of adult volunteers consuming different amounts of vitamin A. American Journal of Clinical Nutrition, 70:874-880.
Herrera MG, Nestel P, El Amin A, Fawzi Ww, Mohamed KA, Weld L (1992) Vitamin A supplementation and child survival. The Lancet, 340:267-271.
Hersh BS, Tambini G, Nogueira AC, Carrasco P, de Quadros CA (2000) Review of regional measles surveillance data in the Americas, 1996-99. The Lancet, 355:1943-1948.
Hill AB (1965) The environment and disease: association or causation? Proceedings of the Royal Society of Medicine, 58:295-300.
Humphrey JH, Agoestina T, Wu L et al. (1996) Impact of neonatal vitamin A supplementation on infant morbidity and mortality. Journal of Pediatrics, 128:489-496.

Hussey GD, Klein M (1990) A randomized, controlled trial of vitamin A in children with severe measles. New England Journal of Medicine, 323: 160-164.
IVACG (1997) Maternal night blindness: extent and associated risk factors. International Life Sciences Institute, Washington, DC.
Katz J, West KP Jr., Khatry SK et al. (2000) Maternal low-dose vitamin A or beta-carotene supplementation has no effect on fetal loss and early infant mortality: a randomized cluster trial in Nepal. American Journal of Clinical Nutrition, 71:1570-1576.
Kjolhede CL, Stallings RY, Dibley MJ, Sadjimin T, Dawiesah S, Padmawati S (1995) Serum retinol levels among preschool children in Central Java: demographic and socioeconomic determinants. International Journal of Epidemiology, 24:399-403.
Krishnan S, Krishnan AD, Mustafa AS, Talwar GP, Ramalingaswami V (1976) Effect of vitamin A and undernutrition on the susceptibility of rodents to a malarial parasite Plasmodium berghei. Journal of Nutrition, 106: 784-791.

Mason JB, Lotfi M, Dalmiya N, Sethuraman K, Dietchler M (2001) The micronutrient report: current progress and trends in the control of vitamin $A$, iodine, and iron deficiencies. The Micronutrient Initiative, Ottawa, ON. http://www.micronutrient.org.
Micronutrient Initiative/UNICEF/Tulane University (1998) Progress towards controlling vitamin A deficiency. The Micronutrient Initiative, Ottawa, ON.
Milton RC, Reddy V, Naidu AN (1987) Mild vitamin A deficiency and childhood morbidity-an Indian experience. American Journal of Clinical Nutrition, 46:827-829.
Muhilal, Permeisih D, Idjradinata YR, Muherdiyantiningsih, Karyadi D (1988) Vitamin A-fortified monosodium glutamate and health, growth, and survival of children: a controlled field trial. American Journal of Clinical Nutrition, 48:1271-1276.

Murray CJ, Lopez AD, eds. (1996a) Global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, Vol 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.
Murray CJ, Lopez AD, eds. (1996b) Global health statistics: a compendium of incidence, prevalence and mortality estimates for over 200 conditions. Global Burden of Disease and Injury, Vol 2. Harvard School of Public Health on behalf of WHO, Cambridge, MA.
Murray CJL, Lopez AD, Mathers CD, Stein C (2001) The global burden of disease 2000 project: aims, methods and data sources. (GPE Discussion Paper No. 36-revised). World Health Organization, Geneva.
Nalubola R, Nestel P (1999) The effect of vitamin A nutriture on health: a review. ILSI Press, Washington, DC.
National Research Council (1989) Recommended Dietary Allowances. U.S. Subcommittee on the tenth edition of the RDAs. National Academy Press, Washington, DC.
Olson JA (1996) Vitamin A. In: Present knowledge in nutrition. 7th edition. Ziegler EE, Filer LJJ, eds. ILSI Press, Washington, DC.
Rahmathullah L, Underwood BA, Thulasiraj RD et al. (1990) Reduced mortality among children in southern India receiving a small weekly dose of vitamin A. New England Journal of Medicine, 323:929-935.

Reingold AL, Phares CR (2001) Infectious diseases. In: International public health: diseases, programs, systems, and policies. Merson MH, Black RE, Mills AJ, eds. Aspen Publishers, Inc., Gaithersburg, MD.
Ribaya-Mercado JD, Mazariegos M, Tang G et al. (1999) Assessment of total body stores of vitamin A in Guatemalan elderly by the deuterated-retinol-dilution method. American Journal of Clinical Nutrition, 69:278-284.
Rosales FJ, Kjolhede C, Goodman S (1996) Efficacy of a single oral dose of 200,000 IU of oil-soluble vitamin A in measles-associated morbidity. American Journal of Epidemiology, 143:413-422.
Rothman KH, Greenland S (1998) Modern epidemiology. 2nd edn. LipponcottRaven Publishers, Philadelphia, PA.

Samba DC, Basco LK, Bleiberg-Daniel F, Lemmonier D, Le Bras J (1992) Absence of effect of retinol on the in vitro development of Plasmodium falciparum. International Journal for Vitamin and Nutrition Research, 62:99-100.

Semba RD, Bloem MW (2002) The anemia of vitamin A deficiency: epidemiology and pathogenesis. European Journal of Clinical Nutrition, 56:271-281.
Semba RD, Miotti PG, Chiphangwi JD et al. (1998) Maternal vitamin A deficiency and infant mortality in Malawi. Journal of Tropical Pediatrics, 44:232-234.

Serghides L, Kain KC (2002) Mechanism of protection induced by vitamin A in falciparum malaria. The Lancet, 359:1404-1406.
Shankar AH (1995) Vitamin A and malaria. American Journal of Clinical Nutrition, 62:842-843.
Shankar AH (2001) Nutritional modulation of immunity and infection. In: Present knowledge in nutrition. 8th edn. Bowman BA, Russell RM, eds. International Life Sciences Institute Press, Washington, DC.

Shankar AH, Genton B, Semba RD et al. (1999) Effect of vitamin A supplementation on morbidity due to Plasmodium falciparum in young children in Papua New Guinea: a randomised trial. The Lancet, 354:203-209.

Sommer A (1982) Nutritional blindness: xerophthalmia and keratomalacia. Oxford University Press, New York.
Sommer A, Davidson FR (2002) Assessment and control of vitamin A deficiency: the Annecy accords. Journal of Nutrition, 132:S2845-2850.
Sommer A, Tarwotjo I, Djunaedi E et al. (1986) Impact of vitamin A supplementation on childhood mortality. A randomised controlled community trial. The Lancet, 1:1169-1173.

Sommer A, Tarwotjo I, Katz J (1987) Increased risk of xerophthalmia following diarrhea and respiratory disease. American Journal of Clinical Nutrition, 45:77-80.

Sommer A, West KP Jr. (1996) Vitamin A deficiency: health, survival and vision. Oxford University Press, New York.

Sommer A, Zeger SL (1991) On estimating efficacy from clinical trials. Statistics in Medicine, 10:45-52.

Stephensen CB (2001) Vitamin A, infection, and immune function. Annual Review of Nutrition, 21:167-192.
Stoltzfus RJ, Jalal F, Harvey PW, Nesheim MC (1989) Interactions between vitamin A deficiency and Plasmodium berghei infection in the rat. Journal of Nutrition, 119:2030-2037.
Sturchler D, Tanner M, Hanck A et al. (1987) A longitudinal study on relations of retinol with parasitic infections and the immune response in children of Kikwawila village, Tanzania. Acta Tropica, 44:213-227.

Tabone MD, Muanza K, Lyagoubi M et al. (1992) The role of interleukin-6 in vitamin A deficiency during Plasmodium falciparum malaria and possible consequences for vitamin A supplementation. Immunology, 75:553-554.

Thurnham DI, Singkamani R (1991) The acute phase response and vitamin A status in malaria. Transactions of the Royal Society of Tropical Medicine and Hygiene, 85:194-199.

Tielsch JM, Rahmathullah L, Katz J, Thulasiraj RD, Coles CL, Devi S (2001) Impact of vitamin A supplementation for newborns on early infant mortality: a community-based, randomized trial in South India. (Presented at the XX International Vitamin A Consultative Group Meeting 2001 in Hanoi, Viet Nam).

Tonascia JA (1993) Meta-analysis of published community trials: impact of vitamin A on mortality. (Proceedings of the Bellagio meeting on vitamin A deficiency and child mortality.) Helen Keller International, New York.

UNICEF (2001) The state of the world's children 2001. UNICEF, New York.
West KP Jr. (2002) Extent of vitamin A deficiency among preschool children and women of reproductive age. Journal of Nutrition, 132:S2857-2866. Erratum in: Journal of Nutrition, 2002;132:3432.

West KP Jr., Katz J, Khatry SK et al. (1999) Double blind, cluster randomised trial of low dose supplementation with vitamin A or beta carotene on mortality related to pregnancy in Nepal. British Medical Journal, 318:570-575. Erratum in British Medical Journal, 1999;318:1386.
West KP Jr., Katz J, Shrestha SR et al. (1995) Mortality of infants $<6$ months of age supplemented with vitamin A: a randomized, double-masked trial in Nepal. American Journal of Clinical Nutrition, 62:143-148.

West KP Jr., Pokhrel RP, Katz J et al. (1991) Efficacy of vitamin A in reducing preschool child mortality in Nepal. The Lancet, 338:67-71.

WHO (1992) International statistical classification of diseases and related health problems, 10th revision. World Health Organization, Geneva.

WHO (1995) Global prevalence of vitamin A deficiency. (MDIS Working Paper No. 2. WHO/NUT/95.3). World Health Organization, Geneva.
WHO (1996) Indicators for assessing vitamin A deficiency and their application in monitoring and evaluating intervention programmes. (WHO/NUT/96.10). World Health Organization, Geneva.
World Bank (1993) World development report 1993: investing in health. Oxford University Press, New York.

## Chapter 5

## Zinc Deficiency

Laura E. Caulfield and Robert E. Black

## Summary

Research conducted during the past 10-15 years suggests that zinc deficiency is widespread and affects the health and well-being of populations worldwide. The objective of this chapter is to quantify the regional and global magnitude, distribution and disease burden implications of zinc deficiency.

We conducted a systematic literature search to identify relevant studies investigating the role of zinc in human health. This search indicated that evidence on the burden of disease related to zinc deficiency would be limited to the results of randomized controlled trials (RCTs) conducted among paediatric populations in developing countries, and would provide information on whether zinc deficiency affects the incidence of diarrhoea, pneumonia and malaria illness among children aged 0-4 years. There are no global estimates of zinc deficiency in paediatric or other populations; however, the International Zinc Nutrition Consultative Group (IZiNCG) has developed a method for estimating the prevalence of inadequate zinc intakes based on the presence and bioavailability of zinc in each country's food supply.

A systematic review of relevant epidemiological research involved meta-analysis from 11 intervention trials. Results of our review indicate that zinc deficiency in children aged $<5$ years increases the risk of incidence for diarrhoeal disease by 1.28 ( $95 \%$ CI 1.10-1.50), pneumonia by 1.52 ( $95 \%$ CI $1.20-1.89$ ) and malaria by 1.56 ( $95 \%$ CI 1.29-1.89). We extended those results as best estimates of the risk of mortality from these causes as a result of zinc deficiency. Following the IZiNCG technique, the global prevalence of zinc deficiency was estimated at $31 \%$, ranging from $4-73 \%$ across subregions. ${ }^{1}$ Based on these estimates, zinc deficiency was estimated to cause 176000 diarrhoea deaths, 406000 pneumonia deaths and 207000 malaria deaths. The associated loss of disability-adjusted life years (DALYs) attributable to zinc deficiency amounts to more than 28 million. The burden of disease due to zinc
deficiency is borne most heavily by countries in Africa, the Eastern Mediterranean and South-East Asia. The burden was approximately equally shared between males and females.

The results of this review suggest that zinc deficiency contributes substantially to the morbidity and mortality of young children throughout the world. Available evidence from RCTs on the strength of the association between zinc deficiency and morbidity from diarrhoea, pneumonia and malaria, combined with the high estimated prevalences of inadequate zinc diets, translates into a significant burden of disease attributable to deficiency of this micronutrient. Evidence relating zinc deficiency to health outcomes in children aged $>5$ years and adult men and women was not available. Given the likely high prevalence of inadequate zinc intakes in these other groups, research is needed to determine the magnitude of the potential health effects.

## 1. Introduction

Zinc is a trace mineral essential to all forms of life because of its fundamental role in gene expression, cell development and replication (Hambridge 2000). Severe or clinical zinc deficiency was defined last century, as a condition characterized by short stature, hypogonadism, impaired immune function, skin disorders, cognitive dysfunction and anorexia (Prasad 1991). Although severe zinc deficiency is considered rare, mild-to-moderate zinc deficiency is likely prevalent throughout the world today (Sandstead 1991). Lack of consensus on indicators of zinc deficiency has hampered efforts to document prevalences of zinc deficiency. Despite this, RCTs of zinc supplementation in areas with habitual low zinc intakes have begun to demonstrate how low dietary intakes of zinc adversely affect child health (Brown et al. 1998a; Zinc Investigators' Collaborative Group 1999, 2000). For this reason, it is important to attempt to quantify the prevalence of zinc deficiency and its contribution to the global burden of disease. The goal of this chapter is to describe the methods used to quantify the magnitude, distribution and disease burden implications of zinc deficiency.

## 2. Nature and definition of the risk factor

Millions of people throughout the world may have inadequate levels of zinc in the diet due to limited access to zinc-rich foods (animal products, oysters and shellfish) and the abundance of zinc inhibitors, such as phytates, common in plant-based diets (Sandstead 1991). Our understanding of the public health importance of inadequate zinc intakes has been hampered by a lack of indicators of zinc status for identifying individuals with zinc deficiency (Wood 2000).

Estimating the prevalence of zinc deficiency is difficult because plasma or serum zinc concentrations, the most widely used indicators of zinc
deficiency at the population level, have not been assessed in national or regional surveys in developing countries. Even in developed countries, such indicators are not used for methodological and cost reasons.

Zinc deficiency is largely related to inadequate intake or absorption of zinc from the diet, although excess losses of zinc during diarrhoea may also contribute (Gibson 1994; WHO 1996). The distinction between intake and absorption is important, because although some intakes of zinc may be acceptable, the levels of inhibitors (e.g. fibre and phytates) in the diet may mean that inadequate amounts of zinc are absorbed. For this reason, zinc requirements for dietary intake are adjusted upward for populations in which animal products, the best sources of zinc, are limited, and in which plant sources of zinc are similarly high in phytates. Because zinc is not well conserved in the body and because zinc deficiency is directly related to dietary zinc intake, an indirect approach to quantify the prevalence of zinc deficiency would be to examine the adequacy of zinc in the diet in various regions throughout the world.

Dietary surveys are conducted in many countries, but few such surveys exist in developing countries (Gibson 1994; Parr 1992; WHO 1996). Even when dietary intake data are available, incomplete information on the content of zinc and its bioavailability in local foods has made calculation of zinc bioavailability problematic.

### 2.1 Prevalence of Zinc deficiency

Several alternative approaches have been taken to estimate the adequacy of the diet in various regions of the world. For example, the World Health Organization (WHO) (1996) used a factorial approach to estimate the average basal and normative zinc requirements (i.e. minimum amount of zinc to cover losses in individuals adapted or not adapted to low usual dietary intakes of zinc) for population subgroups, and then considered issues of zinc bioavailability from the typical diet in various regions of the world to derive estimates of the minimal dietary intake of zinc that would meet these requirements. They then compared the estimated dietary intakes of zinc from 210 dietary surveys conducted throughout the world as a percentage of the minimal dietary intake of zinc, and identified surveys in which the mean intake of zinc did not meet the minimal basal or normative level of intake. Overall, this method led to the conclusion that populations with inadequate intakes of zinc were likely widespread throughout the world, but concentrated in areas of the world consuming plant-based diets in which zinc was only of low to moderate bioavailability. For example, of 148 surveys conducted in populations with presumed highly bioavailable zinc (mostly western Europe and the United States of America), only one dietary survey indicated a mean intake lower than the minimal zinc intake to meet average normative requirements for zinc. In contrast, among 47 surveys conducted in populations with moderate zinc bioavailability, 40 indicated average
intakes lower than the minimal normative zinc intake, and of 15 surveys conducted in populations with low zinc bioavailability, none reported mean intakes greater than the minimal normative zinc intake. Based on this approach, it was concluded that zinc nutrition in developing countries should be a priority area for research, and emphasized the need for studies of dietary intakes and dietary constituents. Although carefully constructed, this analysis could not provide us with the required data on estimated regional prevalences of zinc deficiency.

In 2001, Brown et al. published an approach that used available data on per capita food availability from 172 countries to estimate the prevalence of inadequate zinc intakes worldwide. Subsequently IZiNCG (forthcoming) revised this approach, improving on the method for assessing the bioavailability of zinc in the food supply. We used this second approach because of the improved methodology and more conservative estimates of deficiency, and were provided with estimated prevalences of inadequate intakes of zinc in the diet in each of the 14 subregions ( K . Brown, personal communication). The technique is described briefly below.

Data are available on the per capita food availability in each country based on country-level information on food production, imports and exports, as reported on food balance sheets (FBS) by the Food and Agriculture Organization of the United Nations (FAO) on an annual basis. Per capita availability of energy was estimated from this information, and population estimates for each country in 1998 (FAO 1999). Per capita zinc availability was estimated based on the zinc:energy ratio, which was derived using FAO values for energy for each food and values of the World Food Program (World Food Dietary Assessment System 1997) for the estimated zinc content of each food. To examine the bioavailability of zinc in the diet, IZiNCG gathered published information from studies of zinc absorption from test meals, and conducted a pooled analysis to derive a prediction equation for the percentage absorbable zinc based on information on the zinc, phytate, calcium and total protein content of the diet. This equation and information on the levels of zinc, phytate, calcium and total protein available per capita per meal (assuming three meals per day) calculated from the FBS database were then applied to estimate the percentage absorbable zinc for each of the 172 countries with available FBS data. The proportion of absorbable zinc ranged from $11 \%$ to $22 \%$. Countries were then categorized as having low ( $<14 \%$ ), moderate ( $15-16 \%$ ) or high ( $>17 \%$ ) mean percentage absorbable zinc in the food supply, based on their main staple food (wheat, rice, maize and other cereals, and tubers) and the contribution of animal protein to the energy available in the food supply.

The average daily zinc requirement for each country was estimated by calculating the mean of the recommended zinc intakes for low, average and high bioavailability diets for the various sexes and ages (WHO 1996), weighted by the sex and age distribution of the country's popu-
lation, as obtained from the WISTAT database (WISTAT 1994). To calculate the estimated percentage of the population with inadequate zinc intake, they assumed that the mean intake of food was equal to mean availability of food, and that the standard deviation of intake is $25 \%$ of the mean-i.e. coefficient of variation (CV) $=0.25$ (WHO 1996). Under assumptions of normality, they calculated the proportion of individuals with intakes below the country-level daily zinc requirement, and thus at risk of zinc deficiency due to inadequate zinc intake (IZiNCG forthcoming). Again, this method is conservative when compared to the previously published method using bioavailability estimates, which considered only the phytate:zinc molar ratio in the food supply (Brown et al. 2001).

The method does not account for zinc intake from breast milk or from drinking water. However, breast milk is only an adequate source of zinc in the diets of infants aged less than six months, and zinc intakes from drinking water typically increase total zinc intakes by only $2 \%$ provided a usual water intake of $2 \mathrm{l} /$ day.

A secondary consideration for our calculations was the selection of the appropriate counterfactual (i.e. theoretical minimum exposure) for zinc deficiency. Unlike some other exposures, zero prevalence of zinc deficiency is in theory possible, and would likely result in complete eradication of the zinc disease burden due to this risk factor. Thus, for purposes of the analysis, we designated the theoretical minimum exposure to be as "no inadequate intakes of zinc" or zero prevalence of inadequate intakes of zinc.

Using the method described above, the estimated global prevalence of zinc deficiency is $31 \%$, and ranges from $4 \%$ to $73 \%$ (Table 5.1). The prevalences of zinc deficiency are low ( $4-7 \%$ ) in AMR-A, EUR-A, EURC and WPR-A. Intermediate prevalences of $9-26 \%$ are found in AMRB, EUR-B and WPR-B. High prevalences are found in AMR-D (68\%), throughout South and Central Africa (37-62\%), North Africa and the Eastern Mediterranean region (25-52\%), and South and South-East Asia (34-73\%).

## 3. HeAlth outcomes Considered

To proceed with a systematic investigation of the evidence regarding the role of zinc in human health, we disaggregated and then linked both specific health outcomes and target populations. The considered health outcomes included risk and/or severity of diarrhoea, pneumonia, malaria, measles, cognitive dysfunction, physical impairment, visual impairment or blindness and mortality. The target populations consisted of the following age-sex groups: children aged 0-4 years; children 5-14 years; women $15-44$ years; men $15-44$ years; all adults $\geq 45$ years. Despite this disaggregated approach to defining risk groups, upon review of the literature described below, we concluded that there were few pub-
Table 5.I Estimated prevalence of inadequate zinc intakes by subregion ${ }^{\text {a }}$

| Subregion | Number of countries | Population (millions) | Energy <br> (kcal/d) | $\begin{aligned} & \text { Zinc } \\ & (m g / d) \end{aligned}$ | Phytate:zinc ratio | Available zinc (mg) | Mean \% of adjusted requirement | \% inadequate intakes (95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | 25 | 278.7 | $\begin{aligned} & 2453 \\ & (327) \end{aligned}$ | $\begin{aligned} & 10.0 \\ & (2.1) \end{aligned}$ | $\begin{gathered} 25.8 \\ (2.6) \end{gathered}$ | $\begin{gathered} 1.14 \\ (0.20) \end{gathered}$ | $\begin{gathered} 57.3 \\ (10.8) \end{gathered}$ | $\begin{gathered} 36.5 \\ (26.4-46.6) \end{gathered}$ |
| AFR-E | 20 | 322.6 | $\begin{aligned} & 2075 \\ & (370) \end{aligned}$ | $\begin{gathered} 8.6 \\ (1.6) \end{gathered}$ | $\begin{gathered} 27.9 \\ (3.7) \end{gathered}$ | $\begin{gathered} 0.94 \\ (0.20) \end{gathered}$ | $\begin{gathered} 46.7 \\ (12.4) \end{gathered}$ | $\begin{gathered} 61.6 \\ (49.3-73.9) \end{gathered}$ |
| AMR-A | 3 | 315.7 | $\begin{aligned} & 3514 \\ & (234) \end{aligned}$ | $\begin{aligned} & 12.1 \\ & (1.0) \end{aligned}$ | $\begin{gathered} 12.6 \\ (0.8) \end{gathered}$ | $\begin{gathered} 2.85 \\ (0.37) \end{gathered}$ | $\begin{aligned} & 85.3 \\ & (7.4) \end{aligned}$ | $\begin{gathered} 6.3 \\ (0-16.1) \end{gathered}$ |
| AMR-B | 26 | 418.6 | $\begin{aligned} & 2828 \\ & (244) \end{aligned}$ | $\begin{aligned} & 10.4 \\ & (1.8) \end{aligned}$ | $\begin{gathered} 20.6 \\ (5.7) \end{gathered}$ | $\begin{gathered} 1.66 \\ (1.15) \end{gathered}$ | $\begin{gathered} 65.4 \\ (14.8) \end{gathered}$ | $\begin{gathered} 26.0 \\ (17.9-34.1) \end{gathered}$ |
| AMR-D | 6 | 68.5 | $\begin{aligned} & 2241 \\ & (223) \end{aligned}$ | $\begin{gathered} 7.5 \\ (1.0) \end{gathered}$ | $\begin{gathered} 24.7 \\ (6.6) \end{gathered}$ | $\begin{gathered} 0.91 \\ (0.15) \end{gathered}$ | $\begin{gathered} 44.5 \\ (5.6) \end{gathered}$ | $\begin{gathered} 68.4 \\ (54.7-82.1) \end{gathered}$ |
| EMR-B | 10 | 130.4 | $\begin{aligned} & 2946 \\ & (195) \end{aligned}$ | $\begin{gathered} 8.5 \\ (1.1) \end{gathered}$ | $\begin{gathered} 22.0 \\ (2.7) \end{gathered}$ | $\begin{gathered} 1.12 \\ (0.40) \end{gathered}$ | $\begin{aligned} & 60.9 \\ & (6.1) \end{aligned}$ | $\begin{gathered} 25.2 \\ (21.1-29.3) \end{gathered}$ |
| EMR-D | 9 | 339.7 | $\begin{aligned} & 2544 \\ & (460) \end{aligned}$ | $\begin{gathered} 7.8 \\ (1.9) \end{gathered}$ | $\begin{gathered} 25.0 \\ (3.4) \end{gathered}$ | $\begin{gathered} 0.91 \\ (0.19) \end{gathered}$ | $\begin{array}{r} 52.2 \\ (11.7) \end{array}$ | $\begin{gathered} 51.8 \\ (33.8-69.8) \end{gathered}$ |


| EUR-A | 22 | 409.4 | $\begin{gathered} 3378 \\ (154) \end{gathered}$ | $\begin{gathered} 12.6 \\ (1.1) \end{gathered}$ | $\begin{aligned} & 11.5 \\ & (1.8) \end{aligned}$ | $\begin{gathered} 3.53 \\ (0.93) \end{gathered}$ | $\begin{aligned} & 92.0 \\ & (8.4) \end{aligned}$ | $\begin{gathered} 3.9 \\ (2.8-5.0) \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| EUR-B | 16 | 213.6 | $\begin{gathered} 3073 \\ (433) \end{gathered}$ | $\begin{aligned} & 10.2 \\ & (1.5) \end{aligned}$ | $\begin{gathered} 18.4 \\ (4.6) \end{gathered}$ | $\begin{gathered} 1.63 \\ (0.62) \end{gathered}$ | $\begin{gathered} 73.2 \\ (10.1) \end{gathered}$ | $\begin{gathered} 12.7 \\ (8.9-16.5) \end{gathered}$ |
| EUR-C | 6 | 247.0 | $\begin{aligned} & 3020 \\ & (109) \end{aligned}$ | $\begin{gathered} 11.2 \\ (0.7) \end{gathered}$ | $\begin{gathered} 13.8 \\ (1.1) \end{gathered}$ | $\begin{gathered} 2.18 \\ (0.37) \end{gathered}$ | 83.5 <br> (4.7) | $\begin{gathered} 5.7 \\ (4.4-7.0) \end{gathered}$ |
| SEAR-B | 3 | 285.1 | $\begin{aligned} & 2616 \\ & (203) \end{aligned}$ | $\begin{gathered} 9.1 \\ (1.1) \end{gathered}$ | $\begin{gathered} 26.9 \\ (3.4) \end{gathered}$ | $\begin{gathered} 1.00 \\ (0.06) \end{gathered}$ | $\begin{gathered} 56.9 \\ (6.0) \end{gathered}$ | $\begin{gathered} 33.5 \\ (14.7-52.3) \end{gathered}$ |
| SEAR-D | 6 | 1198.0 | $\begin{aligned} & 2356 \\ & (127) \end{aligned}$ | $\begin{gathered} 7.9 \\ (0.7) \end{gathered}$ | $\begin{aligned} & 27.5 \\ & (0.4) \end{aligned}$ | $\begin{gathered} 0.85 \\ (0.07) \end{gathered}$ | 43.5 <br> (4.8) | $\begin{gathered} 72.5 \\ (62.3-82.7) \end{gathered}$ |
| WPR-A | 4 | 148.9 | $\begin{gathered} 2954 \\ (122) \end{gathered}$ | $\begin{aligned} & \text { II. } 8 \\ & (0.9) \end{aligned}$ | $\begin{gathered} 16.6 \\ (3.4) \end{gathered}$ | $\begin{gathered} 2.13 \\ (1.27) \end{gathered}$ | $\begin{gathered} 91.2 \\ (2.9) \end{gathered}$ | $\begin{gathered} 3.6 \\ (3.2-4.0) \end{gathered}$ |
| WPR-B | 13 | 1498.3 | $\begin{aligned} & 2705 \\ & (159) \end{aligned}$ | $\begin{gathered} 10.7 \\ 0.9) \end{gathered}$ | $\begin{gathered} 19.0 \\ (1.3) \end{gathered}$ | $\begin{gathered} 1.47 \\ (0.23) \end{gathered}$ | $\begin{gathered} 80.5 \\ (8.7) \end{gathered}$ | $\begin{gathered} 8.5 \\ (3.6-13.4) \end{gathered}$ |
| World | 172 | 5874.3 | $\begin{gathered} 2706 \\ (434) \end{gathered}$ | $\begin{gathered} 9.9 \\ (2.0) \end{gathered}$ | $\begin{gathered} 21.3 \\ (6.0) \end{gathered}$ | $\begin{gathered} 1.51 \\ (0.90) \end{gathered}$ | $\begin{gathered} 66.5 \\ (19.4) \end{gathered}$ | $\begin{gathered} 31.3 \\ (26.7-35.9) \end{gathered}$ |
| Source: | eans <br> comin | ions). |  |  |  |  |  |  |

lished studies regarding the contribution of zinc deficiency to morbidity or mortality in any age group other than children aged $0-4$ years. For this reason, the chapter focuses on the contribution of zinc deficiency to the global burden of disease among children aged $0-4$ years.

## 4. Risk factor-Disease relationship

### 4.1 Search strategy

Evidence for the relationship between zinc deficiency and various diseases was gathered through literature searches of published studies. Articles related to zinc deficiency in human populations were identified for review based upon a Medline database search of literature published 1966-2001 in English or with an English abstract. The search was conducted using combinations of the following keywords: zinc, deficiency, mortality, death, morbidity, acute respiratory infection, pneumonia, diarrhoea, measles, malaria, child, pregnancy, infant, neonatal, fetal, premature, congenital, abortion, stillbirth, miscarriage, birth weight, retardation, development, intelligence, cognitive, psychomotor and neurological. Further searches on Medline were conducted using author names and the "related articles" option.

Abstracts were reviewed to select English-language articles that examined the effect of zinc deficiency (as defined by clinical, epidemiological or biochemical assessment) or supplementation/fortification on several health outcomes of interest in human populations. These health outcomes included infant or neonatal, child or adult mortality; incidence or severity of malaria, diarrhoea, measles or acute respiratory infection; prevalence of mental retardation or cognitive dysfunction; and risk prevalence of physical impairment such as hearing or neurological dysfunction. Once copies of articles were obtained, additional publications were identified from the reference lists of those articles. The only studies excluded from review were animal studies and case reports.

### 4.2 Methods for combining risk estimates from individual studies

The most appropriate study design to assess prospective risk for our purposes would be a prospective cohort study. Unfortunately, no such studies were identified despite an exhaustive search. Certainly this relates in part to the fundamental problem of characterizing the zinc status of individuals discussed above. There have been, however, a significant number of RCTs of zinc supplementation examining morbidity outcomes in young children living in developing countries with presumably low intakes of zinc. The studies that were identified and included in our calculations are summarized in Table 5.2.

The studies included were RCTs of zinc supplementation for the prevention of one of three health outcomes: diarrhoea, pneumonia or
Studies contributing data on the effect of zinc deficiency on morbidity among children, aged 0-4 years

| Country (reference) | Study design | Supplement | $\begin{gathered} \text { Age } \\ \text { (months) } \end{gathered}$ | Sample size |  | Enrolment criteria | Outcomes measured |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | Zinc group | Control group |  | Diarrhoea | Pneumonia | Malaria |
| Burkina Faso (Müller et al. 2001) | Continuous supplementation (6 days/ week) over a 6-month period. Malaria defined based on community-based active case detection | 12.5 mg zinc as sulfate | 6-31 | 356 | 353 | Community-based (included for diarrhoea only) | $\checkmark$ | - | $\checkmark$ |
| Ethiopia (Umeta et al. 2000) | Continuous supplementation (6 days/ week) for 6 months | 10 mg zinc as sulfate | 6-12 | 92 | 92 | Stratified based on length-for-age $<-2 S D$ | $\checkmark$ | - | - |
| Gambia (Bates et al. 1993) | Twice weekly supplementation over a I.25-year period. Supplement provided in a fruit-flavoured drink; malaria defined as clinic visit with confirmation by microscopic evaluation | 70 mg zinc as zinc acetate | 6-28 | 55 | 54 | Matched on age and sex | - | - | $\checkmark$ |
| Guatemala (Ruel et al. 1997) | Daily supplementation for 28 weeks | 10 mg zinc as sulfate | 6-9 | 45 | 44 | Community-based | $\checkmark$ | - | - |
| India (Bhandari et al. 2002) | Daily supplementation for 16 weeks | 10 mg zinc to infants; 20 mg zinc to older children; as gluconate | 6-35 | 1241 | 1241 | Community-based | - | $\checkmark$ | - |

Table 5.2

| Country (reference) | Study design | Supplement | $\begin{gathered} \text { Age } \\ \text { (months) } \end{gathered}$ | Sample size |  | Enrolment criteria | Outcomes measured |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | Zinc group | Control group |  | Diarrhoea | Pneumonia | Malaria |
| India (Sazawal et al. 1997, forthcoming) | Daily supplementation for 26 weeks | 10 mg zinc as gluconate, vitamin A, B, D, E | 6-35 | 286 | 293 | Recovered from acute diarrhoea | $\checkmark$ | $\checkmark$ | - |
| Jamaica (Meeks- <br> Gardner et al. 1998) | Daily supplementation for 12 weeks | 5 mg zinc as sulfate, vitamin A, B, C, D | 6-24 | 31 | 30 | Weight-for-height $<-2 S D$ | $\checkmark$ | $\checkmark$ | - |
| Mexico (Rosado et al. 1997) | Continuous supplementation (5 days/ week) for 54 weeks | 20 mg zinc as methionate, half with iron | 18-36 | 97 | 97 | Community-based | $\checkmark$ | - | - |
| Papua New Guinea (Shankar et al. 1997, 2000) | Continuous supplementation (6 days/week) over a 46-7 week period. Malaria defined based on active case detection as well as clinic visits, confirmed by microscopic evaluation | 10 mg zinc as gluconate | 6-60 | 136 | 138 | Community-based | $\checkmark$ | - | $\checkmark$ |
| Peru (Penny et al. 1999) | Daily supplementation for 26 weeks | 10 mg zinc as gluconate | 6-35 | 80 | 79 | Recovered from persistent diarrhoea | $\checkmark$ | $\checkmark$ | - |
| Viet Nam (Ninh et al. 1996) | Daily supplementation for 22 weeks | 10 mg zinc as sulfate | 4-36 | 73 | 73 | Weight-for-age and height-for-age $<-2 S D$ | $\checkmark$ | $\checkmark$ | - |
| $\checkmark$ Measured. <br> - Not measured. |  |  |  |  |  |  |  |  |  |

malaria. In addition, there are numerous RCTs in which zinc supplements are evaluated as therapeutic agents (Zinc Investigators' Collaborative Group 2000), but these are not included. In all, nine studies contributed findings on zinc deficiency and risk of diarrhoea (Meeks-Gardner et al. 1998; Müller et al. 2001; Ninh et al. 1996; Penny et al. 1999; Rosado et al. 1997; Ruel et al. 1997; Sazawal et al. 1997; Shankar et al. 1997; Umeta et al. 2000), five contributed findings on risk of pneumonia (Bhandari et al. 2002; Meeks-Gardner et al. 1998; Ninh et al. 1996; Ruel et al. 1997; Sazawal et al. 1998), and three contributed findings on risk of malaria (Bates et al. 1993; Müller et al. 2001; Shankar et al. 2000).

The data required for this work are the risk of each health outcome associated with zinc deficiency. This was calculated as the inverse of the odds ratio or relative risk estimated from RCTs given two assumptions: subjects in the study population have some level of zinc deficiency, and supplementation with zinc eliminated zinc deficiency. Because the first of these assumptions is likely to be generally but not universally true, and the second assumption is not likely to be true, we view our results as conservative.

### 4.3 Diarrhoea

The results of the nine RCTs with findings on zinc deficiency and diarrhoea incidence are presented in Table 5.3. The results are presented for each individual study as well as a pooled estimate of all nine studies. The pooled estimate was derived using random effects models as described

Table 5.3 Zinc deficiency and risk of diarrhoea incidence

| Country (reference) | Zinc group |  | Control group |  | Relative risk(95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Episode ${ }^{\text {a }}$ | Follow-up ${ }^{\text {b }}$ | Episode ${ }^{\text {a }}$ | Follow-up ${ }^{\text {b }}$ |  |
| Burkina Faso (Müller et al. 2001) | 322 | 49126 | 374 | 47844 | 1.19 (1.03-1.39) |
| Ethiopia (Shankar et al. 1997) | 27 | 16790 | 59 | 16790 | 2.22 (1.39-3.45) |
| Guatemala (Ruel et al. 1997) | 387 | 8482 | 467 | 8361 | 1.22 (1.08-I.4I) |
| India (Sazawal et al. 1997) | 934 | 44866 | 1033 | 45555 | 1.09 (1.00-1.31) |
| Jamaica (Meeks-Gardner et al. 1998) | 39 | 2604 | 37 | 2265 | 1.09 (0.69-I.72) |
| Mexico (Rosado et al. 1997) | 82 | 42322 | 132 | 42751 | 1.59 (1.20-2.13) |
| Papua New Guinea (Umeta et al. 2000) | 63 | 27490 | 77 | 29465 | I. 14 (0.82-I.59) |
| Peru (Penny et al. 1999) | 564 | 13178 | 661 | 13648 | 1.14 (1.01-I.27) |
| Viet Nam (Ninh et al. 1996) | 56 | 11242 | 100 | 1 242 | 1.79 (1.28-2.50) |
| All |  |  |  |  | 1.28 (1.10-1.49) |

[^10]in a previous pooled analysis (Zinc Investigators’ Collaborative Group 1999). Although each study reported the effects of zinc supplementation on the incidence of diarrhoeal illness, we have presented estimates of the relative risk of diarrhoeal disease incidence due to zinc deficiency. As shown, the pooled estimate indicates that the relative risk of diarrhoea in young children due to zinc deficiency is 1.28 ( $95 \%$ CI 1.10-1.49).

In 1999, a pooled analysis was conducted of published RCTs of zinc supplementation to reduce diarrhoea and pneumonia morbidity in young children (Zinc Investigators’ Collaborative Group 1999); at that time, seven of these studies had been published (Meeks-Gardner et al. 1998; Ninh et al. 1996; Penny et al. 1999; Rosado et al. 1997; Ruel et al. 1997; Sazawal et al. 1997; Umeta et al. 2000). The pooled estimate based on these studies indicated that zinc deficiency increases the risk of diarrhoea in young children by 1.33 ( $95 \%$ CI 1.14-1.59).

### 4.4 Pneumonia

The results of the five RCTs with findings on zinc deficiency and pneumonia incidence are presented in Table 5.4. The results are presented for each individual study as well as a pooled estimate across all five studies. As shown, it is estimated that zinc deficiency increased the risk of pneumonia in young children by 1.52 ( $95 \%$ CI 1.20-1.89).

With the exception of the study by Bhandari et al. (2002), the studies had been included in a published pooled analysis of the effects of zinc supplementation on pneumonia (Zinc Investigators' Collaborative Group 1999). The summary estimate from that analysis was of similar magnitude: the relative risk of pneumonia due to zinc deficiency in young children was 1.69 (95\% CI 1.20-2.45).

Table 5.4 Zinc deficiency and risk of pneumonia incidence

| Country (reference) | Zinc group |  | Control group |  | Relative risk(95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Episode ${ }^{\text {a }}$ | Follow-up ${ }^{\text {b }}$ | Episode ${ }^{\text {a }}$ | Follow-up ${ }^{\text {b }}$ |  |
| India (Bhandari et al. 2002) | 88 | 132000 | 118 | 134400 | 1.32 (1.01-1.72) |
| India (Sazawal et al. 1998) | 24 | 44866 | 43 | 45555 | 1.76 (1.08-2.94) |
| Jamaica (Meeks-Gardner | 0 | 2604 | I | 2265 | 3.13 (0.16-100.0) |
| 1998) |  |  |  |  |  |
| Peru (Penny et al. 1999) | 9 | 13178 | 11 | 13648 | 1.22 (0.5I-2.86) |
| Viet Nam (Ninh et al. 1996) | 45 | 11242 | 81 | 11242 | 1.79 (1.25-2.56) |
| All |  |  |  |  | 1.52 (1.20-1.89) |
| a Episode refers to an episode of pneumonia as per the case definition in the individual study. |  |  |  |  |  |
| Follow-up refers to the total number of child-days of follow-up or disease surveillance in each study. |  |  |  |  |  |

### 4.5 Malaria

Three RCTs of zinc supplementation to prevent malaria morbidity were identified. In Papua New Guinea, Shankar et al. (2000) found a $38 \%$ ( $95 \%$ CI $3-60 \%$ ) reduction in malarial attacks based on clinic visits for malaria or fever with confirmed parasitaemia above a predefined threshold. Using an analogous definition, Bates et al. (1993) also found an approximately one-third reduction in malarial attacks, but it was not statistically significant $(P=0.09)$, and the authors concluded there was no effect of zinc deficiency on malarial attack rates. However, Shankar et al. (2000) obtained the original data from the study (Bates et al. 1993) and utilized random effects models to derive a pooled estimate of a $36 \%$ ( $95 \%$ CI $9-55 \%)$ reduction in clinic-based malarial attacks with parasitaemia with zinc supplementation. The paper by Müller et al. (2001) utilized a different methodology to detect malarial illness-communitybased daily surveillance-and thus, examined a different aspect of malarial illness. They found no significant reduction in malarial incidence by treatment, with a zinc treatment odds ratio of 0.98 ( $95 \%$ CI 0.86-1.11). Studies with community-based malaria surveillance largely detect associations with early manifestations of a malaria episode, and on episodes tending to be less severe, whereas studies focusing on clinical malaria episodes detect effects on the progression of a malaria episode, and on cases perceived to be severe enough to warrant a visit to the health centre (Cox et al. 1994). Because of the heterogeneity of outcomes studied, and our desire to focus on malarial morbidity that contributes to the global disease burden estimates (Snow et al. 1999), we excluded the study by Müller et al. (2001) from further analysis. Instead we used Shankar's pooled analysis to derive a relative risk for malaria morbidity associated with zinc deficiency in young children of 1.56 ( $95 \%$ CI 1.29-1.89).

### 4.6 Mortality

Clearly, zinc deficiency contributes to increased risk of incidence for important childhood diseases that are predominant causes of death among children. Direct estimation of the risk of cause-specific mortality among children due to zinc deficiency is not available from the literature. Unless there were an expectation that zinc deficiency reduced casefatality or severity of these diseases, the risk of mortality due to zinc deficiency should be at least equivalent to the risk of disease occurrence due to zinc deficiency. For this reason, we have proposed that the relative risk of mortality related to diarrhoea, pneumonia and malaria associated with zinc deficiency, are $1.28,1.52$ and 1.56 , respectively (Table 5.5). These might well be conservative estimates, given the high likelihood that zinc deficiency increases the risk of severity and death during illness with diarrhoea or pneumonia (Zinc Investigators' Collaborative Group 1999, 2000). There are three pieces of direct evidence that zinc deficiency increases mortality in young children. First, Indian infants

Table 5.5 Estimated effect of zinc deficiency on morbidity and mortality due to diarrhoea, pneumonia and malaria in children aged 0-4 years

|  | Morbidity |  |
| :--- | :---: | :---: |
| IIIness | Relative risk (95\% Cl) | Mortality |
| Diarrhoea | $1.28(1.10-1.49)$ | $1.28(1.10-1.49)$ |
| Pneumonia | $1.52(1.20-1.89)$ | $1.52(1.20-1.89)$ |
| Malaria | $1.56(1.29-1.89)$ | $1.56(1.29-1.89)$ |

born small for gestation who received zinc supplements six days a week were $0.32(95 \%$ CI $0.12-0.89)$ less likely to die during infancy than those receiving the control supplement (Sazawal et al. 2001). Second, Bangladeshi children who received supplements of $20 \mathrm{mg} / \mathrm{d}$ zinc as adjuvant to oral rehydration solution (ORS) during diarrhoea, were half as likely to die than those receiving ORS alone (Baqui et al. unpublished data). Third, zinc supplementation was associated with a marginally significant reduction in all-cause mortality ( 5 deaths vs 12 deaths, $P=0.10$ ) in the study by Müller et al. (2001). With respect to malarial deaths, it should be reiterated that the reduction in malaria attacks attributable to zinc supplementation was based on clinic attack rates, that is, more severe malaria morbidity rather than less severe illness that would be detected through community-based surveillance, as conducted in the study by Müller et al. (2001). Finally, it should be noted that infections themselves cause secondary zinc deficiency because zinc is sequestered by the liver as part of the acute phase response and is thus less available for many cellular functions (Keen et al. 1993). This point provides additional rationale as to why zinc deficiency may be more strongly related to deaths than to illness incidence. This may also partially explain findings that zinc supplementation reduces the duration of diarrhoeal episodes in prospective trials, as well as in trials in which zinc supplements are provided therapeutically (Zinc Investigators' Collaborative Group 1999, 2000).

### 4.7 DISEASE CAUSATION MECHANISMS

The results of this body of research do not define the explicit biological mechanisms through which zinc deficiency increases disease risk in young children. There is no doubt, however, that zinc is a critical nutrient for cell replication and function and thus, critical for normal functioning of all body systems. The first system known to be compromised with even mild deficiencies of zinc is the immune system, reflecting the profound and ubiquitous role zinc plays in immune function. As recently reviewed by Shankar and Prasad (1998), zinc deficiency impairs multiple aspects of immune function, including barrier and non-specific immunity, spe-
cific immune components (lymphyocytes, monocytes and macrophages, neutrophils, natural killer cells), and mediators of immune function such as glucocorticoid and thymulin activity, and cytokine function. Given the multiple roles of zinc in the immune function, it is not surprising that zinc deficiency should confer increased risk of morbidity and mortality due to infectious diseases.

Zinc deficiency also results in reduced growth rates in animals, and there is ample evidence from RCTs that the provision of supplemental zinc to young children can reduce growth faltering in preschool-aged children. Recently, Brown et al. (1998a) published a meta-analysis of more than 52 RCTs investigating the effect of supplemental zinc on anthropometric status or growth in children. The results of their analysis indicate that improvements in zinc intakes achieved with supplemental zinc (at dosages similar to those reported here) improve the weight-for-age and height-for-age $z$-scores by 0.26 SD and 0.22 SD, respectively. It is heuristic to conclude that some of the improved growth is due to zinc-related effects which reduce morbidity incidence and duration/severity.

## 5. Burden of disease estimates

The estimated deaths and DALYs attributable to zinc deficiency are shown in Tables $5.6-5.7$. When examined by region, the burden of zinc

Table 5.6 Deaths in children aged 0-4 years from zinc deficiency, by subregion

|  |  |  |  |
| :--- | :---: | :---: | ---: |
| Subregion | Diarrhoea | Pneumonia |  |
| AFR-D | 17 | 50 | Malaria |
| AFR-E | 47 | 91 | 74 |
| AMR-A | 0 | 0 | 107 |
| AMR-B | 2 | 3 | 0 |
| AMR-D | 3 | 6 | 0 |
| EMR-B | 1 | 2 | 0 |
| EMR-D | 31 | 48 | 0 |
| EUR-A | 0 | 0 | 10 |
| EUR-B | 1 | 3 | 0 |
| EUR-C | 0 | 0 | 0 |
| SEAR-B | 2 | 6 | 0 |
| SEAR-D | 70 | 187 | 0 |
| WPR-A | 0 | 10 | 16 |
| WPR-B | 2 | 406 | 0 |
| World | 176 |  | 0 |


| Table 5.7 | Disease burden attributable to zinc deficiency, by subregion |  |  |  |
| :--- | :---: | :---: | ---: | :---: |
|  | DALYs (000s) |  |  |  |
| Subregion | Diarrhoea | Pneumonia | Malaria |  |
| AFR-D | 604 | 1705 | 2729 |  |
| AFR-E | 1631 | 3105 | 3978 |  |
| AMR-A | 1 | 1 | 0 |  |
| AMR-B | 69 | 143 | 2 |  |
| AMR-D | 109 | 202 | 2 |  |
| EMR-B | 36 | 93 | 1 |  |
| EMR-D | 1071 | 1679 | 371 |  |
| EUR-A | 0 | 0 | 0 |  |
| EUR-B | 19 | 102 | 0 |  |
| EUR-C | 1 | 7 | 0 |  |
| SEAR-B | 84 | 244 | 21 |  |
| SEAR-D | 2456 | 6579 | 560 |  |
| WPR-A | 0 | 0 | 0 |  |
| WPR-B | 61 | 361 | 5 |  |
| World | 6142 | 14223 | 7669 |  |

deficiency and its consequences are borne most heavily by Africa, the Eastern Mediterranean and South-East Asia. This is true for both diarrhoea and pneumonia. The burden of malarial disease attributable to zinc deficiency is borne almost exclusively by those in the African Region.

## 6. Discussion

We estimated that zinc deficiency in children aged $<5$ years caused 176000 diarrhoea deaths, 406000 pneumonia deaths, and 207000 malaria deaths. The associated DALYs attributable to zinc deficiency were more than 28 million. This was because the risks of morbidity and mortality associated with zinc deficiency are relatively high, and because available data suggest that zinc deficiency-as defined by inadequate dietary zinc-is highly prevalent in many parts of the world, and particularly in parts of the world where the majority of deaths due to diarrhoea, pneumonia and malaria occur. This places zinc deficiency as a key factor conferring risk of morbidity and mortality to young children, one that is ostensibly preventable through public health action.

The evidence that zinc deficiency increases incidence risk is strong because it results from RCTs conducted in areas of the world with limited zinc available in the diet and where pneumonia, diarrhoea and malaria are public health problems. The risk estimates for diarrhoea and
pneumonia are particularly strong, based on results from nine and five RCTs, respectively. The estimated increased risk for malaria is less well studied. Our risk estimate was based on two RCTs with similar risk estimates in which clinic attack rates were the outcome of interest. We excluded the only other published RCT in which the effect of supplemental zinc on less severe malaria morbidity (detected through community surveillance) was studied and in which no increased risk of malaria was found. Clearly more research on the relation between zinc deficiency and malaria is needed to substantiate the risk estimates provided here.

The use of experimental designs has allowed for the characterization of the consequences of zinc deficiency, yet much work needs to be done in order to describe the magnitude and distribution of zinc deficiency throughout the world. This is an important task because the placement of zinc deficiency as a major contributor to the global burden of disease rests in its ubiquitous presence throughout the world. The approach used to characterize zinc deficiency was to estimate the proportion of the population living in areas with inadequate zinc in their food supply as determined from food balance sheet data compiled by FAO. This is far from the traditional approach of characterizing nutrient deficiencies using biochemical indicators. As stated earlier, however, there is no clear choice of biochemical indicator for zinc status, and the information available on zinc status based on plasma zinc concentration (the most commonly cited indicator of zinc status) would be limited to restricted samples of study participants conducted in selected regions of the world. To our knowledge, there are few if any national nutrition surveys utilizing biochemical indicators of zinc status and few laboratories capable of conducting trace mineral assays in developing countries due to the difficulties in eliminating contamination.

The advantage of the approach developed by IZiNCG is that a common methodology could be applied across similarly collected data from 172 countries, representing all 14 subregions. Further, because zinc is stored in the body in only limited amounts, and because certain key features of the diet (total energy content, animal protein availability, fibre/phytate content associated with the principal grain) largely determine the bioavailability of the zinc present in the food supply, the adequacy of dietary zinc intakes is an appropriate indicator to approximate the prevalence of zinc deficiency. There are multiple uncertainties inherent in this approach, and key assumptions include the following: (i) that per capita food availability data relates to actual dietary intake and ultimately to zinc status; (ii) that foods not included in the data (e.g. foraged foods) are not rich sources of zinc; (iii) that our knowledge of the composition of foods with respect to zinc, calcium, fibre and phytate is adequate for our purposes; (iv) that characterization of the bioavailability of zinc is correct; and (v) that we have the appropriate zinc requirement distribution and cut-point for defining adequacy of
intake. It is only through further research that we can determine the validity of the assumptions made, and judge the certainty with which we have characterized the prevalences of zinc deficiency in young children.

There are reasons to believe that the estimates presented here are conservative. First, food availability is usually greater than food intake. This effect, however, may be countered by underestimation of intake in the FAO data which do not always account for subsistence production. Second, the groups for which these numbers are likely to be most conservative are women of reproductive age and small children who may be particularly disadvantaged in terms of obtaining the best sources of zinc (animal products) in their diet. Third, dietary intake surveys, which more directly estimate the prevalence of inadequate intakes in populations, typically have estimated higher prevalences of inadequate zinc intakes than those depicted here (Parr 1992, 1996). On the other hand, because breastfeeding is an adequate source of zinc in the diets of infants aged less than six months, the same age when a large proportion of childhood mortality is concentrated, applying the risks to all children aged 0-4 years may result in substantial overestimation of disease burden in some populations.

It is also important to note that maternal or gestational zinc deficiency may affect immunological development in the newborn in ways that compromise immune function throughout the lifespan irrespective of zinc status (Caulfield et al. 1998; Shankar and Prasad 1998). Although this is well demonstrated in certain animal models, it has not been well characterized in humans, but a recent study by Osendarp et al. (2001) provides intriguing evidence of potential long-term effects on infant health in the first year of life due to maternal prenatal zinc deficiency. Fetal accumulation of zinc is a function of maternal zinc status, and therefore, newborn zinc deficiency (assessed as low serum zinc concentration) is likely among women with inadequate dietary zinc intakes during pregnancy (Caulfield et al. 1999). This deficiency is likely transitory for all but the most vulnerable (preterm or low-birth-weight infants) because the zinc content of colostrum is high and some zinc becomes available to the infant as part of the haematological changes accompanying the transition to extrauterine life (WHO 1996). Beginning at around six months of age, however, breast milk intake no longer provides sufficient zinc to meet requirements (Krebs 2000), making zinc-rich complementary foods necessary (Brown et al. 1998b). If zinc-rich sources are not available on a routine basis, zinc deficiency develops over time and persists until changes in the diet are made. Thus, it must be recognized that removal of the global burden of disease due to zinc deficiency will likely require improvements in child and maternal zinc intakes. In this chapter, we have considered only the burden of disease for children aged 0-4 years since no available evidence on the role of zinc in morbidity and mortality among other age groups was available. Thus, it may
be true that zinc deficiency contributes greatly to death and disability among other age groups. Defining these risk relations should be a priority area for future research.

Because low weight-for-age (usually described as underweight) also increases risk of morbidity and mortality from diarrhoea, pneumonia and malaria (see chapter 2), there is overlap between zinc deficiency and underweight as risk factors for morbidity and mortality from these causes in young children. The degree of overlap cannot be quantified empirically, but evidence from studies can provide insight in this issue. In the meta-analysis of zinc RCTs (Zinc Investigators' Collaborative Group 1999), the authors found that the effects of zinc supplementation on risk of diarrhoea and pneumonia across studies did not depend on the underlying level of malnutrition as characterized by the average $z$ score in the study subjects. However, the study by Umeta et al. (2000) did find greater reductions in morbidity in stunted as opposed to nonstunted children. Zinc supplementation has reduced morbidity with no detectable changes in growth rates (Meeks-Gardner et al. 1998). Overall, limited evidence shows that the efficacy of zinc supplementation for reducing morbidity is more substantial than for reducing growth faltering in studies in which both outcomes have been characterized. There are also studies in the literature in which improvements in immune function have been observed with zinc supplementation in well-nourished subjects (e.g. those with adequate weight); thus, although most of the research on the benefits to health from zinc supplements has been conducted in populations similarly characterized by underweight or stunting, the potential reductions in the global burden of disease due to zinc deficiency should not be limited to those who are also underweight or stunted. In the case of malaria, Shankar (2000) describes how zinc deficiency exerts fairly specific effects on malaria morbidity (increasing febrile illness accompanied by hyperparasitaemia), whereas low anthropometric status appears to exert more generalized effects. With these considerations, we conclude that although zinc deficiency and underweight co-exist in many populations, and both are risk factors for the disease outcomes described here, the majority of the disease burden attributable to zinc deficiency is not restricted to children who are generally malnourished.

In summary, the available evidence suggests that zinc deficiency contributes substantially to death and disability throughout the world, and particularly in Africa, the Eastern Mediterranean, and South-East Asia. Because epidemiological evidence is limited to studies conducted among preschool children and this analysis considers only three causes of death and disability, it is clear that further research on the role of zinc in morbidity, mortality and disability due to other causes and in other age groups is urgently needed. Further research is needed to identify indicators of exposure to zinc deficiency as well as effective strategies to reduce zinc deficiency and its consequences.

## 7. PROJECTIONS OF EXPOSURE

Currently, there are no programmes or policy initiatives in place to specifically improve the zinc intakes of human populations. This is true with respect to improving the zinc supply in foods (e.g. interventions to improve intakes of animal products which are good sources of zinc or to reduce phytates in foods which impair zinc absorption). Further, there are no examples of programmes utilizing supplemental zinc to improve zinc status, although some are beginning to use supplemental zinc as adjuvant therapy during diarrhoeal illness, based on research demonstrating the efficacy of zinc in shortening the duration of the current episode and prolonging the time until the next episode in paediatric populations with heavy disease burden. Because of the general lack of policy or programmatic initiatives to address zinc deficiency, we project that the magnitude and distribution of zinc deficiency will remain the same for the years 2010 and 2020, as presented here.

## Acknowledgements

We are grateful for the support of Kenneth H. Brown in providing us with the data on the prevalence of inadequate zinc intakes in 172 countries; Sunil Sazawal for updating the pooled analyses of zinc and risk of diarrhoea and pneumonia, and Anuraj H. Shankar for conducting the pooled analysis of zinc and malaria and for his insights into the various dimensions of malarial morbidity.

This manuscript was supported in part by the Family Health and Child Survival (FHACS) Cooperative Agreement between the United States Agency for International Development, Office of Health and Nutrition, the Johns Hopkins Bloomberg School of Public Health, Department of International Health, and by the World Health Organization.

## Note

1 See preface for an explanation of this term.

## References

Bates CJ, Evans PH, Dardenne M et al. (1993) A trial of zinc-supplementation in young rural Gambian children. British Journal of Nutrition, 69:243255.

Bhandari N, Bahl R, Taneja S et al. (2002) Routine zinc supplementation on pneumonia in children aged 6 months to 3 years: randomised controlled trial in an urban slum. British Medical Journal, 324:1358.

Brown KH, Dewey KG, Allen LH (1998a) Complementary feeding of young children in developing countries: a review of current scientific knowledge. World Health Organization, Geneva.

Brown KH, Peerson JM, Allen LH (1998b) Effect of zinc supplementation on children's growth: a meta-analysis of intervention trials. In: Role of trace elements for health promotion and disease prevention. Sandstrom B, Walter P, eds. Bibliotheca Nutritio et Dieta, 54:76-83.

Brown KH, Wuehler SE, Peerson JM (2001) The importance of zinc in human nutrition and estimation of the global prevalence of zinc deficiency. Food and Nutrition Bulletin, 22:113-125.
Caulfield LE, Zavaleta N, Figueroa A (1999) Adding zinc to prenatal iron and folate supplements improves maternal and neonatal zinc status in a Peruvian population. American Journal of Clinical Nutrition, 69:1257-1263.
Caulfield LE, Zavaleta N, Shankar A, Merialdi M (1998) Potential contribution of maternal zinc supplementation during pregnancy for maternal and child survival. American Journal of Clinical Nutrition, 68:S499508.

Cox MJ, Kum DE, Tavul L et al. (1994) Dynamics of malaria parasitaemia associated with febrile illness in children from a rural village of Madang, Papua New Guinea. Transactions of the Royal Society of Tropical Medicine and Hygiene, 88:191-197.
FAO (1999) FAO food balance sheets: food balance sheet country averages 1990-1997. Food and Agriculture Organization. Available at http:// apps.fao.org/lim500/nph-wrap.pl.
Gibson RS (1994) Zinc nutrition in developing countries. Nutrition Research Reviews, 7:151-173.

Hambridge KM (2000) Human zinc deficiency. Journal of Nutrition, 130:S1344-1349.

IZiNCG (forthcoming) Assessment of zinc status in populations and options for the control of zinc deficiency. (IZiNCG Technical Document No. 1). International Zinc Nutrition Consultative Group (IZiNCG). UNU/IUNS, Cornell University, Ithaca, NY.
Keen CL, Taubeneck MV, Daston GP, Rogers JM, Gershwin E (1993) Primary and secondary zinc deficiency as factor underlying abnormal CNS development. Annals of the New York Academy of Sciences, 678:37-47.

Krebs N (2000) Dietary zinc and iron sources, physical growth and cognitive development of breastfed infants. Journal of Nutrition, 130:S358-360.

Meeks-Gardner J, Witter M, Ramdath D (1998) Zinc supplementation effects on the growth and morbidity of undernourished Jamaican children. European Journal of Clinical Nutrition, 52:34-39.
Müller O, Becher H, van Zweeden AB et al. (2001) Effect of zinc supplementation on malaria and other causes of morbidity in west African children: randomised double blind placebo controlled trial. British Medical Journal, 322:1567-1573.
Ninh X, Thissen P, Collette L, Grard G, Khoi H, Keteislegers M (1996) Zinc supplementation increases growth and circulating insulin-like growth factor 1 (IGF-1) in growth retarded Vietnamese children. American Journal of Clinical Nutrition, 63:514-518.

Osendarp SJ, van Raaij JM, Darmstadt GL, Baqui AH, Hautvast JG, Fuchs GJ (2001) Zinc supplementation during pregnancy and effects on growth and morbidity in low birthweight infants: a randomized placebo controlled trial. The Lancet, 357:1080-1085.

Parr RM (1992) Human dietary intakes of trace elements: a global literature survey mainly for the period 1970-1991: data listing and references. International Atomic Energy Agency, IAEA-NAHRES, Vienna.
Parr RM (1996) Assessment of dietary intakes. In: Trace elements in buman nutrition and health. World Health Organization, Geneva.
Penny M, Person M, Marin M et al. (1999) Randomized community-based trial of the effect of zinc supplementation, with or without other micronutrients, on the duration of persistent diarrhea in Lima, Peru. Journal of Pediatrics, 135:208-217.

Prasad AS (1991) Discovery of human zinc deficiency and studies in an experimental human model. American Journal of Clinical Nutrition, 53:403412.

Rosado J, Lopez P, Monoz E, Martinez H, Allen LH (1997) Zinc supplementation reduced morbidity, but neither zinc nor iron supplementation affected growth or body composition of Mexican preschoolers. American Journal of Clinical Nutrition, 65:13-19.

Ruel M, Rivera J, Santizo M, Lonnerdal B, Brown K (1997) The impact of zinc supplementation on morbidity from diarrhea and respiratory infections among young rural Guatemalan children. Pediatrics, 99:808-813.

Sandstead HH (1991) Zinc deficiency. A public health problem? American Journal of Diseases of Children, 145:853-859.

Sazawal S, Black R, Bhan M, Jalla S, Sinha A, Bhandari N (1997) Efficacy of zinc supplementation in reducing the incidence and prevalence of acute diar-rhea-a community-based, double-blind, controlled trial. American Journal of Clinical Nutrition, 66:413-418.
Sazawal S, Black R, Jalla S, Mazumdar S, Sinha A, Bhan M (1998) Zinc supplementation reduces the incidence of acute lower respiratory infections in infants and preschool children. A double-blind, controlled trial. Pediatrics, 102:1-5.

Sazawal S, Black R, Monon VP et al. (2001) Zinc supplementation in infants born small for gestational age reduces mortality: a prospective randomized controlled trial. Pediatrics, 108:1280-1286.
Shankar AH (2000) Nutritional modulation of malaria morbidity and mortality. Journal of Infectious Diseases, 182:S37-53.
Shankar AH, Genton B, Baisor M et al. (2000) The influence of zinc supplementation on morbidity due to Plasmodium falciparum: a randomized trial in preschool children in Papua New Guinea. American Journal of Tropical Medicine and Hygiene, 62:663-669.

Shankar AH, Genton B, Tamja S et al. (1997) Zinc supplementation can reduce malaria-related morbidity in preschool children. American Journal of Tropical Medicine, 57:A434.

Shankar AH, Prasad AS (1998) Zinc and immune function: the biological basis of altered resistance to infection. American Journal of Clinical Nutrition, 68:S430-434.

Snow RW, Craig M, Deichmann U, Marsh K (1999) Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population. Bulletin of the World Health Organization, 77:624-640.
Umeta M, West CE, Haider J, Deurenberg P, Hautvast JG (2000) Zinc supplementation and stunted infants in Ethiopia: a randomized controlled trial. The Lancet, 355:2021-2026.
WHO (1996) Zinc. In: Trace elements in human nutrition and health. World Health Organization, Geneva.

WISTAT (1994) Women's Indicators and Statistics Database of the United Nations. (CD-ROM, Version 3.) United Nations. Available at http:// www.un.org/Depts/unsd/gender/wistat/topiclist.htm.

Wood RJ (2000) Assessment of marginal zinc status in humans. Journal of Nutrition, 130:S1350-1354.

World Food Dietary Assessment System (1997) (Computer program, Version 2.0.) University of California, Berkeley, CA.

Zinc Investigators' Collaborative Group (1999) Prevention of diarrhea and pneumonia by zinc supplementation in children in developing countries: pooled analysis of randomized controlled trials. Journal of Pediatrics, 135:689-697.
Zinc Investigators' Collaborative Group (2000) Therapeutic effects of zinc in acute and persistent diarrhea in children in developing countries: pooled analysis of randomized controlled trials. American Journal of Clinical Nutrition, 72:1516-1522.

## Chapter 6

# High blood pressure 

Carlene M.M. Lawes, Stephen Vander Hoorn, Malcolm R. Law, Paul Elliott, Stephen MacMahon and Anthony Rodgers

## Summary

This chapter aims to estimate the burden of disease attributable to raised blood pressure in 2000, and the burden avoidable by distributional shifts of blood pressure. The key initial steps were to define the blood pressure variable, choose disease outcomes to be assessed, estimate current global blood pressure levels, choose a theoretical minimum and estimate risk factor-disease relationships.

Blood pressure was defined as systolic blood pressure (SBP) in mmHg. SBP was chosen principally because data are widely available for this index of blood pressure, and it appears to be at least as good a predictor of cardiovascular disease as other indices.

Based on consistent direct, continuous associations in cohort studies and evidence from randomized clinical trials, the outcomes assessed were stroke, ischaemic heart disease (IHD), hypertensive disease and other cardiac disease. Similar data suggested causal relationships with renal failure, but this could not be mapped to a Global Burden of Disease (GBD) study outcome.

Raw SBP data were obtained from studies after a systematic review of population-based surveys, and included data from about 230 surveys and over 650000 participants. Sex-specific associations of SBP with age were estimated for each of the subregions ${ }^{1}$ separately, based on populationweighted (by country) study estimates of mean values. There was moderate variation in the final age-specific and sex-specific estimates of mean blood pressure across the 14 subregions, with the range between the highest and lowest age-specific mean SBP levels across subregions typically being about 20 mmHg . Confidence intervals around mean SBP and standard deviations were calculated to reflect uncertainty. Trends in mean SBP over time were extrapolated from published epidemiological data.

The theoretical minimum of SBP was estimated to be mean 115 mmHg and standard deviation 6 mmHg (usual SBP), for all age, sex and subregional groups. The main basis for this estimate was the level of SBP down to which epidemiological relationships with cardiovascular disease outcomes are observed. This theoretical minimum is also consistent with the levels of SBP in populations with little or no cardiovascular disease. Furthermore, recent clinical trial data have indicated reductions in stroke after lowering blood pressure by about 10 mmHg SBP in those with mean SBP 125 mmHg .

Data on the risk factor-disease relationship were obtained from the Asia-Pacific Cohort Studies Collaboration (APCSC), an individual participant meta-analysis that combined data from 37 prospective observational cohorts involving over 425000 individuals with $2-27$ years of follow-up (mean 7 years), and in total over 3 million person-years of observation. Data from this meta-analysis were complemented by data from other overviews and large cohort studies. The main findings were direct, positive and continuous associations of usual SBP with the risks of all end-points of interest. The risks were similar by sex and subregion, except there was possibly a stronger association of blood pressure with stroke in Asian compared to non-Asian populations, due partly to the higher proportion of haemorrhagic strokes. Overall, each 10 mmHg below-usual SBP was associated with $38 \%$ ( $95 \%$ CI 37-39\%) lower stroke risk and a $26 \% ~(95 \%$ CI $24-29 \%)$ lower risk of IHD. Each 10 mmHg below-usual SBP was associated with $46 \%$ ( $95 \%$ CI $40-51 \%$ ) and $18 \% ~(95 \%$ CI $15-20 \%)$ lower risks of hypertensive disease and other cardiac disease. Few data were available for the GBD end-point "other cardiac disease", so, given some uncertainty about causality and the varying composition for this end-point around the world, the relative risks were halved for this outcome. There was attenuation of proportional associations with age for all these outcomes.

Data on risk reversibility came from meta-analyses of randomized controlled trials of blood pressure lowering. In total 23 trials were reviewed, which included over 71000 participants allocated a variety of blood-pressure-lowering agents or placebo. The mean overall SBP reduction was 10 mmHg , and a total of 2632 strokes and 3693 IHD events were observed. Overall the trials confirmed the size of the reductions expected from epidemiological relationships in middle age. However, the reductions in risk in old age were larger than expected from the cohort study data.

We used the observational epidemiological data for relative risk estimates and the trial overviews for the time frame of risk reversibility. For example, in middle age the epidemiology suggests a 10 mmHg increase in SBP is associated with about $40 \%$ more stroke and $25 \%$ more IHD. The trials suggested that within 3-5 years of lowering SBP by 10 mmHg , most or all of this increased risk for stroke and hypertensive disease and
approximately two thirds for IHD and other cardiovascular disease are reversed.

The foregoing methods and assumptions allowed estimates of burden attributable to mean SBP levels of more than 115 mmHg . The analysis showed that:

- forty-nine per cent of IHD is attributable to SBP $>115 \mathrm{mmHg}$ worldwide (range of $40-64 \%$ by subregion), which translates into 2991000 deaths and 28.2 million disability-adjusted life years (DALYs) in the year 2000;
- sixty-two per cent of stroke is attributable to SBP $>115 \mathrm{mmHg}$ worldwide (range of $52-78 \%$ by subregion), which translates into 3149000 deaths and 27.8 million DALYs in the year 2000;
- seventy-six per cent of hypertensive disease is attributable to SBP $>115 \mathrm{mmHg}$ worldwide (range of $64-90 \%$ by subregion), which translates into 663000 deaths and 5.4 million DALYs in the year 2000; and
- fourteen per cent of other cardiovascular disease is attributable to SBP $>115 \mathrm{mmHg}$ worldwide (range of $8-22 \%$ by subregion), which translates into 338000 deaths and 2.8 million DALYs in the year 2000.

Worldwide, 7.1 million deaths (about $12.8 \%$ of the total) and 64.3 million DALYs ( $4.4 \%$ of the total) were estimated to be due to nonoptimal blood pressure. The proportion of DALYs was lower than the proportion of deaths as most blood pressure-related cardiovascular disease events occurred in middle or old age, so the years of life lost due to premature mortality and years of life lost due to disability is a smaller percentage of the total DALYs than of the total deaths. Overall, the results suggest that a considerable proportion of cardiovascular disease is related to non-optimal blood pressure, and this translates into an important portion of deaths and years of life lost to deaths and disability worldwide. Approximately one third of the blood pressure-related attributable burden occurred in developed subregions, one third in lower mortality developing subregions (AMR-B, EMR-B, EUR-B, EUR-C, SEAR-B, WPR-B) and a further third in high mortality developing subregions (AFR-D, AFR-E, AMR-D, EMR-D, SEAR-D).

The estimates of burden attributable to non-optimal blood pressure are approximately double those from the World Health Organization (WHO) GBD study in 1990. However, the 1990 estimates did not correct for regression dilution bias, which means there was underestimation of risk of cardiovascular disease by a factor of about two. This therefore explains the almost doubling of burden attributable to blood pressure in current estimates.

## 1. Introduction

### 1.1 Defining the blood pressure variable

Blood pressure is a measure of the force that the circulating blood exerts on the walls of the main arteries. The pressure wave transmitted along the arteries with each heartbeat is easily felt as the pulse-the highest (systolic) pressure is created by the heart contracting and the lowest (diastolic) pressure is measured as the heart fills.

Blood pressure is described as a continuous variable as it is commonly reported in this manner, with mean and standard deviation values. Relative risk values for the risk factor-disease relationship are also available for this format. The standard unit for measuring blood pressure is mmHg , which may be applied to SBP, diastolic blood pressure (DBP), or alternative measurements such as mean arterial pressure and pulse pressure ( PP ).

Historically, many classification systems and treatment recommendations placed more emphasis on DBP, as elevated DBP was thought to confer greater risk for cardiovascular disease than elevated SBP (Kannel 1999, 2000; Lloyd-Jones et al. 1999). There is, however, now evidence for a paradigm shift to consider SBP as well as DBP (Black 1999b). Both the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC-VI) (Anonymous 1997) in 1997 and the 1999 guidelines from the WHO-International Society of Hypertension (WHO/ISH) (Anonymous 1999) now agree that both SBP and DBP should be used to classify hypertension (Black 1999a).

Prospective observational studies have provided data on whether, over the long term, there appears to be an association between blood pressure and disease end-points. Therefore, comparisons of the strength of associations for both DBP and SBP may be made. Data published from the Framingham study over the past 30 years have suggested that cardiovascular consequences do not necessarily derive principally from DBP (Kannel et al. 1969). While DBP may be a better predictor of cardiovascular disease in those aged $<45$ years, SBP is a better predictor of stroke and cardiovascular disease in those aged $>60$ years (Kannel et al. 1970, 1971). Overall, the risk of cardiovascular events was greater in the presence of isolated systolic hypertension than diastolic hypertension. For each standard deviation increase in mean SBP, the cardiovascular disease risk increased by $40-50 \%$, whereas for DBP the increment was $30-35 \%$. (This persisted after adjusting for age, and occurred in both men and women.) Combined systolic and diastolic hypertension carried only marginally greater risk than isolated systolic hypertension (Kannel 1996).

Other prospective studies corroborate these results with evidence that, in both sexes, the overall association between blood pressure and cardiovascular end-points is stronger for SBP than DBP (Franklin et al.

1999, 2001; Lichtenstein et al. 1985; Miall 1982; Mitchell et al. 1997; Miura et al. 2001; Sesso et al. 2000; Stamler et al. 1993). Within each level of DBP, a higher SBP was related to increased risk of IHD in a continuous and graded fashion. There was also an increase in risk with higher DBP within each SBP level, but in this latter scenario, the increases were not as steep or consistent (Stamler et al. 1993). Subjects who subsequently died from IHD were better identified by their SBP than by their DBP. Overall, SBP was a better predictor of outcome (Lichtenstein et al. 1985).

There are some data to suggest that PP (the difference between SBP and DBP) is a good predictor of cardiovascular risk (Abernethy et al. 1986; Black 1999b; Domanski et al. 1999; Frohlich 2000). However, it has not been consistently shown to be superior to SBP (Abernethy et al. 1986; Franklin 1999; Franklin et al. 1999, 2001; Sesso et al. 2000), and it was not a better predictor than SBP in the Asia-Pacific Cohort Studies Collaboration overview (APCSC 2003b); further, estimates of relative risk are available for SBP rather than PP. SBP was therefore used in preference to PP for the purposes of these analyses.

### 1.2 Disease outcomes

There are a number of potential disease outcomes associated with blood pressure. Blood pressure is generally accepted to have a role in accelerating atherosclerosis of the blood vessels and thereby influencing cardiovascular disease. Atherosclerosis is believed to start with "fatty streaks" on the intimal surface of blood vessels. Over time the intima is invaded by "foam cells" (lipid-laden macrophages) and plaques develop (Warlow et al. 1996). These plaques may be complicated by platelet adhesion, activation and aggregation, and formation of a thrombus. Many ischaemic cardiovascular disease events are due to "atherothromboembolism", where a fibrous plaque may obstruct blood vessel lumen resulting in infarction, or the thrombotic component of the plaque may break down and embolize (Warlow et al. 1996). It is proposed that haemorrhagic events (e.g. strokes) are due to degenerative atherosclerotic changes in blood vessels, which may cause formation of microaneurysms and fibroid necrosis, both of which may weaken the artery wall and lead to rupture (Leppala et al. 1999; MacKenzie 1996; Swales 1994; Warlow et al. 1996). Potential outcomes considered for blood pressure are as follows.

## Stroke

Data from a variety of prospective cohort studies and overviews have demonstrated a strong, continuous temporal association between blood pressure and stroke (Anonymous 1998; APCSC 2003a; MacMahon and Rodgers 1993a; MacMahon et al. 1990; PSC 1995). Further, a causal association is biologically plausible and clinical trials have demonstrated reversibility (Collins et al. 1990; Neal and MacMahon 1999; Neal et al.
2000). More recent prospective studies have also assessed the association between blood pressure (BP) and stroke subtypes, such as haemorrhagic and ischaemic stroke. However, the GBD classification system for diseases and injuries has only one category for total stroke (end-point 108 G4 cerebrovascular disease, International Statistical Classification of Diseases, ninth revision [ICD-9] codes 430-438). Thus, it was not possible to analyse stroke subtypes individually in this work. Epidemiological studies show a positive association between blood pressure and fatal, non-fatal and total strokes, and total strokes were included in the analyses.

## ISCHAEMIC HEART DISEASE

As with stroke, a variety of prospective cohort studies and overviews have demonstrated a strong, continuous temporal association between blood pressure and IHD (APCSC 2003a; MacMahon and Rodgers 1993a; MacMahon et al. 1990); a causal association is biologically plausible and clinical trials have demonstrated reversibility (Collins et al. 1990; Neal and MacMahon 1999; Neal et al. 2000). The GBD classification system for diseases and injuries has a category for total IHD (endpoint 107 G3 ischaemic heart disease, ICD-9 codes 410-414).

## RENAL DISEASE

Cohort study data on mortality in association with hypertension do not often include renal disease as an end-point (Whelton and Klag 1989). However, the largest cohort to date, the Multiple Risk Factor Intervention Trial (MRFIT) study assessed the development of end-stage renal disease in men followed for 16 years. This study identified a strong, graded relation between both systolic and DBP and end-stage renal disease, independent of other disease associations (Klag et al. 1996). There were several limitations in this study, but many of these could potentially be addressed in analyses. Only baseline blood pressure measurements were taken, which would underestimate associations substantially, but correction factors could be applied. No women were included, but the proportional effect is the same in males and females for other disease associations, so it could be assumed to be consistent here. No information was available on anti-hypertensive agents, which could weaken the association, but would not negate it. The major limitation was that renal function was not assessed at baseline in all participants (Klag et al. 1996). This makes the issue of temporality difficult to establish, as renal disease could predispose to high blood pressure rather than the reverse.

Additional limited, but consistent, data about the relationship between blood pressure and renal failure are available from other prospective observational studies and hypertension treatment trials (MacMahon 1994; Whelton and Klag 1989). While this evidence is suggestive, it is not conclusive (Whelton and Klag 1989).

The above evidence is suggestive of an etiological role for blood pressure in the progression of renal failure. However, there is a further major consideration in choosing disease outcomes for this project. All disease outcomes must map to GBD end-points, and there is no end-point that is equivalent to renal failure. The "nephritis and nephrosis" category (end-point 121 J1, ICD-9 codes 580-589) includes a wide range of more diverse renal conditions. Therefore, apart from the more restricted "hypertensive renal disease" incorporated in the GBD end-point, hypertensive disease (below) analyses of renal failure could not be conducted.

## Hypertensive disease

The GBD classification system end-point "hypertensive disease" (endpoint 106 G2, ICD-9 codes 401-405) includes essential hypertension, hypertensive heart disease, hypertensive renal disease, hypertensive heart and renal disease and secondary hypertension. To some extent this is not an ideal disease outcome category. The evidence suggests continuous associations between blood pressure and disease end-points, so partitioning off those with hypertension is somewhat artificial as they are more likely to be part of a continuum rather than a distinct group. It would be preferable to have more specific disease categories that were not specifically labelled as "hypertensive". If any disease is selectively diagnosed in people with high blood pressure, the strong association with blood pressure becomes self-fulfilling. However, the choice of end-points for analyses is limited to those classified in the GBD classification system for diseases. As there are direct and positive associations with blood pressure and these outcomes, it is necessary to include this end-point, despite its shortcomings.

## Other Cardiovascular Disease

The final end-point included in the GBD classification system associated with blood pressure is "other cardiac disease". This compromises a miscellany of cardiovascular conditions including heart failure, pulmonary heart disease, diseases of the pericardium and endocardium, conduction disorders, cardiac dysrhythmias, diseases of arteries, arterioles and capillaries, and diseases of veins and lymphatics. Numerically, one of the most important is likely to be heart failure. Data from the Framingham study suggest that for those with hypertension, the incidence of heart failure is increased sixfold relative to those who are not hypertensive (Stokes et al. 1989). The risk of congestive heart failure for highest:lowest blood pressure quintiles was two-three times (Kannel and Belanger 1991). This positive association was also apparent in analyses of data from the APCSC (APCSC secretariat, personal communication, 2001), and the risk of heart failure appears to rise continuously and steeply with increasing levels of blood pressure (Stokes et al. 1989). Clinical trials also suggest reversibility of heart failure with blood pressure lowering (Neal et al. 2000). Data from APCSC also indicate strong
positive associations with all cardiovascular deaths that are not due to stroke, IHD or hypertensive disease.

Ideally, the disease end-points would have more specific categories rather than this heterogeneous group. Unfortunately, available data do not allow for analyses by specific diagnostic category within this group. Due to the substantial number of deaths that were likely to be associated with blood pressure, it was not appropriate to drop this category altogether. Analyses, therefore, included this end-point, but as discussed later in this chapter, relative risk estimates were modified to take into account the varying composition and uncertainty of causality of all diseases within this "other cardiovascular disease" group.

### 1.3 Theoretical minimum

## LOWEST RELATIVE RISK OF END-POINTS——OBSERVATIONAL STUDIES

APCSC combined data from 37 prospective observational cohorts involving over 425000 individuals with 2-27 years of follow-up (mean 7 years), and in total over 3 million person-years of observation (APCSC 1999). Analyses of the relationship between SBP and relative risk of cardiovascular disease in the Asia-Pacific region illustrate a continuous log linear relationship for both end-points. The lowest demonstrated relative risk for stroke and IHD occurred at approximately 115 mmHg (Figure 6.1), which suggests that this would be an appropriate level for a theoretical minimum.

Figure 6.I Usual SBP and risk of stroke and IHD in the Asia-Pacific region


[^11]
## RECENT DATA FROM BLOOD-PRESSURE-LOWERING TRIALS

Recently available data from a large trial of blood pressure lowering (PROGRESS, involving 6105 participants with cerebrovascular disease) indicated benefits of blood pressure lowering even in the group with baseline SBP $<140 \mathrm{mmHg}$ (mean 125 mmHg ) (Progress Collaborative Group 2001).

## LOW BLOOD PRESSURE POPULATIONS

An alternative source of data relevant to setting a theoretical minimum comes from studies on populations with little or no cardiovascular disease, based on research since the 1930s and 1940s (Poulter and Sever 1994). These include populations that are relatively isolated and have preserved their lifestyle for many generations. In many of these populations there is no, or very low prevalence of cardiovascular disease, including studies of autopsy findings (Poulter and Sever 1994). A further feature common to these populations is low blood pressure, and no or limited increase in blood pressure with age. This pattern is consistent in isolated populations throughout the world (Table 6.1).

These studies do have limitations. In most studies, there was no discussion of how age was assessed in these remote populations; in some surveys, other criteria were used when age was unknown, such as physical appearance, number and age of children, personal knowledge of interpreters and calendars of local events such as initiation at puberty (Carvalho et al. 1989; Mann et al. 1964). Despite this, the data suggest that "low blood pressure" populations do exist (mean SBP $\leq 115 \mathrm{mmHg}$ ) where hypertension is absent or rare (Carvalho et al. 1989; He et al. 1991a, 1991b; Poulter and Sever 1994). Cardiovascular disease also tends to have a very low prevalence in these populations (Poulter and Sever 1994). Typically these populations have diets low in salt, cholesterol and fat (particularly animal fat), a lifestyle requiring heavy physical labour, and an absence of obesity (Barnes 1965; Carvalho et al. 1989; Connor et al. 1978; He et al. 1991a, 1991b; Page et al. 1974; Poulter and Sever 1994; Sever et al. 1980). Low salt intake appears to be particularly relevant to the blood pressure differences. There is also evidence from studies conducted in a variety of settings (Poulter and Sever 1994), such as Africa (Poulter et al. 1988, 1990), China (He et al. 1991a, 1991b) and the Pacific (Joseph et al. 1983; Salmond et al. 1985, 1989), that blood pressure levels rise after migration to more urbanized "acculturated" settings and the slope of the age-BP relationship increases. In addition to changes in blood pressure, body mass index and heart rate tend to increase (Poulter and Sever 1994). Pre-migration data suggest that these changes are not due to selective migration (Poulter et al. 1988). Instead, it is likely that factors such as dietary changes of increased intake of sodium, animal protein, fat and processed foods, and decreased intake of potassium and vegetable protein are important (He et al. 1991a, 1991b; Poulter and Sever 1994; Poulter et al. 1988).

| Country or area (reference) | Population | Mean SBP levels | Patterns with age |
| :---: | :---: | :---: | :---: |
| Africa |  |  |  |
| Tanganyika ${ }^{\text {a }}$ (Mann et al. 1964) | Masai tribe (mostly men) with virtually no cardiovascular disease | $\begin{aligned} & 120-125 \mathrm{mmHg} \text { SBP } \\ & >160 \mathrm{mmHg} \text { rare } \end{aligned}$ | No age-related rise in BP |
| Southern Africa (Sever et al. 1980) | Tribal and urban Xhosa people | $\begin{aligned} & 120-130 \mathrm{mmHg} \\ & \text { rural } 130-140 \mathrm{mmHg} \\ & \text { urban } \end{aligned}$ |  |
| Kenya (Carvalho et al. 1989) | Rural villagers in remote areas (INTERSALT study) | $108-114 \mathrm{mmHg}$ in males and females |  |
| Americas |  |  |  |
| Mexico (Connor et al. 1978) | Tarahumara Indians who lived in the mountains | $105-110 \mathrm{mmHg}$ | No age-related rise in BP |
| Brazil (Carvalho et al. 1989) | Yanamamo Indians of the Amazon rain forest and Xingu Indians (INTERSALT study) | $90-105 \mathrm{mmHg}$ | Virtually no increase in BP with age |
| Western Pacific |  |  |  |
| China (He et al. 1991a, 1991b) | Rural Yi farmers in isolated areas | $100-110 \mathrm{mmHg}$ | BP does not tend to rise after puberty |
| Papua New Guinea (Barnes 1965) | The isolated Bomai and Yongamuggi people | $105-110 \mathrm{mmHg}$ for males and females | No age-related rise in BP |
| Papua New Guinea (Carvalho et al. 1989) | Rural villagers in remote areas (INTERSALT study) | $105-110 \mathrm{mmHg}$ | No age-related rise in BP |
| Solomon Islands (Page et al. 1974) | Six tribal societies with varying levels of acculturation | $110-120 \mathrm{mmHg}$, lowest in least acculturated groups | No age-related rise in BP |

[^12]Data from these studies indicate that populations do exist that have mean SBP levels of about 115 mmHg or even lower. The final decision on theoretical minimum, based on all the data, was that it would be set at 115 mmHg . This level correlated with data from cohort studies on the lowest level where accurate data are available on the association between SBP and cardiovascular disease.

## Standard deviation around the theoretical minimum

The choice of standard deviation around the theoretical minimum was based on examining the relationship of the standard deviation and mean of SBP using all available data obtained from the review of blood pressure surveys discussed in the following section (Figure 6.2). From this illustration, a distribution with a mean of 115 mmHg would typically

Figure 6.2 Association between mean SBP and standard deviation

have a standard deviation of 12 mmHg (baseline) which, when correcting for regression dilution bias, would be equivalent to 6 mmHg "usual".

## 2. Risk factor exposure

### 2.1 Data on blood pressure levels

Data on global blood pressure levels were collated from three major sources. The first source was data from the MONICA (Anonymous 1989b) and INTERSALT (Anonymous 1989a) studies. These studies collected blood pressure data from a variety of world regions. The MONICA study involved surveys at 39 collaborating centres in 22 countries carried out between 1979 and 1987 (Anonymous 1989b). The INTERSALT study involved surveys at 52 collaborating centres in 32 countries carried out between 1985 and 1987 (Anonymous 1989a). These two studies provide important information about blood pressure patterns, but do not provide a truly global overview of blood pressure
distributions. The MONICA study included populations that were predominantly European in origin, neither study included populations from the Eastern Mediterranean Region, and there were limited data from South-East Asia (other than India) and Africa. It was therefore necessary to include additional blood pressure data for comparative risk assessment (CRA) analyses.

The second major source of data were obtained through a literature search using Medline and the key words "blood pressure", "hypertension", "survey", "health survey" and "cross sectional survey". Studies were reviewed and included in analyses if they fulfilled the following criteria:

- conducted from 1980 onwards;
- included randomly-selected or representative participants;
- included a sample size of over 1000 in developed countries (a smaller sample size was acceptable in other countries if other criteria were fulfilled);
- described sample size and age group of participants;
- presented mean values of blood pressure by age and sex; and
- utilized a standard protocol for blood pressure measurement.

The final sources of data were personal communications with researchers and study investigators. The authors had access to data from APCSC-a collaboration involving 37 cohorts in the Asia-Pacific region. This collaboration includes prospective cohort studies undertaken in the Asia-Pacific region with at least 5000 person-years of follow-up recorded or planned, and data on date of birth or age, sex and blood pressure have been collated (APCSC 1999). Blood pressure data from eligible studies that had been collected from 1980 onwards were included in the SBP database. Data were also available from a previous global review of blood pressure levels for the last round of GBD, and these data were also included where the above criteria were fulfilled (P. Elliott, personal communication, 2001). Finally, authors of surveys/studies were contacted and age-specific and sex-specific data requested, where they had not been published in this format.

Many studies have not published data in the format required for this project (e.g. age-specific and sex-specific mean SBP level), and unfortunately time and resource constraints limited attempts to obtain all of these data from researchers. It was also very difficult to obtain results of surveys that have, for example, been published only in local/national reports but not in peer-reviewed journals. Access to these data would have been greatly improved if these reports were more widely available in electronic format, such as on the Internet. Details of studies currently included are presented in Table 6.2.
Table 6.2 Studies currently included in blood pressure data review

| Subregion | Country or area | Study (reference) | Year published | Sample size | Age range (years) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | Gambia | van der Sande et al. (1997) | 1997 | 6041 | $16-\geq 76$ |
|  | Ghana | Nyarko et al. (1994) | 1994 | 79 | 20-50 |
|  | Liberia | Giles et al. (1994) | 1994 | 3588 | 20->55 |
|  | Mauritius | Nan et al. (1991) | 1991 | 4905 | 25-74 |
|  | Nigeria | Akinkug (P. Elliot, personal communication, 2001) | 1980-1994 | 1411 | 30-260 |
|  | Nigeria | Bunker et al. (1992) | 1992 | 559 | 25-54 |
|  | Nigeria | Idahosa (1985) | 1985 | 1450 | 15->70 |
|  | Nigeria | Idahosa (1987) | 1983 | 1225 | 20-64 |
|  | Nigeria | Idahosa and llawole (1984) | 1984 | 585 | 18-54 |
|  | Nigeria | Ogunlesi et al. (1991) | 1991 | 404 | 18-54 |
|  | Nigeria | Okesina et al. (1999) | 1999 | 500 | $11-\geq 50$ |
|  | Nigeria | Oviasu and Okupa (1980a, 1980b) | 1980 | 1263 | $30-\geq 60$ |
|  | Senegal | Astagneau et al. (1992) | 1992 | 2300 | 15-86 |
|  | Senegal | Lang et al. (1988) | 1988 | 1862 | 15->54 |
|  | Seychelles | Bovet et al. (1991) | 1991 | 1081 | 25-64 |
|  | Sierra Leone | Lisk et al. (1999) | 1999 | 1204 | 15->75 |
|  |  |  |  | 28457 |  |
| AFR-E | Democratic Republic of the Congo | M'Buyamba-Kabangu et al. (1986, 1987) | 1987 | 987 | 10-60 |
|  | Ethiopia | Pauletto et al. (1994) | 1994 | 4869 | 15->65 |
|  | Kenya | INTERSALT study (Anonymous 1989a) | 1988 | 176 | 20-59 |
|  | Kenya | Poulter et al. (1984) | 1984 | 1737 | 17-265 |
|  | Malawi | Simmons et al. (1986) | 1983 | 988 | 15-65 |
|  | South Africa | CORIS study (P. Elliot, personal communication, 2001) | 1980-1994 | 5260 | $30-\geq 60$ |
|  | South Africa | Mollen (P. Elliot, personal communication, 2001) | 1980-1984 | 1429 | 30->60 |
|  | South Africa | Morbid (P. Elliot, personal communication, 2001) | 1980-1994 | 779 | $30-\geq 60$ |
|  | South Africa | Seedat (P. Elliot, personal communication, 2001) | 1980-1994 | 2800 | $30-\geq 60$ |

Table 6.2 Studies currently included in blood pressure data review (continued)

| Subregion | Country or area | Study (reference) | Year published | Sample size | Age range (years) |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | South Africa | Seedat et al. (1982) | 1982 | 1982 | <20-70 |
|  | South Africa | Steyn et al. (1985) | 1985 | 976 | 15-64 |
|  | South Africa | Steyn et al. (1996) | 1996 | 986 | 15-64 |
|  | United Republic of Tanzania | Edwards et al. (2000) | 2000 | 1695 | $15-\geq 55$ |
|  | United Republic of Tanzania | Kitange et al. (1993) | 1993 | 1670 | 15-19 |
|  | United Republic of Tanzania | Swai et al. (1993) | 1993 | 7272 | $15-\geq 65$ |
|  | Zimbabwe | Allain and Matenga (T. Allain, personal communication, 2001) | 1997 | 270 | $60-\geq 80$ |
|  | Zimbabwe | Hunter et al. (2000) | 2000 | 515 | $15-\geq 45$ |
|  | Zimbabwe | INTERSALT study (Anonymous 1989a) | 1988 | 195 | 20-59 |
|  | Zimbabwe | Mufunda et al. (2000) | 2000 | 775 | $25-\geq 55$ |
|  |  |  |  | 35361 |  |
| AMR-A | Canada | Joffres et al. (1992) | 1992 | 20582 | 18-74 |
|  | Halifax | MONICA study (Anonymous 1989b) | 1989 | 857 | 25-64 |
|  | Labrador | INTERSALT study (Anonymous 1989a) | 1988 | 161 | 20-59 |
|  | St Johns | INTERSALT study (Anonymous 1989a) | 1988 | 200 | 20-59 |
|  | USA | Sprafka et al. (1990) | 1990 | 4641 | 25-74 |
|  | USA | Hutchinson et al. (1997) | 1997 | 15743 | 45-64 |
|  | USA NHANES | Burt et al. (1995); NHANES web site (NCHS 2002) | 1995 | 13037 | 18-74 |
|  | Chicago | INTERSALT study (Anonymous 1989a) | 1988 | 196 | 20-59 |
|  | Goodman | INTERSALT study (Anonymous 1989a) | 1988 | 186 | 20-59 |
|  | Goodman | INTERSALT study (Anonymous 1989a) | 1988 | 198 | 20-59 |
|  | Hawaii | INTERSALT study (Anonymous 1989a) | 1988 | 187 | 20-59 |
|  | Jackson | INTERSALT study (Anonymous 1989a) | 1988 | 184 | 20-59 |
|  | Jackson | INTERSALT study (Anonymous 1989a) | 1988 | 199 | 20-59 |
|  | Stanford | MONICA study (Anonymous 1989b) | 1989 | 1504 | 25-64 |
|  |  |  | 57875 |  |  |


| AMR-B | Argentina | Delena (P. Elliot, personal communication, 2001) | 1980-1994 | 891 | 30-260 |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Argentina | INTERSALT study (Anonymous 1989a) | 1988 | 200 | 20-59 |
|  | Bahamas | Nassau (P. Elliot, personal communication, 2001) | 1980-1994 | 1673 | 30-260 |
|  | Barbados | Foster et al. (1993) | 1980-1993 | 464 | 30-260 |
|  | Barbados | Ischib (P. Elliot, personal communication, 2001) | 1980-1994 | 694 | 30-260 |
|  | Belize | Simmons et al. (1983) | 1983 | 1268 | 20-69 |
|  | Brazil | Costa et al. (1990); P. Elliot, personal communication, 2001 | 1990 | 2897 | 30-260 |
|  | Brazil | P. Elliot, personal communication, 2001; Ribeiro and Ribeiro (1986); Ribeiro et al. (198I) | 1981-1986 | 3114 | 30-260 |
|  | Brazil | Fuchs et al. (2001); F. D. Fuchs, personal communication, 2001 | 2001 | 1174 | 18-260 |
|  | Xingu | INTERSALT study (Anonymous 1989a) | 1988 | 198 | 20-59 |
|  | Yanomamo | INTERSALT study (Anonymous 1989a) | 1988 | 195 | 20-59 |
|  | Chile | Barrios (P. Elliot, personal communication, 2001) | 1980-1994 | 888 | 30-260 |
|  | Chile | L. Jadue, personal communication, 2001; Jadue et al. (1999) | 1999 | 3113 | 25-64 |
|  | Colombia | INTERSALT study (Anonymous 1989a) | 1988 | 191 | 20-59 |
|  | Jamaica | Miall (P. Elliot, personal communication, 2001) | 1980-1994 | 1974 | 30- 260 |
|  | Jamaica | R. Wilks et al. personal communication, 2001 | 1980 onwards | 2085 | 25-74 |
|  | Mexico | Gonzalez-Villalpando et al. (I999); C. Gonzalez, personal communication, 2001 | 1999 | 2281 | 35-64 |
|  | Mexico | INTERSALT study (Anonymous 1989a) | 1988 | 172 | 20-59 |
|  | Mexico | Rosenthal et al. (1989) | 1989 | 645 | 19-25 |
|  | Mexico | L. Yamamoto, personal communication, 2001 | 1996 | 825 | 20-90 |
|  | Paraguay | Ramirez et al. (1995) | 1995 | 9585 | 18-74 |
|  | Saint Lucia | Khaw and Rose (1982) | 1982 | 359 | 15-265 |
|  | Trinidad and Tobago | INTERSALT study (Anonymous 1989a) | 1988 | 176 | 20-59 |
|  | Trinidad and Tobago | Tobago (P. Elliot, personal communication, 2001) | 1980-1994 | 1284 | 30-260 |
|  | Uruguay | Latir (P. Elliot, personal communication, 2001) | 1980-1994 | 512 | $30-260$ |
|  |  |  |  | 36858 |  |

Table 6.2 Studies currently included in blood pressure data review (continued)

| Subregion | Country or area | Study (reference) | Year published | Sample size | Age range (years) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AMR-D | Ecuador | Cornejo et al. (2000) | 2000 | 10605 | $19-\geq 98$ |
|  |  |  |  | 10605 |  |
| EMR-B | Iran (Islamic Republic of) | SarrafZadegan and AminiNik (1997) | 1997 | 6532 | 19-70 |
|  | Jordan | H. Jaddou, personal communication, 2001 | 2000 | 2276 | $25-\geq 70$ |
|  | Saudi Arabia | Khalid et al. (1994) | 1994 | 1093 | 10-76 |
|  | Saudi Arabia | Soyannwo et al. (1998) | 1998 | 5305 | $0-\geq 70$ |
|  | Tunisia | Ghannem and Hadj-Fredj (1997); H. Ghannem, personal communication, 200I | 1997 | 957 | $20-\geq 70$ |
|  | Tunisia | Ghannem et al. (2001); H. Ghannem, personal communication, 2001 | 2001 | 1569 | 13-19 |
|  |  |  |  | 17732 |  |
| EMR-D | Egypt | Ashour et al. (1995); Ibrahim et al. (1995) | 1995 | 6600 | $25-\geq 70$ |
|  |  |  |  | 6600 |  |
| EUR-A | Belgium |  |  |  |  |
|  | Charleroi | INTERSALT study (Anonymous 1989a) | 1988 | 157 | 20-59 |
|  | Charleroi | MONICA study (Anonymous 1989b) | 1989 | 678 | 25-64 |
|  | Ghent | INTERSALT study (Anonymous 1989a) | 1988 | 200 | 20-59 |
|  | Ghent | MONICA study (Anonymous 1989b) | 1989 | 1021 | 25-64 |
|  | Belgium; Luxembourg | MONICA study (Anonymous 1989b) | 1989 | 1934 | 25-64 |
|  | Former Czechoslovakia | Czech (P. Elliot, personal communication, 2001) | 1980-1994 | 2345 | $30-\geq 60$ |
|  | Denmark | Andersen et al. (1994) | 1994 | 6009 | 17 |
|  | Denmark | INTERSALT study (Anonymous 1989a) | 1988 | 199 | 20-59 |
|  | Glostrup | MONICA study (Anonymous 1989b) | 1989 | 3785 | 25-64 |




## 1980-1994


응ㅇㅇㅇㅇㅇㅇ응


FINNIII (P. Elliot, personal communication, 200I) Puska et al. (I993)


INTERSALT study (Anonymous 1989a) MONICA study (Anonymous 1989b) INTERSALT study (Anonymous 1989a) MONICA study (Anonymous 1989b) MONICA study (Anonymous 1989b)
MONICA study (Anonymous 1989b) INTERSALT study (Anonymous 1989a) MONICA study (Anonymous 1989b) O
 Heinemann et al. (1995) Hoffmeister et al. (1994) MONICA study (Anonymous 1989b) MONICA study (Anonymous 1989b) MONICA study (Anonymous 1989b) INTERSALT study (Anonymous 1989a)


 (e686I snowkuouv) Kpnas $17 \forall S$ Y $\exists \perp \mathrm{NI}$ MONICA study (Anonymous 1989b) Fogari et al. (1997)
 Vaccarino et al. (1995)
Finland
Finland
Finland
Joensuu
Kuopio
Turku
Turku; Loimaa
N. Karelia
France
Bas-Rhin
Lille
Former German Berlin-Lichtenberg
Cottbus county Cottbus county
Halle county Karl-Marx-Stadt Germany Germany
Augsburg (Rural) Augsburg (Urban) Bernried
Bremen
Heidelberg Rhein-Neckar Iceland Iceland 츺 त 츤
Studies currently included in blood pressure data review (continued)

|  |  |  |
| :--- | :--- | :--- | :--- |
| Subregion | Study (reference) | Age range |
| (years) |  |  |


|  | United Kingdom | Mann et al. (1988) | 1988 | 12092 | 25-59 |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | England, | Bajekal et al. (1999) | 1999 | 12555 | $16-275$ |
|  | Birmingham | INTERSALT study (Anonymous 1989a) | 1988 | 200 | 20-59 |
|  | Northern Ireland, |  |  |  |  |
|  | Belfast | INTERSALT study (Anonymous 1989a) | 1988 | 199 | 20-59 |
|  | Belfast | MONICA study (Anonymous 1989b) | 1989 | 2358 | 25-64 |
|  | Scotland | Hawth (P. Elliot, personal communication, 2001) | 1980-1994 | 2286 | 30-260 |
|  | Scotland, | Smith et al. (1989) | 1989 | 10359 | 40-59 |
|  | Glasgow | MONICA study (Anonymous 1989b) | 1989 | 1247 | 25-64 |
|  | Wales, |  |  |  |  |
|  | South Wales | INTERSALT study (Anonymous 1989a) | 1988 | 199 | 20-59 |
|  | Ireland | MacAuley et al. (1996) | 1996 | 603 | 16-74 |
|  | Ireland | Shelley et al. (1991, 1995) | 1991-1995 | 1433 | 30-260 |
|  |  |  |  | 191304 |  |
| EUR-B | Poland | Davis et al. (1994) | 1994 | 729 | 46-52 |
|  | Krakow | INTERSALT study (Anonymous 1989a) | 1988 | 200 | 20-59 |
|  | Warsaw | INTERSALT study (Anonymous 1989a) | 1988 | 200 | 20-59 |
|  | Warsaw | MONICA study (Anonymous 1989b) | 1989 | 2646 | 25-64 |
|  | Uzbekistan | King et al. (1998) | 1998 | 1956 | 35- 265 |
|  | Former Yugoslavia | MONICA study (Anonymous 1989b) | 1989 | 1575 | 25-64 |
|  |  |  |  | 7306 |  |
| EUR-C | Hungary | INTERSALT study (Anonymous 1989a) | 1988 | 200 | 20-59 |
|  | Budapest | MONICA study (Anonymous 1989b) | 1989 | 1512 | 25-64 |
|  | Lithuania | Bosma et al. (1994) | 1994 | 2149 | 45-70 |
|  | Russian Federation | Puska et al. (1993) | 1993 | 837 | 25-64 |
|  | The former Soviet Union ${ }^{\text {a }}$ | INTERSALT study (Anonymous 1989a) | 1988 | 194 | 20-59 |
|  | Kaunas | MONICA study (Anonymous 1989b) | 1988 | 1462 | 25-64 |
|  | Moscow | MONICA study (Anonymous 1989b) | 1989 | 1425 | 25-64 |
|  | Moscow | MONICA study (Anonymous 1989b) | 1989 | 2404 | 25-64 |
|  | Novosibirsk | MONICA study (Anonymous 1989b) | 1989 | 2820 | 25-64 |
|  | Novosibirsk | MONICA study (Anonymous 1989b) | 1989 | 1623 | 25-64 |
|  |  |  |  | 14626 |  |

Studies currently included in blood pressure data review (continued)

| Subregion | Country or area | Study (reference) | Year published | Sample size | Age range (years) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| SEAR-B | Sri Lanka | Mendis et al. (1988) | 1988 | 9405 | 20-84 |
|  | Sri Lanka | Mohidn (P. Elliot, personal communication, 2001) | 1980-1994 | 704 | $30-\geq 60$ |
|  |  |  | 10109 |  |  |
| SEAR-D | India | B.V. Babu, personal communication, 200I; Y.S. Kusuma, personal communication, 2001 | 1993 onwards | 295 | $15-\geq 55$ |
|  | India | B.V. Babu, personal communication, 200 I; Y.S. Kusuma, personal communication, 2001 | 2001 | 1094 | $20-\geq 60$ |
|  | India | Gilberts et al. (1994) | 1994 | 10260 | $20-\geq 70$ |
|  | India | Gupta et al. (1995) | 1995 | 2212 | $20-\geq 70$ |
|  | India | India (P. Elliot, personal communication, 2001) | 1980-1994 | 2078 | $30-\geq 60$ |
|  | India | Misra et al. (2001); A. Misra, personal communication, 2001 | 2001 | 679 | 14-25 |
|  | India | Singh et al. (1997a) | 1997 | 1806 | 25-64 |
|  | India | Singh et al. (1997b) | 1997 | 1935 | $25-\geq 64$ |
|  | India | Singh et al. (1998) | 1997 | 2310 | 25-64 |
|  | India | Reddy (P. Elliot, personal communication, 2001) | 1980-1994 | 5574 | $30-\geq 60$ |
|  | Ladakh | INTERSALT study (Anonymous 1989a) | 1988 | 200 | 20-59 |
|  | New Delhi | INTERSALT study (Anonymous 1989a) | 1988 | 199 | 20-59 |
|  |  |  | 19408 |  |  |
| WPR-A | Australia | APCSC-Busselton (APCSC secretariat, personal communication, 200I) | 1990 onwards | 976 | $15-\geq 70$ |
|  | Australia | APCSC-Perth (APCSC secretariat, personal communication, 2001) | 1990 onwards | 6456 | $15-\geq 70$ |
|  | Australia | Gliksman et al. (1990) | 1990 | 2284 | 9-15 |
|  | Australia | Jamrozik and Hockey (1989); P. Elliot, personal | 1989 | 9222 | $30-\geq 60$ |


Table 6.2 Studies currently included in blood pressure data review (continued)

| Subregion | Country or area | Study (reference) | Year published | Sample size | Age range (years) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| WPR-B | China | APCSC-Anzhen (APCSC secretariat, personal communication, 2001) | 1980 onwards | 8378 | $30-\geq 70$ |
|  | China | APCSC-Anzhen02 (APCSC secretariat, personal communication, 2001) | 1980 onwards | 4152 | 30-69 |
|  | China | APCSC-East Beijing (APCSC secretariat, personal communication, 2001) | 1980 onwards | 1198 | $5-\geq 70$ |
|  | China | APCSC-Hong Kong SAR (APCSC secretariat, personal communication, 2001) | 1980 onwards | 998 | $60-\geq 70$ |
|  | China | APCSC-Hong Kong SAR (APCSC secretariat, personal communication, 2001) | 1980 onwards | 2019 | $\geq 70$ |
|  | China | APCSC-Seven Cities cohort (APCSC secretariat, personal communication, 2001) | 1980 onwards | 37635 | $15-\geq 70$ |
|  | China | APCSC-Six Cohorts (APCSC secretariat, personal communication, 2001) | 1980 onwards | 19387 | 30-59 |
|  | China | APCSC-Tianjin (APCSC secretariat, personal communication, 2001) | 1980 onwards | 9335 | $15-\geq 70$ |
|  | China | He et al. (1991a); Klag et al. (1993) | 1991-1993 | 14055 | $15-\geq 65$ |
|  | China | PRC-USA study (Huang et al. 1994; People's Republic of China-United States Cardiovascular and Cardiopulmonary Epidemiology Research Group 1992) | 1992-1994 | 8885 | 35-54 |


| 1988 | 200 | $20-59$ |
| :--- | ---: | :---: |
| 1989 | 1672 | $25-64$ |
| 1988 | 200 | $20-59$ |
| 1989 | 200 | $20-59$ |
| 1980 onwards | 7004 | $0-\geq 70$ |
| 1988 | 181 | $20-59$ |
| 1988 | 198 | $20-59$ |
| 1994 | 21240 | $30-\geq 70$ |
| 1988 | 162 | $20-59$ |
| 1985 | 1455 | $20-\geq 55$ |
| 1994 | 203 | $20-86$ |
| $1980-1994$ | 7454 | $30-\geq 60$ |
| 1983 | 1039 | $20-60$ |
| $1980-1994$ | 253 | $30-\geq 60$ |
| $1980-1994$ | 249 | $30-\geq 60$ |
|  |  |  |


Russia in the original publication.

Figure 6.3 SBP data coverage expressed as total study sample size, by geographical region


Figure 6.3 illustrates data coverage by geographical region. Data from about 230 studies (total sample size of over 660000 participants) have been included.

The subregions with the most data-studies including a total of more than 50000 participants-were AMR-A, EUR-A, WPR-A and WPR-B. All other subregions had data on a total of $>2000$ participants.

Approximately half of the studies from which data were obtained utilized random sampling of individuals or households (54\%) -including stratified random sampling (see Appendix A). Forty-six per cent of studies obtained samples by other methods such as house-to-house or workplace surveys. Response rates were documented in half of the studies. Of those studies that did provide these data, the response rate was $>80 \%$ in $42 \%$ of studies, between $50-80 \%$ in $54 \%$ of studies, and documented to be $<50 \%$ in only five studies. For completeness, full documentation of sampling method, response rate and blood pressure measuring techniques is presented in Appendix A. A summary is given in Table 6.3.

### 2.2 Methodology to estimate mean and standard deviation of blood pressure

Study summary data were taken from all data sources (i.e. summary statistics and not individual participant data were available for analysis). Mean SBP, standard deviation, sample size and age ranges from all data sources were extracted and entered into an Excel database. Scatter plots of these SBP data for each subregion, utilizing the midpoints in study age categories, are presented in Figures 6.4-6.7.

Table 6.3 Blood pressure measuring techniques of studies included in review

|  | Yes (\%) | No (\%) | Not stated (\%) |
| :---: | :---: | :---: | :---: |
| Trained and certified staff | 95 | NA | 5 |
| Sphygmomanometer |  |  | 3 |
| Standard/mercury | 92 | NA | NA |
| Electronic | 5 | NA | NA |
| Multiple cuff sizes used | 49 | 28 | 23 |
| 2-3 BP measures taken after $\geq 5$ mins rest | 92 | 2 | 6 |

NA Not applicable.

The SBP data obtained from the literature came from studies that included different age ranges and were presented broken down by different age categories from that required for the CRA project. For this reason a method was needed that made complete use of all the available data. First, exploratory data analysis techniques were utilized to assess the general shape of association as well as to check for data errors. Secondly, models were fit to the data and country-level blood pressure estimates were made. Finally, subregional estimates were obtained by pooling across the country-level estimates. This approach is explained in greater detail below.

## SBP-AGE RELATIONSHIPS

The data from the surveys were used to assess the shape of the association of SBP with age for all data combined and for each subregion and sex separately using Splus software. At this stage no assumptions were made about the shape of association, and therefore non-parametric methods were applied (i.e. the generalized additive model). When the shape of the association was examined using all of the blood pressure data, there appeared to be an approximately linear association from the age of about $30-70$ years in males and females (Figure 6.8). When the analyses were repeated by subregion, the association increased approximately linearly from age 30 years for females ( 20 years for males) until approximately 70 years in all subregions, despite variation in the slope and overall starting SBP level.

In many subregions, SBP levels appeared to flatten for those aged $>70$ years, although there was some variation between the subregions in the point at which the association levelled off. In a few subregions, particularly the "A" subregions, there was no obvious levelling out of SBP in the 70-79 year age group; however, data for those aged $\geq 80$ years were particularly scarce, so a flattening out of SBP is still possible in these subregions. The shape of the association in those aged $>70$ years within each subregion was influenced by the fact that there were considerably less

Figure 6.4 Mean SBP-age relationship for input data by subregion, for females


Figure 6.5 Mean SBP-age relationship for input data by subregion, for males


Figure 6.6 The relationship of standard deviation of mean SBP with age by subregion, for females


Figure 6.7 The relationship of standard deviation of mean SBP with age by subregion, for males


Figure 6.8 The SBP-age association for all studies combined

survey data available for females and males aged $>70$ years. The literature clearly indicates that SBP tends to rise linearly until the eighth or ninth decade (Franklin et al. 1997; Kannel 1996, 1999; Whelton 1994; Whelton et al. 1994), after which it levels out (Franklin et al. 1997; Kannel 1996, 1999; Whelton 1994; Whelton et al. 1994). Therefore, it was assumed that in all subregions SBP levels increased linearly to 70 years of age in both males and females, after which (i.e. the $\geq 70$ year groups) levels flattened out. This approach was appropriate given the limitations of the data for those aged $>70$ years, and the information from the literature indicating that this pattern is widespread. It was also more conservative than assuming that the levels of SBP continue to increase with age.

The trend with age in industrialized subregions such as AMR-A, EURA, WPR-A corresponds with evidence in the literature, which shows a consistency of relationship between age and blood pressure in surveys in industrialized countries (Kannel 1999; Whelton et al. 1994). The literature suggests that between the third, fourth and seventh decades the rate of rise in SBP is approximately $0.6-0.8 \mathrm{mmHg} /$ year for women and approximately $0.3-0.5 \mathrm{mmHg} /$ year for men in industrialized populations
(Whelton et al. 1994). By the seventh decade women tend to have levels of SBP equal or exceeding those in men (Whelton et al. 1994). These findings concur with those for industrialized subregions in the current analyses. There were similar patterns in most developing countries, that is, an age-related rise in blood pressure which is consistently higher for systolic than DBP. SBP tends to rise linearly, and the age-related rise in blood pressure during adulthood is steeper for women than men (Kannel 1999; Whelton et al. 1994). However, the size of the age-related increase in blood pressure varies by subregion.

## Country-level estimates

After gathering all country-level studies, we obtained country-level estimates for countries where data were available. There were studies available that were nationally representative and these were used solely to derive the respective country-level estimates. In these cases any additional studies from those countries were not utilized as they only contributed additional heterogeneity to the best estimates obtained from the nationally representative data. There were eight countries where this was possible: Australia, Canada, Egypt, Germany, Japan, Paraguay, the United Kingdom of Great Britain and Northern Ireland and the United States of America. There were sufficient data to assess SBP in 61 additional countries and for these a regression approach was required so that all remaining study-level data could be described as well as extrapolated within countries where data did not cover all age groups. The following model was applied to both males and females:

$$
Y_{i j}=\beta_{1 i} \text { subregion }+\beta_{2 i} \text { age }+\beta_{3 i} \text { subregion } \times \text { age }+\beta_{4 j} \text { country }
$$

This equation expresses how the study summary data were modelled to estimate blood pressure levels for each country separately. SBP was regressed against several explanatory covariates: age was included as a continuous variable, linear between 30 and 70 as described in the previous section. Subregion was categorical and the "subregion $x$ age" interaction term included allowing for differing slopes of SBP with age between subregions. Country was included as a dummy variable and in this way the resulting beta coefficients can be used to assess the intercountry variation after controlling for age and subregional differences. The $i$ and $j$ subscripts denote the $i$ th subregion and $j$ th country within each subregion. Finally, the model was weighted by the sample size collected within each age group in each study to make the mean blood pressure estimates.

The modelling process is illustrated in Figure 6.9 where study level data within EUR-A were available for 18 of the 26 countries. The figure also shows, for the purpose of illustration only, the overall relationship of blood pressure with age in this subregion. It is clear from this figure that there is significant variation around the overall relationship with age

Figure 6.9 SBP data from studies in EUR-A

and also that a great deal of this variation is due to the differing patterns of blood pressure between countries. This occurred in most of the other subregions, especially where there were a large number of countries. In this example, each of the 18 countries within EUR-A was estimated to have the same overall slope within this subregion but at differing absolute levels.

Mean SBP was estimated for each country for each of the five relevant GBD age groups ( $>30$ years) by using the fitted regression model to predict the best estimates within each country at the midpoints of each age group. A similar approach was used to estimate the standard deviations for each age, sex and country combination.

## Subregional level estimates

Estimates for each of the 14 subregions were made by weighting across the country-level estimates by their respective population sizes. For example, the following estimates for females were made for countries within WPR-A (Table 6.4). The weight given to each country within this subregion is equal to the percentage of total subregional population for each of the three respective countries.

Table 6.4 Estimates of age-specific mean SBP levels ( mmHg ) for females in WPR-A by country and pooled for the whole subregion

|  | Age group (years) |  |  |  |
| :--- | ---: | ---: | ---: | ---: |
| Country | $30-44$ | $45-59$ | $60-69$ | $70-79$ |
| Australia/New Zealand | 116.7 | 127.3 | 136 | 139.5 |
| Percentage of subregion | $14 \%$ | $11 \%$ | $8 \%$ | $9 \%$ |
| Japan | 121.0 | 133.7 | 144.3 | 148.6 |
| Percentage of subregion | $79 \%$ | $85 \%$ | $88 \%$ | $88 \%$ |
| Singapore | 113.9 | 126.1 | 136.3 | 140.3 |
| Percentage of subregion | $4 \%$ | $2 \%$ | $1 \%$ | $1 \%$ |
| WPR-A total | 120 | 133 | 144 | 148 |

Table 6.5 Estimates of mean SBP levels $(\mathrm{mmHg})$ by subregion, sex and age

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  |  | Males |  |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | 123 | 136 | 146 | 150 | 150 | 127 | 135 | 141 | 144 | 144 |
| AFR-E | 121 | 131 | 140 | 143 | 143 | 124 | 130 | 135 | 137 | 137 |
| AMR-A | 114 | 127 | 138 | 143 | 149 | 122 | 129 | 137 | 141 | 144 |
| AMR-B | 115 | 130 | 142 | 147 | 147 | 122 | 131 | 138 | 141 | 141 |
| AMR-D | 117 | 129 | 139 | 143 | 143 | 123 | 131 | 136 | 139 | 139 |
| EMR-B | 126 | 137 | 147 | 150 | 150 | 125 | 132 | 139 | 143 | 143 |
| EMR-D | 121 | 135 | 146 | 150 | 150 | 123 | 131 | 138 | 141 | 141 |
| EUR-A | 122 | 136 | 147 | 151 | 151 | 130 | 138 | 144 | 147 | 147 |
| EUR-B | 122 | 141 | 154 | 161 | 161 | 128 | 140 | 149 | 153 | 153 |
| EUR-C | 125 | 143 | 158 | 164 | 164 | 129 | 138 | 146 | 149 | 149 |
| SEAR-B | 120 | 130 | 139 | 142 | 142 | 122 | 131 | 138 | 141 | 141 |
| SEAR-D | 117 | 126 | 132 | 135 | 135 | 118 | 127 | 134 | 137 | 137 |
| WPR-A | 120 | 133 | 144 | 148 | 148 | 127 | 137 | 145 | 148 | 148 |
| WPR-B | 115 | 127 | 137 | 141 | 141 | 117 | 126 | 133 | 136 | 136 |

Country-level estimates were possible in 69 countries (with 61 of these derived from the model described above) for both sexes representing all 14 subregions. For most of the subregions there was reasonable data coverage based on the literature review. However, there were some subregions for which very few data were found and estimation at the country level proved difficult. Therefore, high level of coverage, as in the example above, was not possible for all of the subregions. The two South-East Asian subregions are an example of this: data on only one country in each of these two subregions were obtained (N.B. India was the country

Table 6.6 Estimates of SBP standard deviations by subregion, sex and age

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  |  | Males |  |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | 20 | 25 | 29 | 31 | 31 | 17 | 21 | 25 | 27 | 27 |
| AFR-E | 13 | 18 | 22 | 23 | 23 | 13 | 16 | 19 | 20 | 20 |
| AMR-A | 14 | 19 | 22 | 24 | 24 | 14 | 17 | 20 | 21 | 21 |
| AMR-B | 15 | 21 | 26 | 28 | 28 | 16 | 20 | 23 | 24 | 24 |
| AMR-D | 15 | 19 | 23 | 24 | 24 | 15 | 18 | 21 | 22 | 22 |
| EMR-B | 15 | 17 | 19 | 20 | 20 | 15 | 19 | 22 | 24 | 24 |
| EMR-D | 15 | 19 | 23 | 24 | 24 | 15 | 18 | 21 | 22 | 22 |
| EUR-A | 15 | 19 | 22 | 23 | 23 | 14 | 18 | 21 | 23 | 23 |
| EUR-B | 16 | 23 | 28 | 31 | 31 | 16 | 21 | 26 | 28 | 28 |
| EUR-C | 17 | 23 | 29 | 31 | 31 | 16 | 21 | 25 | 27 | 27 |
| SEAR-B | 15 | 20 | 23 | 25 | 25 | 13 | 18 | 22 | 24 | 24 |
| SEAR-D | 14 | 18 | 20 | 21 | 21 | 14 | 19 | 23 | 24 | 24 |
| WPR-A | 15 | 18 | 21 | 22 | 22 | 15 | 18 | 21 | 22 | 22 |
| WPR-B | 16 | 22 | 27 | 29 | 29 | 14 | 19 | 24 | 25 | 25 |

Note: Reported standard deviations are derived from baseline SBP (i.e. one-off measures), which when corrected for regression dilution bias were used in all calculations on the "usual" scale.
for which estimation was possible in SEAR-D). It was assumed that in these data-sparse subregions, any country-level estimates belonging to those subregions were representative of the respective overall subregional blood pressure levels. Estimates of SBP distribution are presented in Tables 6.5 and 6.6.

### 2.3 Methodology to estimate uncertainty around mean and standard deviation SBP

The level of uncertainty around the means and standard deviations also had to be estimated. The approach described in the previous section to estimate the means and standard deviations makes as complete use as possible of all available data. However, there will always be increased uncertainty associated when generalizing from the data collected. Assessing the uncertainty in SBP estimation involved considering three main sources:

1. countries with nationally representative survey data;
2. countries that had multiple studies that were heterogeneous; and
3. countries with little or no data.

First, for the eight countries where very reliable large-scale nationally representative data were available, the standard errors were taken
directly from those study data. This uncertainty reflects, almost entirely, sampling variation in those studies and was generally small.

Second, for each of the 61 countries for which data were modelled, standard errors were obtained from the model predictions made for each country, age and sex combination. These standard errors reflected the variability of the data at the study level within each country. Significant interstudy variation within a particular country resulted in a larger standard error for that country. Conversely, in cases where studies consistently provided similar blood pressure levels, standard errors were very small.

Finally, an additional type of uncertainty was incorporated to reflect that when pooling country-level estimates to the subregional level there were varying degrees of missing information between the subregions. The "A" subregions all had excellent coverage with greater than $95 \%$ of the populations in these subregions represented in the available study data. On the other hand, only data on Sri Lanka were available in SEAR-B with no study data available on all remaining countries in that subregion, providing only $6 \%$ coverage. To allow for this varying coverage, an additional factor was introduced that depended on the proportion of total subregional population covered. The regression modelling approach described earlier provided a way of summarizing how much variability between countries was unaccounted for by age, sex and subregional differences. For example, the country coefficients ranged approximately $20-25 \mathrm{mmHg}$ for females across all the countries included in the model. The degree of inter-country variation provided an indication of uncertainty in countries without data and therefore provided a basis from which to determine a suitable level of uncertainty within each subregion. One way to achieve this was to assume that the proportion of the subregion without any data had uncertainty equal to the intercountry variation. In effect, missing countries are assigned an uncertainty equivalent to choosing a country at random from the distribution of country effects (i.e. similar to a random effects approach).

The uncertainty for countries with complete, partial and no data, each obtained as described above, were then used in the relationship for pooled standard error of the mean weighted by population size. The following example illustrates how this was achieved for females aged 30-44 years in AMR-A.

```
\(\operatorname{Var}(\) Pooled mean SBP in AMR-A \()=\)
    \(w_{1}^{2} \operatorname{var}(\) mean SBP - USA \()+w_{2}^{2} \operatorname{var}(\) mean SBP - Canada \()+w_{3}^{2}\)
    var(mean SBP - Cuba)
```

where $w_{1}, w_{2}$ and $w_{3}$ are the weights equal to proportion of total subregional population in each of the three countries in AMR-A. In this example, variance estimates (i.e. square of the standard error) were derived directly from existing estimates for the United States and Canada, as there were data available for these two countries. There were no data available for Cuba and therefore the final term depended on all the other
Values of $95 \% \mathrm{Cl}$ around mean SBP $(\mathrm{mmHg})$ by subregion, sex and age

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  |  | Males |  |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 30-40 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | (119-128) | (131-141) | (141-152) | (144-157) | (144-157) | (123-132) | (130-140) | (136-146) | (138-150) | (138-150) |
| AFR-E | (116-125) | (127-136) | (135-144) | (137-148) | (137-148) | (121-128) | (126-134) | (130-140) | (132-142) | (132-142) |
| AMR-A | (113-115) | (126-128) | (137-139) | (141-144) | (148-150) | (122-123) | (128-130) | (136-138) | (140-142) | (142-145) |
| AMR-B | (112-118) | (127-132) | (139-145) | (144-150) | (144-150) | (120-124) | (129-133) | (136-140) | (139-144) | (139-144) |
| AMR-D | (110-125) | (121-137) | (130-148) | (133-152) | (133-152) | (116-130) | (123-138) | (128-145) | (129-148) | (129-148) |
| EMR-B | (122-129) | (134-140) | (143-150) | (146-154) | (146-154) | (122-128) | (129-135) | (136-143) | (139-146) | (139-146) |
| EMR-D | (114-128) | (127-142) | (138-154) | (142-159) | (142-159) | (116-130) | (124-139) | (130-147) | (132-150) | (132-150) |
| EUR-A | (121-124) | (134-137) | (145-148) | (150-153) | (150-153) | (129-131) | (137-139) | (143-146) | (145-148) | (145-148) |
| EUR-B | (116-128) | (135-147) | (148-161) | (153-168) | (153-168) | (122-134) | (134-147) | (142-156) | (145-160) | (145-160) |
| EUR-C | (121-128) | (139-146) | (154-162) | (159-169) | (159-169) | (126-132) | (135-142) | (142-151) | (145-154) | (145-154) |
| SEAR-B | (112-128) | (121-139) | (129-149) | (131-153) | (131-153) | (113-130) | (121-140) | (128-148) | (130-151) | (130-151) |
| SEAR-D | (115-119) | (123-128) | (129-135) | (131-139) | (131-139) | (116-121) | (125-129) | (131-138) | (134-14I) | (134-141) |
| WPR-A | (119-121) | (132-134) | (142-145) | (146-150) | (146-150) | (126-128) | (136-137) | (144-145) | (147-149) | (147-149) |
| WPR-B | (113-116) | (126-128) | (135-139) | (139-143) | (139-143) | (116-118) | (125-127) | (131-135) | (134-138) | (134-138) |

Table 6.8 Values of $95 \% \mathrm{Cl}$ around the standard deviation of mean SBP by subregion, sex and age

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  |  | Males |  |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 30-40 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | (18-22) | (22-29) | (23-37) | (24-40) | (24-40) | (15-18) | (20-24) | (23-30) | (24-32) | (24-32) |
| AFR-E | (11-18) | (14-22) | (17-29) | (17-32) | (17-32) | (10-16) | (13-20) | (13-25) | (14-27) | (14-27) |
| AMR-A | (12-17) | (17-2\|) | (20-26) | (21-28) | (21-28) | (12-15) | (16-19) | (18-23) | (18-25) | (18-25) |
| AMR-B | (12-17) | (18-2\|) | (22-27) | (23-30) | (23-30) | (14-18) | (17-20) | (19-24) | (20-26) | (20-26) |
| AMR-D | (12-19) | (15-23) | (18-30) | (18-33) | (18-33) | (12-18) | (15-22) | (15-27) | (16-29) | (16-29) |
| EMR-B | (12-17) | (12-22) | (11-26) | (10-28) | (10-28) | (14-16) | (17-21) | (20-25) | (21-27) | (21-27) |
| EMR-D | (12-19) | (15-23) | (18-30) | (18-33) | (18-33) | (12-18) | (15-22) | (15-27) | (16-29) | (16-29) |
| EUR-A | (14-15) | (18-20) | (21-23) | (22-25) | (22-25) | (14-15) | (18-19) | (20-23) | (22-24) | (22-24) |
| EUR-B | (13-19) | (20-25) | (23-34) | (24-38) | (24-38) | (13-18) | (18-24) | (20-30) | (21-34) | (21-34) |
| EUR-C | (15-18) | (22-25) | (27-32) | (29-34) | (29-34) | (15-17) | (20-23) | (23-27) | (25-29) | (25-29) |
| SEAR-B | (12-18) | (17-23) | (20-28) | (21-30) | (21-30) | (11-16) | (15-21) | (17-26) | (18-29) | (18-29) |
| SEAR-D | (13-16) | (16-19) | (18-23) | (19-25) | (19-25) | (13-16) | (17-21) | (19-26) | (20-28) | (20-28) |
| WPR-A | (14-16) | (17-19) | (19-23) | (20-24) | (20-24) | (14-16) | (17-19) | (19-22) | (20-23) | (20-23) |
| WPR-B | (14-17) | (20-23) | (24-29) | (25-32) | (25-32) | (13-15) | (18-20) | (22-26) | (23-28) | (23-28) |

countries in the regression model and the variance in this case was equal to 20 mmHg . For subregions with low population coverage, this approach would result in much larger uncertainty. Having calculated the overall variance of the mean for a subregion, it was then converted into a $95 \%$ CI. A similar approach was taken to estimate the uncertainty for blood pressure standard deviations. Results are shown in Tables 6.7 and 6.8.

## 3. Blood pressure-disease relationships

Data on the relationship between blood pressure and disease outcomes come from two main types of study (MacMahon 1994). Prospective observational studies provide data from which the effects of prolonged blood pressure differences can be estimated (MacMahon et al. 1990). Trials provide data about the effects of short-term blood pressure reduction (Collins et al. 1990) or risk reversal. Results from prospective observational studies are discussed in this section.

### 3.1 Risk accumulation

A large number of observational studies have been conducted in a variety of settings that examine the association between blood pressure and disease. However, the results of many of these individual studies are limited. Many observational studies had small sample sizes, or an insufficient number of end-points and therefore lacked the power required to provide reliable estimates of associations for different population subgroups (e.g. sex and age groups) and/or specific diagnostic categories (MacMahon et al. 1990). Individual studies have not always provided information on the direction of the association at lower blood pressure levels, making it difficult to assess whether the observed association is continuous or has a threshold level. In addition, these studies frequently do not standardize the size of the association for bias and confoundingin particular regression dilution bias (MacMahon et al. 1990). This bias occurs when associations are calculated from "one-off" measures of blood pressure (i.e. measures taken once on one occasion) rather than "usual" blood pressure. (There is further discussion of this bias in the next section.) Overviews of observational studies, or meta-analyses of trials, overcome many of the size limitations of individual studies. At the time of writing, four major overviews have been conducted, and their main design features are summarized in Table 6.9. In contrast to the first three overviews, which utilized tabular data, the APCSC (1999) involved analyses of individual participant data. Therefore, it could more reliably adjust for confounding and provide more reliable risk estimates.

Of the four overviews, all included the stroke end-point, but only two included analyses for the IHD end-point (MacMahon et al. 1990; APCSC 2003a). Only APCSC included analyses for hypertensive disease and other cardiovascular disease (APCSC 2003a). Further advantages with this study are that all analyses were based on individual participant
Table 6.9 Characteristics of major cohort study overviews

| Study characteristics | MacMahon et al. (1990) | Prospective Studies Collaboration (PSC 1995) | Eastern Collaborative Research Group (Anonymous 1998) | Asia-Pacific Cohort Studies Collaboration (APCSC 2003a) |
| :---: | :---: | :---: | :---: | :---: |
| Aims | To determine the strength of the association between BP and stroke and IHD throughout BP range | To assess the relationship between BP and total blood cholesterol and stroke, and determine how the strength of the relationship between BP and stroke varied with age | To assess the relationship between BP and total blood cholesterol and stroke in Asian populations, and determine whether the strength of the relationship varied with type of stroke | To produce subregion, age- and sex-specific blood pressure associations for stroke (including subtypes), IHD and total cardiovascular disease |
| Number of cohort studies included | 9 | 45 | 18 | 37 |
| Population included | Mostly from USA, but also from Europe, Puerto Rico and Hawaii (largely Caucasian) | Asia, Australia, Europe, Hawaii, the Middle East, USA | China (13 cohorts), Japan (5) | Australia (4), mainland China (14), Hong Kong SAR (I), Japan (I2), New Zealand (I), Republic of Korea (1), Singapore (2), Taiwan, China (2) |
| Number of participants | 420000 | 450000 | 124774 | 425251 |
| \% male | 96\% | 61\% | 61\% | 57\% |
| Age range (years) | 25-84 | 15-99 | 18-98 | 20-107 |
| Mean age at baseline | - | - | 48 years | 47 years |
| Follow-up | Range 6-25 years | Range 5-30 years | Range 2-17 years | Range 2-27 years |
|  | Mean 10 years | Mean 15 years | Mean 9 years | Mean 7 years |
| - No data. |  |  |  |  |

data and provided age-specific analyses of all end-points. APCSC was therefore used as the primary data source for relative risk estimates of all end-points.

## ANALYTICAL ISSUES

Regression dilution bias has particular relevance in observational studies of blood pressure, and must be accounted for in analyses. This bias occurs as baseline or one-off measures of blood pressure are subject to random fluctuations, due partly to the measurement process and partly to any real but temporary deviations at the baseline visit from the usual blood pressure level (MacMahon et al. 1990). Therefore, baseline blood pressure values have a wider distribution than the "usual" blood pressure values. With repeated measures there is a "regression to the mean" of values (MacMahon 1994) whereby an initially extreme observation tends to become less extreme with replication (Strachan and Rose 1991). This imprecision in measurement not only influences distribution, but will also affect the association with disease outcomes (MacMahon et al. 1990) (Figure 6.10).

The baseline distribution has a shallower slope on the curve relation of blood pressure to relative risk of stroke and IHD, which means there is an underestimation of the strength of the association between blood pressure and disease incidence ("regression dilution bias") (MacMahon 1994). If not corrected for, this bias systematically dilutes the apparent importance of blood pressure and can result in systematic and substantial underestimation of risk of disease with usual blood pressure (MacMahon et al. 1990).

The size of the dilution is directly related to the extent to which blood pressure measurements are subject to regression to the mean. Several

Figure 6.10 Effects of regression dilution bias on the association between SBP and relative risk of cardiovascular disease

major meta-analyses conducted in recent years aimed to address these limitations with correction for regression dilution bias (Anonymous 1998; APCSC 2003a; MacMahon et al. 1990; PSC 1995). It is possible to use repeated measures of blood pressure to obtain an estimate of the attenuation factor in order to correct for this bias in the analysis. Data from cohort studies suggest that for DBP, the relationship with stroke/IHD is about $60 \%$ steeper if 4 -year mean blood pressure values are used rather than mean values at baseline (Garcia-Palmieri and Costas 1986; Paul et al. 1963). As a result, a correction factor of 1.6 was used in early meta-analyses (MacMahon et al. 1990). However, other studies suggest that the association with usual blood pressure is twice as strong as that with baseline DBP (Aromaa 1981; Hughes and Pocock 1992) and a correction factor of 2.1 overall has been applied in the subsequent prospective studies collaboration (PSC) and Eastern meta-analyses (Anonymous 1998; PSC 1995). The APCSC meta-analyses primarily used SBP rather than DBP. The overall attenuation factor was 1.84 (APCSC 2003a). This concurs with other un-published data (PSC secretariat, personal communication) and these later meta-analyses are the most reliable source of estimates for risk accumulation.

A summary of the main results of the four prospective study overviews is presented in Table 6.10. While SBP is primarily being used to classify blood pressure in these analyses, some analyses have only published risk estimates in terms of DBP; therefore data for both indices have been presented for completeness. The risk estimates do appear slightly different across overviews, partly explained by different age distributions within the overviews.

## Stroke

There is evidence of a positive association between blood pressure and all types of stroke (Anonymous 1998; APCSC 2003a; MacMahon et al. 1990; PSC 1995). All of the overviews have demonstrated heterogeneity among studies in the strength of the association between blood pressure and stroke risk, and a major factor contributing to this variability is attenuation with age (APCSC 2003a; PSC 1995).

The slope of the association between relative risk of stroke (plotted on a log scale) and mean usual blood pressure is roughly constant, implying a log-linear relationship (Figure 6.11). This means that the relative difference in risk associated with an absolute difference in usual blood pressure is similar at all levels of blood pressure, at least within the range studied. There is no good evidence of a threshold level below which lower levels of blood pressure were no longer associated with lower relative risks of stroke (down to SBP 115 mmHg ). Likewise, there was no evidence of a threshold level above which the relative risk of stroke increases much more rapidly (Anonymous 1998; APCSC 2003a; MacMahon et al. 1990; PSC 1995). The strength of the overall relation was not altered by restricting analyses to those with and without a
Table 6.10 Summary results of major cohort study overviews

| Study end-points | MacMahon et al. (1990) | Prospective Studies Collaboration (PSC 1995) | Eastern Collaborative Research Group (Anonymous 1998) | Asia-Pacific Cohort Studies Collabration (APCSC 2003a) |
| :---: | :---: | :---: | :---: | :---: |
| Stroke end-points | 843 strokes 599 (71\%) fatal | 13397 participants were recorded as having had a stroke | 1798 strokes 995 (55\%) fatal $39 \%$ had data on subtype of which $42 \%$ haemorrhagic | 4355 strokes <br> 60\% fatal <br> 1335 were haemorrhagic <br> I 347 were ischaemic |
| BP-stroke association | $5 \mathrm{mmHg} \downarrow$ DBP RR $=0.66$ <br> (34\% reduction in stroke risk) Association does not differ by sex, or fatal/non-fatal | Log linear continuous relation $5 \mathrm{mmHg} \downarrow \mathrm{DBP} \mathrm{RR}=0.73$ (95\% Cl 0.68-0.75) <br> ( $27 \%$ reduction in stroke risk) Association differs by age but not by sex | hip with no threshold BP level $\begin{aligned} & 5 \mathrm{mmHg} \downarrow \text { DBP RR }=0.56 \\ & (95 \% \mathrm{Cl} 0.53-0.59) \end{aligned}$ <br> (44\% reduction in stroke risk) Association differs by stroke subtype | $5 \mathrm{mmHg} \downarrow$ DBP or $10 \mathrm{mmHg} \downarrow$ SBP RR $=0.62$ ( $0.6 \mathrm{I}-0.63$ ) <br> ( $38 \%$ reduction in stroke risk) Association did not differ by sex, or fatal/non-fatal events, attenuated by age |
| IHD end-points | 4856 IHD events 4260 (88\%) fatal | - | - | 3560 IHD events 56\% fatal |
| BP-IHD association | Log linear continuous relationship $5 \mathrm{mmHg} \downarrow$ DBP RR $=0.79$ ( $21 \%$ reduction in IHD risk) Association does not differ by sex | - - |  | Log linear continuous relationship $5 \mathrm{mmHg} \downarrow$ DBP or $10 \mathrm{mmHg} \downarrow$ SBP RR $=0.74$ (0.7I-0.76) ( $25 \%$ reduction in IHD risk) Association did not differ by sex, or fatal/non-fatal events, attenuated by age |

[^13]Figure 6.II Usual SBP and risk of stroke in the Asia-Pacific region


Source: APCSC, personal communication, 2001.
history of cardiovascular disease at baseline (APCSC 2003a; PSC 1995).

There has been no evidence in any of the overviews that the strength of association between blood pressure and stroke varies by sex (Figure 6.12). This is consistent for all age subgroups, so it is not necessary to have sex-specific estimates for CRA.

Age has an important influence on the size of the SBP-stroke relationship, as the associations are steeper for those in younger age groups (Figure 6.13). While the proportional change in stroke risk per unit change in blood pressure is less extreme in old age than in middle age, the relationship remains positive for all age groups (APCSC 2003a; PSC 1995).

The different age distributions within the four overviews therefore explain the slightly different relative risk estimates. This limits the ability to directly compare the overall relative risk estimates across the overviews, as it is only appropriate to compare age-specific results. It is therefore necessary to use age-specific risk estimates for blood pressure, provided by the APCSC data. There were no statistically significant differences in the age-specific relative risk for blood pressure and stroke between subgroups, or for stroke subtypes in APCSC analyses (2003a).

Risk estimates for the association between blood pressure and total stroke are shown in Table 6.11. These estimates of relative risk were vir-

Figure 6.12 Usual SBP and risk of stroke event in males and females


Source: APCSC, personal communication, 2001.

Figure 6.13 Usual SBP and risk of stroke by age


[^14]Table 6.II Relative risk of total stroke event with a 10 mmHg decrease in usual SBP

| Age group (years) | Relative risk $(95 \%$ Cl) | Events |
| :--- | :---: | ---: |
| $30-44$ | $0.42(0.38-0.47)$ | 249 |
| $45-59$ | $0.50(0.49-0.52)$ | 1602 |
| $60-69$ | $0.64(0.62-0.66)$ | 1349 |
| $70-79$ | $0.73(0.70-0.76)$ | 1250 |
| $\geq 80$ | $0.83(0.78-0.88)$ | 738 |

Source: APCSC, personal communication, 2001.

Figure 6.14 Usual SBP and risk of IHD in the Asia-Pacific region


Source: APCSC, personal communication, 2001.
tually unchanged when analyses adjusted for cholesterol, smoking, alcohol or past history of cardiovascular disease in the APCSC cohorts that provided these additional data.

## Ischaemic heart disease

Two overviews have published data on the association between blood pressure and risk of IHD. As with stroke, the relationship between blood pressure and IHD was log-linear with no apparent threshold levels of blood pressure (Figure 6.14).

The slope of the association of blood pressure with IHD is less than that for stroke, each 10 mmHg decrease in usual SBP was associated with about a $25 \%$ lower risk of IHD. However, in most studies the absolute effects of IHD tended to be greater because IHD was so much more common than stroke (MacMahon et al. 1990).

The blood pressure IHD association was attenuated with age (Figure 6.15), and did not differ significantly between males and females (APCSC 2003a; PSC secretariat, personal communication). A comparison of the strength of the SBP association with IHD between Asia and Australasia demonstrated that there were no differences between the age-specific relative risks in the two subgroups.

Age-specific relative risk estimates for IHD from the APCSC cohort are presented in Table 6.12. These estimates are consistent with unpublished data from PSC. They were virtually unchanged when analyses adjusted for potential confounders such as cholesterol, smoking, alcohol or past history of cardiovascular disease in the APCSC cohorts that provided these additional data.

## Hypertensive disease

The APCSC project provides the only age-specific estimates for this endpoint, and each 10 mmHg decrease in usual SBP was associated with

Figure 6.15 Usual SBP and risk of IHD by age


[^15]Table 6.12 Relative risk of IHD event with a 10 mmHg decrease in usual SBP

| Age group (years) | Relative risk $(95 \%$ CI) | Events |
| :--- | :---: | ---: |
| $30-44$ | $0.52(0.42-0.65)$ | 95 |
| $45-59$ | $0.60(0.57-0.64)$ | 727 |
| $60-69$ | $0.75(0.72-0.79)$ | 779 |
| $70-79$ | $0.80(0.76-0.84)$ | 844 |
| $\geq 80$ | $0.94(0.88-1.01)$ | 601 |

Source: APCSC, personal communication, 2001.

Table 6.13 Relative risk of death from hypertensive disease with a 10 mmHg decrease in usual SBP

| Age group (years) | Relative risk $(95 \% \mathrm{Cl})$ | Events |
| :--- | :---: | :---: |
| $30-44$ | $0.16(0.07-0.37)$ | 2 |
| $45-59$ | $0.40(0.24-0.67)$ | 7 |
| $60-69$ | $0.57(0.46-0.7 \mathrm{I})$ | 28 |
| $70-79$ | $0.65(0.53-0.8 \mathrm{I})$ | 48 |
| $\geq 80$ | $0.63(0.50-0.8 \mathrm{I})$ | 47 |

Source: APSCS, personal communication, 2001.
$46 \% ~(95 \%$ CI $40-51 \%$ ) lower risk of hypertensive disease. There were no statistically significant differences between males and females or Asia and Australasia for this end-point. As with other cardiovascular endpoints, there was attenuation with age. Age-specific relative risk estimates are shown in Table 6.13.

## Other cardiovascular end-points

The category "other cardiac disease" includes end-points such as heart failure. Overall, each 10 mmHg decrease in usual SBP was associated with $18 \% ~(95 \%$ CI $15-20 \%$ ) lower risk of other cardiac disease. The age-specific APCSC associations for this composite end-point are shown in Table 6.14.

There were no statistically significant differences in associations by sex or subregion, although there were relatively few data to assess this reliably. This end-point was defined as all cardiovascular deaths other than those due to stroke, IHD or hypertensive disease. Numerically most of these were ascribed to heart failure. Given some uncertainty about causality and the varying composition for this end-point around the world, the relative risk reductions were halved for this outcome to avoid any potential overestimation of effects.

Table 6.14 Relative risk of death from "other cardiac disease" with a 10 mmHg decrease in usual SBP

| Age group (years) | Relative risk $(95 \%$ CI) | Events |
| :--- | :---: | ---: |
| $30-44$ | $0.66(0.45-0.96)$ | 32 |
| $45-59$ | $0.76(0.69-0.84)$ | 258 |
| $60-69$ | $0.83(0.78-0.89)$ | 500 |
| $70-79$ | $0.90(0.84-0.96)$ | 551 |
| $\geq 80$ | $0.92(0.85-1.00)$ | 507 |

Source: APCSC, personal communication, 2001.

### 3.2 RISK REVERSIBILITY

As already alluded to, prospective observational studies provide data on the association between blood pressure and stroke and IHD, but they do not provide data on the impact of blood pressure lowering on these outcomes (MacMahon et al. 1990). From prospective studies alone it is not possible to tell whether outcomes are reversible, or whether the association reflects, to some extent, irreversible cumulative effects of blood pressure differences that have persisted for years (Collins and Peto 1994).

The results from randomized clinical trials do provide data on reversibility. They are relevant to assessing how rapidly, and to what extent, the epidemiologically expected reductions in stroke and IHD are produced by lowering blood pressure (Collins and Peto 1994; Collins et al. 1990). They therefore provide estimates of the proportion of the potential long-term benefit from a particular blood pressure difference that may be expected within a few years of blood pressure lowering. This section will discuss how results from trials can be utilized in CRA analyses.

## Data sources on risk reversal

A variety of trials have studied the impact of blood pressure lowering. However, as with prospective studies, individual studies usually lack sufficient power to reliably detect moderate changes in events (Collins and Peto 1994; Collins et al. 1990; He and Whelton 2000). To accurately detect small but potentially important differences in risk reduction of cardiovascular disease (e.g. 10-15\%), it is necessary for trials to record many hundreds of end-points. At least 1000 events would be necessary to reliably detect a $15 \%$ difference in risk of myocardial infarction (MacMahon and Neal 2000). Since most have not achieved this, overviews are necessary to provide more accurate estimates of the impact of blood pressure lowering on stroke and IHD. Results from these overviews will contribute to estimates of "risk reversibility".

A variety of overviews of trials have been published. These include:

- an overview of 14 randomized trials including nearly 37000 individuals and almost 190000 patient years of follow-up (Collins and Peto 1994; Collins et al. 1990). This overview was subsequently updated including three additional trials (MacMahon and Rodgers 1993a, 1993b). The updated review included 17 trials (four large, 13 small) of pharmacological treatment including over 47000 individuals with over 12000 aged over 60 years, mean follow-up 4.9 years (MacMahon and Rodgers 1993a, 1993b).
- a review by He and Whelton (1999) included 10 trials, a total of 18542 participants, and an average follow-up of four years (He and Whelton 1999).
- Psaty et al. (1997) included 18 long-term randomized trials in their review, which included over 48000 patients followed up for an average of five years in the United States, Europe, Scandinavia, Australia and Japan (Psaty et al. 1997).
- the analyses from the Blood Pressure Lowering Treatment Trialists' Collaboration (BPLT) included data from 15 trials, 74696 individuals, mean age 62 years and $53 \%$ male (Neal et al. 2000).
- a review by Pahor et al. (2000) included nine trials and 27743 individuals (Pahor et al. 2000).
- an updated meta-analysis conducted on all trials included in previous overviews.


## Estimates of risk reversibility

## Stroke

Overviews of randomized controlled trials have all confirmed the reversibility of stroke with blood pressure lowering (Collins and Peto 1994; Collins et al. 1990; He and Whelton 1999; MacMahon and Rodgers 1993a, 1993b; Neal et al. 2000; Pahor et al. 2000; Psaty et al. 1997). The cohort study overviews already discussed suggested that a 10 mmHg lower SBP (or a 5 mmHg lower DBP) was associated with about a $30-40 \%$ lower risk of stroke. A major overview of clinical trials has demonstrated a quantitatively similar result with this level of blood pressure reduction (Collins and Peto 1994; Collins et al. 1990; MacMahon and Rodgers 1993a, 1993b). Analyses of additional trials in subsequent published overviews have had similar results (He and Whelton 1999; Neal et al. 2000; Psaty et al. 1997). The updated meta-analysis categorized trials of drugs vs placebo by drug classes, and found consistent results for each class of drug vs placebo and all trials combined (Table 6.15).

These results suggest that potentially all of the excess risk of stroke due to blood pressure can be reversed. The trial overviews also concur with cohort studies in that the reductions were similar for fatal and
Table 6.I5 Clinical trials comparing a blood pressure-lowering drug with placebo or no treatment

| Trial | Participants |  | Stroke |  | IHD |  | Mean age (years) | \% male | Net reduction ${ }^{\text {a }}$ |  | RR (95\% CI) |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | A | C | A | C | A | C |  |  | SBP | DBP | Stroke | IHD |
| Beta blocker and/or diuretic vs placebo or no treatment |  |  |  |  |  |  |  |  |  |  |  |  |
| ANBPS (Anonymous 1980) | 1721 | 1706 | 13 | 22 | 33 | 33 | 50 | 63 | - | 6 | 0.59 (0.30-1.16) | 0.99 (0.61-1.60) |
| Barraclough (Anonymous 1973) ${ }^{\text {b }}$ | 58 | 58 | 0 | 0 | 1 | 2 | 56 | 43 | 15 | 10 | - | 0.50 (0.05-5.36) |
| Carter (1970) | 49 | 48 | 10 | 21 | 2 | 2 | NA | 57 | - | - | 0.47 (0.25-0.88) | 0.98 (0.14-6.68) |
| Coope and Warrender (1986) | 419 | 465 | 20 | 39 | 35 | 38 | 69 | 31 | 18 | 11 | 0.57 (0.34-0.96) | 1.02 (0.66-1.59) |
| EWPHE (Amery et al. 1985) | 416 | 424 | 32 | 48 | 48 | 59 | 72 | 30 | 21 | 8 | 0.68 (0.44-1.04) | 0.83 (0.58-1.18) |
| HDFP (Anonymous 1979a, 1979b, 1982a, 1982b, 1984) | 5485 | 5455 | 102 | 158 | 275 | 343 | 51 | 55 | - | 6 | 0.64 (0.50-0.82) | 0.80 (0.68-0.93) |
| HSCSG III III (Anonymous 1974) ${ }^{\text {b }}$ | 233 | 219 | 43 | 52 | 7 | 12 | 59 | 41 | 25 | 12 | 0.78 (0.54-1.11) | 0.55 (0.22-1.37) |
| MRC older (Anonymous 1992) | 2183 | 2213 | 101 | 134 | 128 | 159 | 70 | 42 | 14 | 7 | 0.76 (0.59-0.98) | 0.82 (0.65-1.02) |
| MRC young (Anonymous 1985) | 8700 | 8654 | 60 | 109 | 222 | 234 | 52 | 52 | 11.5 | 6 | 0.55 (0.40-0.75) | 0.94 (0.79-1.13) |
| Oslo (Helgeland 1980; Leren and Helgeland 1986) | 406 | 379 | 0 | 5 | 14 | 10 | 45 | 100 | 17 | 10 | 0.08 (0.00-1.53) | 1.31 (0.59-2.91) |
| SHEP (Anonymous 1991) ${ }^{\text {b }}$ | 2365 | 2371 | 105 | 162 | 104 | 142 | 72 | 43 | 12 | 4 | 0.65 (0.51-0.83) | 0.73 (0.57-0.94) |
| STOP H (Dahlof et al. 1991) | 812 | 815 | 30 | 55 | 31 | 32 | 76 | 37 | 20 | 8 | 0.55 (0.35-0.85) | 0.97 (0.60-1.58) |
| USPHS (Smith 1977) ${ }^{\text {b }}$ | 193 | 196 | 1 | 6 | 15 | 18 | 44 | 80 | 15 | 10 | 0.17 (0.02-1.39) | 0.85 (0.44-1.63) |
| VA (Anonymous 1967, 1970) ${ }^{\text {b }}$ | 254 | 257 | 6 | 23 | 11 | 15 | 51 | 100 | 17 | 10 | 0.26 (0.11-0.64) | 0.74 (0.35-1.58) |
| VA-NHL BI (Anonymous 1978; Perry 1977) | 508 | 504 | 0 | 0 | 8 | 5 | 38 | 81 | - | 7 | - | 1.59 (0.52-4.82) |
| Wolff and Lindeman (1966) | 45 | 42 | 2 | 1 | 0 | 0 | 49 | 32 | - | 20 | 1.87 (0.18-19.84) | - |
| Summary ${ }^{\text {c }}$ | 23847 | 23806 | 525 | 835 | 934 | 1104 | 59 | 49 | 13 | 6 | 0.63 (0.57-0.70) | 0.85 (0.78-0.92) |

ACE inhibitors vs placebo

| HOPE (Yusuf et al. 2000) | 4654 | 4652 | 156 | 226 | 459 | 570 | 66 | 73 | 3 | 1 | 0.69 (0.57-0.84) | 0.81 (0.72-0.91) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| PART 2 (MacMahon et al. 2000) | 308 | 309 | 7 | 4 | 24 | 35 | 61 | 82 | 6 | 4 | 1.76 (0.52-5.94) | 0.69 (0.42-1.13) |
| PROGESS (Progress Collaborative Group 2001) | 3051 | 3054 | 307 | 420 | 115 | 154 | 64 | 70 | 9 | 4 | 0.73 (0.64-0.84) | 0.75 (0.59-0.95) |
| QUIET (Cashin-Hemphill et al. 1999) | 878 | 872 | I | 1 | 48 | 54 | 58 | 82 | - | - | 0.99 (0.06-15.85) | 0.88 (0.61-1.29) |
| SCAT (Teo et al. 2000) | 229 | 231 | 2 | 9 | 8 | 13 | 61 | 89 | 4 | 3 | 0.22 (0.05-1.03) | 0.62 (0.26-1.47) |
| Summary ${ }^{\text {c }}$ | 9111 | 9118 | 473 | 660 | 654 | 826 | 65 | 73 | 5 | 2 | 0.72 (0.64-0.80) | 0.79 (0.72-0.87) |
| Calcium antagonists vs placebo |  |  |  |  |  |  |  |  |  |  |  |  |
| PREVENT (Pitt et al. 2000) | 417 | 408 | 5 | 5 | 19 | 20 | 57 | 80 | 5 | 4 | 0.98 (0.29-3.35) | 0.93 (0.50-1.72) |
| Syst-Eur (Staessen et al. 1997) | 2398 | 2297 | 49 | 80 | 60 | 76 | 70 | 33 | 11 | 5 | 0.59 (0.41-0.83) | 0.76 (0.54-1.06) |
| Summary ${ }^{\text {c }}$ | 2815 | 2705 | 54 | 85 | 79 | 96 | 68 | 40 | 10 | 5 | 0.61 (0.44-0.85) | 0.79 (0.59-1.06) |

Key: A, active treatment group; C, control group.
No data.
NA Not applicable.
Net BP reduction = blood pressure difference between the randomized groups (i.e. BP change in treatment group minus BP change in placebo group).
These trials also included older classes of drugs such as methyldopa and alkaloid
Summary mean age, \% male and net SBP and DBP change are weighted by study size for those studies that provided data on net changes in both SBP and DBP. Weighting by number of IHD events made little material difference.
non-fatal strokes, and for males and females (He and Whelton 1999). The recently completed PROGRESS trial also demonstrated similar relative risk reductions for those classified as hypertensive and non-hypertensive at the start of the trial (Progress Collaborative Group 2001). In contrast to the treatment effect demonstrated in drug vs placebo trials, the treatment effect of one drug class vs another was relatively minor. Another important piece of information from the trial overviews is the time frame of this risk reversal: the data suggest that these reductions in risk of stroke occurred after only 2-3 years of blood pressure treatment (He and Whelton 1999; MacMahon and Rodgers 1993a).

## Ischaemic heart disease

Overviews of randomized controlled trials have also confirmed the reversibility of IHD with blood pressure lowering (Collins and Peto 1994; Collins et al. 1990; He and Whelton 1999; MacMahon and Rodgers 1993a, 1993b; Neal et al. 2000; Pahor et al. 2000; Psaty et al. 1997). The cohort studies suggested that a 10 mmHg lower SBP (or a 5 mmHg lower DBP) was associated with about a $20-25 \%$ lower risk of IHD. Overviews of clinical trials have demonstrated that this magnitude of blood pressure reduction is associated with about a $15-20 \%$ reduction in IHD (Collins et al. 1990; He and Whelton 1999; MacMahon and Rodgers 1993a), which was confirmed by the updated meta-analysis (Table 6.15).

There were similar reductions for fatal and non-fatal strokes, for males and females (He and Whelton 1999), and for those with and without hypertension (Progress Collaborative Group 2001). As with stroke, these reductions in risk occurred after 2-3 years of blood pressure treatment (He and Whelton 1999; MacMahon and Rodgers 1993a). These data suggest that approximately two thirds of blood pressure related risk of IHD (not all) was reversed in this time frame.

The data indicate that within $3-5$ years, 10 mmHg lower SBP resulted in reversal of most or all of the epidemiologically expected risk for stroke events and approximately two thirds of IHD. Likewise, it will be assumed that most or all of the expected risk for hypertensive disease and approximately two thirds of other cardiovascular disease will be reversed in this time frame. The time frames for reversibility from trials are important for estimating avoidable burden, but the trial relative risk estimates will not be used. There are a variety of problems with the generalizability of trial relative risk data, which make it less appropriate than prospective study data. These limitations include the fact that many trials only used mortality as an end-point (Linjer and Hansson 1997) and they were not powered for adequate subgroup analyses (e.g. age and sex subgroups). Overall, using relative risks from cohort studies will lead to conservative estimates of relative risk and avoidable burden because the clinical trial data have not to date demonstrated the age attenuation in relative risks that is observed in the epidemiological data.

In summary, data from the APCSC will be used for risk accumulation complemented by data from other overviews such as PSC. The main findings were direct, positive and continuous associations of usual SBP with the risks of all end-points of interest. The risks were similar by sex and subregion, except for a stronger association of blood pressure with stroke in Asian compared to non-Asian populations, due partly to the higher proportion of haemorrhagic strokes. Few data were available for the GBD end-point "other cardiac disease" and so given some uncertainty about causality and the varying composition for this end-point around the world, the relative risks will be halved for this outcome. There was attenuation of proportional associations with age for all these outcomes.

Data on risk reversibility come from overviews of all unconfounded randomized trials of blood pressure lowering. For the CRA analyses, it is proposed to use the observational epidemiological data for risk accumulation and use the trial overviews for the time frame of risk reversibility. For example, in middle age the epidemiology suggests that a 10 mmHg increase in SBP is associated with about $40 \%$ more stroke and $25 \%$ more IHD. The trials suggest that within $3-5$ years of lowering SBP by 10 mmHg , most or all of this increased risk for stroke and hypertensive disease is reversed and approximately two thirds for IHD and cardiovascular disease.

## 4. Results

### 4.1 Attributable fraction

The "attributable fraction" refers to the proportion of disease burden that would theoretically not have occurred if the population distribution of blood pressure had been equal to that of the theoretical minimum (mean SBP of 115 mmHg ). The attributable fractions for each of the endpoints increased initially with age in males and females, but then declined for the oldest age groups, reflecting the balance between increasing mean SBP, and decreasing relative risks with age.

## Stroke

Globally, $62 \%$ of stroke is attributable to $\mathrm{SBP}>115 \mathrm{mmHg}$ (range of $52-78 \%$ by subregion). Subregions with the lowest attributable fractions included AFR-E, AMR-D, SEAR-D and EMR-D. In contrast, EUR-B and EUR-C had the highest values. In most subregions the attributable fraction for males and females were similar (within $5 \%$ ).

## ISCHAEMIC HEART DISEASE

Globally, $49 \%$ of IHD is attributable to $\mathrm{SBP}>115 \mathrm{mmHg}$ (range of $40-64 \%$ by subregion). Subregions with the lowest attributable fractions included AMR-A, AMR-D, SEAR-D and WPR-B. In contrast, EUR-B and EUR-C had the highest values.

## Hypertensive disease

Globally, $76 \%$ of hypertensive disease is attributable to SBP $>115 \mathrm{mmHg}$ (range of $64-90 \%$ by subregion). Subregions with the lowest attributable fractions included AFR-E, EMR-D, AMR-D and SEAR-D. In contrast, EUR-B and EUR-C had the highest values.

## OTHER CARDIOVASCULAR DISEASE

Globally, $14 \%$ of other cardiovascular disease is attributable to SBP $>115 \mathrm{mmHg}$ (range of $8-22 \%$ by subregion). Subregions with the lowest attributable fractions included AMR-D, EMR-D, SEAR-D and WPR-B. In contrast, EUR-B and EUR-C had the highest values.

### 4.2 Attributable burden

For the year 2000, WHO estimated that there were 55.9 million deaths and 1455 million DALYs worldwide; about $22 \%$ of these deaths and $7 \%$ of these DALYs were due to ischaemic heart disease and stroke. The burden of disease was distributed across the developed and developing world (e.g. about one quarter of all stroke deaths and one third of IHD deaths occurred in the least developed subregions), and predominantly affected those aged $>60$ years.

The "attributable burden" refers to the number of deaths or DALYs that would theoretically not have occurred if the population distribution of blood pressure had been equal to that of the theoretical minimum (mean SBP of 115 mmHg ). In total, the attributable burden equated to 3.1 million stroke deaths, just under 3.0 million IHD deaths, about 0.7 million hypertensive disease deaths and 0.3 million other cardiovascular disease deaths. For DALYs these figures were 27.8 million stroke DALYs, 28.2 million IHD DALYs, 5.4 million hypertensive disease DALYs and 2.8 million other cardiovascular disease DALYs (Table 6.16).

Worldwide this means that 7.1 million deaths (about $12.8 \%$ of the total) and 64.3 million DALYs ( $4.4 \%$ of the total) were estimated to be due to non-optimal blood pressure. The proportion of DALYs was lower than the proportion of deaths, as most blood pressure-related cardiovascular disease events occurred in middle or old age, so the years of life lost due to premature mortality and years of life lost due to disability is a smaller percentage of the total DALYs than of the total deaths.

The age group with the greatest attributable deaths for both males and females was 70-79 years. The number of attributable deaths then declined, reflecting the smaller denominator population, and the smaller total number of events in the oldest age group. For DALYs, this decline occurred sooner, as there are less years of life lost and years of life lived with disability with advancing age. For example, $48 \%$ of excess IHD deaths were in those aged $\geq 70$ years, compared to $24 \%$ of excess DALYs.

For each end-point, the attributable deaths were higher for males than females for the age categories $30-44,45-59$ and $60-69$ years. In
Table 6.16 Attributable deaths and DALYs for systolic blood pressure $>115 \mathrm{mmHg}$, by cardiocvascular end-point and subregion

| Subregion | Deaths (000s) |  |  |  | DALYs (000s) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Stroke | Ischaemic heart disease | Hypertensive disease | Other cardiovascular disease | Stroke | Ischaemic heart disease | Hypertensive disease | Other cardiovascular disease |
| AFR-D | 93 | 82 | 20 | 19 | 982 | 893 | 197 | 204 |
| AFR-E | 91 | 67 | 21 | 15 | 1014 | 760 | 214 | 173 |
| AMR-A | 104 | 203 | 37 | 26 | 824 | 1548 | 247 | 164 |
| AMR-B | 133 | 129 | 53 | 18 | 1376 | 1303 | 425 | 141 |
| AMR-D | 14 | 13 | 10 | 3 | 145 | 132 | 87 | 23 |
| EMR-B | 28 | 71 | 27 | 6 | 300 | 811 | 230 | 68 |
| EMR-D | 101 | 176 | 43 | 15 | 987 | 1922 | 387 | 183 |
| EUR-A | 263 | 290 | 56 | 70 | 1729 | 2079 | 263 | 381 |
| EUR-B | 210 | 263 | 58 | 40 | 1839 | 2320 | 447 | 272 |
| EUR-C | 513 | 603 | 35 | 35 | 4134 | 5239 | 312 | 333 |
| SEAR-B | 115 | 100 | 48 | 9 | 1185 | 1050 | 458 | 103 |
| SEAR-D | 451 | 650 | 49 | 36 | 4262 | 7080 | 539 | 446 |
| WPR-A | 91 | 51 | 7 | 11 | 736 | 400 | 33 | 63 |
| WPR-B | 944 | 291 | 198 | 35 | 8315 | 2664 | 1601 | 247 |
| World | 3149 | 2991 | 663 | 338 | 27829 | 28201 | 5440 | 2801 |

contrast, attributable deaths were higher for females in the oldest age groups: 70-79 and $\geq 80$ years. There was a similar trend for attributable DALYs. This was due to older females having higher mean SBP (and therefore greater attributable fraction) than males, and there are also a larger number of cardiovascular deaths occurring in older females compared to older males.

## Stroke

The subregion with the highest attributable stroke burden was WPR-B ( 944000 deaths, 8.3 million DALYs), followed by EUR-C ( 513000 deaths, 4.1 million DALYs) and SEAR-D ( 451000 deaths, 4.3 million DALYs).

## ISCHAEMIC HEART DISEASE

SEAR-D had the highest attributable IHD burden ( 650000 deaths, 7.1 million DALYs), followed by EUR-C ( 603000 deaths, 5.2 million DALYs), EUR-A (290000 deaths, 2.1 million DALYs) and WPR-B (291000 deaths, 2.6 million DALYs).

## Hypertensive disease

The attributable burden for hypertensive disease, was highest in WPRB (198000 deaths, 1.6 million DALYs), followed by EUR-B and A for deaths ( 58000 and 56000, respectively) and SEAR-D and B for DALYs ( 539000 and 458000 , respectively).

## Other cardiovascular disease

The highest attributable burden for other cardiovascular disease was in EUR-A ( 70000 deaths, 381000 DALYs), SEAR-D (36000 deaths, 446000 DALYs) and EUR-B ( 40000 deaths, 272000 DALYs).

Overall, approximately one third of attributable DALYs occurred in the most developed subregions ( 23.3 million), a further third (20.3 million) in low mortality developing subregions and a final third (20.7 million) in high mortality developing subregions. A higher proportion of these deaths and DALYs were from IHD than stroke in developed and high mortality developing subregions, but more were stroke than IHD in low mortality developing subregions.

These results indicate where most of the worldwide attributable cardiovascular disease burden occurred, and provide any given subregion with an indication of absolute size of the attributable burden. However, the age structure, population size and the number of estimated events occurring in a subregion, influences the ranking of subregions by absolute number of attributable deaths and DALYs. The relative impact of the attributable DALYs indicates that between 3-33\% of all deaths and $1-17 \%$ of all DALYs across the subregions were attributable to nonoptimal blood pressure. Approximately $22 \%$ of all deaths and $11 \%$ of
all DALYs were attributable to non-optimal blood pressure in developed subregions, $14 \%$ and $5.0 \%$, respectively, in low mortality developing subregions, and $7.5 \%$ and $2.5 \%$, respectively, in high mortality developing subregions.

## 5. Discussion

### 5.1 Attributable deaths and DALYs

The analyses in this chapter suggest that globally, a substantial proportion of cardiovascular disease is attributable to non-optimal blood pressure, defined as mean SBP $>115 \mathrm{mmHg}$. Overall, about two thirds of stroke, one half of IHD, and about three quarters of hypertensive disease were attributable to non-optimal blood pressure. The attributable fractions were higher in the more developed parts of the world than the least developed regions, as would be expected given the higher blood pressure levels.

Worldwide, 7.1 million deaths (about $12.8 \%$ of the total) and 64.3 million DALYs $(4.4 \%$ of the total) were estimated to be due to nonoptimal blood pressure. Overall, the results suggest that a considerable proportion of cardiovascular disease is related to non-optimal blood pressure and this translates into an important portion of deaths and years of life lost to deaths and disability worldwide.

### 5.2 Attributable deaths and DALYs by subregion

In absolute terms, most of the excess burden of cardiovascular disease occurred in the populous subregions of WPR-B, SEAR-D and EUR-C; each subregion accounted for approximately $15-20 \%$ of worldwide attributable deaths and DALYs. The relative impact that attributable deaths and DALYs have in different subregions may also be calulcated as the proportion of all deaths and DALYs attributable to non-optimal SBP (i.e. mean $>115 \mathrm{mmHg}$ ) within each specific subregion. Overall, between $3-33 \%$ of all deaths and $1-17 \%$ of all DALYs across the subregions were attributable to non-optimal blood pressure. Excess cardiovascular disease burden was highest in European subregions where mean blood pressure levels were highest.

### 5.3 Comparison with the first WHO GBD study

The WHO GBD study in 1990 estimated that $8.7 \%$ of all deaths were due to stroke and $12.4 \%$ were due to IHD worldwide, compared to $9.5 \%$ and $12.5 \%$ in the current 2000 WHO estimates. Worldwide, $2.8 \%$ of all DALYs were from stroke in 1990 and $2.9 \%$ in 2000, and for IHD these percentages were $3.4 \%$ in 1990 and $3.8 \%$ in 2000 . These percentages relate to worldwide estimates, and will vary across subregions. The 2000 WHO estimates indicated that about $5.8 \%$ of DALYs were due to stroke in developed subregions, which compares to $5.0 \%$, and
the percentage of DALYs attributed to IHD across these studies was $9.2 \%$ (2000), compared to $9.0 \%$ in 1990. (Murray and Lopez 1996).

In terms of attributable burden, both GBD studies have estimated the total number of deaths and DALYs attributable to non-optimal blood pressure. A key difference is that the current estimates of burden attributable to blood pressure are approximately double those of the 1990 estimates. Worldwide $5.8 \%$ of deaths and $1.4 \%$ of DALYs were attributable to blood pressure in 1990, compared to $12.8 \%$ of deaths and $4.4 \%$ of DALYs in these analyses.

Potential reasons for the increases in the estimates include differences in the estimates of blood pressure prevalence; the theoretical minimum or comparison blood pressure group; the relative risk estimates; and in the way that DALYs were estimated. There were some differences in the studies used to estimate blood pressure prevalence and the comparison blood pressure level ( 110 vs 115 mmHg ), but these would not have been sufficient to account for an approximate doubling of the burden attributable to blood pressure. A particularly important difference however, relates to the relative risk estimates used. The 1990 estimates did not correct for regression dilution bias. As discussed earlier, this bias occurs when relative risk estimates are based on one-off measures of blood pressure, which are subject to random fluctuations. Correcting for this bias with re-measurement data results in "regression to the mean" of blood pressure values, and a steeper association between blood pressure and disease outcomes. Without correction for this bias there will be systematic underestimation of risk of cardiovascular disease by a factor of about two. This therefore explains the almost doubling of burden attributable to blood pressure in the current estimates, and also indicates that the current estimates are the most comprehensive, accurate and robust.

## 6. Estimating future exposure

We used the following basic steps for estimating future exposure, with each step to some extent dictating the following step.

1. The literature was reviewed to assess what is already known about trends in blood pressure in different populations.
2. Regions were categorized into broad groupings based on likely direction of changes in blood pressure levels over time.
3. The future trends in blood pressure distribution were estimated.

### 6.1 Literature review of blood pressure trends over time

There is considerable evidence in the literature that changes in risk factors such as blood pressure occur over time. For example, a variety of studies from different world regions and populations have demonstrated increases and decreases in blood pressure levels.

## Established market economies (AMR-A, EUR-A, WPR-A)

A large number of studies have documented reductions in mean blood pressure levels in populations in "A" subregions over the past 30 years. These studies have included populations in countries such as the United Kingdom (Capewell et al. 1999; McCarron et al. 2001), New Zealand and Australia (Bonita 1993; Bonita and Beaglehole 1986, 1987; Capewell et al. 2000; Dobson et al. 1999), the United States (Burt et al. 1995; Garraway and Whisnant 1987; McGovern et al. 1996), Finland (Tuomilehto et al. 1991; Vartiainen et al. 1994), Iceland (Sigfusson et al. 1991), Switzerland (Wietlisbach et al. 1997) and Japan (Okayama et al. 1993; Sakata and Labarthe 1996; Shimamoto et al. 1989; Ueshima et al. 2000).

The populations surveyed, the methodologies used, and the time periods and durations covered varied greatly between studies, however, mean SBP levels have decreased by up to $5-10 \mathrm{mmHg}$ for a 10 -year period (Bonita 1993; Bonita and Beaglehole 1986, 1987; Capewell et al. 1999, 2000; McCarron et al. 2001; Shimamoto et al. 1989; Sigfusson et al. 1991; Vartiainen et al. 1994; Ueshima et al. 2000). However, in more recent studies (from the mid-1980s onwards) these decreases have been less than 5 mmHg per 10 years (Dobson et al. 1999; McGovern et al. 1996; Okayama et al. 1993; Sakata and Labarthe 1996; Wietlisbach et al. 1997). Data also indicate that the reduction in mean SBP levels has been greatest in the oldest age groups (e.g. mean SBP levels from national Japanese survey data declined by about 2 mmHg in those aged $30-39$ and $40-49$ years between 1980 and 1990 , but by $3-5 \mathrm{mmHg}$ in those aged $>70$ years) (Sakata and Labarthe 1996).

Low mortality subregions (AMR-B, EMR-B, EUR-B, EUR-C, SEAR-B, WPR-B)
These subregions are quite varied and there are little reliable time trend data. Survey data from the Czech Republic demonstrated only small changes in blood pressure between 1985 and 1992 (Bobak et al. 1997). In China, there is some evidence of increasing blood pressure levels in the past 10 years (Chow and Muyinda 1999; Wu et al. 1996), and in the Republic of Korea, between 1984 and 1999 there were only small changes in the prevalence of hypertension (Suh 2001).

```
High mortality subregions (AFR-D, AFR-E, AMR-D, EMR-D,
SEAR-D)
```

Very little reliable longitudinal survey data are available in these subregions. However, data from a variety of more "traditional" populations with lower blood pressure have demonstrated that as those populations become more industrialized, risk factor profiles such as blood pressure change. This is particularly evident in migration studies where there is evidence from studies conducted in a variety of settings (Poulter and Sever 1994), such as Africa (Poulter et al. 1988, 1990), China (He et al.

1991a, 1991b) and the Pacific (Joseph et al. 1983; Salmond et al. 1985, 1989), that blood pressure levels rise after migration to more urbanized settings. Pre-migration data suggest that these changes are not due to selective migration (Poulter et al. 1988). Instead, it is likely that factors such as dietary changes of increased intake of sodium, animal protein, fat, and processed foods, and decreased intake of potassium, and vegetable protein are important (He et al. 1991a, 1991b; Poulter and Sever 1994; Poulter et al. 1988). Obviously, the lifestyle and blood pressure changes in an entire population will be slower than those that occur in migrant groups.

## Potential categories of blood pressure change

Data presented above indicate that subregions may be classified into three broad groups, and changes in blood pressure over the next 10,20 and 30 years may be based on trends already documented. Assuming that the absolute increases and decreases in SBP are similar for males and females but greater in the older age groups, the following general scenario was considered.

These changes suggest that:

- the A subregions will experience small decreases in mean SBP levels over the next 30 years, which is consistent with recent literature (Dobson et al. 1999; McGovern et al. 1996; Okayama et al. 1993; Sakata and Labarthe 1996; Wietlisbach et al. 1997). However, with time these reductions attenuate.
- in the case of $B$ and $C$ subregions, there is no change initially in mean SBP levels, but in the following 10-year and 20-year periods there are small decreases in SBP as improvements in lifestyle occur.
- in the case of D and E subregions, small increases in mean SBP levels occur over the entire 30-year period; however these are greater at the beginning (2000-2010) and then start to lessen and level out (See Table 6.17).

Due to the limited data available, these estimates are susceptible to a high degree of uncertainty. There is a wide range of factors that could influence future risk factor levels, and it is difficult to capture these factors even with sophisticated modelling. A decision was therefore made to use a relatively simple, but transparent method for the purposes of these analyses.

Estimates of mean blood pressure levels over time
Tables 6.18-6.20 summarize mean SBP levels by subregion, sex and age in 2010, 2020 and 2030.

Table 6.I7 Scenario for changes in mean SBP levels $(\mathrm{mmHg})$ by age and subregion over the next IO, 20 and 30 years

|  | $30-44$ years | $45-59$ years | $60-69$ years | $70-79$ years | $\geq 80$ years |
| :--- | :---: | :---: | :---: | :---: | :---: |
| A subregions |  |  |  |  |  |
| 2000 to 2010 | -1.5 | -2.5 | -3.5 | -4.5 | -4.5 |
| 2010 to 2020 | -1.0 | -2.0 | -3.0 | -3.5 | -3.5 |
| 2020 to 2030 | -0.5 | -1.5 | -2.5 | -3.0 | -3.0 |
| B and C subregions |  |  |  |  |  |
| 2000 to 2010 | 0 | 0 | 0 | 0 | 0 |
| 2010 to 2020 | -1.0 | -2.0 | -3.0 | -4.0 | -4.0 |
| 2020 to 2030 | -1.5 | -2.5 | -3.5 | -4.5 | -4.5 |
| $D$ and E subregions |  |  |  |  |  |
| 2000 to 2010 | 1.0 | 2.0 | 3.0 | 4.0 | 4.0 |
| 2010 to 2020 | 0.5 | 1.5 | 2.5 | 3.5 | 3.5 |
| 2020 to 2030 | 0.5 | 1.5 | 2.5 | 3.5 | 3.5 |

Table 6.I8 Estimates of mean SBP $(\mathrm{mmHg})$ by subregion, sex and age in 2010

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  | Males |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ |
| AFR-D | 124 | 138 | 149 | 154 | 128 | 137 | 144 | 148 |
| AFR-E | 122 | 133 | 143 | 147 | 125 | 132 | 138 | 141 |
| AMR-A | 112.5 | 124.5 | 134.5 | 138.5 | 120.5 | 126.5 | 133.5 | 136.5 |
| AMR-B | 115 | 130 | 142 | 147 | 122 | 131 | 138 | 141 |
| AMR-D | 118 | 131 | 142 | 147 | 124 | 133 | 139 | 143 |
| EMR-B | 126 | 137 | 147 | 150 | 125 | 132 | 139 | 143 |
| EMR-D | 122 | 137 | 149 | 154 | 124 | 133 | 141 | 145 |
| EUR-A | 120.5 | 133.5 | 143.5 | 146.5 | 128.5 | 135.5 | 140.5 | 142.5 |
| EUR-B | 122 | 141 | 154 | 161 | 128 | 140 | 149 | 153 |
| EUR-C | 125 | 143 | 158 | 164 | 129 | 138 | 146 | 149 |
| SEAR-B | 120 | 130 | 139 | 142 | 122 | 131 | 138 | 141 |
| SEAR-D | 118 | 128 | 135 | 139 | 119 | 129 | 137 | 141 |
| WPR-A | 118.5 | 130.5 | 140.5 | 143.5 | 125.5 | 134.5 | 141.5 | 143.5 |
| WPR-B | 115 | 127 | 137 | 141 | 117 | 126 | 133 | 136 |

[^16]Table 6.19 Estimates of mean SBP $(\mathrm{mmHg})$ by subregion, sex and age in 2020

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  | Males |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ |
| AFR-D | 124.5 | 139.5 | 151.5 | 157.5 | 128.5 | 138.5 | 146.5 | 151.5 |
| AFR-E | 122.5 | 134.5 | 145.5 | 150.5 | 125.5 | 133.5 | 140.5 | 144.5 |
| AMR-A | 111.5 | 122.5 | 131.5 | 135 | 119.5 | 124.5 | 130.5 | 133 |
| AMR-B | 114 | 128 | 139 | 143 | 121 | 129 | 135 | 137 |
| AMR-D | 118.5 | 132.5 | 144.5 | 150.5 | 124.5 | 134.5 | 141.5 | 146.5 |
| EMR-B | 125 | 135 | 144 | 146 | 124 | 130 | 136 | 139 |
| EMR-D | 122.5 | 138.5 | 151.5 | 157.5 | 124.5 | 134.5 | 143.5 | 148.5 |
| EUR-A | 119.5 | 131.5 | 140.5 | 143 | 127.5 | 133.5 | 137.5 | 139 |
| EUR-B | 121 | 139 | 151 | 157 | 127 | 138 | 146 | 149 |
| EUR-C | 124 | 141 | 155 | 160 | 128 | 136 | 143 | 145 |
| SEAR-B | 119 | 128 | 136 | 138 | 121 | 129 | 135 | 137 |
| SEAR-D | 118.5 | 129.5 | 137.5 | 142.5 | 119.5 | 130.5 | 139.5 | 144.5 |
| WPR-A | 117.5 | 128.5 | 137.5 | 140 | 124.5 | 132.5 | 138.5 | 140 |
| WPR-B | 114 | 125 | 134 | 137 | 116 | 124 | 130 | 132 |

a The same SBP levels apply to those aged $\geq 80$ years.

Table 6.20 Estimates of mean SBP $(\mathrm{mmHg})$ by subregion, sex and age in 2030

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  | Males |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ |
| AFR-D | 125 | 141 | 154 | 161 | 129 | 140 | 149 | 155 |
| AFR-E | 123 | 136 | 148 | 154 | 126 | 135 | 143 | 148 |
| AMR-A | 111 | 121 | 129 | 132 | 119 | 123 | 128 | 130 |
| AMR-B | 112.5 | 125.5 | 135.5 | 138.5 | 119.5 | 126.5 | 131.5 | 132.5 |
| AMR-D | 119 | 134 | 147 | 154 | 125 | 136 | 144 | 150 |
| EMR-B | 123.5 | 132.5 | 140.5 | 141.5 | 122.5 | 127.5 | 132.5 | 134.5 |
| EMR-D | 123 | 140 | 154 | 161 | 125 | 136 | 146 | 152 |
| EUR-A | 119 | 130 | 138 | 140 | 127 | 132 | 135 | 136 |
| EUR-B | 119.5 | 136.5 | 147.5 | 152.5 | 125.5 | 135.5 | 142.5 | 144.5 |
| EUR-C | 122.5 | 138.5 | 151.5 | 155.5 | 126.5 | 133.5 | 139.5 | 140.5 |
| SEAR-B | 117.5 | 125.5 | 132.5 | 133.5 | 119.5 | 126.5 | 131.5 | 132.5 |
| SEAR-D | 119 | 131 | 140 | 146 | 120 | 132 | 142 | 148 |
| WPR-A | 117 | 127 | 135 | 137 | 124 | 131 | 136 | 137 |
| WPR-B | 112.5 | 122.5 | 130.5 | 132.5 | 114.5 | 121.5 | 126.5 | 127.5 |
| a The same SBP levels apply to those aged $\geq 80$ years. |  |  |  |  |  |  |  |  |

## Acknowledgements

Particular thanks are due to the Asia-Pacific Cohort Studies Collaboration secretariat, and all study collaborators for allowing us to access and analyse their data for estimates of relative risk. The Prospective Studies Collaboration assisted with analyses and provided unpublished data on relative risk estimates that could be compared with those presented here.

The authors are grateful to Varsha Parag and Derrick Bennett for biostatistical support, including analyses of data from the Asia-Pacific Cohort Studies Collaboration. Angela Hurley provided valuable research assistance and Clarissa Gould-Thorpe lent much appreciated secretarial support.

## Note

1 See the preface for an explanation of this term.

## References

Abernethy J, Borhani NO, Hawkins CM et al. (1986) Systolic blood pressure as an independent predictor of mortality in the hypertension detection and follow-up program. American Journal of Preventive Medicine, 2:123-132.
Amery A, Birkenhager W, Brixko P et al. (1985) Mortality and morbidity results from the European Working Party on High Blood Pressure in the Elderly trial. The Lancet, 1:1349-1354.
Andersen LB (1994) Blood pressure, physical fitness and physical activity in 17-year-old Danish adolescents. Journal of Internal Medicine, 236:323-329.

Anonymous (1967) Effects of treatment on morbidity in hypertension. Results in patients with diastolic blood pressures averaging 115 through 129 mmHg . Journal of the American Medical Association, 202:1028-1034.
Anonymous (1970) Effects of treatment on morbidity in hypertension. II. Results in patients with diastolic blood pressure averaging 90 through 114 mmHg . Journal of the American Medical Association, 213:1143-1152.
Anonymous (1973) Control of moderately raised blood pressure. Report of a cooperative randomized controlled trial. British Medical Journal, 3:434-436.
Anonymous (1974) Effect of antihypertensive treatment on stroke recurrence. Hypertension-Stroke Cooperative Study Group. Journal of the American Medical Association, 229:409-418.

Anonymous (1978) Evaluation of drug treatment in mild hypertension: VANHLBI feasibility trial. Plan and preliminary results of a two-year feasibility trial for a multicenter intervention study to evaluate the benefits versus the disadvantages of treating mild hypertension. Prepared for the Veterans Administration-National Heart, Lung, and Blood Institute Study Group for evaluating treatment in mild hypertension. Annals of the New York Academy of Sciences, 304:267-292.

Anonymous (1979a) Five-year findings of the hypertension detection and followup program. I. Reduction in mortality of persons with high blood pressure, including mild hypertension. Hypertension Detection and Follow-up Program Cooperative Group. Journal of the American Medical Association, 242: 2562-2571.
Anonymous (1979b) Five-year findings of the hypertension detection and followup program. II. Mortality by race-sex and age. Hypertension Detection and Follow-up Program Cooperative Group. Journal of the American Medical Association, 242:2572-2577.

Anonymous (1980) The Australian therapeutic trial in mild hypertension. Report by the Management Committee. The Lancet, 1:1261-1267.
Anonymous (1981) Distribution of some risk factors for atherosclerosis in nine Italian population samples. The research group ATS-RF2 of the Italian National Research Council. American Journal of Epidemiology, 113: 338-346.
Anonymous (1982a) The effect of treatment on mortality in "mild" hypertension: results of the hypertension detection and follow-up program. New England Journal of Medicine, 307:976-980.
Anonymous (1982b) Five-year findings of the hypertension detection and follow-up program. III. Reduction in stroke incidence among persons with high blood pressure. Hypertension Detection and Follow-up Program Cooperative Group. Journal of the American Medical Association, 247:633-638.
Anonymous (1984) Effect of stepped care treatment on the incidence of myocardial infarction and angina pectoris. 5-year findings of the hypertension detection and follow-up program. Hypertension, 6:I198-206.
Anonymous (1985) MRC trial of treatment of mild hypertension: principal results. Medical Research Council Working Party. British Medical Journal, 291:97-104.
Anonymous (1989a) The INTERSALT co-operative research group. Appendix tables. Centre-specific results by age and sex. Journal of Human Hypertension, 3:331-407.

Anonymous (1989b) The WHO MONICA project. A worldwide monitoring system for cardiovascular diseases: cardiovascular mortality and risk factors in selected communities. World Health Statistics, A:27-149.

Anonymous (1991) Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. Final results of the systolic hypertension in the elderly program (SHEP). SHEP Cooperative Research Group. Journal of the American Medical Association, 265: 3255-3264.

Anonymous (1992) Medical Research Council trial of treatment of hypertension in older adults: principal results. MRC Working Party. British Medical Journal, 304:405-412.

Anonymous (1997) The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. Archives of Internal Medicine, 157:2413-2446.

Anonymous (1998) Blood pressure, cholesterol, and stroke in eastern Asia. Eastern Stroke and Coronary Heart Disease Collaborative Research Group. The Lancet, 352:1801-1807.

Anonymous (1999) 1999 World Health Organization-International Society of Hypertension guidelines for the management of hypertension. Guidelines Subcommittee. Journal of Hypertension, 17:151-183.
Aromaa A (1981) Blood pressure level, hypertension and five-year mortality in Finland. Acta Medica Scandinavica-Supplementum, 646:43-50.
Ashour Z, Ibrahim MM, Appel LJ, Ibrahim AS, Whelton PK (1995) The Egyptian National Hypertension Project (NHP). Design and rationale. The NHP Investigative Team. Hypertension, 26:880-885.

APCSC (1999) Determinants of cardiovascular disease in the Asia-Pacific region: protocol for a collaborative overview of cohort studies. Cardiovascular Disease Prevention, 2:281-289.

APCSC (2003a) Blood pressure and cardiovascular disease in the Asia-Pacific region. Journal of Hypertension, 21:707-716.

APCSC (2003b) Blood pressure indices as risk factors for cardiovascular disease in the Asia Pacific region: a pooled analysis. Hypertension, 42:69-75.
Asplund-Carlson A, Carlson LA (1994) Studies in hypertriglyceridaemia. 1. Serum triglyceride distribution and its correlates in randomly selected Swedish middle-aged men. Journal of Internal Medicine, 236:57-64.
Astagneau P, Lang T, Delarocque E, Jeannee E, Salem G (1992) Arterial hypertension in urban Africa: an epidemiological study on a representative sample of Dakar inhabitants in Senegal. Journal of Hypertension, 10:1095-1101.

Baba S, Pan WH, Ueshima H et al. (1991) Blood pressure levels, related factors, and hypertension control status of Japanese and Americans. Journal of Human Hypertension, 5:317-332.

Bajekal M, Boreham R, Erens B et al. (1999) Health survey for England '98. The Department of Health, London.

Barnes R (1965) Comparisons of blood pressure and blood cholesterol levels in New Guineans and Australians. Medical Journal of Australia, 1:611-617.
Black HR (1999a) Isolated systolic hypertension in the elderly: lessons from clinical trials and future directions. Journal of Hypertension, 17:S49-54.
Black HR (1999b) The paradigm has shifted, to systolic blood pressure. Hypertension, 34:386-387.

Bobak M, Skodova Z, Pisa Z, Poledne R, Marmot M (1997) Political changes and trends in cardiovascular risk factors in the Czech Republic, 1985-92. Journal of Epidemiology and Community Health, 51:272-277.
Bonita R (1993) Stroke trends in Australia and New Zealand: mortality, morbidity, and risk factors. Annals of Epidemiology 3:529-533.

Bonita R, Beaglehole R (1986) Does treatment of hypertension explain the decline in mortality from stroke? British Medical Journal, 292:191-192.

Bonita R, Beaglehole R (1987) The decline in stroke mortality: the limited role of antihypertensive therapy. New Zealand Medical Journal, 100:454-456.

Bosma H, Appels A, Sturmans F, Grabauskas G, Gostautas A (1994) Differences in mortality and coronary heart disease between Lithuania and the Netherlands: results from the WHO Kaunas-Rotterdam Intervention Study (KRIS). International Journal of Epidemiology, 23:12-19.
Bovet P, Shamlaye C, Kitua A, Riesen WF, Paccaud F, Darioli R (1991) High prevalence of cardiovascular risk factors in the Seychelles (Indian Ocean). Arteriosclerosis and Thrombosis, 11:1730-1736.
Bullen C, Simmons G, Trye P, Lay-Yee R, Bonita R, Jackson R (1998) Cardiovascular disease risk factors in 65-84 year old men and women: results from the Auckland University heart and health study 1993-4. New Zealand Medical Journal, 111:4-7.

Bunker CH, Ukoli FA, Nwankwo MU et al. (1992) Factors associated with hypertension in Nigerian civil servants. Preventive Medicine, 21:710-722.

Burt VL, Culter JA, Higgins M et al. (1995) Trends in the prevalence, awareness, treatment, and control of hypertension in the adult US population. Data from the health examination surveys, 1960 to 1991. Hypertension, 26:60-69.
Capewell S, Beaglehole R, Seddon M, McMurray J (2000) Explanation for the decline in coronary heart disease mortality rates in Auckland, New Zealand, between 1982 and 1993. Circulation, 102:1511-1516.
Capewell S, Morrison CE, McMurray JJ (1999) Contribution of modern cardiovascular treatment and risk factor changes to the decline in coronary heart disease mortality in Scotland between 1975 and 1994. Heart 81:380-386.
Carter AB (1970) Hypotensive therapy in stroke survivors. The Lancet, 1: 485-489.

Carvalho JJ, Baruzzi RG, Howard PF et al. (1989) Blood pressure in four remote populations in the INTERSALT study. Hypertension, 14:238-246.
Cashin-Hemphill L, Holmvang G, Chan RC, Pitt B, Dinsmore RE, Lees RS (1999) Angiotensin-converting enzyme inhibition as antiatherosclerotic therapy: no answer yet. Quiet investigators. Quinapril ischemic event trial. American Journal of Cardiology, 83:43-74.
Chow CM, Muyinda MB (1999) Cardiovascular diseases in the People's Republic of China: the patterns and challenges. ProCor. Available at http://www. procor.org/procor-hma/msg00406.html.
Collins R, Peto R (1994) Antihypertensive drug therapy: Effects on stroke and coronary heart disease. In: Textbook of hypertension. Swales J, ed. Blackwell Scientific Publications, Oxford.

Collins R, Peto R, MacMahon S et al. (1990) Blood pressure, stroke, and coronary heart disease. Part 2, short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. The Lancet, 335:827-838.

Connor WE, Cerqueira MT, Connor RW, Wallace RB, Malinow MR, Casdorph HR (1978) The plasma lipids, lipoproteins, and diet of the Tarahumara Indians of Mexico. American Journal of Clinical Nutrition, 31:11311142.

Coope J, Warrender TS (1986) Randomised trial of treatment of hypertension in elderly patients in primary care. British Medical Journal, 293:1145-1151.

Cornejo C, Vinueza R, Moscoso F et al. (2000) Prevalence of arterial hypertension in the urban adult population of Ecuador: Quito, Guayaquil, and Cuenca-the PREHTAE study. Cardiovascular Prevention, 3:47-58.
Costa E, Rose GA, Klein CH et al. (1990) Salt and blood pressure in Rio Grande do Sul, Brazil. Bulletin of the Pan American Health Organization, 24:159176.

Cox DR (1972) Regression models and life tables (with discussion). Journal of the Royal Statistical Society, 34:187.
Dahlof B, Lindholm LH, Hansson L, Schersten B, Ekbom T, Wester PO (1991) Morbidity and mortality in the Swedish trial in old patients with hypertension (STOP-hypertension). The Lancet, 338:1281-1285.

Davis CE, Pajak A, Rywik S et al. (1994) Natural menopause and cardiovascular disease risk factors. The Poland and US collaborative study on cardiovascular disease epidemiology. Annals of Epidemiology, 4:445-448.

Dobson AJ, McElduff P, Heller R, Alexander H, Colley P, D’Este K (1999) Changing patterns of coronary heart disease in the hunter region of New South Wales, Australia. Journal of Clinical Epidemiology, 52:761771.

Domanski MJ, Davis BR, Pfeffer MA, Kastantin M, Mitchell GF (1999) Isolated systolic hypertension: prognostic information provided by pulse pressure. Hypertension, 34:375-380.

Edwards R, Unwin N, Mugusi F et al. (2000) Hypertension prevalence and care in an urban and rural area of Tanzania. Journal of Hypertension, 18:145-152.
Flight RJ, McKenzie-Pollock M, Hamilton MA, Salmond CE, Stokes YM (1984) The health status of fourth form students in Northland. New Zealand Medical Journal, 97:1-6.

Fogari R, Zoppi A, Marasi G et al. (1997) The epidemiology of resting heart rate in a male working population: association with blood pressure, age, smoking habits and other cardiovascular risk factors. Journal of Cardiovascular Risk, 4:209-213.
Foster C, Rotimi C, Fraser H et al. (1993) Hypertension, diabetes, and obesity in Barbados: findings from a recent population-based survey. Ethnicity and Disease, 3:404-412.

Franklin SS (1999) Ageing and hypertension: the assessment of blood pressure indices in predicting coronary heart disease. Journal of Hypertension, 17:S29-36.

Franklin SS, Gustin Wt, Wong ND et al. (1997) Hemodynamic patterns of agerelated changes in blood pressure. The Framingham heart study. Circulation, 96:308-315.

Franklin SS, Khan SA, Wong ND, Larson MG, Levy D (1999) Is pulse pressure useful in predicting risk for coronary heart disease? The Framingham heart study. Circulation, 100:354-360.

Franklin SS, Larson MG, Khan SA et al. (2001) Does the relation of blood pressure to coronary heart disease risk change with aging? The Framingham heart study. Circulation, 103:1245-1249.
Frohlich ED (2000) Recognition of systolic hypertension for hypertension. Hypertension, 35:1019-1020.
Fuchs SC, Petter JG, Accordi MC et al. (2001) Establishing the prevalence of hypertension. Influence of sampling criteria. Arquivos Brasileiros de Cardiologia, 76:445-452.
Garcia-Palmieri MR, Costas R Jr. (1986) Risk factors of coronary heart disease: a prospective epidemiologic study in Puerto Rico. In: Progress in cardiology. Vol 14. Goodwin JF, Yu PH, eds. Lea and Febiger, Philadelphia, PA.
Garraway WM, Whisnant JP (1987) The changing pattern of hypertension and the declining incidence of stroke. Journal of the American Medical Association, 258:214-217.

Ghannem H, Darioli R, Limam K et al. (2001) Epidemiology of cardiovascular risk factors among schoolchildren in Sousse, Tunisia. Journal of Cardiovascular Risk, 8:87-91.

Ghannem H, Hadj-Fredj A (1997) Prevalence of cardiovascular risk factors in the urban population of Sousse in Tunisia. Journal of Public Health Medicine, 19:392-396.
Gilberts EC, Arnold MJ, Grobbee DE (1994) Hypertension and determinants of blood pressure with special reference to socioeconomic status in a rural South Indian community. Journal of Epidemiology and Community Health, 48: 258-261.

Giles WH, Pacque M, Greene BM et al. (1994) Prevalence of hypertension in rural West Africa. American Journal of the Medical Sciences, 308:271-275.

Gliksman MD, Dwyer T, Wlodarczyk J (1990) Differences in modifiable cardiovascular disease risk factors in Australian schoolchildren: the results of a nationwide survey. Preventive Medicine, 19:291-304.
Gofin J, Kark JD, Abramson JH, Epstein L (1995) Trends in blood pressure levels over time in middle-aged and elderly Jerusalem residents. European Heart Journal, 16:1988-1994.
Gonzalez-Villalpando C, Stern MP, Haffner SM, Gonzalez Villapando ME, Gaskill S, Rivera Martinez D (1999) Prevalence of hypertension in a Mexican population according to the sixth report of the joint national committee on prevention, detection, evaluation and treatment of high blood pressure. Journal of Cardiovascular Risk, 6:177-181.
Gupta R, Guptha S, Gupta VP, Prakash H (1995) Prevalence and determinants of hypertension in the urban population of Jaipur in western India. Journal of Hypertension, 13:1193-1200.
Haheim LL, Holme I, Hjermann I, Leren P (1993) The predictability of risk factors with respect to incidence and mortality of myocardial infarction and total mortality. A 12-year follow-up of the Oslo study, Norway. Journal of Internal Medicine, 234:17-24.

He J, Klag MJ, Whelton PK et al. (1991a) Migration, blood pressure pattern, and hypertension: the Yi migrant study. American Journal of Epidemiology, 134:1085-1101.

He J, Tell GS, Tang YC, Mo PS, He GQ (1991b) Effect of migration on blood pressure: the Yi people study. Epidemiology, 2:88-97.
He J, Whelton PK (1999) Elevated systolic blood pressure as a risk factor for cardiovascular and renal disease. Journal of Hypertension-Supplement, 17:S7-13.
He J, Whelton PK (2000) Selection of initial antihypertensive drug therapy. The Lancet, 356:1942-1943.
Heinemann L, Barth W, Hoffmeister H (1995) Trend of cardiovascular risk factors in the East German population 1968-1992. Journal of Clinical Epidemiology, 48:787-795.
Helgeland A (1980) Treatment of mild hypertension: a five year controlled drug trial. The Oslo study. American Journal of Medicine, 69:725-732.
Hoffmeister H, Mensink GB, Stolzenberg H (1994) National trends in risk factors for cardiovascular disease in Germany. Preventive Medicine, 23:197-205.

Hughes K, Yeo PP, Lun KC et al. (1990) Cardiovascular diseases in Chinese, Malays, and Indians in Singapore. II. Differences in risk factor levels. Journal of Epidemiology and Community Health, 44:29-35.
Huang Z, Wu X, Stamler J et al. (1994) A north-south comparison of blood pressure and factors related to blood pressure in the People's Republic of China: a report from the PRC-USA collaborative study of cardiovascular epidemiology. Journal of Hypertension, 12:1103-1112.

Hughes MD, Pocock SJ (1992) Within-subject diastolic blood pressure variability: implications for risk assessment and screening. Journal of Clinical Epidemiology, 45:985-998.

Hunter JM, Sparks BT, Mufunda J, Musabayane CT, Sparks HV, Mahomed K (2000) Economic development and women's blood pressure: field evidence from rural Mashonaland, Zimbabwe. Social Science and Medicine, 50: 773-795.

Hutchinson RG, Watson RL, Davis CE et al. (1997) Racial differences in risk factors for atherosclerosis. The ARIC study. Atherosclerosis risk in communities. Angiology, 48:279-290.

Ibrahim MM, Rizk H, Appel LJ et al. (1995) Hypertension prevalence, awareness, treatment, and control in Egypt. Results from the Egyptian national hypertension project (NHP). NHP investigative team. Hypertension, 26: 886-890.

Idahosa PE (1985) Blood pressure pattern in urban Edos. Journal of Hypertension—Supplement, 3:S379-381.
Idahosa PE (1987) Hypertension: an ongoing health hazard in Nigerian workers. American Journal of Epidemiology, 125:85-91.

Idahosa PE, Ilawole M (1984) Towards Nigeria's first national blood pressure survey: a pilot study of arterial blood pressure in members of the national youth service corps. Journal of Hypertension, Supplement, 2:S209-211.

Jadue L, Vega J, Escobar MC et al. (1999) Risk factors for non communicable diseases: methods and global results of the CARMEN program basal survey. Revista Medica de Chile, 127:1004-1013.
Jamrozik K, Hockey R (1989) Trends in risk factors for vascular disease in Australia. Medical Journal of Australia, 150:14-18.
Jeng JS, Lee TK, Chang YC et al. (1998) Subtypes and case-fatality rates of stroke: a hospital-based stroke registry in Taiwan (SCAN-IV). Journal of the Neurological Sciences, 156:220-226.

Joffres MR, Hamet P, Rabkin SW, Gelskey D, Hogan K, Fodor G (1992) Prevalence, control and awareness of high blood pressure among Canadian adults. Canadian heart health surveys research group. Canadian Medical Association Journal, 146:1997-2005.
Joseph JG, Prior IA, Salmond CE, Stanley D (1983) Elevation of systolic and diastolic blood pressure associated with migration: the Tokelau island migrant study. Journal of Chronic Diseases, 36:507-516.
Kannel WB (1996) Blood pressure as a cardiovascular risk factor: prevention and treatment. Journal of the American Medical Association, 275:15711576.

Kannel WB (1999) Historic perspectives on the relative contributions of diastolic and systolic blood pressure elevation to cardiovascular risk profile. American Heart Journal, 138:205-210.

Kannel WB (2000) Elevated systolic blood pressure as a cardiovascular risk factor. American Journal of Cardiology, 85:251-255.

Kannel WB, Belanger AJ (1991) Epidemiology of heart failure. American Heart Journal, 121:951-957.
Kannel WB, Gordon T, Schwartz MJ (1971) Systolic versus diastolic blood pressure and risk of coronary heart disease. The Framingham study. American Journal of Cardiology, 27:335-346.
Kannel WB, Schwartz MJ, McNamara PM (1969) Blood pressure and risk of coronary heart disease: the Framingham study. Diseases of the Chest, 56: 43-52.
Kannel WB, Wolf PA, Verter J, McNamara PM (1970) Epidemiologic assessment of the role of blood pressure in stroke. The Framingham study. Journal of the American Medical Association, 214:301-310.

Khalid ME, Ali ME, Ahmed EK, Elkarib AO (1994) Pattern of blood pressures among high and low altitude residents of southern Saudi Arabia. Journal of Human Hypertension, 8:765-769.
Khaw KT, Rose G (1982) Population study of blood pressure and associated factors in St Lucia, West Indies. International Journal of Epidemiology, 11:372-377.
Kim JS, Jones DW, Kim SJ, Hong YP (1994) Hypertension in Korea: a national survey. American Journal of Preventive Medicine, 10:200-204.

King H, Abdullaev B, Djumaeva S, Nikitin V, Ashworth L, Dobo MG (1998) Glucose intolerance and associated factors in the Fergana valley, Uzbekistan. Diabetic Medicine, 15:1052-1062.

King H, Collins A, King LF et al. (1985) Blood pressure in Papua New Guinea: a survey of two highland villages in the Asaro valley. Journal of Epidemiology and Community Health, 39:215-219.
Kitange HM, Swai AB, Masuki G, Kilima PM, Alberti KG, McLarty DG (1993) Coronary heart disease risk factors in sub-Saharan Africa: studies in Tanzanian adolescents. Journal of Epidemiology and Community Health, 47:303-307.

Klag MJ, He J, Whelton PK, Chen JY, Qian MC, He GQ (1993) Alcohol use and blood pressure in an unacculturated society. Hypertension, 22:365-370.
Klag MJ, Whelton PK, Randall BL et al. (1996) Blood pressure and end-stage renal disease in men. New England Journal of Medicine, 334:13-18.

Lang T, Pariente P, Salem G, Tap D (1988) Social, professional conditions and arterial hypertension: an epidemiological study in Dakar, Senegal. Journal of Hypertension, 6:271-276.
Leppala JM, Virtamo J, Fogelholm R, Albanes D, Heinonen OP (1999) Different risk factors for different stroke subtypes: association of blood pressure, cholesterol, and antioxidants. Stroke, 30:2535-2540.
Leren P, Helgeland A (1986) Oslo hypertension study. Drugs, 31:S41-45.
Lichtenstein MJ, Shipley MJ, Rose G (1985) Systolic and diastolic blood pressures as predictors of coronary heart disease mortality in the Whitehall study. British Medical Journal, 291:243-245.

Lindeberg S, Nilsson-Ehle P, Terent A, Vessby B, Schersten B (1994) Cardiovascular risk factors in a Melanesian population apparently free from stroke and ischaemic heart disease: the Kitava study. Journal of Internal Medicine, 236:331-340.

Linjer E, Hansson L (1997) Underestimation of the true benefits of antihypertensive treatment: an assessment of some important sources of error. Journal of Hypertension, 15:221-225.
Lisk DR, Williams DE, Slattery J (1999) Blood pressure and hypertension in rural and urban Sierra Leoneans. Ethnicity and Disease, 9:254-263.
Lloyd-Jones DM, Evans JC, Larson MG, O’Donnell CJ, Levy D (1999) Differential impact of systolic and diastolic blood pressure level on JNC-VI staging. Joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. Hypertension, 34:381-385.

MacAuley D, McCrum EE, Stott G et al. (1996) Physical activity, physical fitness, blood pressure, and fibrinogen in the Northern Ireland health and activity survey. Journal of Epidemiology and Community Health, 50:258-263.

MacKenzie JM (1996) Intracerebral haemorrhage. Journal of Clinical Pathology, 49:360-364.

MacMahon S (1994) Blood pressure and risks of cardiovascular disease. In: Textbook of hypertension. Swales J, ed. Blackwell Scientific Publications, Oxford.

MacMahon S, Neal B (2000) Differences between blood-pressure-lowering drugs. The Lancet, 356:352-353.
MacMahon S, Peto R, Cutler J et al. (1990) Blood pressure, stroke, and coronary heart disease. Part I, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. The Lancet, 335:765-774.
MacMahon S, Rodgers A (1993a) The effects of anti-hypertensive treatment on vascular disease: reappraisal of the evidence in 1994. Journal of Vascular Medicine and Biology, 4:265-271.
MacMahon S, Rodgers A (1993b) The effects of blood pressure reduction in older patients: an overview of five randomized controlled trials in elderly hypertensives. Clinical and Experimental Hypertension (New York), 15: 967-978.

MacMahon S, Sharpe N, Gamble G et al. (2000) Randomized, placebocontrolled trial of the angiotensin-converting enzyme inhibitor, ramipril, in patients with coronary or other occlusive arterial disease. PART-2 collaborative research group. Prevention of atherosclerosis with ramipril. Journal of the American College of Cardiology, 36:438-443.
Mann GV, Shaffer RD, Anderson RS, Sandstead HH (1964) Cardiovascular disease in the Masai. Journal of Atherosclerosis, 4:289-312.
Mann JI, Lewis B, Shepherd J et al. (1988) Blood lipid concentrations and other cardiovascular risk factors: distribution, prevalence, and detection in Britain. British Medical Journal Clinical Research Ed., 296:1702-1706.

Masia R, Pena A, Marrugat J et al. (1998) High prevalence of cardiovascular risk factors in Gerona, Spain, a province with low myocardial infarction incidence. REGICOR investigators. Journal of Epidemiology and Community Health, 52:707-715.

M'Buyamba-Kabangu JR, Fagard R, Lijnen P et al. (1986) Epidemiological study of blood pressure and hypertension in a sample of urban Bantu of Zaire. Journal of Hypertension, 4:485-491.
M’Buyamba-Kabangu JR, Fagard R, Staessen J, Lijnen P, Amery A (1987) Correlates of blood pressure in rural and urban Zaire. Journal of Hypertension, 5:371-375.

McCarron P, Okasha M, McEwen J, Smith GD (2001) Changes in blood pressure among students attending Glasgow University between 1948 and 1968: analyses of cross sectional surveys. British Medical Journal, 322:885-889.
McGovern PG, Pankow JS, Shahar E et al. (1996) Recent trends in acute coronary heart disease-mortality, morbidity, medical care, and risk factors. The Minnesota Heart Survey Investigators. New England Journal of Medicine, 334:884-890.

Mendis S, Ranasinghe P, Dharmasena BD (1988) Prevalence of hypertension in Sri Lanka. A large population study in the central province. Public Health, 102:455-462.
Miall WE (1982) Systolic or diastolic hypertension-which matters most? Clinical and Experimental Hypertension-Part A, Theory and Practice, 4:11211131.

Ministry of Health (1999) NZ food: NZ people. Key results of the 1997 national nutrition survey. Ministry of Health, Wellington.

Misra A, Sharma R, Pandey RM, Khanna N (2001) Adverse profile of dietary nutrients, anthropometry and lipids in urban slum dwellers of northern India. European Journal of Clinical Nutrition, 55:727-734.

Mitchell GF, Moye LA, Braunwald E et al. (1997) Sphygmomanometrically determined pulse pressure is a powerful independent predictor of recurrent events after myocardial infarction in patients with impaired left ventricular function. SAVE investigators. Survival and Ventricular Enlargement. Circulation, 96:4254-4260.

Miura K, Dyer AR, Greenland P et al. (2001) Pulse pressure compared with other blood pressure indexes in the prediction of 25 -year cardiovascular and allcause mortality rates: The Chicago Heart Association Detection Project in Industry Study. Hypertension, 38:232-237.
Mufunda J, Scott LJ, Chifamba J et al. (2000) Correlates of blood pressure in an urban Zimbabwean population and comparison to other populations of African origin. Journal of Human Hypertension, 14:65-73.
Murray CJL, Lopez AD, eds. (1996) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, Vol. 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Nan L, Tuomilehto J, Dowse G et al. (1991) Prevalence and medical care of hypertension in four ethnic groups in the newly-industrialized nation of Mauritius. Journal of Hypertension, 9:859-866.
NCHS (2002) The National Health and Nutrition Examination Survey (NHANES) III. National Center for Health Statistics. Available at http://www.cdc.gov/nchs/nhanes.htm.
Neal B, MacMahon S (1999) An overview of 37 randomised trials of blood pressure lowering agents among 270000 individuals. World Health OrganizationInternational Society of Hypertension Blood Pressure Lowering Treatment Trialists' Collaboration. Clinical and Experimental Hypertension (New York), 21:517-529.

Neal B, MacMahon S, Chapman N (2000) Effects of ace inhibitors, calcium antagonists, and other blood-pressure-lowering drugs: results of prospectively designed overviews of randomised trials. Blood Pressure Lowering Treatment Trialists' Collaboration. The Lancet, 356:1955-1964.
Nyarko NK, Adubofour KO, Ofei F, Pobee JO, Owusu SK (1994) Serum lipids and lipoprotein in adult Ghanaians. Journal of Internal Medicine, 236: 251-253.

Ogunlesi A, Osotimehin B, Abbiyessuku F et al. (1991) Blood pressure and educational level among factory workers in Ibadan, Nigeria. Journal of Human Hypertension, 5:375-380.

Okayama A, Ueshima H, Marmot MG, Nakamura M, Kita Y, Yamakawa M (1993) Changes in total serum cholesterol and other risk factors for cardio-
vascular disease in Japan 1980-1989. International Journal of Epidemiology, 22:1038-1047.

Okesina AB, Oparinde DP, Akindoyin KA, Erasmus RT (1999) Prevalence of some risk factors of coronary heart disease in a rural Nigerian population. East African Medical Journal, 76:212-216.
Oviasu VO, Okupa FE (1980a) Arterial blood pressure and hypertension in Benin in the equatorial forest zone of Nigeria. Tropical and Geographical Medicine, 32:241-244.
Oviasu VO, Okupa FE (1980b) Relation between hypertension and occupational factors in rural and urban Africans. Bulletin of the World Health Organization, 58:485-489.

Page LB, Damon A, Moellering RC, Jr. (1974) Antecedents of cardiovascular disease in six Solomon Islands societies. Circulation, 49:1132-1146.

Pahor M, Psaty BM, Alderman MH et al. (2000) Health outcomes associated with calcium antagonists compared with other first-line antihypertensive therapies: a meta-analysis of randomised controlled trials. The Lancet, 356: 1949-1954.

Patrick RC, Prior IA, Smith JC, Smith AH (1983) Relationship between blood pressure and modernity among Ponapeans. International Journal of Epidemiology, 12:36-44.
Paul O, Lepper M, Phelan W, Dupertuis G, MacMillan A, McKean H (1963) A longitudinal study of coronary heart disease. Circulation, 28:20-31.
Pauletto P, Caroli M, Pessina AC, Dal Palu C (1994) Hypertension prevalence and age-related changes of blood-pressure in semi-nomadic and urban Oromos of Ethiopia. European Journal of Epidemiology, 10:159-164.

People's Republic of China-United States Cardiovascular and Cardiopulmonary Epidemiology Research Group (1992) An epidemiological study of cardiovascular and cardiopulmonary disease risk factors in four populations in the People's Republic of China. Baseline report from the PRC-USA collaborative study. Circulation, 85:1083-1096.
Perry HM, Jr. (1977) Treatment of mild hypertension. Preliminary results of a two-year feasibility trial. Circulation Research, 40:I180-187.
Pitt B, Byington RP, Furberg CD et al. (2000) Effect of amlodipine on the progression of atherosclerosis and the occurrence of clinical events. Circulation, 102:1503-1510.

Poulter N, Khaw KT, Hopwood BE et al. (1984) Blood pressure and associated factors in a rural Kenyan community. Hypertension, 6:810-813.
Poulter N, Sever P (1994) Blood pressure in other populations. A. Low blood pressure populations and the impact of rural-urban migration. In: Textbook of hypertension. Swales J, ed. Blackwell Scientific Publications, Oxford.
Poulter NR, Khaw KT, Hopwood BE et al. (1990) The Kenyan Luo migration study: observations on the initiation of a rise in blood pressure. British Medical Journal, 300:967-972.

Poulter NR, Khaw KT, Sever PS (1988) Higher blood pressures of urban migrants from an African low-blood pressure population are not due to selective migration. American Journal of Hypertension, 1:S143-145.

Progress Collaborative Group (2001) Randomised trial of a perindoprilbased blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack. The Lancet, 358:1033-1041.

Prospective Studies Collaboration (PSC) (1995) Cholesterol, diastolic blood pressure, and stroke: 13000 strokes in 45000 people in 45 prospective cohorts. The Lancet, 346:1647-1653.
Psaty BM, Smith NL, Siscovick DS et al. (1997) Health outcomes associated with antihypertensive therapies used as first-line agents. A systematic review and meta-analysis. Journal of the American Medical Association, 277:739-745.

Puska P, Matilainen T, Jousilahti P et al. (1993) Cardiovascular risk factors in the Republic of Karelia, Russia, and in North Karelia, Finland. International Journal of Epidemiology, 22:1048-1055.

Ramirez MO, Pino CT, Furiasse LV, Lee AJ, Fowkes FG (1995) Paraguayan national blood pressure study: prevalence of hypertension in the general population. Journal of Human Hypertension, 9:891-897.
Ribeiro AB, Ribeiro MB (1986) Epidemiological and demographic considerations. Hypertension in underdeveloped countries. Drugs, 31:S23-28.
Ribeiro MB, Ribeiro AB, Neto CS et al. (1981) Hypertension and economic activities in São Paulo, Brazil. Hypertension, 3:II-233-237.
Rosenthal J (1989) The epidemiology of blood pressure in young Mexican adults. Journal of Hypertension, 7:355-360.

Sakata K, Labarthe DR (1996) Changes in cardiovascular disease risk factors in three Japanese national surveys 1971-1990. Journal of Epidemiology, 6:93-107.

Salmond CE, Joseph JG, Prior IA, Stanley DG, Wessen AF (1985) Longitudinal analysis of the relationship between blood pressure and migration: the Tokelau Island migrant study. American Journal of Epidemiology, 122: 291-301.

Salmond CE, Prior IA, Wessen AF (1989) Blood pressure patterns and migration: a 14-year cohort study of adult Tokelauans. American Journal of Epidemiology, 130:37-52.
SarrafZadegan N, AminiNik S (1997) Blood pressure pattern in urban and rural areas in Isfahan, Iran. Journal of Human Hypertension, 11:425-428.

Seedat YK, Seedat MA, Hackland DB (1982) Prevalence of hypertension in the urban and rural Zulu. Journal of Epidemiology and Community Health, 36:256-261.

Sesso HD, Stampfer MJ, Rosner B et al. (2000) Systolic and diastolic blood pressure, pulse pressure, and mean arterial pressure as predictors of cardiovascular disease risk in men. Hypertension, 36:801-807.
Sever PS, Gordon D, Peart WS, Beighton P (1980) Blood-pressure and its correlates in urban and tribal Africa. The Lancet, 2:60-64.

Shelley E, Daly L, Collins C et al. (1995) Cardiovascular risk factor changes in the Kilkenny health project. A community health promotion programme. European Heart Journal, 16:752-760.
Shelley E, Daly L, Kilcoyne D, Graham I, Mulcahy R (1991) Risk factors for coronary heart disease: a population survey in county Kilkenny, Ireland, in 1985. Irish Journal of Medical Science, 160:S22-28.

Shimamoto T, Komachi Y, Inada H et al. (1989) Trends for coronary heart disease and stroke and their risk factors in Japan. Circulation, 79:503-515.
Sigfusson N, Sigvaldason H, Steingrimsdottir L et al. (1991) Decline in ischaemic heart disease in Iceland and change in risk factor levels. British Medical Journal, 302:1371-1375.

Simmons D (1983) Blood pressure, ethnic group, and salt intake in Belize. Journal of Epidemiology and Community Health, 37:38-42.
Simmons D, Barbour G, Congleton J et al. (1986) Blood pressure and salt intake in Malawi: an urban rural study. Journal of Epidemiology and Community Health, 40:188-192.

Singh RB, Beegom R, Ghosh S et al. (1997a) Epidemiological study of hypertension and its determinants in an urban population of North India. Journal of Human Hypertension, 11:679-685.
Singh RB, Beegom R, Mehta AS et al. (1998) Prevalence and risk factors of hypertension and age-specific blood pressures in five cities: a study of Indian women. NKP Salve Institute of Medical Sciences, Nagpur, India. Five city study group. International Journal of Cardiology, 63:165-173.
Singh RB, Sharma JP, Rastogi V, Niaz MA, Singh NK (1997b) Prevalence and determinants of hypertension in the Indian social class and heart survey. Journal of Human Hypertension, 11(1):51-56.
Smith WM (1977) Treatment of mild hypertension: results of a ten-year intervention trial. Circulation Research, 40:198-105.

Smith WC, Tunstall-Pedoe H, Crombie IK, Tavendale R (1989) Concomitants of excess coronary deaths--major risk factor and lifestyle findings from 10359 men and women in the Scottish Heart Health Study. Scottish Medical Journal, 34:550-555.
Soyannwo MA, Kurashi NY, Gadallah M et al. (1998) Blood pressure pattern in Saudi population of Gassim. African Journal of Medicine and Medical Sciences, 27:107-116.

Sprafka JM, Burke GL, Folsom AR, Luepker RV, Blackburn H (1990) Continued decline in cardiovascular disease risk factors: results of the Minnesota Heart Survey, 1980-1982 and 1985-1987. American Journal of Epidemiology, 132:489-500.
Staessen J, Fagard R, Thijs L, Celis H et al. (1997) Randomised double-blind comparison of placebo and active treatment for older patients with isolated systolic hypertension. The Lancet, 350:757-764.
Stamler J, Stamler R, Neaton JD (1993) Blood pressure, systolic and diastolic, and cardiovascular risks. US population data. Archives of Internal Medicine, 153:598-615.

Steyn K, Fourie J, Lombard C, Katzenellenbogen J, Bourne L, Jooste PL (1996) Hypertension in the black community of the Cape Peninsula, South Africa. East African Medical Journal, 73:758-763.

Steyn K, Jooste PL, Langenhoven ML et al. (1985) Coronary risk factors in the coloured population of the Cape Peninsula. South African Medical Journal, 67:619-625.

Stokes J 3rd., Kannel WB, Wolf PA, D’Agostino RB, Cupples LA (1989) Blood pressure as a risk factor for cardiovascular disease. The Framingham study30 years of follow-up. Hypertension, 13:I13-18.
Strachan D, Rose G (1991) Strategies of prevention revisited: Effects of imprecise measurement of risk factors on the evaluation of "high-risk" and "population-based" approaches to prevention of cardiovascular disease. Journal of Clinical Epidemiology, 44:1187-1196.

Suh I (2001) Cardiovascular mortality in Korea: a country experiencing epidemiologic transition. Acta Cardiologica, 56:75-81.

Swai AB, McLarty DG, Kitange HM et al. (1993) Low prevalence of risk factors for coronary heart disease in rural Tanzania. International Journal of Epidemiology, 22:651-659.
Swales J (1994) Textbook of hypertension. Blackwell Scientific Publications, Oxford.

Teo KK, Burton JR, Buller CE et al. (2000) Long-term effects of cholesterol lowering and angiotensin-converting enzyme inhibition on coronary atherosclerosis: the Simvastatin/Enalapril coronary atherosclerosis trial (SCAT). Circulation, 102:1748-1754.

Tuomilehto J, Bonita R, Stewart A, Nissinen A, Salonen JT (1991) Hypertension, cigarette smoking, and the decline in stroke incidence in eastern Finland. Stroke, 22:7-11.
Ueshima H, Zhang XH, Choudhury SR (2000) Epidemiology of hypertension in China and Japan. Journal of Human Hypertension, 14:765-769.
Vaccarino V, Borgatta A, Gallus G, Sirtori CR (1995) Prevalence of coronary heart disease risk factors in northern-Italian male and female employees. European Heart Journal, 16:761-769.
van der Sande MA, Bailey R, Faal H et al. (1997) Nationwide prevalence study of hypertension and related non-communicable diseases in the Gambia. Tropical Medicine and International Health, 2:1039-1048.

Vartiainen E, Jousilahti P, Alfthan G, Sundvall J, Pietinen P, Puska P (2000) Cardiovascular risk factor changes in Finland, 1972-1997. International Journal of Epidemiology, 29:49-56.

Vartiainen E, Puska P, Pekkanen J, Tuomilehto J, Jousilahti P (1994) Changes in risk factors explain changes in mortality from ischaemic heart disease in Finland. British Medical Journal, 309:23-27.
Warlow C, Dennis M, VanGijn J et al. (1996) Stroke. A practical guide to management. Blackwell Science Ltd, Oxford.

Whelton P, He J, Klag M (1994) Blood pressure in westernised populations. In: Textbook of hypertension. Swales J, ed. Blackwell Scientific Publications, Oxford.

Whelton PK (1994) Epidemiology of hypertension. The Lancet, 344:101-106.
Whelton PK, Klag MJ (1989) Hypertension as a risk factor for renal disease. Review of clinical and epidemiological evidence. Hypertension, 13:I19-27.
Wietlisbach V, Paccaud F, Rickenbach M, Gutzwiller F (1997) Trends in cardiovascular risk factors (1984-1993) in a Swiss region: results of three population surveys. Preventive Medicine, 26:523-533.
Wolff FW, Lindeman RD (1966) Effects of treatment in hypertension. Results of a controlled study. Journal of Chronic Diseases, 19:227-240.
Wu X, Huang Z, Stamler J et al. (1996) Changes in average blood pressure and incidence of high blood pressure 1983-1984 to 1987-1988 in four population cohorts in the People's Republic of China. The PRC-USA Cardiovascular and Cardiopulmonary Epidemiology Research Group. Journal of Hypertension, 14:1267-1274.
Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G (2000) Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The heart outcomes prevention evaluation study investigators. New England Journal of Medicine, 342:145-153.
Appendix A: Sampling methods, response rate and measuring techniques in blood pressure studies included
IN THE REVIEW
Africa

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | Gambia | van der Sande et al. (1997) | Multi-stage stratified cluster sampling based on the 1993 census | 94.9\% | ?Trained \& certified staff <br> Automated electronic sphg ?cuff <br> 2 measures taken ?position after ?mins rest, ?arm |
|  | Ghana | Nyarko et al. (1994) | Employees from the University of Ghana were randomly selected | Not stated | Trained \& certified staff ?sphg ?cuff <br> ?measure taken ?position after ?mins rest, ?arm |
|  | Liberia | Giles et al. (1994) | House-to-house survey of adults in rural area | 84\% | ?Trained \& certified staff <br> Automated digital sphg ?cuff <br> 2 measures taken sitting after 5 mins rest, $L$ arm |
|  | Mauritius | Nan et al. (1991) | Population clusters were selected randomly by 2stage process based on census | $\begin{aligned} & 83.4-88.9 \% \\ & (M \text { and } F) \end{aligned}$ | ?Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Nigeria | Akinkug (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
|  | Nigeria | Bunker et al. (1992) | Sample of urban civil servants from six ministries of Bendel State | Not stated | Trained \& certified staff Standard sphg various cuffs 3 measures taken sitting after 5 mins rest, ?arm |
|  | Nigeria | Idahosa (1985) | Sample of males from the Edos tribal group in Benin | Not stated | Trained \& certified staff Automated sphg various cuffs 3 measures taken sitting after 5 mins rest, ?arm |

Africa (continued)

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Nigeria | Idahosa (1987) | Cross sectional survey in Bendel, Nigeria | Not stated | Trained \& certified staff <br> Automated digital sphg one cuff <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Nigeria | Idahosa and Llwawole (1984) | Survey of Corps members stationed in Bendel State | Not stated | Trained \& certified staff Standard sphg one cuff 3 measures taken sitting after 5 mins rest, R arm |
|  | Nigeria | Ogunlesi et al. (1991) | Sample taken from male employees at a factory in Ibadan, Nigeria | Not stated | Trained \& certified staff Automated sphg various cuffs ?measure taken sitting after 5 mins rest, R arm |
|  | Nigeria | Okesina et al. (1999) | Houses in villages selected randomly, then individuals randomly selected | Not stated | ?Trained \& certified staff Standard sphg ?cuff ?measures taken sitting after ?mins rest, ?arm |
|  | Nigeria | Oviasu and Okupa (1980a, 1980b) | Cross sectional survey of workers in state capital secretariat | 95\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, $L$ arm |
|  | Senegal | Astagneau et al. (1992) | Random inclusion of households using a town map | 96\% <br> households then 86\% individuals | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ?arm |
|  | Senegal | Lang et al. (1988) | Sample taken from crosssectional survey of employees of six companies in Dakar, Senegal | 97\% | Trained \& certified staff Standard mercury sphg ?cuff ?measure taken sitting after 5 mins rest, ?arm |
|  | Seychelles | Bovet et al. (1991) | Age and sex strata random sampling from national census | 84-89\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |


|  | Sierra Leone | Lisk et al. (1999) | Multistage random sampling from census | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, $L$ arm |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-E | Democratic <br> Republic of the Congo | M'Buyamba-Kabangu et al. (I986, I987) | Rural population sampled from one village and compared to random sample of urban population | Not stated | Trained \& certified staff Standard aneroid sphg one cuff 5 measures taken sitting after 5 mins rest, R arm |
|  | Ethiopia | Pauletto et al. (1994) | From the outpatient department of a rural hospital | Not stated | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Kenya | INTERSALT study (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Kenya | Poulter et al. (1984) | Random from census | Not stated | Trained \& certified staff <br> Standard sphg ?cuff <br> 2 measures taken ?position after 5 mins rest, ?arm |
|  | Malawi | Simmons et al. (1986) | Two cross sectional surveys of an urban estate and "typical" rural villages | 96.9\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | South Africa | CORIS study (P. Elliot, personal communication, 2001) | A stratified sample by age and sex from census data | Not stated | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | South Africa | Mollen (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff <br> Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |

Africa (continued)

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | South Africa | Morbid (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
|  | South Africa | Seedat (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
|  | South Africa | Seedat et al. (1982) | House-to-house study of urban and rural population | Not stated | Trained \& certified staff Standard aneroid sphg various cuffs 3 measures taken sitting after 5 mins rest, ?arm |
|  | South Africa | Steyn et al. (1985) | A stratified sample by age and sex from census data | Not stated | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | South Africa | Steyn et al. (1996) | A stratified proportional sample from target population using census data | Not stated | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | United Republic of Tanzania | Edwards et al. (2000) | Randomized cluster sampling based on households from censusall adults in house | $\begin{aligned} & \text { 80-85\% (Shari) } \\ & \text { 64-79\% (Ilala) } \end{aligned}$ | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |


|  | United Republic of Tanzania | Kitange et al. (1993) | Whole population in four villages, random sample in four | 60-94\% | ?Trained \& certified staff <br> Standard sphg ?cuff <br> 2 measures taken sitting after 5 mins rest, R arm |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | United Republic of Tanzania | Swai et al. (1993) | Community based survey of eight villages in three areas in rural United Republic of Tanzania | 90.9-96\% | Trained \& certified staff <br> Standard mercury sphg ?cuff <br> 2 measures taken sitting after ?mins rest, R arm |
|  | Zimbabwe | Allain and Matenga (T. Allain, personal communication, 2001) | Stratified sampling | Not stated | Trained \& certified staff Standard sphg ? various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Zimbabwe | Hunter et al. (2000) | From health facilities | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Zimbabwe | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Zimbabwe | Mufunda et al. (2000) | Random sampling of household clusters, then sampled adults according to "probability proportionate to size"over sampling of older age groups | 56\% of potentially eligible, 79\% of those contacted | Trained \& certified staff Automated and standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | ails not given. |  |  |  |  |

Americas

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AMR-A | Canada | Joffres et al. (1992) | A probability sample was selected from the health insurance registries in each province | 64-69\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Halifax | MONICA study <br> (Anonymous 1989b) | Sample from public health service register | 4I-63\% | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Labrador | INTERSALT study (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | St Johns | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | USA | Sprafka et al. (1990) | Clusters of 1000 households in the metropolitan area were randomly selected | 68.1\% | Trained \& certified staff RZ sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | USA | Hutchinson et al. (1997) | Random sample from study centres | Not stated | Trained \& certified staff Standard sphg various cuffs 3 measures taken sitting after 5 mins rest, R arm |
|  | USA (NHANES III) | Burt et al. (1995); NHANES web site (NCHS 2002) | Stratified multistage probability sample designs | 77\% | Trained \& certified staff Standard sphg various cuffs 3 measures taken sitting after 5 mins rest, R arm |
|  | Chicago | INTERSALT study (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Goodman | INTERSALT study (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |


| Goodman | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample <br> representative of population | Not stated |
| :--- | :--- | :--- | :--- |

Americas (continued)

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Barbados | Ischib (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
|  | Belize | Simmons et al. (1983) | Sample of seven largest villages in Belize road, Cayo district, using electoral roll | 92\% | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 10 mins rest, L arm |
|  | Brazil | Costa et al. (1990) | Random survey of residents in Rio Grande do Sul | 94.4\% | ?Trained \& certified staff <br> Standard sphg ?cuff <br> ?2 measures taken sitting after 5 mins rest, ?arm |
|  | Brazil | Fuchs et al. (2001); F.D. Fuchs, personal communication, 2001 | Representative survey of urban population in Porto Alegre | 93\% | Trained \& certified staff <br> Aneroid sphg one cuff <br> 2 measures taken sitting after 5 mins rest, ? arm |
|  | Brazil | Ribeiro and Ribeiro (1986); Ribeiro et al. (198I) | Random selection of individuals from city of São Paulo | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Xingu | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Yanomamo | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff <br> RZ sphg various cuffs <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Chile | Barrios (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |


| Chile | L. Jadue, personal communication, 2001; Jadue et al. (1999) | Stratified sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
| :---: | :---: | :---: | :---: | :---: |
| Colombia | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Jamaica | Miall (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
| Jamaica | R. Wilks et al. personal communication, 2001 | Stratified sampling from Sparush Town, Jamaica | 64\% | Trained \& certified staff Standard sphg various cuff sizes 2 measures taken sitting after 5 mins rest, R arm |
| Mexico | Gonzalez-Villalpando et al. (1999); C. <br> Gonzalez, personal communication, 2001 | Random sample of population from a selected area-a complete household survey was performed | 80.2\% | Trained \& certified staff RZ sphg various cuffs 3 measures taken sitting after 5 mins rest, $R$ arm |
| Mexico | INTERSALT study (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Mexico | Rosenthal (1989) | Random sample of students enrolled in Mexico City college system | 97\% | Trained \& certified staff Standard sphg one cuff 3 measures taken sitting after 5 mins rest, R arm |
| Mexico | L. Yamamoto, personal communication, 2001 | Random stratified sample in Mexico city | 87\% | Trained \& certified staff Standard sphg various cuff sizes 2 measures taken sitting after 5 mins rest, ? arm |

Americas (continued)

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Paraguay | Ramirez et al. (1995) | Multistage sampling process, within country | Not stated | ?Trained \& certified staff <br> Aneroid sphg ?cuff <br> 2 measures taken sitting after 10 mins rest, either arm |
|  | Saint Lucia | Khaw and Rose (1982) | A random sample, stratified by age and sex from 2 of the 7 valleys, based on census data | 97\% | ?Trained \& certified staff RZ sphg ?cuff <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Trinidad and Tobago | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Trinidad and Tobago | Tobago (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff <br> Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
|  | Uruguay | Latir (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
| AMR-D | Ecuador | Cornejo et al. (2000) | Multistage stratified sample in three cities in Ecuador | 99\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 10 mins rest, R arm |

[^17]Eastern Mediterranean

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
| EMR-B | Iran (Islamic Republic of) | SarrafZadegan and AminiNik (1997) | Random cluster sampling | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 10 mins rest, R arm |
|  | Jordan | Jaddou HY (H. Jaddou, personal communication, 2001) | Stratified sample from the population | 62\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Saudi Arabia | Khalid et al. (1994) | Recruited from two areas with the assistance of the head of tribes | Not stated | Trained \& certified staff Standard sphg ?cuff 3 measures taken sitting after 10 mins rest, $R$ arm |
|  | Saudi Arabia | Soyannwo et al. (1998) | Cross-sectional house-tohouse population survey | 73.6\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Tunisia | Ghannem and HadjFredj (1997); H. Ghannem, personal communication, 2001 | Random stratified sample of adult urban population using household survey | 71.1\% | Trained \& certified staff Electronic sphg various cuffs 3 measures taken sitting after 15 mins rest, ?arm |
|  | Tunisia | Ghannem et al. (2001); H. Ghannem, personal communication, 2001 | Random stratified sample of schoolchildren from the urban region of Sousse | 95.4\% | Trained \& certified staff Electronic sphg various cuffs 2 measures taken sitting after 10 mins rest, ?arm |
| EMR-D | Egypt | Ashour et al. (1995); Ibrahim et al. (1995) | Multistage probability sample of clusters of households | 85\% | Trained \& certified staff standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| ? Specific details not given. |  |  |  |  |  |

Europe

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
| EUR-A | Belgium, Charleroi | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Charleroi | MONICA study <br> (Anonymous 1989b) | Sample from population register | 59\% | Trained \& certified staff RZ sphg various cuffs <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Ghent | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Ghent | MONICA study (Anonymous 1989b) | Sample from public health service register | 54-57\% | Trained \& certified staff <br> RZ sphg various cuffs <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Belgium; Luxembourg | MONICA study <br> (Anonymous 1989b) | Sample from electoral register | 54-54\% | Trained \& certified staff RZ sphg one cuff 2 measures taken lying after 5 mins rest, either arm |
|  | Former Czechoslovakia | Czech (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
|  | Denmark | Andersen (1994) | Sample from schools | Not stated | ?Trained \& certified staff <br> Standard sphg ?cuff <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Denmark | INTERSALT study <br> (Anonymous 1989a) | Community sample | Not stated | Trained \& certified staff RZ sphg various cuffs <br> 2 measures taken sitting after 5 mins rest, R arm |


| Glostrup | MONICA study <br> (Anonymous 1989b) | Sample from population register | 79-80\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| :---: | :---: | :---: | :---: | :---: |
| Finland | Puska et al. (1993) | Stratified random sample from population registers | 68-81\% | Trained \& certified staff Standard sphg ?cuff 2 measures taken sitting after 15 mins rest, R arm |
| Finland | Vartiainen et al. (2000) | Sample from population register | 70-85\% | Trained \& certified staff ?sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |
| Joensuu | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Kuopio | MONICA study <br> (Anonymous 1989b) | Sample from population register | 85\% | Trained \& certified staff <br> Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
| Turku | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Turku/Loimaa | MONICA study <br> (Anonymous 1989b) | Sample from population register | 86\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
| N. Karelia | MONICA study <br> (Anonymous 1989b) | Sample from population register | 80\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
| France, Bas-Rhin | MONICA study <br> (Anonymous 1989b) | Sample from population register | 86\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |

Europe (continued)

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Lille | MONICA study <br> (Anonymous 1989b) | Sample from communities \& electoral roll | 65-80\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Former German Democratic Republic | INTERSALT study (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Berlin-Lichtenberg | MONICA study (Anonymous 1989b) | Sample from national X ray screening register | 70-80\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Cottbus county | MONICA study <br> (Anonymous 1989b) | Sample from national X ray screening register | 77\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Halle county | MONICA study (Anonymous 1989b) | Sample from national X ray screening register | 88\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Karl-Marx-Stadt | MONICA study (Anonymous 1989b) | Sample from national X ray screening register | 90\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Germany | Heinemann et al. (1995) | Random sample from urban and rural population | Not stated | Trained \& certified staff <br> Standard sphg ?cuff <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Germany | MONICA study (Anonymous 1989b) | Sample from national X ray screening register | 72\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Augsburg (Rural) | MONICA study (Anonymous 1989b) | Sample from population register | 82-84\% | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |


| Augsburg (Urban) | MONICA study (Anonymous 1989b) | Sample from population register | 76-80\% | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| :---: | :---: | :---: | :---: | :---: |
| Bernried | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Bremen | MONICA study (Anonymous 1989b) | Sample from resident register | 71\% | Trained \& certified staff <br> RZ sphg various cuffs <br> 2 measures taken sitting after 5 mins rest, $R$ arm |
| Heidelberg | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Rhein-Neckar | MONICA study <br> (Anonymous 1989b) | Sample from population register | 74\% | Trained \& certified staff <br> Automatic/standard sphg one cuff <br> 2 measures taken sitting after 5 mins rest, R arm |
| Iceland | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Iceland | MONICA study <br> (Anonymous 1989b) | Sample from national roster | 76\% | Trained \& certified staff <br> Standard sphg one cuff <br> 2 measures taken sitting after 5 mins rest, either arm |
| Italy | Fogari et al. (1997) | All men belonging to the same working community | Not stated | Trained \& certified staff Standard sphg various cuffs 3 measures taken sitting after 5 mins rest, R arm |
| Italy | Nine populations (Anonymous 198I) | Random sample of nine populations from eight areas | 62\% | Trained \& certified staff Standard sphg ?cuff 2 measures taken sitting after 4 mins rest, R arm |
| Italy | Vaccarino et al. (1995) | All employees of IBM asked to participate | 45\% | Trained \& certified staff ?sphg ?cuff <br> 2 measures taken sitting after 5 mins rest, R arm |

Europe (continued)

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Bassiano | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Brianza | MONICA study <br> (Anonymous 1989b) | Sample from population register | 70-71\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Friuli | MONICA study <br> (Anonymous 1989b) | Sample from regional health roll | 80-82\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Gubbio | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Latina | MONICA study <br> (Anonymous 1989b) | Sample from electoral rolls | 76\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Mirano | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Naples | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff <br> RZ sphg various cuffs <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Israel | Gofin et al. (1995) | All residents in a geographically defined area asked to participate | 85\% | Trained \& certified staff <br> Standard sphg ?cuff <br> 3 measures taken sitting after 5 mins rest, R arm |
|  | Israel | Kark (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |


| Malta | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| :---: | :---: | :---: | :---: | :---: |
| Malta | MONICA study (Anonymous 1989b) | Sample from electoral rolls | 62\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
| Netherlands | Bosma et al. (1994) | Sample taken from a 10 year follow-up to the Kaunus-Rotterdam Intervention Study (KRIS) | Not stated | Trained \& certified staff ?sphg ?cuff ?measures taken sitting after ?mins rest, ?arm |
| Netherlands | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Norway | Trumso (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
| Portugal | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Slovenia | Gradsk (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
| Spain | Masia et al. (1998) | Two-stage population random sample stratified by age from census data | 72.7\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, ?arm |
| Catalonia | MONICA study (Anonymous 1989b) | Sample from population register | 76-79\% | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |

Europe (continued)

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Manresa | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, ?arm |
|  | Torrejon | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Sweden | Asplund-Carlson and Carlson (1994) | Selected at random | 63\% | Trained \& certified staff standard sphg ?cuff ?measure taken sitting after no rest, ? arm |
|  | N. Sweden | MONICA study (Anonymous 1989b) | Sample from population register | 84-86\% | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Switzerland, Vaud; Fribourg | MONICA study (Anonymous 1989b) | Sample from population register | 61-69\% | Trained \& certified staff RZ sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | United Kingdom | British Regional Heart Study (P. Elliot, personal communication, 2001) | Random sample from general practice registers | 77\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | United Kingdom | Mann et al. (1988) | Opportunistic case finding from GP lists and random sampling from age-sex registers | 73\% | Trained \& certified staff standard sphg ?cuff <br> 2 measures taken sitting after 5 mins rest, ?arm |
|  | England | Bajekal et al. (1999) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |


|  | Birmingham | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Northern Ireland, Belfast | MONICA study <br> (Anonymous 1989b) | Sample from GP lists | 57-70\% | Trained \& certified staff RZ sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Scotland | Hawth (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff <br> Standard sphg one cuff <br> 2 measures taken sitting after 5 mins rest, ? arm |
|  | Scotland | Smith et al. (1989) | Sample taken from selected districts then random sampling from GP lists | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Scotland, Glasgow | MONICA study <br> (Anonymous 1989b) | Sample from GP lists | 50-64\% | Trained \& certified staff <br> RZ sphg one cuff <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Wales, South Wales | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Ireland | MacAuley et al. (1996) | Random sample | 70\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, $L$ arm |
|  | Ireland | Shelley et al. (1995, 1991) | Random sample from electoral roll in two communities | 70-75\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| EUR-B | Poland | Davis et al. (1994) | A stratified random sample of the population | Not stated | Trained \& certified staff Standard sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Krakow | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |

Europe (continued)

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Warsaw | INTERSALT study (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Warsaw | MONICA study <br> (Anonymous 1989b) | Sample from electoral register | 74\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Uzbekistan | King et al. (1998) | Sample of semi-rural and urban population taken from house-to-house census | Not stated | Trained \& certified staff <br> Standard mercury sphg ?cuff <br> ?measure taken sitting after 10 mins rest, R arm |
|  | Former Yugoslavia | MONICA study <br> (Anonymous 1989b) | Sample from population register | 82\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| EUR-C | Hungary | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Budapest | MONICA study <br> (Anonymous 1989b) | Sample from population register | 75-80\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, either arm |
|  | Lithuania | Bosma et al. (1994) | Sample taken from a 10 year follow-up to the Kaunus-Rotterdam Intervention Study (KRIS) | Not stated | Trained \& certified staff ?sphg ?cuff <br> ?measure taken sitting after ?mins rest, ?arm |


| Russian Federation | Puska et al. (1993) | Stratified random sample from population registers | 77-92\% | Trained \& certified staff Standard sphg ?cuff 2 measures taken sitting after 15 mins rest, R arm |
| :---: | :---: | :---: | :---: | :---: |
| The former Soviet Union ${ }^{\text {a }}$ | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff <br> RZ sphg various cuffs <br> 2 measures taken sitting after 5 mins rest, R arm |
| Kaunas | MONICA study <br> (Anonymous 1989b) | Sample from GP lists | 70\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Moscow | MONICA study <br> (Anonymous 1989b) | Sample from GP lists | 78\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Moscow | MONICA study <br> (Anonymous 1989b) | Sample from GP lists | 67\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
| Novosibirsk | MONICA study <br> (Anonymous 1989b) | Sample from electoral list | 69\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
| Novosibirsk | MONICA study <br> (Anonymous 1989b) | Sample from electoral list | 73\% | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
| ? Specific details not given. |  |  |  |  |
| a Russia in the original publication. |  |  |  |  |

South-East Asia

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
| SEAR-B | Sri Lanka | Mendis et al. (1988) | Everybody visiting patients at a teaching hospital in Sri Lanka | Not stated | Trained \& certified staff Standard sphg ?cuff I measure taken sitting after 5 mins rest, ?arm |
|  | Sri Lanka | Mohidn (P. Elliot, personal communication, 2001) | Not stated | Not stated | Trained \& certified staff ?standard sphg ?cuff <br> 2 measures taken sitting after 5 mins rest, ?arm |
| SEAR-D | India | B.V. Babu, personal communication, 200I; Y.S. Kusuma, personal communication, 2001 | Random sampling in the state of Andhra Pradesh | 98\% | Trained \& certified staff standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | India | B.V. Babu, personal communication, 200I; Y.S. Kusuma, personal communication, 2001 | Multistage sampling from rural and urban areas | 99\% | Trained \& certified staff standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | India | Gilberts et al. (1994) | Cross-sectional survey conducted door-to-door in a rural south Indian community | Not stated | Trained \& certified staff standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | India | Gupta et al. (1995) | Randomly chosen wards from different areas of the city | $\begin{aligned} & 57.3-87.9 \% \\ & \text { (F-M) } \end{aligned}$ | Trained \& certified staff standard sphg ?cuff 2 measures taken sitting after 5 mins rest, ?arm |


Western Pacific

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
| WPR-A | Australia | APCSC-Busselton (APCSC secretariat, personal communication, 2001) | A stratified random sample representative of population | Not stated | Trained \& certified staff standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Australia | APCSC-Perth (APCSC secretariat, personal communication, 2001) | Stratified random sampling of Perth metropolitan area | >70\% | Trained \& certified staff standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Australia | Gliksman et al. (1990) | Two-stage probability sampling of all school children in Australia | Not stated | Trained \& certified staff standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Australia | Jamrozik and Hockey (1989); P. Elliot, personal communication, 2001 | Random sample from electoral register | 75\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Newcastle | MONICA study <br> (Anonymous 1989b) | Sample from electoral register | 68-82\% | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Perth | MONICA study <br> (Anonymous 1989b) | Sample from electoral register | 81-84\% | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Japan | 1990 National Survey (Sakata and Labarthe 1996) | Randomly selected from National Health Survey districts | 81.5\% | Trained \& certified staff standard sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |


|  | APCSC-Aito Town <br> (APCSC secretariat, <br> personal <br> communication, | Random sample of Aito <br> town residents | 75\% |
| :--- | :--- | :--- | :--- |

Western Pacific (continued)

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate |
| :---: | :---: | :--- | :--- | :--- |


|  | Singapore | APCSC-Kinmen (APCSC secretariat, personal communication, 2001) | All registered residents aged $\geq 50$ years in two townships invited to participate | 77.4\% | Trained \& certified staff standard sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Singapore | Hughes et al. (1990) | Random samples of census districts, units, houses, then individuals (weighted) | 52-66\% | Trained \& certified staff <br> Standard sphg one cuff <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | Singapore | Singapore National Health survey (P. Elliot, personal communication, 2001) | Random selection of households from national database, then random selection of individuals | 45-65\% | Trained \& certified staff Standard sphg one cuff 3 measures taken sitting after 5 mins rest, R arm |
|  | Singapore | APCSC-Singapore <br> Heart Survey <br> (APCSC secretariat, personal communication, 2001) | Random sampling of population of Singapore | Not stated | Trained \& certified staff standard sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |
| WPR-B | China | APCSC-Anzhen (APCSC secretariat, personal communication, 2001) | Stratified random sampling and cluster sampling in Beijing area | 85\% | Trained \& certified staff standard sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | China | APCSC-Anzhen02 (APCSC secretariat, personal communication, 2001) | Cluster sampling | Not stated | Trained \& certified staff standard sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |

Western Pacific (continued)

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | China | APCSC-East Beijing (APCSC secretariat, personal communication, 2001) | Cluster sampling in individuals in East district of Beijing | Not stated | Trained \& certified staff standard sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | China | APCSC-Hong Kong SAR (APCSC secretariat, personal communication, 2001) | Stratified random sampling in Hong Kong SAR | 85\% | Trained \& certified staff standard sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | China | APCSC-Hong Kong SAR (APCSC secretariat, personal communication, 2001) | Stratified random sampling in Hong Kong SAR | 85\% | Trained \& certified staff standard sphg ?cuff <br> 2 measures taken sitting after 5 mins rest, R arm |
|  | China | APCSC-Seven Cities cohort (APCSC secretariat, personal communication, 2001) | Study samples selected from Chinese populations of farmers, workers and fishermen | Not stated | Trained \& certified staff Standard sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | China | APCSC-Six Cohorts (APCSC secretariat, personal communication, 2001) | Cluster sampling of six cohorts in villages, including farmers of Shanxi, Shaanxi, Guangxi, Jiangsu province, minors of Hebei province and fishermen of Zhejiang | $\geq 85 \%$ | Trained \& certified staff standard sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |


| China | APCSC-Tianjin <br> (APCSC secretariat, <br> personal <br> communication, | A stratified random sample <br> drawn from urban areas of <br> the city | Not stated |
| :--- | :--- | :--- | :--- |

Western Pacific (continued)

| Subregion | Country or area | Study (reference) | Sampling methods | Response rate | BP measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Taiwan, China | APCSC-CVDFACTS/ <br> Two Townships (APCSC secretariat, personal communication, 2001) | Two townships in Taiwan, China | Not stated | Trained \& certified staff standard sphg ?cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | Taiwan, China | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Republic of Korea | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, R arm |
|  | Republic of Korea | Kim et al. (1994) | Randomly selected districts then households and individuals | 81.8\% | Trained \& certified staff Standard sphg various cuffs 2 measures taken sitting after 5 mins rest, $L$ arm |
|  | Papua New Guinea | INTERSALT study <br> (Anonymous 1989a) | A stratified random sample representative of population | Not stated | Trained \& certified staff RZ sphg various cuffs 2 measures taken sitting after 5 mins rest, $R$ arm |
|  | Papua New Guinea | King et al. (1985) | Purposive cluster samples from six chosen communities | 60-85\% | Trained \& certified staff RZ sphg ?cuff <br> 2 measures taken sitting after 10 mins rest, R arm |


|  | Papua New Guinea | Lindeberg et al. (1994) | Randomly selected within specific age range during a period of seven weekssome non-randomized subjects included due to low participation | $45-63 \%$ <br> overall <br> 40-59\% for <br> BP | Trained \& certified staff <br> Standard sphg various cuffs <br> 3 measures taken sitting after ?min rest, ?arm |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Pacific | PACIFIC (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
|  | Pacific | Patrick et al. (1983) | A census of all households was conducted in three communities selected for sampling | 90.2\% | Trained \& certified staff <br> Zero-muddler sphg various cuffs <br> 2 measures taken sitting after ?min rest, ? arm |
|  | Pacific | PUKAPK (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
|  | Pacific | RARTNG (P. Elliot, personal communication, 2001) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, ? arm |
| Key: | Specific details not given. Sphg, sphygmomanometer; RZ, | om zero. |  |  |  |

## Chapter 7

# High cholesterol 

Carlene M.M. Lawes, Stephen Vander Hoorn, Malcolm R. Law and Anthony Rodgers

## Summary

Cholesterol is a fat-like substance, found in the blood stream and also in bodily organs and nerve fibres. While there are different etiological roles for various types of cholesterol, such as high and low density lipoprotein, the large majority of descriptive and epidemiological data are available only for total cholesterol levels. Therefore, in this analysis, cholesterol was defined as total serum cholesterol expressed in millimoles per litre of blood ( $\mathrm{mmol} / \mathrm{l}$ ) a continuous variable with mean and standard deviation.

The primary outcomes assessed were ischaemic heart disease (Global Burden of Disease [GBD] study end-point 107) and non-fatal stroke (end-point 108). Ischaemic heart disease (IHD) was chosen on the basis of clear and consistent positive associations observed in cohort studies and evidence of reversibility in clinical trials of cholesterol lowering treatments. Cholesterol is positively associated with ischaemic stroke, but has a qualitatively different association with haemorrhagic stroke. As endpoints must all be mapped to the GBD classification system for disease, total stroke was used in the analyses. However, application of relative risk estimates to the 14 subregions ${ }^{1}$ were adjusted to reflect differences in stroke subtypes. Cholesterol has been observed to be inversely associated with a number of other outcomes such as cancer and chronic respiratory disease. However, evidence suggests that these associations are due to the effects of disease on cholesterol, rather than vice versa. Consequently these outcomes were not included.

Raw cholesterol data were obtained from studies after a systematic review of population-based surveys, which included about 160 surveys and almost 640000 participants. Sex-specific associations of cholesterol with age were estimated for each of the subregions separately, based on country-weighted study estimates of mean values. There was moderate variation in the final age- and sex-specific estimates of mean cholesterol
across the 14 subregions, with the range between the highest and lowest age-specific mean cholesterol levels typically being about $2 \mathrm{mmol} / \mathrm{l}$.

A theoretical-minimum-risk distribution of cholesterol (i.e. one that would yield the lowest population risk of adverse health outcomes) was taken as 3.8 standard deviation $0.6 \mathrm{mmol} / \mathrm{l}$ (usual) for all age, sex and subregional groups. The main basis for this estimate was the level of cholesterol down to which epidemiological relationships with cardiovascular disease outcomes are observed, and clinical trial data showing benefits from cholesterol lowering among those with below-average cholesterol levels. This theoretical minimum was also consistent with the levels of cholesterol in populations with little cardiovascular disease.

For the comparative risk assessment (CRA) analyses, observational epidemiological data were used to estimate the hazard ratios for the risk factor-disease relationship and trial meta-analyses for the time frame of risk reversibility. Data on hazard ratios for IHD were included from an overview of the ten largest observational studies conducted in populations from industrialized countries. This overview included data from 494804 participants followed for 7-23 years, among whom 18811 IHD events were observed. There was evidence of differences in the strength of the association by age but no difference between males and females. These data were closely similar in size and shape to the associations seen in an overview of 29 cohorts involving 353065 participants from the Asia-Pacific region. In this individual participant meta-analysis, 2937 strokes were observed as well as 2838 IHD events. Overall, each $0.6 \mathrm{mmol} / \mathrm{l}$ lowering of usual cholesterol was associated a $27 \%$ reduction in IHD. A $1 \mathrm{mmol} / / \mathrm{lowering}$ of cholesterol was associated with a $13 \%$ (range $6-19 \%$ ) reduction in total stroke, predominantly due to the effect on ischaemic stroke.

An overview of 49 trials of cholesterol lowering indicated that risk reductions in IHD of $11 \%, 24 \%, 33 \%$ and $36 \%$ are associated with a $1 \mathrm{mmol} / \mathrm{l}$ reduction in cholesterol after $1,2,3-5$ and $>5$ years, respectively. Taking into account the size of the reduction in cholesterol, there was no clear difference in effects according to how the cholesterol reduction was achieved, either by diet or any one of several classes of drugs. These data indicated that in middle age, the risks associated with high cholesterol are reversed within a few years of cholesterol lowering. A meta-analysis of cholesterol lowering trials also confirmed a reduction in risk of stroke with cholesterol lowering.

The foregoing methods and assumptions allowed estimates of burden attributable to cholesterol levels of more than $3.8 \mathrm{mmol} / \mathrm{l}$. Overall, $56 \%$ of IHD mortality and disease burden was attributable to cholesterol $>3.8 \mathrm{mmol} / \mathrm{l}$ worldwide (range of $44-68 \%$ by subregion). This translated into 3609000 deaths in the year 2000 . Also $32 \%$ of ischaemic stroke was attributable to cholesterol $>3.8 \mathrm{mmol} / \mathrm{l}$ worldwide (range of $25-45 \%$ by subregion), which translated into 805000 deaths in the year 2000 .

Worldwide, 4.4 million deaths (about $7.9 \%$ of the total) and 40.4 million disability-adjusted life years (DALYs) $(2.8 \%$ of the total) were estimated to be due to non-optimal cholesterol. The proportion of DALYs was lower than the proportion of deaths as most cholesterolrelated cardiovascular disease events occurred in middle or old age, so the years of life lost due to premature mortality and years of life lost due to disability is a smaller percentage of the total DALYs than of the total deaths. Overall, the results suggest that a considerable proportion of cardiovascular disease is attributable to non-optimal cholesterol, defined as mean cholesterol $>3.8 \mathrm{mmol} / 1$, and this translates into an important portion of deaths and years of life lost to deaths and disability worldwide. Approximately $40 \%$ of the cholesterol-related attributable burden occurred in developed subregions, $20 \%$ in low mortality developing subregions (AMR-B, EMR-B, EUR-B, EUR-C, SEAR-B, WPR-B), and a further $40 \%$ in high mortality developing subregions (AFR-D, AFR-E, AMR-D, EMR-D, SEAR-D).

In absolute terms, most of the excess burden of cardiovascular disease occurred in the populous regions, as would be expected. The relative impact of attributable deaths and DALYs-calculated as the proportion of all deaths and DALYs attributable to non-optimal cholesterol (i.e. mean $>3.8 \mathrm{mmol} / \mathrm{l}$-was highest in the European subregions, where mean cholesterol levels were highest.

## 1. Introduction

Cholesterol is a fat-like substance, found in the blood stream and also in bodily organs and nerve fibres. Most cholesterol in the body is made by the liver from a wide variety of foods, but especially from saturated fats, such as those found in animal products. A diet high in saturated fat content, heredity, and various metabolic conditions such as type II diabetes, influence an individual's level of cholesterol.

### 1.1 Choice of exposure variable

Cholesterol was described as a continuous variable, as it is commonly reported in this manner with mean and standard deviation values. The System Internationale (SI) unit for measuring cholesterol is millimoles per litre of blood ( $\mathrm{mmol} / \mathrm{l}$ ), but in some studies, particularly those from the United States of America and Japan, it is reported in milligrams per decilitre of blood (conversion factor $1 \mathrm{mg} / \mathrm{dl}=0.02586 \mathrm{mmol} / \mathrm{l}$ ). Total cholesterol was chosen in preference to other potential measures of risk associated with blood lipids (e.g. high density lipoprotein [HDL cholesterol] and low density lipoprotein [LDL cholesterol]) as more data and information were available on both risk factor levels and the risk-factor-disease relationship (relative risk values) for total cholesterol.

### 1.2 Disease outcomes

A number of diseases are associated with non-optimal cholesterol levels. Cholesterol is thought to amplify and accelerate atherosclerosis, and influence IHD and ischaemic stroke events, but the exact mechanisms are unclear. It has been proposed that cholesterol, particularly LDL cholesterol which accounts for about $60 \%$ of total cholesterol in the circulation, is taken up by macrophages. When cholesterol levels are high, macrophages take up more cholesterol than they can metabolize and become "foam cells". These cells are important in the early stages of atheromatous plaque formation (Bronner et al. 1995; Gorelick et al. 1997; Warlow et al. 1996).

By contrast, there is no positive association between cholesterol and haemorrhagic stroke, and some research indicates that those people with lower levels of cholesterol are at greater risk of haemorrhagic stroke (Leppala et al. 1999; Neaton et al. 1992; Yano et al. 1994). The mechanisms are unclear, but low levels of cholesterol may weaken the endothelium of small intracranial vessels which, in combination with high blood pressure, may rupture, and/or result in "osmotic fragility" of red blood cells, thereby increasing the risk of haemorrhage (Bronner et al. 1995; Iso et al. 1989; Tanaka et al. 1982; Warlow et al. 1996). Potential outcomes considered as affected by cholesterol are described below.

## ISCHAEMIC HEART DISEASE

Data from prospective cohort studies have demonstrated a strong, continuous temporal association between cholesterol and IHD (APCSC 2003; Law et al. 1994b, 1994c; Lewington and MacMahon 1999; Neaton and Wentworth 1992; Neaton et al. 1992). Further, a causal association is biologically plausible and clinical trials have demonstrated reversibility (Bucher et al. 1998; LaRosa et al. 1999; Law et al. 1994b; Pignone et al. 2000; Ross et al. 1999). The GBD classification system for diseases and injuries has a category for total IHD (end-point 107 G3 ischaemic heart disease, the International Statistical Classification of Diseases, ninth revision [ICD-9] codes 410-414) (Bucher et al. 1998; LaRosa et al. 1999; Pignone et al. 2000; Ross et al. 1999).

## Stroke

Data on the association between cholesterol and stroke are more complex. Most major cohort studies and overviews have shown that cholesterol is positively associated with ischaemic stroke (Anonymous 1998a; Neaton et al. 1992; PSC 1995). However, there appears to be a qualitatively different association with haemorrhagic stroke, with some studies observing a negative association (APCSC 2003; Neaton et al. 1992) and others a null association with this outcome (Suh et al. 2001). There are no specific GBD end-points for stroke subtypes, only for total
stroke (end-point 108 G4 cerebrovascular disease, ICD-9 codes 430-438). However, application of RR estimates to the subregions will reflect differences in stroke subtypes, as discussed later in this chapter.

## Other disease and injury outcomes

Cholesterol has been observed to be inversely associated with a number of other outcomes such as cancer and chronic respiratory disease (Law et al. 1994a). However, evidence suggests that these associations are due to the effects of disease on cholesterol, rather than vice versa (Law et al. 1994a). Consequently these outcomes were not included.

## 2. Risk FACTOR EXPOSURE

### 2.1 Data on cholesterol levels

Data on global cholesterol levels were collated from three major sources. The first source was the MONICA study (Anonymous 1989), which collected cholesterol data from 39 collaborating centres in 22 countries and was carried out between 1979 and 1987. This study provides important information about cholesterol patterns, but does not provide a truly global overview of cholesterol distributions. It included populations that were predominantly European in origin, and did not include populations from the Eastern Mediterranean, South-East Asian or African Regions. Further, MONICA data were collected in the early 1980s, and more recent data are also available now. We therefore found it necessary to include additional cholesterol data for this analysis.

The second major source of data was a literature search using Medline and the key words "cholesterol", "survey" "health survey" and "crosssectional survey". Studies were reviewed, and included in analyses if they fulfilled the following criteria:

- conducted from 1980 onwards;
- included randomly selected or representative participants;
- had a sample size of over 1000 —however, a smaller sample size was acceptable in specific regions, or age groups where data were limited, if the study fulfilled the other criteria;
- described sample size and age group of participants;
- presented mean values of cholesterol by age and sex; and
- utilized a standard protocol for cholesterol measurement.

The final source of data was personal communications with researchers and study investigators. The authors had access to data from the Asia-Pacific Cohort Studies Collaboration (APCSC), a collaboration involving 37 cohorts in the Asia-Pacific region, which includes at least 5000 person-years of follow-up recorded or planned. Data on date of
birth or age, sex, and cholesterol have been recorded and collated (APCSC 1999). Cholesterol data from eligible studies that had been collected from 1980 onwards were included in the cholesterol database. In addition, authors of surveys/studies were contacted and age and sexspecific data requested, where these had not been available in the published format.

Many studies did not publish data in the format required for this project (e.g. omitting age- and sex-specific mean cholesterol levels), and unfortunately time and resource constraints limited attempts to obtain all of these data from researchers. It was also very difficult to obtain results of surveys that have only been published in local/national reports but not in peer-reviewed journals. Access to these data would have been greatly improved had these reports been more widely available in electronic formats such as on the Internet. Details of studies currently included are presented in Table 7.1.

In total, approximately 1900 abstracts were reviewed. Data from about 160 surveys (total sample size of almost 640000 participants) have been included. Figure 7.1 illustrates data coverage by geographical region.

No data were available for AMR-D comprising Bolivia, Ecuador, Guatemala, Haiti, Nicaragua and Peru, and very limited data (studies totalling $<2000$ participants) for SEAR-B and EMR-D. Data on more than 50000 participants were obtained for AMR-A, EUR-A, WPR-A and WPR-B, owing to many studies being available and large sample sizes.

Figure 7.I Cholesterol data coverage expressed as total study sample size, by geographical region


Table 7.I Studies currently included in the cholesterol data review

| Subregion | Country or area | Study (reference) | Sample size | Age range (years) |
| :---: | :---: | :---: | :---: | :---: |
| AFR-D | Ghana | Nyarko et al. (1994) | 79 | 20-50 |
|  | Nigeria | Erasmus et al. (1994) | 417 | 11->60 |
|  | Nigeria | Okesina et al. (1999) | 500 | $11-\geq 50$ |
|  | Seychelles | Bovet et al. (1991) | 1055 | 25-64 |
|  |  |  | 2051 |  |
| AFR-E | South Africa | Oelofse et al. (1996) | 986 | 15-64 |
|  | South Africa | Steyn et al. (1987, 1985) | 976 | 15-64 |
|  | United Republic of Tanzania | Kitange et al. (1993) | 1409 | 15-19 |
|  | United Republic of Tanzania | Swai et al. (1993) | 7272 | $15-\geq 65$ |
|  | Zimbabwe | Allain et Matenga (T. Allain et al. personal communication, 2001) | 261 | 60- $\geq 80$ |
|  |  |  | 10904 |  |
| AMR-A | Canada | Connelly et al. (1992) | 1169 | 42-59 |
|  | Canada | Lupien et al. (1985) | 1169 | 42-59 |
|  | Halifax | MONICA study (Anonymous 1989) | 837 | 25-64 |
|  | USA | Abbott et al. (1997) | 971 | 71-93 |
|  | USA | Brown et al. (1993) | 14521 | 45-64 |
|  | USA | Burke et al. (1991) | 4545 | 25-74 |
|  | USA | Donker et al. (1997) | 1411 | 7-11 |
|  | USA | Eisenberg et al. (1986) | 2477 | 35-64 |
|  | USA | Ettinger et al. (1992) | 4814 | $65-\geq 85$ |
|  | USA | Ferrara et al. (1997) | 2339 | $50-\geq 80$ |
|  | USA | Hutchinson et al. (1997) | 15743 | 45-64 |
|  | USA | Johnston et al. (1993) | 7836 | 20- $\geq 75$ |
|  | USA | Sprafka et al. (1990) | 4641 | 25-74 |
|  | USA | Srinivasan et al. (1991) | 4020 | 5-26 |
|  | USA | Wallace and Colsher (1992) | 1959 | $71-\geq 95$ |
|  | USA | Yano et al. (1986) | 1363 | 60- $\geq 75$ |
|  | Stanford | MONICA study (Anonymous 1989) | 1430 | 25-64 |
|  |  |  | 86995 |  |
| AMR-B | Brazil | INCLEN (Anonymous 1992b) | 406 | 35-65 |
|  | Chile | INCLEN (Anonymous 1992b) | 399 | 35-65 |
|  | Chile | Jadue et al. (1999); L. Jadue, personal communication, 2001 | 1591 | 25-64 |
|  | Colombia | INCLEN (Anonymous 1992b) | 200 | 35-65 |

Table 7.I Studies currently included in the cholesterol data review (continued)

| Subregion | Country or area | Study (reference) | Sample size | Age range (years) |
| :---: | :---: | :---: | :---: | :---: |
|  | Dominican Republic | Aono et al. (1997) | 2000 | 20-76 |
|  | Jamaica | Wilks et al. (2001) | 1134 | 25-74 |
|  | Mexico | Gonzalez et al. (1999, 2001) | 2251 | 35-64 |
|  | Mexico | Posadas-Romero et al. $(1995)$ | 33660 | 20-70 |
|  | Mexico | Yamamoto et al. (2001) | 825 | 20-90 |
|  |  |  | 42466 |  |
| EMR-B | Bahrain | al-Mahroos et al. (2000) | 2090 | 40-59 |
|  | Jordan | H. Jaddou, personal communication, 2001 | 2273 | $25-\geq 70$ |
|  | Kuwait | Olusi et al. (1997) | 751 | $1-\geq 70$ |
|  | Saudi Arabia | al-Nuaim et al. (1996, 1997) | 4539 | $15-\geq 60$ |
|  | Saudi Arabia | al-Nuaim (1997) | 2960 | 25-64 |
|  | Saudi Arabia | al Shammari et al. (1994) | 1005 | <35->65 |
|  | Saudi Arabia | Mitwalli et al. (1994) | 966 | $<25-\geq 55$ |
|  | Tunisia | Ghannem et al. (2001); <br> H. Ghannem, personal communication, 2001 | 1497 | 13-19 |
|  |  |  | 16081 |  |
| EMR-D | Pakistan | Molla et al. (1990) | 634 | 4-59 |
|  |  |  | 634 |  |
| EUR-A | Belgium | Kesteloot et al. (1987) | 18090 | <20->55 |
|  | Charleroi | MONICA study (Anonymous 1989) | 646 | 25-64 |
|  | Ghent | MONICA study <br> (Anonymous 1989) | 1260 | 25-64 |
|  | Denmark, Glostrup | MONICA study (Anonymous 1989) | 3780 | 25-64 |
|  | Finland | Lakka and Salonen (1992) | 2492 | 42-60 |
|  | Finland | Myllkangas et al. (1995) | 2210 | 45-64 |
|  | Finland | Nikkila and Heikkinen (1990) | 535 | 85 |
|  | Finland | Nissinen et al. (1987) | 6523 | 25-64 |
|  | Finland | Puska et al. (1993) | 1398 | 25-64 |
|  | Finland | Vartiainen et al. (2000) | 10025 | 30-59 |
|  | Finland | Vikari et al. (1985) | 3654 | 3-18 |
|  | Kuopio | MONICA study (Anonymous 1989) | 2789 | 25-64 |
|  | Turku; Loimaa | MONICA study (Anonymous 1989) | 3283 | 25-64 |
|  | N. Karelia | MONICA study (Anonymous 1989) | 3138 | 25-64 |

Table 7.I Studies currently included in the cholesterol data review (continued)

| Subregion | Country or area | Study (reference) | Sample size | Age range (years) |
| :---: | :---: | :---: | :---: | :---: |
| France, |  |  |  |  |
|  | Haute-Garonne | MONICA study (Anonymous 1989) | 1265 | 25-64 |
|  | Lille | MONICA study (Anonymous 1989) | 1436 | 25-64 |
| Former German |  |  |  |  |
|  | Democratic Republic, Berlin-Lichtenberg | MONICA study <br> (Anonymous 1989) | 1197 | 25-64 |
|  | Cottbus county | MONICA study (Anonymous 1989) | 1377 | 25-64 |
|  | Germany | Heinemann et al. (1995) | 1939 | 25-64 |
|  | Germany | Hoffmeister et al. (1994) | 15436 | 25-69 |
|  | Germany | Herman et al. (1988) | 1696 | 25-69 |
|  | Germany | MONICA study (Anonymous 1989) | 1015 | 25-64 |
|  | Augsburg (Rural) | MONICA study (Anonymous 1989) | 2109 | 25-64 |
|  | Augsburg (Urban) | MONICA study (Anonymous 1989) | 1667 | 25-64 |
|  | Bremen | MONICA study (Anonymous 1989) | 1625 | 25-64 |
|  | Rhein-Neckar | MONICA study (Anonymous 1989) | 3066 | 25-64 |
|  | Iceland | MONICA study (Anonymous 1989) | 1743 | 25-64 |
|  | Israel | Eisenberg et al. (1986) | 1588 | 35-64 |
|  | Israel | Greenland et al. (1993) | 1200 | 9-18 |
|  | Italy | Cesana et al. (1989) | 1387 | $25-\geq 55$ |
|  | Italy | Nine populations (Anonymous 1981) | 6699 | 20-59 |
|  | Italy | Salvaggio et al. (Law and Wald 1994; Salvaggio et al. 1991) | 8953 | 18-65 |
|  | Italy | Vaccarino et al. (1995) | 3401 | $<30-\geq 50$ |
|  | Brianza | MONICA study (Anonymous 1989) | 1647 | 25-64 |
|  | Friuli | MONICA study (Anonymous 1989) | 1849 | 25-64 |
|  | Latina | MONICA study (Anonymous 1989) | 1773 | 25-64 |
|  | Netherlands | Bosma et al. (1994) | 3015 | 45-70 |
|  | Netherlands | Vershuren et al. (1994) | 41622 | 20-59 |
|  | Norway | Graff-lverson et al. (1998) | 7523 | 40-54 |
|  | Norway | Thune et al. (1998) | 6307 | 20-49 |
|  | Spain | Masia et al. (1998) | 1670 | 24-74 |
|  | Catalonia | MONICA study (Anonymous 1989) | 2544 | 25-64 |

Table 7.I Studies currently included in the cholesterol data review (continued)

| Subregion | Country or area | Study (reference) | Sample size | Age range (years) |
| :---: | :---: | :---: | :---: | :---: |
|  | Sweden | Asplund-Carlson and Carlson (1994) | 1564 | 40-50 |
|  | Sweden | Rosengren et al. (2000) | 798 | 50 |
|  | Gothenburg | MONICA study (Anonymous 1989) | 1354 | 25-64 |
|  | Switzerland, Ticino | MONICA study (Anonymous 1989) | 1483 | 25-64 |
|  | Vaud; Fribourg | MONICA study (Anonymous 1989) | 1582 | 25-64 |
|  | United Kingdom | Brown et al. (1994) | 3939 | $70-\geq 80$ |
|  | England | Bajekal et al. (1999) | 11162 | $16-\geq 75$ |
|  | England | Razay et al. (1992) | 1218 | 40-69 |
|  | Northern Ireland, Belfast | MONICA study (Anonymous 1989) | 2327 | 25-64 |
|  | Scotland | Smith et al. (1989) | 10359 | 40-59 |
|  | Scotland, Glasgow | MONICA study (Anonymous 1989) | 1143 | 25-64 |
|  | Multiple sites in EUR-A | Kafatos et al. (1991) | 2114 | 70-79 |
|  |  |  | 237704 |  |
| EUR-B | Poland, Tarnobrzeg Voivodship | MONICA study (Anonymous 1989) | 2700 | 25-64 |
|  | Warsaw | MONICA study (Anonymous 1989) | 2591 | 25-64 |
|  | Turkey | Mahley et al. (1995); R. Mahley, personal communication, 2001 | 8882 | $20-\geq 70$ |
|  | Turkey | Onat et al. (1992) | 3687 | $30-\geq 70$ |
|  |  |  | 17860 |  |
| EUR-C | Former Czechoslovakia | MONICA study (Anonymous 1989) | 2552 | 25-64 |
|  | Estonia | Olferev et al. (1990, 1991) | 2936 | 20-54 |
|  | Hungary | Biro et al. (1996) | 2559 | $18-\geq 60$ |
|  | Hungary | Kafatos et al. (1991) | 42 | 70-79 |
|  | Budapest | MONICA study (Anonymous 1989) | 1486 | 25-64 |
|  | Pecs | MONICA study (Anonymous 1989) | 1584 | 25-64 |
|  | Lithuania | Bosma et al. (1994) | 2149 | 45-70 |
|  | Russian Federation | Puska et al. (1993) | 837 | 25-64 |
|  | The former Soviet Union ${ }^{\text {a }}$, Kaunas | MONICA study (Anonymous 1989) | 1462 | 25-64 |
|  | Novosibirsk C | MONICA study (Anonymous 1989) | 2538 | 25-64 |

Table 7.I Studies currently included in the cholesterol data review (continued)

| Subregion | Country or area | Study (reference) | Sample size | Age range (years) |
| :---: | :---: | :---: | :---: | :---: |
|  | Novosibirsk I | MONICA study (Anonymous 1989) | 1436 | 25-64 |
|  |  |  | 19581 |  |
| SEAR-B | Indonesia | INCLEN (Anonymous 1992b) | 210 | 35-65 |
|  | Thailand | Bhuripanyo et al. (1993) | 911 | $30-\geq 60$ |
|  | Thailand | INCLEN (Anonymous 1992b) | 416 | 35-65 |
|  |  |  | 1537 |  |
| SEAR-D | India | Chadha et al. (1997) | 2124 | 25-64 |
|  | India | Misra et al. (2001) | 508 | 20-70 |
|  | India | Misra et al. (2001) | 652 | 14-25 |
|  | India | Reddy et al. (1994) | 380 | $40-\geq 70$ |
|  |  |  | 3664 |  |
| WPR-A | Australia | APCSC-Busselton (APCSC secretariat, personal communication, 2001) | 976 | $15-\geq 70$ |
|  | Australia | APCSC-Perth (APCSC secretariat, personal communication, 2001) | 6456 | $15-\geq 70$ |
|  | Australia | Bennett and Magnus (1994) | 6096 | 25-64 |
|  | Australia | Boulton et al. (1995) | 856 | 8-9 |
|  | Australia | Glikman et al. (1990) | 1743 | 9-15 |
|  | Australia | Simons et al. (1991) | 3182 | 60->80 |
|  | Australia | van Beurden et al. (1991) | 9238 | 18-98 |
|  | Newcastle | MONICA study (Anonymous 1989) | 2396 | 25-64 |
|  | Perth | MONICA study (Anonymous 1989) | 1758 | 25-64 |
|  | Japan | 1990 National Survey (Sakata and Labarthe 1996) | 7836 | $30-\geq 70$ |
|  | Japan | APCSC-Aito Town (APCSC secretariat, personal communication, 2001) | 1720 | $15-\geq 70$ |
|  | Japan | APCSC-Akabane (APCSC secretariat, personal communication, 2001) | 1834 | $15-\geq 70$ |
|  | Japan | APCSC-Ohasama (APCSC secretariat, personal communication, 2001) | 2240 | $30-\geq 60$ |
|  | Japan | Choudhury et al. (1994) | 832 | 35-59 |
|  | Japan | Okayama et al. (1993) | 5921 | 30-69 |
|  | Japan | Serum Lipid Survey (Anonymous 1996) | 33234 | 4-99 |
|  | New Zealand | APCSC-Fletcher challenge (APCSC secretariat, personal communication, 2001) | 10462 | $15-\geq 70$ |

Table 7.I Studies currently included in the cholesterol data review (continued)

| Subregion | Country or area | Study (reference) | Sample size | Age range (years) |
| :---: | :---: | :---: | :---: | :---: |
|  | New Zealand | Bullen et al. (1998) | 986 | 65-84 |
|  | New Zealand | Flight et al. (1984) | 323 | 14-15 |
|  | New Zealand | Mann et al. (1991) | 2363 | 18-64 |
|  | New Zealand | National Survey (Ministry of Health 1999) | 3223 | $15-\geq 80$ |
|  | Singapore | APCSC-Singapore Heart Survey (APCSC secretariat, personal communication, 2001) | 2450 | $15-\geq 70$ |
|  | Singapore | Hughes et al. (1990) | 2058 | 18-69 |
|  |  |  | 109750 |  |
| WPR | China | APCSC-Anzhen02 (APCSC secretariat, personal communication, 200I) | 4152 | 30-69 |
|  | China | APCSC-Hong Kong SAR (APCSC secretariat, personal communication, 2001) | 2019 | $\geq 70$ |
|  | China | APCSC-Seven Cities cohort (APCSC secretariat, personal communication, 2001) | 37635 | $15-\geq 70$ |
|  | China | APCSC-Six Cohorts (APCSC secretariat, personal communication, 2001) | 19387 | 30-59 |
|  | China | INCLEN (Anonymous 1992b) | 1188 | 35-65 |
|  | China | Tao et al. (1992) | 4280 | 35-54 |
|  | China | Tian et al. (1995) | 631 | 15-64 |
|  | China | Yang et al. (1986) | 1054 | $15-\geq 55$ |
|  | China | Zhuang et al. (1986) | 4072 | 0-102 |
|  | Beijing | MONICA study (Anonymous 1989) | 1673 | 25-64 |
|  | Hong Kong SAR | Fong et al. (1994) | 696 | $20-\geq 60$ |
|  | Hong Kong SAR | Woo et al. (1997) | 1010 | 25-74 |
|  | Taiwan, China | APCSC-CVDFACTS/Two <br> Townships (APCSC <br> secretariat, personal <br> communication, 2001) | 7004 | $0-\geq 70$ |
|  | Papua New Guinea | Lindeberg et al. (1994) | 166 | 20-86 |
|  | Papua New Guinea | Scrimgeour et al. (1989) | 121 | 17-59 |
|  | Philippines | INCLEN (Anonymous 1992b) | 274 | 35-65 |
|  | 90057 |  |  |  |

[^18]Table 7.2 Cholesterol measuring techniques of studies included in this review

|  | Yes (\%) | Not stated (\%) |
| :--- | :---: | :---: |
| Trained staff | 94 | 6 |
| Approved laboratory ${ }^{\text {a }}$ | 94 | 6 |
| Fasting sample |  | 28 |
| All | 47 | NA |
| Some | 4 | NA |
| None | 21 | NA |
| Storage of sample | 16 | 43 |
| Refrigeration | 30 | NA |
| Deep freeze | 11 | NA |
| Storage not necessary | 56 | NA |
| Analysis method | 6 | 34 |
| Enzymatic | 4 | NA |
| Extraction | NA |  |
| Other | NA |  |

NA Not applicable.
a Laboratory associated with a hospital, university or research institution.

Approximately half of the studies utilized random sampling of individuals or households, including stratified random sampling; the other half used methods such as house-to-house or workplace surveys. Response rates were documented in $72 \%$ of the studies, of which response rate was $>80 \%$ in $32 \%$ of them, between $50 \%$ and $80 \%$ (including those where response was $>80 \%$ in some subcategories) in $63 \%$ of them, and documented to be lower than $50 \%$ in only six studies. For completeness, full documentation of sampling method, response rate and cholesterol measuring techniques are presented in Appendix A, which is intended only for those who require more data on individual studies. A summary is given in Table 7.2.

In most studies, a laboratory was associated with a hospital, university or research institution, and the staff appeared to have received appropriate training. Whether the blood sample was taken after the individual had fasted has an impact on lipid subfractions such as triglycerides rather than on total cholesterol, but this information also provides an indication of the consistency with which samples were taken, and the degree of detail given in publications. The majority of analyses were performed using newer enzymatic techniques, rather than the older extraction analyses.

### 2.2 Methodology to estimate mean and standard deviation of Cholesterol data

Mean cholesterol, standard deviation, sample size and age ranges from all reviewed data sources were extracted and entered into an Excel database. Scatter plots of these cholesterol data for each subregion, utilizing the midpoints in study age categories, are presented below in Figures 7.2-7.5.

The cholesterol data obtained from published studies including various age ranges were presented by age categories that were different from those required for the CRA project. For this reason, a method was needed which would make complete use of the available data and then combine them to produce estimates for subregions and age groups. The first step was to assess the general shape of association using all the available data. Second, associations were estimated for each subregion separately where sufficient data were available. For subregions with limited data, the association was estimated using complete data available from other subregions. Finally, these subregional age-cholesterol associations for males and females were used to predict subregional cholesterol levels.

It should be noted that no step specifically estimates mean cholesterol at the country level; this approach is explained in greater detail below.

## Age-Cholesterol relationships

The data from the literature were first used to assess the shape of the association of cholesterol with age using SPLUS software. At this stage no assumptions were made about the shape of association and therefore non-parametric methods were applied.

Figures 7.2 and 7.3 demonstrate that there were variable amounts of data available among the different subregions, so initially analyses were limited to those subregions with the most data spread across all age groups. The six subregions thus examined were AMR-A, AMR-B, EMRB, EUR-A, WPR-A and WPR-B. Previous analyses had demonstrated that the shape of the age-cholesterol relationship may vary depending on absolute level of cholesterol (APCSC secretariat, personal communication, 2001). Therefore, analyses of pooled data from the six subregions were stratified by overall cholesterol levels into plausible low $(<4.7 \mathrm{mmol} / \mathrm{l})$, medium ( $4.7-5.5 \mathrm{mmol} / \mathrm{l}$ ) and high ( $>5.5 \mathrm{mmol} / \mathrm{l}$ ) groups (Figure 7.6), as we describe below using the example of WPR-A.

The purpose of determining the overall shape of the association between cholesterol and age was to enable extrapolation to other subregions where only limited data were available. The method utilized for subregions with sparse data is detailed below. The associations between age and cholesterol for both females and males were non-linear. The association for females increased between the ages of 30 and 65 years, and then fell slightly afterwards. For males, the association increased

Figure 7.2 Scatter plots of mean cholesterol (mmol/l) against age by subregion, for females


Figure 7.3 Scatter plots of mean cholesterol (mmol/l) against age by subregion, for males


Figure 7.4 Scatter plots of mean cholesterol standard deviation (mmol/l) against age by subregion, for females


Figure 7.5 Scatter plots of mean cholesterol standard deviation (mmol/l) against age by subregion, for males


Figure 7.6 The cholesterol-age association from pooled data of six subregions ${ }^{\text {a }}$ for populations with high, medium and low cholesterol levels


Females
a AMR-A, AMR-B, EMR-B, EUR-A, WPR-A and WPR-B.
Note: Overall cholesterol levels-Low $=<4.7 \mathrm{mmol} / /$; Medium $=4.7-5.5 \mathrm{mmol} / /$; High $=>5.5 \mathrm{mmol} / \mathrm{l}$.
between the ages of 30 and 50 years, and then flattened before declining slightly in older age. Very little data were available in the populations aged $>80$ years, and therefore it is unclear whether cholesterol continues to decrease beyond the age of 70-79 years. These patterns were consistent among high, medium and low cholesterol subgroups; however, the gradient of the increase was greatest in the high cholesterol groups, and least in the low cholesterol groups.

Having established the non-linear association between age and mean cholesterol, the second step involved estimating the shape and level of this association for males and females in each subregion separately. The methodology differed by subregion depending on the available data. There were sufficient data to estimate the associations for males and females in six of the 14 subregions (AMR-A, AMR-B, EUR-A, EMR-B,

Figure 7.7 Combining cholesterol data (mmol/l) from country studies in WPR-A


WPR-A and WPR-B). For example, for males in WPR-A, data were available from four countries (Australia, Japan, New Zealand and Singapore). The scatter plot in Figure 7.7 presents these cholesterol data at the country level.

Cholesterol data differed among and between these countries, with the size of the plot circles proportional to the population size and study size where more than one study was available within a country (Figure 7.7). Cholesterol levels were consistently lower for Japan across all age groups. The curve of best fit through these data weighted by study size within each country and by country population size within each subregion was estimated (i.e. a weighted regression line). Clearly, the fitted curve is weighted mostly to the Japanese studies due to Japan having the largest population size.

Although there was significant variation in the overall cholesterol levels between the different subregions, the shape of association across age groups was consistent for the six subregions where sufficient data were available (Figure 7.8).

For six of the remaining subregions (AFR-D, AFR-E, EUR-B, EUR-C, SEAR-B and SEAR-D), insufficient data were available to draw the age-cholesterol association accurately. Within these subregions, the available cholesterol data were pooled together to give age standardized mean levels weighted by country. That is, a single mean cholesterol level was estimated for each subregion with sparse data for males and females separately. Based on this level, one of three "age-cholesterol curves" was

Figure 7.8 The association of cholesterol with age for males and females in six subregions

chosen (Figure 7.6), and this curve was then shifted (up or down) to correspond with the calculated age adjusted mean cholesterol level for that subregion. Using this methodology, the "low cholesterol" curve (overall cholesterol $<4.7 \mathrm{mmol} / \mathrm{l}$ ) was used for AFR-D and AFR-E, the "middle cholesterol" curve (overall cholesterol $4.7-5.5 \mathrm{mmol} / \mathrm{l}$ ) was used for EUR-B, SEAR-B and SEAR-D, and the "high cholesterol" curve (overall cholesterol $>5.5 \mathrm{mmol} / \mathrm{l}$ ) was used for EUR-C.

Of the two remaining subregions, no data were collected in this review for AMR-D, so data from AMR-B were used to estimate the mean cholesterol levels. In the one study for EMR-D, which included only 634 individuals, the overall mean cholesterol level was skewed by particularly high cholesterol levels in the oldest age group, compared to age patterns observed in other subregions. This was thought not to be representative of that subregion, therefore data from EMR-B were used to estimate cholesterol levels for EMR-D.

## Estimates of mean cholesterol and standard deviation

Estimates of mean cholesterol for each of the age groups were made using the age-cholesterol associations outlined above for each subregion and sex separately. Utilizing these associations, age-specific estimates for each subregion were obtained by using midpoints of each GBD age group and predicting the subregional average cholesterol levels. Cholesterol values for those aged $>80$ years were assumed to be the same as the $70-79$ age group due to the extremely limited availability of cholesterol levels in the populations studied with a mean age $>80$ years. (Figures 7.2 and 7.3 demonstrate that very few subregions had any data for those with a mean age of $>80$ years.) A similar approach was used to estimate the standard deviations for each age, sex and subregion category. Current mean cholesterol results are presented in Table 7.3, and standard deviations are in Table 7.4.

### 2.3 UnCERTAINTY OF MEAN AND STANDARD DEVIATION OF Cholesterol levels

The uncertainty of the means and standard deviations of population distributions of cholesterol also had to be estimated. The approach described in the previous section to estimate the means and standard deviations makes as complete use as possible of all available data.

Table 7.3 Estimates of mean cholesterol levels (mmol/l) by subregion, sex and age

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  |  | Males |  |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | 4.6 | 5.1 | 5.2 | 5.2 | 5.2 | 4.5 | 4.7 | 4.7 | 4.7 | 4.7 |
| AFR-E | 4.6 | 5.1 | 5.2 | 5.2 | 5.2 | 4.5 | 4.7 | 4.7 | 4.7 | 4.7 |
| AMR-A | 4.8 | 5.6 | 6.0 | 5.9 | 5.9 | 5.2 | 5.5 | 5.5 | 5.2 | 5.2 |
| AMR-B | 4.8 | 5.4 | 5.4 | 5.2 | 5.2 | 5.3 | 5.0 | 4.9 | 4.8 | 4.8 |
| AMR-D | 4.8 | 5.4 | 5.4 | 5.2 | 5.2 | 5.3 | 5.0 | 4.9 | 4.8 | 4.8 |
| EMR-B | 4.6 | 5.1 | 5.4 | 5.6 | 5.6 | 4.8 | 4.9 | 5.1 | 5.4 | 5.4 |
| EMR-D | 4.6 | 5.1 | 5.4 | 5.6 | 5.6 | 4.8 | 4.9 | 5.1 | 5.4 | 5.4 |
| EUR-A | 5.4 | 6.3 | 6.7 | 6.5 | 6.5 | 5.9 | 6.1 | 6.1 | 5.9 | 5.9 |
| EUR-B | 4.6 | 5.3 | 5.5 | 5.4 | 5.4 | 5.0 | 5.1 | 5.2 | 5.2 | 5.2 |
| EUR-C | 5.4 | 6.2 | 6.6 | 6.5 | 6.5 | 5.5 | 5.8 | 5.8 | 5.7 | 5.7 |
| SEAR-B | 4.0 | 4.5 | 4.6 | 4.6 | 4.6 | 5.0 | 5.1 | 5.2 | 5.2 | 5.2 |
| SEAR-D | 4.9 | 5.7 | 5.8 | 5.7 | 5.7 | 4.8 | 5.0 | 5.0 | 5.0 | 5.0 |
| WPR-A | 4.9 | 5.5 | 5.6 | 5.4 | 5.4 | 5.2 | 5.2 | 5.1 | 5.0 | 5.0 |
| WPR-B | 4.3 | 4.8 | 5.1 | 5.1 | 5.1 | 4.5 | 4.6 | 4.7 | 4.8 | 4.8 |

Table 7.4 Estimates of cholesterol standard deviations (mmol/l) by subregion, sex and age

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  |  | Males |  |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | 1.0 | 1.0 | 1.0 | 1.1 | 1.1 | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 |
| AFR-E | 1.0 | 1.0 | 1.0 | 1.1 | 1.1 | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 |
| AMR-A | 0.9 | I.I | 1.1 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.9 | 0.9 |
| AMR-B | 1.1 | I.I | 1.2 | 1.2 | 1.2 | 1.1 | 1.1 | 1.1 | I.I | 1.1 |
| AMR-D | I.I | 1.1 | 1.2 | 1.2 | 1.2 | 1.1 | 1.1 | 1.1 | I.I | I.I |
| EMR-B | 1.4 | 1.4 | 1.4 | 1.2 | 1.2 | 1.4 | 1.4 | 1.3 | 1.1 | I.I |
| EMR-D | 1.4 | 1.4 | 1.4 | 1.2 | 1.2 | 1.4 | 1.4 | 1.3 | 1.1 | 1.1 |
| EUR-A | 1.0 | 1.2 | 1.2 | 1.1 | 1.1 | 1.2 | 1.2 | 1.1 | 1.1 | 1.1 |
| EUR-B | 1.0 | 1.1 | 1.1 | 1.1 | I.I | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| EUR-C | 1.1 | 1.1 | 1.2 | 1.2 | 1.2 | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 |
| SEAR-B | 0.8 | 0.9 | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| SEAR-D | 1.0 | 1.1 | 1.1 | 1.1 | 1.1 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| WPR-A | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.8 | 0.8 |
| WPR-B | 1.0 | 1.1 | 1.3 | 1.2 | 1.2 | 1.1 | 1.1 | 1.2 | 1.1 | 1.1 |

Note: Reported standard deviations are derived from baseline cholesterol (i.e. one-off measures), which when corrected for regression dilution bias were used in all calculations on the "usual" scale.

However, there will always be errors associated with generalizing from the data collected to the subregional level. From the perspective of uncertainty, subregions can be divided into three main categories:

1. subregions where most of the countries have nationally representative surveys;
2. subregions with a lower coverage of countries and/or countries for which well designed surveys exist but on sub-national populations; and
3. subregions with countries where few studies and little or no data existed.

A subregion in which a national study is available for most of the countries would be expected to have low uncertainty (e.g. AMR-A and WPR-A). The remaining 12 subregions can be placed into either category 2 or 3 , depending on the number of high-quality studies available. For these, subregional means and standard deviations were estimated from available data. Examining the heterogeneity observed between
different countries in these subregions provided a basis to quantify the uncertainty associated with data quality as well as extrapolation (see Figures 7.2 and 7.3). There may be a variety of causes underlying the heterogeneity, including sampling error and real sub-population differences. However, a subregion whose estimate is based on data from sub-national studies that suggest a wider range of mean cholesterol levels would have greater uncertainty. An example of this occurred in WPR-B, where, in the absence of a single national study for China, an estimate had to be based on all eligible studies even though there were significant differences between studies. Conversely, in WPR-A, heterogeneity is due in greater part to real population differences (i.e. Japan and Australia/New Zealand have substantially different cholesterol levels).

A simulation approach was taken to incorporate the uncertainty introduced in estimating mean cholesterol levels to reflect both sampling error as well as the variation of studies within countries. For each subregion separately, the simulation proceeded by reiterating the same weighted regression approach described earlier (i.e. weighted regression of mean cholesterol on mean age). Each iteration involved re-sampling of data points 10000 times by treating each study as a random effect (i.e. two components of variation: study sampling error plus variation of study within country). The regression analysis was then re-iterated to obtain uncertainty distributions.

For subregions where insufficient data were available to assess the second component of variation accurately between countries, data were utilized from the data-rich subregions. Specifically, the average intercountry variation observed in AMR-A, AMR-B, EUR-A, EMR-B, WPRA and WPR-B was applied to the remaining eight subregions.

Similar simulations were utilized to estimate the uncertainty intervals around the estimated standard deviations, the only difference being that the sampling distribution for the standard deviations was based on a chisquare in place of a normal distribution. Results from these simulations are provided in Tables 7.5 and 7.6.

### 2.4 Theoretical-minimum-risk exposure distribution

To make a judgment on the theoretical minimum cholesterol level, the evidence of the association between cholesterol and relative risk of cardiovascular end-points and data from "low cholesterol" populations were examined.

## LOWEST RELATIVE RISK OF END-POINTS

A variety of prospective studies have examined the association between cholesterol and cardiovascular end-points. There is strong evidence that increased cholesterol is associated with increased risk of IHD, and the relationship is approximately linear on a logarithmic scale (Law 1999; Law et al. 1994b).
Table 7.5 Values of $95 \% \mathrm{Cl}$ for mean cholesterol ( $\mathrm{mmol} / \mathrm{l}$ ) by subregion, sex and age

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  |  | Males |  |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | (4.0-5.2) | (4.5-5.7) | (4.5-5.9) | (4.4-6.0) | (4.4-6.0) | (3.9-5.1) | (4.1-5.3) | (4.0-5.4) | (3.9-5.5) | (3.9-5.5) |
| AFR-E | (4.0-5.2) | (4.5-5.7) | (4.5-5.9) | (4.4-6.0) | (4.4-6.0) | (3.9-5.1) | (4.1-5.3) | (4.0-5.4) | (3.9-5.5) | (3.9-5.5) |
| AMR-A | (4.5-5.1) | (5.5-5.7) | (5.9-6.1) | (5.7-6.1) | (5.7-6.1) | (4.9-5.5) | (5.4-5.6) | (5.4-5.6) | (5.0-5.4) | (5.0-5.4) |
| AMR-B | (4.1-5.5) | (5.1-5.7) | (5.2-5.6) | (5.0-5.4) | (5.0-5.4) | (4.6-6.0) | (4.7-5.3) | (4.7-5.1) | (4.6-5.0) | (4.6-5.0) |
| AMR-D | (4.1-5.5) | (5.1-5.7) | (5.2-5.6) | (5.0-5.4) | (5.0-5.4) | (4.6-6.0) | (4.7-5.3) | (4.7-5.1) | (4.6-5.0) | (4.6-5.0) |
| EMR-B | (4.0-5.2) | (4.5-5.7) | (4.8-6.0) | (5.0-6.2) | (5.0-6.2) | (4.2-5.4) | (4.3-5.5) | (4.5-5.7) | (4.8-6.0) | (4.8-6.0) |
| EMR-D | (4.0-5.2) | (4.5-5.7) | (4.8-6.0) | (5.0-6.2) | (5.0-6.2) | (4.2-5.4) | (4.3-5.5) | (4.5-5.7) | (4.8-6.0) | (4.8-6.0) |
| EUR-A | (5.1-5.7) | (6.0-6.6) | (6.4-7.0) | (6.2-6.8) | (6.2-6.8) | (5.6-6.2) | (5.8-6.4) | (5.8-6.4) | (5.6-6.2) | (5.6-6.2) |
| EUR-B | (4.2-5.0) | (5.0-5.6) | (5.2-5.8) | (5.1-5.7) | (5.1-5.7) | (4.6-5.4) | (4.8-5.4) | (4.9-5.5) | (4.9-5.5) | (4.9-5.5) |
| EUR-C | (4.8-6.0) | (5.6-6.8) | (5.9-7.3) | (5.7-7.3) | (5.7-7.3) | (5.1-5.9) | (5.5-6.1) | (5.5-6.1) | (5.4-6.0) | (5.4-6.0) |
| SEAR-B | (3.4-4.6) | (3.9-5.1) | (3.9-5.3) | (3.8-5.4) | (3.8-5.4) | (4.4-5.6) | (4.5-5.7) | (4.5-5.9) | (4.4-6.0) | (4.4-6.0) |
| SEAR-D | (4.5-5.3) | (5.4-6.0) | (5.5-6.1) | (5.4-6.0) | (5.4-6.0) | (4.2-5.4) | (4.4-5.6) | (4.3-5.7) | (4.2-5.8) | (4.2-5.8) |
| WPR-A | (4.6-5.2) | (5.2-5.8) | (5.3-5.9) | (5.0-5.8) | (5.0-5.8) | (4.9-5.5) | (4.9-5.5) | (4.8-5.4) | (4.6-5.4) | (4.6-5.4) |
| WPR-B | (3.9-4.7) | (4.5-5.1) | (4.8-5.4) | (4.8-5.4) | (4.8-5.4) | (4.1-4.9) | (4.3-4.9) | (4.4-5.0) | (4.5-5.1) | (4.5-5.1) |

Values of $95 \% \mathrm{Cl}$ for standard deviation of cholesterol (mmol/l) by subregion, sex and age

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  |  | Males |  |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.9-1.3) | (0.9-1.3) | (0.7-I.1) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) |
| AFR-E | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.9-1.3) | (0.9-1.3) | (0.7-I.I) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) |
| AMR-A | (0.8-1.0) | (1.0-1.2) | (1.0-1.2) | (0.9-I.I) | (0.9-I.1) | (0.9-1.1) | (0.9-1.1) | (0.9-1.1) | (0.8-1.0) | (0.8-1.0) |
| AMR-B | (1.0-1.2) | (1.0-1.2) | (1.1-1.3) | (1.0-1.4) | (1.0-1.4) | (1.0-1.2) | (1.0-1.2) | (1.0-1.2) | (0.9-1.3) | (0.9-1.3) |
| AMR-D | (1.0-1.2) | (1.0-1.2) | (1.1-1.3) | (1.0-1.4) | (1.0-1.4) | (1.0-1.2) | (1.0-1.2) | (1.0-1.2) | (0.9-1.3) | (0.9-1.3) |
| EMR-B | (1.2-1.6) | (1.1-1.7) | (1.2-1.6) | (1.0-1.4) | (1.0-1.4) | (1.2-1.6) | (1.1-I.7) | (1.1-1.5) | (0.9-1.3) | (0.9-1.3) |
| EMR-D | (1.2-1.6) | (1.1-1.7) | (1.2-1.6) | (1.0-1.4) | (1.0-1.4) | (1.2-1.6) | (1.1-I.7) | (1.1-1.5) | (0.9-1.3) | (0.9-1.3) |
| EUR-A | (0.9-1.1) | (0.9-1.5) | (1.0-1.4) | (1.0-1.2) | (1.0-1.2) | (1.1-1.3) | (0.9-1.5) | (0.9-1.3) | (1.0-1.2) | (1.0-1.2) |
| EUR-B | (0.8-1.2) | (0.9-1.3) | (0.9-I.3) | (0.9-1.3) | (0.9-1.3) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) |
| EUR-C | (0.9-I.3) | (0.9-1.3) | (1.0-1.4) | (1.0-1.4) | (1.0-1.4) | (0.9-1.3) | (0.9-1.3) | (0.9-1.3) | (0.9-1.3) | (0.9-1.3) |
| SEAR-B | (0.6-1.0) | (0.7-1.1) | (0.7-I.1) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) |
| SEAR-D | (0.8-1.2) | (0.9-1.3) | (0.9-I.3) | (0.9-1.3) | (0.9-1.3) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) | (0.8-1.2) |
| WPR-A | (0.7-1.1) | (0.8-1.0) | (0.8-1.0) | (0.8-1.0) | (0.8-I.0) | (0.7-I.1) | (0.8-1.0) | (0.8-1.0) | (0.7-0.9) | (0.7-0.9) |
| WPR-B | (0.7-I.3) | (0.9-1.3) | (1.2-I.4) | (1.0-1.4) | (1.0-1.4) | (0.8-I.4) | (0.9-1.3) | (1.1-I.3) | (0.9-1.3) | (0.9-1.3) |

Figure 7.9 The association between cholesterol and IHD in the Shanghai Factory Workers Study


Source:
Reprinted, by permission of the publisher, from Chen et al. (1991). Serum cholesterol concentration and coronary heart disease in population with low cholesterol concentrations. British Medical Journal, 303:276-282.

In populations from industrialized countries (e.g. Europe and North America), this relationship has seldom been demonstrated below a cholesterol level of $4.5-5.0 \mathrm{mmol} / 1$ (Chen et al. 1991). However, in other populations such as China, data suggest that this relationship may continue, without a threshold, below cholesterol levels of about $4.0 \mathrm{mmol} / \mathrm{l}$ (Chen et al. 1991; Law and Wald 1994; Law et al. 1994b) (Figure 7.9). Based on this evidence it would be feasible to set a theoretical minimum of usual cholesterol of at least $3.8 \mathrm{mmol} / \mathrm{l}$.

Another source of data is clinical trials of cholesterol lowering from recent meta-analyses. Trials included in these overviews achieved substantial reduction in cholesterol with no adverse effects. The reductions are summarized in Table 7.7. These reductions of cholesterol also occurred in the lowest cholesterol group, i.e. those with baseline mean cholesterol of about $5.0 \mathrm{mmol} / \mathrm{l}$. The aggregated data from trials therefore indicate that a substantial reduction in cholesterol is possible, across different baseline cholesterol levels, without risk of substantial adverse effects.

Further, the recently completed large-scale Heart Protection study (Heart Protection Study Collaborative Group 2002a, 2002b) demonstrated that cholesterol lowering in those with average or below average

Table 7.7 Reduction in cholesterol achieved by trials of statins

| Trial | Baseline mean <br> cholesterol (mmol/I) | Mean cholesterol <br> reduction (\%) | Final mean <br> cholesterol (mmol/I) |
| :--- | :---: | :---: | :---: |
| LRC | 7.5 | 9 | 6.9 |
| HHS | 7.4 | 10 | 6.7 |
| WOSCOPS | 7.0 | 20 | 5.6 |
| 4S | 6.8 | 26 | 5.0 |
| AFCAPS/TexCAPS | 5.7 | 19 | 4.6 |
| LIPID | 5.6 | 18 | 4.6 |
| CARE | 5.4 | 20 | 4.3 |

Source: LaRosa et al. (1999) and Pignone et al. (2000).
total cholesterol levels was still associated with reduced risk of cardiovascular disease. This trial included over 20000 participants and found no evidence of a threshold level below which lowering cholesterol did not produce a lower risk.

## LOW CHOLESTEROL POPULATIONS

An alternative source of data relevant to setting a theoretical minimum comes from examining mean cholesterol levels in studies of so-called "unacculturated" populations with little cardiovascular disease and low cholesterol levels (Poulter and Sever 1994). Unacculturated generally refers to those populations that are relatively isolated and have preserved their lifestyle over many generations.

An overview of typical cholesterol values in men aged 45-64 years in different populations noted that while mean cholesterol ranged from 5.5 to $7.0 \mathrm{mmol} / \mathrm{l}$ in many industrialized populations, it was much lower in hunter-gatherer societies where mean values were as low as 3.0$3.5 \mathrm{mmol} / \mathrm{l}$ (Law 1999; Law and Wald 1994). Data from individual studies around the world that have focused on these "low cholesterol" populations are summarized in Table 7.8. In the Shanghai study, there was a strongly positive and apparently independent relation between cholesterol concentration and death from IHD, even at these lower levels (Chen et al. 1991) (Figure 7.10).

Typically, the data show that there is a very low prevalence of cardiovascular disease in these populations, which exhibit an absence of obesity due to diets low in salt, cholesterol and fat (particularly animal fat), and a lifestyle requiring heavy physical labour (Barnes 1965; Carvalho et al. 1989; Connor et al. 1978; He et al. 1991a, 1991b; Page et al. 1974; Poulter and Sever 1994; Sever et al. 1980). There is also evidence of low blood pressure levels, and no age-related rise in either cholesterol or blood pressure levels. Data from these studies indicate that many of these populations have mean cholesterol levels of about

Table 7.8 Cholesterol levels in low cholesterol populations

| Country (reference) | Population | Mean cholesterol levels | Patterns with age |
| :--- | :--- | :--- | :--- |
| Africa |  |  |  |
| Tanganyika <br> et al. I964) | Masai tribe (mostly <br> males) with virtually no <br> cardiovascular disease. | $3.0-3.7 \mathrm{mmol} / \mathrm{l} ;$ <br> $>5.2 \mathrm{mmol} / \mathrm{I}$ rare | No evidence of <br> increasing <br> cholesterol with <br> advancing age |
| Americas |  |  |  |
| Mexico (Casdorph <br> I972; Connor et al. | Tarahumara Indians <br> who lived in the <br> mountains | $3.0-3.5 \mathrm{mmol} / \mathrm{I}$ | No age-related |
| I978) |  |  |  |

$3.8 \mathrm{mmol} / \mathrm{l}$ or lower. Therefore, setting a theoretical minimum at 3.8 $\mathrm{mmol} / \mathrm{l}$ is justified based on all of the data.

## Safety of low cholesterol levels

Concerns have been expressed over whether lowering cholesterol may be harmful, and therefore setting a theoretical minimum too low may have adverse consequences. Some data have suggested a negative association between cholesterol and haemorrhagic stroke (Iso et al. 1989); however, a recent analysis has suggested that this is in fact more likely to be a null association (Suh et al. 2001). It should also be noted that even if there were a small increase in risk of haemorrhagic stroke for people with very low cholesterol levels, this would be outweighed overall by the benefits resulting from lower risk of IHD (Bucher et al. 1998; Law et al. 1994a).

There is no strong consistent evidence that low or reduced cholesterol concentrations increase mortality from non-cardiovascular diseases such as cancer or infection (Law et al. 1994a). Overviews of trials have not
detected increases in causes of non-cardiovascular disease mortality (Hebert et al. 1997; LaRosa et al. 1999) despite achieving relatively large decreases in cholesterol. In many cases, the apparent association is actually due to disease causing low cholesterol rather than vice versa (Law et al. 1994a). A recent meta-analysis of trials found no convincing evidence that the risk of "non-illness"-related mortality (deaths from suicide, accident or trauma) was strongly associated with cholesterol lowering (Muldoon et al. 2001).

Finally, data relating to low cholesterol levels come from research on those with heterozygous familial hypobetalipoproteinaemia. These individuals have cholesterol levels as low as $2.0-3.0 \mathrm{mmol} / \mathrm{l}$ and a prolonged life expectancy, as coronary artery disease is avoided-but no recognized adverse effects from the low cholesterol (Law 1999).

## Standard deviation of the theoretical-minimum-Risk EXPOSURE DISTRIbUTION

The choice of standard deviation around the theoretical minimum was based on examining the relationship of the standard deviation and mean of cholesterol using all available data from our review of the literature (Figure 7.10). From this illustration, a distribution with a mean of $3.8 \mathrm{mmol} / 1$ would typically have a standard deviation of $0.9 \mathrm{mmol} / 1$ (baseline), which equates to $0.6 \mathrm{mmol} / 1$ (usual).

## 3. Cholesterol-Disease relationships

Data on the relationship between cholesterol and disease outcomes come from two main types of studies (MacMahon 1994). Prospective observational studies provide data from which the effects of prolonged cholesterol differences can be estimated (MacMahon et al. 1990), that is, hazard ratios. Trials provide data about the effects of short-term cholesterol reduction (Collins et al. 1990), or risk reversal. Results from prospective observational studies will be considered first.

### 3.1 Data sources for hazard ratios

Observational studies have been conducted in a variety of settings that examine the association between cholesterol and disease. However, the results of many of these individual studies are limited. Many observational studies have small sample sizes or an insufficient number of end-points, and therefore lack the power required to provide reliable estimates of associations for different population subgroups (e.g. sex and age groups) and/or specific diagnostic categories (MacMahon et al. 1990). Individual studies do not always provide information on the direction of the association at lower cholesterol levels, making it difficult to assess whether the observed association is continuous or has a threshold level. In addition, these studies frequently do not standardize the size of the association for bias and confounding-in particular,

Figure 7.10 Association between mean cholesterol and standard deviation for review data

Table 7.9 Characteristics of major cholesterol cohort study overviews

| Study characteristics | Law et al. (1994b) | Prospective Studies Collaboration (PSC 1995) | Eastern Collaborative Research Group (Anonymous 1998a) | Asia-Pacific Cohort Studies Collaboration (APCSC secretariat, personal communication, 2001) |
| :---: | :---: | :---: | :---: | :---: |
| Aims | To estimate by how much and how quickly a given reduction in cholesterol concentration will reduce the risk of IHD | To assess the relationship between BP and total blood cholesterol and stroke, and determine how the strength of the relationship between BP and stroke varied with age | To assess the relationship between BP and total blood cholesterol and stroke in Asian populations, and determine whether the strength of the relationship varied with type of stroke | To produce region, age- and sex-specific blood pressure and cholesterol associations for stroke (including subtypes), IHD, and total cardiovascular disease |
| Number of studies included | 10 | 45 | 18 | 29 |
| Regions included | England, Scotland, Europe, Hawaii and Israel | Asia, Australia, Europe, Hawaii, the Middle East and USA | China (13 cohorts), Japan (5) | Australia (3), mainland China (9), Hong Kong SAR (I), Japan (I2), New Zealand (I), Republic of Korea (I), Singapore (I), Taiwan, China (I) |
| Number of participants | 494804 | 450000 | 124774 | 353065 |
| \% male | 100\% | 61\% | 61\% | 57\% |
| Age range (years) | 35-84 | 15-99 | 18-98 | 20-107 |
| Mean age at baseline | Not available | Not available | 48 years | 47 years |
| Follow-up | Range of 7-23 years | Range of 5-30 years, mean 15 years | Range unknown, mean 9 years | Range of <I-29 years, mean 7 years |
| BP Blood pressure. |  |  |  |  |

Of the four overviews, only two included analyses for the IHD end-point; Law et al. (1994b) and APCSC (APCSC secretariat, personal communication, 2001). The former will be used as the primary data source for relative risk estimates of IHD due to the greater sample size. Three overviews included data on cholesterol and stroke (Anonymous 1998a; APCSC secretariat, personal communication, 2001; PSC 1995). However, only APCSC analyses were based on individual participant data and provided age-specific analyses of total stroke and stroke subtypes. It will therefore be the primary data source for relative risk estimates of stroke.

### 3.2 Analysis issues

An important factor that must be accounted for with estimates of the association between cholesterol and cardiovascular end-points from observational data is regression dilution bias. This bias occurs because baseline or one-off measures of cholesterol are subject to random fluctuations, due partly to the measurement process, and partly to any real but temporary deviations at the baseline from the usual cholesterol level (MacMahon et al. 1990). Therefore, baseline cholesterol values have a wider distribution than the "usual" cholesterol values. With repeated measures there is a "regression to the mean" of values (MacMahon 1994), whereby an initially extreme observation tends to become less abnormal with replication (Strachan and Rose 1991).

This imprecision in measurement not only influences distribution, but also affects the association with disease outcomes (MacMahon et al. 1990). Figure 7.11 illustrates the effect of regression dilution

Figure 7.1I Effects of regression dilution bias on the association between blood pressure and relative risk of cardiovascular disease


Systolic blood pressure ( mmHg )
bias for blood pressure, but the same would be true for cholesterol. The baseline distribution has a shallower slope than usual cholesterol on the curve-relation of cholesterol to relative risk of disease. Analyses have suggested that correcting for regression dilution bias increases the slope of the association by as much as $61 \%$ (Law et al. 1994c). If this bias is not corrected, the strength of the association between cholesterol and disease incidence is underestimated ("regression dilution bias") (MacMahon 1994). This systematically dilutes the apparent importance of cholesterol and can result in systematic and substantial underestimation of risk of disease with usual cholesterol (MacMahon et al. 1990).

The size of the dilution is directly related to the extent to which cholesterol measurements are subject to regression to the mean. Several major meta-analyses conducted in recent years aimed to address these limitations with correction for regression dilution bias (Anonymous 1998a; APCSC 1999; PSC 1995). It is possible to use repeated measures on cholesterol to obtain an estimate of the attenuation factor in order to correct for this bias in the analysis. An attenuation factor of 1.82 was calculated from remeasurement data in the Prospective Studies Collaboration (PSC) (1995), and 1.90 in the Eastern study (Anonymous 1998a). In the APCSC, analyses of remeasurement data from cohorts estimated that the correction factor was 1.8. (APCSC secretariat, personal communication, 2001). Here, use of the term "usual cholesterol" indicates that the association between cholesterol and the disease end-point has been corrected for regression dilution bias.

The surrogate dilution effect is another factor that should be considered specifically in cholesterol studies. This effect arises because observational studies underestimate the effect of lower cholesterol relative to the trials (Law et al. 1994c). In a cohort study, a $1 \mathrm{mmol} / /$ lower cholesterol concentration is associated with about $0.67 \mathrm{mmol} / 1$ of LDL, but in cholesterol lowering trials, a $1 \mathrm{mmol} / /$ lower total cholesterol is usually comparable to $1 \mathrm{mmol} / /$ lower LDL. Law et al. (1994c) adjusted for both regression dilution bias and the surrogate dilution effect in their analyses, and the final correction factor was 1.61.

### 3.3 Estimates of cholesterol-disease relationships

A summary of the results of the four prospective study overviews is presented in Table 7.10, and each of these end-points will now be considered in detail.

## Cholesterol and IHD

Analyses from the APCSC overview (APCSC secretariat, personal communication, 2001) and that by Law et al. (1994b) demonstrated that the relative risk of IHD increased with increasing cholesterol level, and that the association was roughly linear when plotted on a log scale (Figure 7.12 and 7.13).

Figure 7.12 Usual cholesterol level and risk of IHD in the Asia-Pacific region


Source: APCSC data (APCSC secretariat, personal communication, 2001).

This means that the proportional difference in risk associated with a given absolute difference in usual cholesterol is similar at all levels of cholesterol, at least within the range studied. This continuous association showed no evidence of a threshold level of cholesterol, below which lower levels of cholesterol are no longer associated with lower relative risks of IHD (down to almost $4.0 \mathrm{mmol} / \mathrm{l}$ ). Further, there was no evidence of an upper threshold level above which the relative risk of IHD increased much more rapidly (Anonymous 1998a; APCSC secretariat, personal communication, 2001; Law et al. 1994c; PSC 1995).

Overall, the analyses in both the APCSC (APCSC secretariat, personal communication, 2001) and by Law et al. (1994b) suggested that a $0.6 \mathrm{mmol} / \mathrm{l}$ lower cholesterol was associated with $24-27 \%$ lower risk of IHD. There was no evidence of a difference in the size of the association between males and females or fatal and non-fatal IHD end-points.

Age has an important influence on the size of the cholesterol-endpoint relationship, as the associations are steeper for those in younger age groups. (This also largely accounts for the apparent heterogeneity between the cholesterol-IHD associations in the cohorts in Figure 7.13.) This limits the ability to compare directly the overall relative risk estimates across the overviews, as it is only appropriate to compare age-specific results. The gradient of the cholesterol-IHD association varies with age at baseline, the gradient being much steeper for younger
Table 7.10 Summary results of major cohort study overviews

| Study end-points | Law et al. (1994b) | Prospective Studies Collaboration (PSC 1995) | Eastern Collaborative Research Group (Anonymous 1998a) | Asia-Pacific Cohort Studies Collaboration (APCSC secretariat, personal communication, 2001) |
| :---: | :---: | :---: | :---: | :---: |
| IHD end-points | I881। IHD events recorded | - | - | 2838 IHD events, <br> I 607 (57\%) fatal |
|  | $0.6 \mathrm{mmol} / \mathrm{I} \downarrow$ cholesterol associated with overall reduction of IHD of $27 \%$ from 10 studies combined | - | - | $0.6 \mathrm{mmol} / \mathrm{I} \downarrow$ cholesterol $R R=0.76(95 \% \text { CI } 0.73-0.79)$ <br> (24\% reduction in IHD) |
| Stroke end-points | - | 13397 participants were recorded as having had a stroke | I 798 strokes, 995 (55\%) fatal, $39 \%$ had data on subtype of which $42 \%$ haemorrhagic | 2937 strokes, 56\% fatal |
|  | - | $\mathrm{I} \mathrm{mmol} / \mathrm{I} \uparrow$ cholesterol $\mathrm{RR}=0.98$ ( $95 \% \mathrm{Cl} 0.94-\mathrm{I} .0 \mathrm{I}$ ) | $0.6 \mathrm{mmol} / \mathrm{I} \downarrow$ cholesterol RR $=$ 0.92 (95\% CI 0.72-I.I7) <br> ( $8 \%$ reduction in stroke) | I mmol/I $\downarrow$ cholesterol RR $=0.87$ (95\% Cl 0.8I-0.94) <br> ( $13 \%$ reduction in stroke) |

[^19]- No data.

Figure 7.13 Incidence of ischaemic heart disease and usual cholesterol level in 10 cohort studies


[^20]age groups. While the proportional change in IHD risk per unit change in cholesterol is less extreme in old age than in middle age, the relationship remains positive for all age groups (APCSC secretariat, personal communication, 2001; Law et al. 1994b). It is therefore necessary to use age-specific estimates for hazard ratios in all cholesterol analyses for CRA.

The results of both overviews are relatively consistent, but due to the larger number of participants and end-points in the review by Law et al. (1994b), these relative risk estimates, transformed for application to GBD age groups, were used in the analyses of this chapter. Table 7.11 presents relative risks from the overview by Law et al. (1994b) transformed into GBD age groups for a $1 \mathrm{mmol} / \mathrm{l}$ difference in usual cholesterol.

Table 7.II Relative risk of IHD associated with a I mmol/I difference in usual cholesterol

| Age group (years) | Relative risk $(95 \% \mathrm{Cl})$ |
| :--- | :---: |
| $30-44$ | $0.5 \mathrm{I}(0.36-0.73)$ |
| $45-59$ | $0.50(0.45-0.56)$ |
| $60-69$ | $0.70(0.62-0.79)$ |
| $70-79$ | $0.77(0.69-0.86)$ |
| $\geq 80$ | $0.75(0.67-0.84)$ |

Source: Modified from overview by Law et al. (1994b).

## Cholesterol and stroke

Few data are available on the association between cholesterol and stroke, and the relationship is more complex. Many individual cohorts and the PSC have found no association between cholesterol and risk of stroke death (PSC 1995). However, recent meta-analyses and other studies have reported that this apparent null association may be principally due to a quantitatively different association between cholesterol and the two major types of stroke-haemorrhagic and ischaemic. A variety of studies have attempted to clarify these associations. Table 7.12 summarizes some of the largest cohort studies undertaken, with several hundred recorded stroke events by subtype.

Several of these studies showed evidence of a positive association between cholesterol and risk of ischaemic stroke (Anonymous 1998a; Benfante et al. 1994; Iso et al. 1989; Jorgensen et al. 1995; Leppala et al. 1999; Neaton et al. 1992, 1993; Yano et al. 1994). Other individual studies failed to find an association between cholesterol and ischaemic stroke (Nakayama et al. 1997; Szatrowski et al. 1984; Tanaka et al. 1982), but the extent of misclassification of stroke subtype was uncertain in these studies. The association of cholesterol with haemorrhagic stroke is less clear, with some analyses suggesting either a negative (Anonymous 1998a; Iso et al. 1989; Leppala et al. 1999; Neaton et al. 1992, 1993; Yano et al. 1989, 1994) or null association (Iribarren et al. 1996; Suh et al. 2001).

Analyses from the APCSC show the different association of cholesterol with stroke subtypes (Figure 7.14); a positive association for ischaemic stroke and a negative/null association for haemorrhagic stroke. In the CRA analyses, data from the APCSC overview were used, given the size and availability on individual data.

The age-specific relative risks of haemorrhagic and ischaemic stroke with a $1 \mathrm{mmol} / 1$ difference in usual cholesterol are presented in Table 7.13.

Figure 7.14 Usual cholesterol level and risk of haemorrhagic and ischaemic stroke


Haemorrhagic

Ischaemic

Source: APCSC data (APCSC secretariat, personal communication, 2001).

Overall, APCSC data indicated that the association of cholesterol with stroke varies with age at baseline, the gradient being much steeper for younger age groups. As discussed in relation to IHD, this means that the proportional change in stroke risk per unit change in cholesterol is less extreme in old age than in middle age (APCSC secretariat, personal communication, 2001). It is therefore necessary to use age-specific estimates for hazard ratios in all cholesterol analyses for CRA. There was no evidence of differences in the association between males and females, so sex-specific estimates are not necessary.

Despite having data from the APCSC, it is difficult to incorporate the different cholesterol-stroke subtype associations into CRA analyses as the only GBD end-point for stroke relates to total stroke. It is not possible to conduct analyses for ischaemic and haemorrhage stroke independently, and it is not appropriate simply to apply the relative risk for total stroke events to all 14 subregions, since there is evidence that the relative proportions of haemorrhagic and ischaemic stroke vary among them (Anderson et al. 1993; Bamford et al. 1990; Chen et al. 1992; Giroud et al. 1991; Hung 1993; Jeng et al. 1998; Nakayama et al. 1997; Suzuki et al. 1987; Thrift et al. 2001; Wu et al. 1992). Analyses of total stroke must therefore be modified to reflect these differing proportions, and the different types of association between cholesterol and stroke
Table 7.12 Data from major cohort studies on the association between cholesterol and risk of stroke subtype

| Study | Participants | Ischaemic stroke | Haemorrhagic stroke |
| :---: | :---: | :---: | :---: |
| The American MRFIT study (Iso et al. 1989; Neaton et al. 1992, 1993) | $350000 \text { males }$ <br> At 6 years, 230 fatal strokes had occurred At 12 years, 765 stroke deaths had occurred | Risk of death was significantly increased with increasing cholesterol at both points of follow-up | Risk of haemorrhagic stroke death highest in those with lowest cholesterol (if DBP $\geq 90 \mathrm{mmHg}$ ) at 6 and 12 years |
| American Kaiser Permanente Study (Iribarren et al. 1996) | 61756 individuals <br> At 16 years 386 haemorrhagic strokes recorded | - | No relationship among those aged 40-64 years Negative association between low cholesterol and risk of haemorrhagic stroke in elderly males |
| British Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (Leppala et al. 1999) | 28519 males aged 50-69 years At 8 years 1057 males had a stroke | Cholesterol was associated with increased risk of cerebral infarction at levels $\geq 7 \mathrm{mmol} / \mathrm{l}$ | Cholesterol was inversely associated with the risk of intracerebral haemorrhage (Leppala et al. 1999) |
| Copenhagen Stroke Study (Jorgensen et al. 1995) | Included over I 000 males who had suffered a stroke | Cholesterol was positively associated with risk of ischaemic stroke | - |
| A cohort study in Sweden (Gatchev et al. 1993) | 54000 participants At 20 years, 347 haemorrhagic strokes had been reported | - | Relative risk for cerebral haemorrhage increased with decreasing cholesterol level in females, but the risk function was U-shaped in males |
| Honolulu Heart <br> Program (Benfante <br> et al. 1994; Yano <br> et al. 1989, 1994) | Over 6000 Japanese males in Hawaii followed for 15 years ( 252 ischaemic strokes) Over 7000 of these males followed for 18 years (116 haemorrhagic strokes) | A continuous increase in isch. stroke rates with increasing levels of cholesterol | An inverse association between cholesterol and the risk of haemorrhagic stroke, at least in males with cholesterol in the lowest quintile |
| Korea Medical Insurance Corporation study (Suh et al. 2001) | 114793 males in the Republic of Korea, aged 35-59 years followed for 6 years | - | No overall association between the relative risk of intracerebral haemorrhage and total cholesterol |
| Eastern Stroke and CHD <br> Collaborative project <br> (Anonymous 1998a) | Overview of 18 cohort studies in China and Japan, II studies, >60000 participants included data on stroke subtype (494 non-haemorrhagic and 404 haemorrhagic strokes) | A positive association between cholesterol and non-haemorrhagic stroke | An inverse association between cholesterol and haemorrhagic stroke (Anonymous 1998a) |

[^21]Table 7.13 Relative risk of stroke with a I mmol/l difference in usual cholesterol

|  | Risk of stroke event |  |
| :--- | :---: | :---: |
| Age group (years) | Haemorrhagic stroke <br> relative risk (95\% Cl) | Ischaemic stroke <br> relative risk (95\% Cl) |
| $30-44$ | $1.03(0.64-1.64)$ | $0.66(0.34-1.26)$ |
| $45-59$ | $1.09(0.92-1.28)$ | $0.66(0.56-0.77)$ |
| $60-69$ | $1.13(0.90-1.4 \mathrm{I})$ | $0.77(0.65-0.92)$ |
| $70-79$ | $1.08(0.84-1.40)$ | $0.92(0.76-1.10)$ |
| $\geq 80$ | $1.43(0.99-2.07)$ | $0.84(0.70-1.00)$ |

Source: APCSC data (APCSC secretariat, personal communication, 2001).
subtype. It was therefore necessary to estimate subregion- and agespecific proportions of ischaemic and haemorrhagic fatal and non-fatal strokes so that weighted RR values could be applied. (An assumption was made that these are the same for males and females.)

Several steps were undertaken to estimate stroke subtype proportions, and these are outlined in Appendix B. They included using data from "gold-standard" studies of stroke incidence, and other observational data on stroke subtypes by subregion. Having estimated the relative percentage of stroke subtypes by age, sex and subregion, the APCSC values of RR for ischaemic stroke were smoothed and then weighted as per these percentages. (Smoothed RRs for the five age groups were 0.61 , $0.69,0.76,0.82$ and 0.90 .) No association was found between cholesterol and haemorrhagic stroke.

### 3.4 Risk Reversal

As already alluded to, prospective observational studies provide data on the association between cholesterol and stroke and IHD, but they do not provide data on the impact of cholesterol lowering on these outcomes over time (MacMahon et al. 1990). From prospective studies alone, it is not possible to tell whether outcomes are reversible, or whether the association reflects, to some extent, irreversible cumulative effects of cholesterol differences that have persisted for years (Collins and Peto 1994).

The results from randomized clinical trials do provide data on reversibility. They are relevant to assessing how rapidly, and to what extent, the epidemiologically expected reductions in stroke and IHD are produced by lowering cholesterol levels (Collins and Peto 1994; Collins et al. 1990). They therefore provide estimates of the proportion of the potential long-term benefit from a particular cholesterol difference that may be expected within a few years of cholesterol lowering.

## Data sources on risk reversal

A variety of trials have studied the impact of cholesterol lowering. However, as with blood pressure lowering trials, individual studies usually lacked sufficient power to reliably detect moderate changes in events (Collins and Peto 1994; Collins et al. 1990; He and Whelton 2000). To accurately detect small but potentially important differences in risk reduction of cardiovascular disease (e.g. 10-15\%), it is necessary for trials to record many hundreds of end-points. Since most have not achieved this, overviews are necessary to provide more accurate estimates of the impact of cholesterol lowering on stroke and IHD. Results from these overviews will contribute to estimates of "risk reversibility".

A number of cholesterol lowering trials as well as major metaanalyses of such trials have been undertaken (Atkins et al. 1993; Blauw et al. 1997; Bucher et al. 1998; Byington et al. 2001; Crouse et al. 1997, 1998; Hebert et al. 1995, 1997; Holme 1990; LaRosa et al. 1999; Law et al. 1994b; Pignone et al. 2000; Ross et al. 1999; M. Law, personal communication, 2001). Most of these trials assessed the impact of lowering cholesterol on IHD, but data are also available on stroke. Results of an updated meta-analysis of all trials included in these previous overviews are given in Tables 7.14-7.16.

## Estimates of Risk reversal

## Cholesterol and risk of IHD

Trials of cholesterol lowering may be broadly grouped into "dietary interventions", "non-statin drugs" (e.g. fibrates), and "statin drugs". Many individual trials and a number of meta-analyses (including the updated meta-analysis) have analysed the association between cholesterol lowering and IHD. Overall, the trials have demonstrated a significant reduction in risk of IHD for those treated.

The current and previous meta-analyses of the early trials demonstrated a risk reduction in those treated of about $10-15 \%$, with dietary interventions or non-statin drugs such as fibrates or resins (Atkins et al. 1993; Bucher et al. 1998; Holme 1990). Later trials with statins, that achieved larger reductions in cholesterol, produced risk reductions of up to 25-30\% (Table 7.14) (Bucher et al. 1998; LaRosa et al. 1999; Pignone et al. 2000; Ross et al. 1999). One recently published overview included diet, fibrates, resins and statins (Bucher et al. 1998), but most recent meta-analyses have limited their analyses to statins. The review suggested that the effects of statins on risk reduction could be greater than other agents, but cautioned that analyses were based on between-study treatment comparisons rather than within study comparisons. Therefore, any apparent differences in drug efficacy may have been due to other factors such as differences in study design or populations. For example, participants in statin trials tended to be older, with higher event rates than on other trials, which may partly be responsible for differences in trial
Table 7.14 Clinical trials of cholesterol lowering by dietary interventions

| Trial | Participants |  | Stroke |  | IHD |  | Mean age (years) | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | Cholesterol reduction | Follow-up | RR (95\% CI) |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | A | C | A | C | A | C |  |  |  |  | Stroke | IHD |
| DART (Burr et al. 1989; <br> Law et al. 1994b) | 1018 | 1015 | - | - | 132 | 144 | 56.6 | 0.0 | 0.3 | 2.0 | - | 0.91 (0.73-I.14) |
| Hjermann et al. (1981) | 604 | 628 | 3 | 3 | 19 | 36 | 45.0 | 0.0 | >1.0 | 6.5 | 1.04 (0.21-5.13) | 0.55 (0.32-0.95) |
| Los Angeles (Dayton et al. 1968) | 424 | 422 | 13 | 22 | 45 | 67 | 65.5 | 0.0 | 0.9 | 8.0 | 0.59 (0.30-1.15) | 0.67 (0.47-0.95) |
| MRC (Anonymous 1968) | 199 | 194 | 2 | 0 | 45 | 51 | - | 0.0 | 1.00 | 3.7 | 4.87 (0.24-100.90) | 0.86 (0.61-1.22) |
| MRC (Research Committee to the Medical Council 1965) | 123 | 129 | - | - | 43 | 44 | - | 0.0 | 0.5 | 6.0 | - | 1.02 (0.73-1.44) |
| Minnesota (Frantz et al. 1975, 1989; Law et al. 1994b) | 4922 | 4853 | - | - | 131 | 121 | 50.0 | 51.5 | 0.7 | 5.0 | - | 1.07 (0.84-1.36) |
| Oslo (Leren 1970) | 206 | 206 | $7^{\text {a }}$ | $5^{\text {a }}$ | 79 | 94 | 45.0 | 0.0 | 1.1 | 11.0 | 1.40 (0.45-4.34) | 0.84 (0.67-1.06) |
| STARS (Law et al. 1994b; Watts et al. 1992) | 60 | 30 | 0 | 1 | 3 | 5 | 51.0 | 0.0 | 1.1 | 3.0 | 0.17 (0.01-4.04) | 0.30 (0.08-1.17) |
| St Mary's (Law et al. 1994b; Rose et al. 1965) | 28 | 52 | 0 | 0 | 8 | 11 | 56.8 | - | 0.6 | 2.0 | - | 1.45 (0.67-3.17) |
| Sydney (Law et al. 1994b; <br> Woodhill et al. 1978) | 221 | 237 | - | - | 37 | 24 | 49.0 | 0 | 0.3 | 5.0 | - | 1.65 (1.02-2.67) |
|  | 7805 | 7766 | 25 | 34 | 542 | 597 | $50.9{ }^{\text {b }}$ | 32.5 | $0.7{ }^{\text {b }}$ | $5.6{ }^{\text {b }}$ | 0.80 (0.48-1.32) | 0.92 (0.82-1.02) |
| Key: A, active treatment group; C, <br> - No data. <br> a Only fatal events reported. <br> b "Weighted mean" totals (weight | rol gro by per | n-years | follow |  |  |  |  |  |  |  |  |  |

Table 7.15 Clinical trials of cholesterol lowering by other non-statin interventions

| Trial | Participants |  | Stroke |  | IHD |  | Mean age (years) | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | Cholesterol reduction | Follow-up | RR (95\% CI) |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | A | C | A | C | A | C |  |  |  |  | Stroke | IHD |
| Acheson and Hutchison ${ }^{\text {a }}$ (1972) | 47 | 48 | 25 | 27 | - | - | - | 31.6 | 0.6 | 3.5 | 0.95 (0.66-1.36) | - |
| CLAS (Blankenhorn et al. 1987) | 94 | 94 | 0 | 0 | 1 | 5 | 54.2 | 0.0 | 1.3 | 2.0 | - | 0.20 (0.02-1.68) |
| Coronary Drug Project <br> (Anonymous 1975) | 2222 | 2789 | 231 | 311 | 596 | 839 | 54.0 | 0.0 | 0.6 | 6.0 | 1.01 (0.73-1.41) | 0.89 (0.82-0.97) |
| Dorr et al. (1978) | 1149 | 1129 | $0^{\text {c }}$ | $1{ }^{\text {c }}$ | 19 | 31 | 53.9 | 52.0 | 0.5 | 2.1 | 0.33 (0.01-8.03) | 0.60 (0.34-1.06) |
| FATS (Brown et al. 1990; Law et al. 1994b) | 94 | 52 | 0 | 0 | 2 | 0 | 46.7 | 0.0 | 1.5 | 2.5 | - | 2.79 (0.14-57.03) |
| Gross and Figueredo (1973) | 23 | 29 | $0^{c}$ | $1{ }^{\text {c }}$ | 1 | 0 | 57.1 | 71.2 | 0.6 | 3.0 | 0.42 (0.02-9.78) | 3.75 (0.16-87.98) |
| Helsinki (Frick et al. 1987) | 2051 | 2030 | $6^{\text {c }}$ | 4 | 56 | 84 | 47.3 | 0.0 | 0.7 | 5.0 | 1.48 (0.42-5.25) | 0.66 (0.47-0.92) |
| Lipid Research Clinic (Anonymous 1984) | 1906 | 1900 | 17 | 14 | 155 | 187 | 47.0 | 0.0 | 0.7 | 7.4 | 1.21 (0.60-2.45) | 0.83 (0.67-1.01) |
| McCaughan ${ }^{\text {b }}$ (Law et al. 1994b; McCaughan 198I) | 88 | 30 | 0 | 0 | 2 | 3 | 49.8 | 0.0 | 0.7 | 1.0 | - | 0.23 (0.04-1.30) |
| Miettinen (Miettinen et al. 1985) | 612 | 610 | 0 | 8 | 19 | 9 | 48.0 | 0.0 | 0.5 | 5.0 | 0.06 (0.00-1.01) | 2.10 (0.96-4.61) |


| Newcastle (Anonymous 197Ib) | 244 | 253 | - | - | 57 | 94 | 52.0 | 19.5 | 0.6 | 5.0 | - | 0.63 (0.48-0.83) |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| NHLBI (Brensike et al. 1984; | 59 | 57 | - | - | 8 | 11 | 46.1 | 19.0 | 0.9 | 5.0 | - | $0.70(0.30-1.62)$ |
| Law et al. 1994b) |  |  |  |  |  |  |  |  |  |  |  |  |

Key: A, active treatment group; C, control group. No data.
This was not strictly a randomized trial but is included in many meta-analyses. Exclusion of this trial does not significantly alter the overall RR and $95 \% \mathrm{Cl}$. This was a multi-factorial trial but is included in many meta-analyses. Exclusion of this trial does not significantly alter the overall RR and $95 \% \mathrm{Cl}$. Only fatal events reported.
"Weighted mean" totals (weighted by person-years of follow-up).
Table 7.16 Clinical trials of cholesterol lowering by statin interventions

| Trial | Participants |  | Stroke |  | IHD |  | Mean age (years) | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | Cholesterol reduction | Follow-up | RR (95\% CI) |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | A | C | A | C | A | C |  |  |  |  | Stroke | IHD |
| 4S (Anonymous 1994b, 1995) | 2221 | 2223 | 70 | 98 | 464 | 691 | 59.0 | 18.6 | 1.8 | 5.4 | 0.71 (0.53-0.97) | 0.67 (0.61-0.74) |
| ACAPS (Furberg et al. 1994a) | 460 | 459 | 0 | 5 | 5 | 9 | 61.7 | 48.5 | 0.9 | 2.8 | 0.09 (0.01-1.64) | 0.55 (0.19-1.64) |
| AFCAPS/TEXCAPS (Downs et al. 1998) | 3304 | 3301 | - | - | 57 | 95 | 58.3 | 15.0 | 1.1 | 5.2 | - | 0.60 (0.43-0.83) |
| CARE (Sacks et al. 1996) | 2081 | 2078 | 54 | 78 | 212 | 274 | 59.0 | 13.8 | 1.1 | 5.0 | - | 0.77 (0.65-0.9) |
| CCAIT (Crouse et al. 1998; Waters et al. 1993; Waters et al. 1994) | 165 | 164 | 2 | 0 | 7 | 7 | 53.0 | 19.0 | 1.0 | 2.0 | 4.97 (0.24-102.74) | 0.99 (0.36-2.77) |
| EXCEL (Bradford et al. 1990, 1991; Crouse et al. 1998; Shear et al. 1992) | 6582 | 1663 | 10 | 1 | 47 | 18 | 56.0 | 41.0 | 1.6 | 0.9 | 2.53 (0.32-19.72) | 0.66 (0.38-1.13) |
| HPS (Heart Protection Study Collaborative Group 2002a, 2002b) | 10269 | 10267 | 444 | 585 | 898 | 1212 | - | 24.7 | 1.3 | 5.0 | 0.76 (0.67-0.86) | 0.74 (0.68-0.80) |
| KAPS (Salonen et al. 1995) | 224 | 223 | 2 | 4 | 3 | 8 | 57.5 | 0.0 | 1.2 | 3.0 | 0.50 (0.09-2.69) | 0.37 (0.10-1.39) |
| KLIS (Anonymous 2000) | 2219 | 1634 | 47 | 41 | 65 | 47 | 58.0 | 0.0 | 0.5 | 5.0 | 0.84 (0.56-1.28) | 1.02 (0.70-1.47) |


| LIPID (Anonymous 1998b; White et al. 2000) | 4512 | 4502 | 169 | 204 | 557 | 715 | 62.0 | 16.8 | 1.0 | 6.1 | 0.83 (0.68-1.01) | 0.78 (0.70-0.86) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| LRT (Weintraub et al. 1994) | 203 | 201 | 0 | 1 | 14 | 5 | 62.0 | 28.0 | 2.0 | 4.0 | 0.33 (0.01-8.05) | 2.77 (1.02-7.55) |
| MAAS (Anonymous 1994a; Crouse et al. 1998; Dumont 1993) | 193 | 188 | 1 | 2 | 11 | 7 | 55.3 | 11.8 | 1.5 | 4.0 | 0.49 (0.04-5.33) | 1.53 (0.61-3.86) |
| MARS (Blankenhorn et al. 1993) | 134 | 136 | 0 | 0 | 22 | 31 | 58.0 | 9.0 | 1.8 | 2.2 | - | 0.72 (0.44-I.18) |
| PLACI (Byington et al. 1995; Pitt et al. 1995) | 206 | 202 | 0 | 2 | 8 | 17 | 57.0 | 22.5 | 1.3 | 3.0 | 0.20 (0.01-4.06) | 0.46 (0.20-1.05) |
| PLAC2 (Crouse et al. 1992, 1995, 1998; Furberg et al. 1994b) | 75 | 76 | 1 | 3 | 4 | 10 | 61.7 | 14.6 | 1.3 | 3.0 | 0.34 (0.04-3.17) | 0.41 (0.13-I.24) |
| PMSG (Anonymous 1993) | 530 | 532 | 0 | 3 | 0 | 7 | 55.0 | 23.3 | 1.2 | 0.5 | 0.11 (0.01-2.07) | 0.07 (0.00-1.17) |
| REGRESS (Jukema et al. 1995) | 450 | 435 | 3 | 5 | 8 | 13 | 55.7 | 0.0 | 1.3 | 2.0 | 0.58 (0.14-2.41) | 0.59 (0.25-1.42) |
| Sahni (Law et al. I994b; Sahni et al. 1991) | 79 | 78 | 0 | 0 | 3 | 4 | 60.2 | 30.5 | 0.7 | 1.0 | - | 0.74 (0.17-3.20) |
| WOSCOPS (Anonymous 1992a; Shepherd et al. 1995) | 3302 | 3293 | 46 | 51 | 174 | 248 | 55.2 | 0.0 | 1.1 | 4.9 | 0.90 (0.61-1.34) | 0.70 (0.58-0.84) |


|  | 37209 | 31655 | 849 | 1083 | 2599 | 3418 | $58.8{ }^{\text {a }}$ | 14.6\% | $1.2^{\text {a }}$ | $5.1{ }^{\text {a }}$ | 0.77 (0.70-0.84) | 0.73 (0.70-0.77) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Key: | A, active treatment group; C, control group. |  |  |  |  |  |  |  |  |  |  |  |
|  | No data. |  |  |  |  |  |  |  |  |  |  |  |
|  | "Weighted mean" totals (weighted by p | son-year | f follo | -up). |  |  |  |  |  |  |  |  |

results (Bucher et al. 1998). Another review is more relevant in addressing this issue. It examined the proportional reduction in risk from each trial titrated to a cholesterol reduction of $0.6 \mathrm{mmol} / \mathrm{l}$ as a yardstick. There was a dose-response relationship with bigger reductions in risk in trials attaining bigger cholesterol reductions, and statins tended to achieve greater reduction in total cholesterol. It was clear however, that for each method of lowering cholesterol (different drugs, diet and surgery), the proportional reduction in risk was close to that expected from the cholesterol reduction, suggesting little difference in relative effect of agents (Law et al. 2003).

Evidence from meta-analyses suggested that there were similar benefits from cholesterol lowering for males and females, and fatal and nonfatal events, but there was no evidence of age attenuation (LaRosa et al. 1999; Law et al. 1994b). However, trials are often not powered for such age subgroup analysis and often include only a narrow age band. On the whole, no substantial evidence has been presented for a significant difference in risk reduction of IHD between trials among those individuals with or without a prior history of cardiovascular disease (Atkins et al. 1993; Crouse et al. 1997, 1998; Hebert et al. 1997; Holme 1990; LaRosa et al. 1999; Law et al. 1994b).

Further important results from meta-analyses of cholesterol reduction trials relate to the time frame of the achieved benefits. A major overview published by Law et al. (1994b) included 28 trials of diet and non-statin drugs, 45254 individuals and 4421 deaths from IHD. It demonstrated that most of the benefits were evident after 5 years of sustained cholesterol lowering. In the first two years, a $0.6 \mathrm{mmol} / \mathrm{l}$ reduction in cholesterol was associated with a $7 \%(95 \%$ CI $0-14 \%)$ lower risk of IHD, but after 5 years, it was $25 \%$ ( $15-35 \%$ ) (Law et al. 1994b). An updated overview which included more cholesterol lowering trials including statins has confirmed these patterns with a $1.0 \mathrm{mmol} / \mathrm{l}$ reduction in LDL cholesterol associated with a $11 \% ~(95 \%$ CI $4-18 \%)$ lower risk of IHD at year 1. At 2 years this increased to $24 \%(17-30 \%)$, at $3-5$ years it was $33 \%(28-37 \%)$, and after 5 years it was $36 \%$ ( $26-45 \%$ ). (Law et al. 2003).

Most other reviews have not reported events according to duration of treatment, and instead only give an "average" risk reduction for the entire period of follow-up. A risk reduction of $20 \%$ after 5 years of follow-up, for example, for a given age group is a combined risk reduction for $1,2,3,4$ and 5 years, rather than a true representation of the risk reduction for a cohort of people after 5 years of treatment, which would be greater than $20 \%$. These data are a vital component to interpreting the time frames and extent of reversibility. The majority of the benefit of cholesterol lowering was achieved after 5 years (Law et al. 1994b). At this time, virtually all of the risk predicted by the observational studies had been reversed. There is no evidence that the effects of
cholesterol lowering on risk of IHD differ between males and females (Byington et al. 2001; LaRosa et al. 1999; Law et al. 1994b).

## Cholesterol and risk of stroke

Early observational studies, trials and overviews did not find an association between cholesterol and stroke. Part of the reason for this was that the analyses involved total stroke, and "concealed" differences in the type of association between cholesterol and each stroke subtype (Crouse et al. 1998). Early trials also only achieved small reductions in cholesterol and included few, mostly fatal, stroke events, which made the influence of cholesterol lowering difficult to detect (Crouse et al. 1998; Law 1999). Trials also tended to be of short duration, so that the long-term reduction in stroke was "diluted" by the absence of an early effect (Law 1999).

Data from individual trials are presented in Tables 7.14-7.16. As with IHD end-points, there were similar risk reduction in males and females and no evidence of age attenuation (Byington et al. 2001). However, trials were often not powered for this subgroup analysis and included only a narrow age-band.

The larger, more recent trial overviews have found a positive association between cholesterol and total stroke, and the size of the risk reduction in more recent overviews was about 20-25\% (Blauw et al. 1997; Bucher et al. 1998; Crouse et al. 1997, 1998; Hebert et al. 1997; Ross et al. 1999). Overviews have also noted a stronger association with non-fatal strokes (which comprise a higher proportion of ischaemic strokes) than fatal strokes (which comprise a greater proportion of haemorrhagic strokes) (Atkins et al. 1993; Blauw et al. 1997; Byington et al. 2001; Ross et al. 1999). One of the most recent overviews performed subgroup analysis and suggested different associations between cholesterol and haemorrhagic and ischaemic strokes (Byington et al. 2001), as predicted by the observational data.

In summary, the data from clinical trials of cholesterol lowering indicated that after 5 years, a reduction of approximately $0.6 \mathrm{mmol} / \mathrm{l}$ of cholesterol resulted in reversal of most or all of the epidemiologically expected risk for IHD (about 27\%). Interpretation is more difficult for stroke as many studies have combined ischaemic and haemorrhagic strokes into one category, when the association of cholesterol with ischaemic stroke differs from that of haemorrhagic stroke. Even when stroke subtype is specified, there will likely be differing rates of misclassification of stroke subtype between studies. However, more recent trial overviews appear to concur with the epidemiological data and suggest that total stroke risk can be reversed, but associations differ by stroke subtype. Unfortunately, analyses have not specifically analysed results by duration, so similar time frames for the cholesterol-IHD association will have to be used. These time frames for reversibility from trials will be
used in CRA analyses, but the relative risk estimates from cohort studies will be used in preference to those from trials for both stroke and IHD. Cohort studies are much larger than the trials, their statistical power is greater, and they are also better able to examine associations across a wide range of cholesterol values (Law 1999). Furthermore, age-specific estimates are more readily available.

## 4. Results

### 4.1 Attributable fraction

The "attributable fraction" refers to the proportion of disease burden that would theoretically not have occurred if the population distribution of cholesterol had been equal to that of the theoretical minimum (mean cholesterol $=3.8 \mathrm{mmol} / \mathrm{l}$ ).

## Stroke

Globally, $32 \%$ of ischaemic stroke was attributable to a total cholesterol level $>3.8 \mathrm{mmol} / \mathrm{l}$ (range of $25-45 \%$ by subregion). Subregions with the lowest attributable fractions included AFR-D, AFR-E and WPRB. In contrast, EUR-C and EUR-A had the highest values. In most subregions, the attributable fractions were slightly lower for males than females.

## ISCHAEMIC HEART DISEASE

Globally, $56 \%$ of IHD was attributable to a total cholesterol level $>3.8 \mathrm{mmol} / \mathrm{l}$ (range of $44-68 \%$ by subregion). Subregions with the lowest attributable fractions included AFR-D, SEAR-B and WPR-B. In contrast, EUR-C and EUR-A had the highest values.

### 4.2 Attributable disease burden

For the year 2000, the World Health Organization (WHO) estimated that there were 55.9 million deaths and 1455 million DALYs worldwide; approximately $22 \%$ of these deaths and $7 \%$ of these DALYs were due to ischaemic heart disease and stroke. The burden of disease was distributed across the developed and developing world (e.g. about one quarter of all stroke deaths and one third of IHD deaths occurred in the least developed regions), and predominantly affected those aged $>60$ years.

The "attributable burden" refers to the number of deaths or DALYs that would theoretically not have occurred if the population distribution of cholesterol had been equal to that of the theoretical minimum (mean cholesterol $=3.8 \mathrm{mmol} / \mathrm{I})$. In total, the attributable burden equated to 805000 stroke deaths, and about 3.6 million IHD deaths. For DALYs, these figures were 8.3 million stroke DALYs, and 32.1 million IHD DALYs (Table 7.17).

Table 7.I7 Attributable deaths and DALYs for cholesterol $>3.8 \mathrm{mmol} / \mathrm{l}$ by subregion and cardiovascular end-point

| Subregion | Stroke | IHD | Stroke | IHD |
| :---: | :---: | :---: | :---: | :---: |
|  | Deaths (000s) |  | DALYs (000s) |  |
| AFR-D | 18 | 68 | 219 | 738 |
| AFR-E | 21 | 68 | 267 | 767 |
| AMR-A | 33 | 317 | 377 | 2086 |
| AMR-B | 31 | 136 | 448 | 1425 |
| AMR-D | 3 | 15 | 47 | 150 |
| EMR-B | 7 | 75 | 89 | 836 |
| EMR-D | 26 | 188 | 287 | 2037 |
| EUR-A | 96 | 451 | 779 | 2600 |
| EUR-B | 45 | 235 | 474 | 1984 |
| EUR-C | 176 | 729 | 1637 | 5683 |
| SEAR-B | 18 | 94 | 224 | 1016 |
| SEAR-D | 145 | 851 | 1542 | 9548 |
| WPR-A | 20 | 58 | 219 | 389 |
| WPR-B | 165 | 322 | 1724 | 2847 |
| World | 805 | 3609 | 8332 | 32105 |

Worldwide this means that 4.4 million deaths (about $7.9 \%$ of the total) and 40.4 million DALYs $(2.8 \%$ of the total) were estimated to be due to non-optimal cholesterol. The proportion of DALYs was lower than the proportion of deaths, as most cholesterol-related cardiovascular disease events occurred in middle or old age, so the years of life lost due to premature mortality and years of life lost due to disability is a smaller percentage of the total DALYs than of the total deaths.

The age group with the greatest attributable deaths for both males and females was $70-79$ years. The number of attributable deaths then declined, reflecting the smaller denominator population, and the smaller total number of events in the oldest age group. For DALYs, this decline occurred sooner, as there are less years of life lost and years of life lived with disability with advancing age.

For each end-point, the attributable deaths were higher for males than females for the age groups 30-44 and 45-59 years. In contrast, attributable deaths were higher for females in the oldest age groups, $70-79$ and $\geq 80$ years. There was a similar trend for attributable DALYs. This was due to older females having higher mean cholesterol (and therefore greater attributable fraction) than males, and there are also a larger
number of cardiovascular deaths occurring in older females compared to older males.

## Stroke

The subregions with the highest attributable ischaemic stroke burden were EUR-C ( 176000 deaths, 1.6 million DALYs), WPR-B ( 165000 deaths, 1.7 million DALYs) and SEAR-D ( 145000 deaths, 1.5 million DALYs).

ISCHAEMIC HEART DISEASE
The subregions with the highest attributable IHD burden were SEAR-D ( 851000 deaths, 9.6 million DALYs) and EUR-C ( 729000 deaths, 5.7 million DALYs).

Overall, approximately $40 \%$ of attributable DALYs occurred in the most developed subregions ( 16.2 million), $40 \%$ ( 15.6 million) in high mortality developing subregions, and $20 \%$ ( 8.6 million) in low mortality developing subregions. A higher proportion of these deaths and DALYs were from IHD than stroke in all subregions.

These results indicate where most of the worldwide attributable cardiovascular disease burden occurred, and provide any given subregion with an indication of absolute size of the attributable burden. However, the age structure, population size, and the number of estimated events occurring in a subregion, influences the ranking of subregions by absolute number of attributable deaths and DALYs. The relative impact of the attributable DALYs indicates that between about $2-25 \%$ of all deaths and $0.5-12 \%$ of all DALYs across the subregions were attributable to non-optimal cholesterol. Approximately $16 \%$ of all deaths and $8 \%$ of all DALYs were attributable to excess cardiovascular disease in developed subregions, and about $5 \%$ and $2 \%$ in high and low mortality developing subregions, respectively.

## 5. Discussion

### 5.1 Attributable deaths and disease burden

The analyses in this chapter suggest that globally, a substantial proportion of cardiovascular disease is attributable to non-optimal cholesterol, defined as mean cholesterol $>3.8 \mathrm{mmol} / \mathrm{l}$. Overall, about one third of ischaemic stroke and over half of IHD was attributable to non-optimal cholesterol. The attributable fractions were higher in the more developed parts of the world than the least developed regions, as would be expected given the higher cholesterol levels.

Worldwide, 4.4 million deaths (about $7.9 \%$ of the total) and 40.4 million DALYs ( $2.8 \%$ of the total) were estimated to be due to nonoptimal cholesterol. Overall, the results suggest that a considerable proportion of cardiovascular disease is related to non-optimal cholesterol
and this translates into an important portion of deaths and years of life lost to deaths and disability worldwide.

### 5.2 Attributable deaths and disease burden by subregion

In absolute terms, most of the excess burden of cardiovascular disease occurred in subregions WPR-B, SEAR-D and EUR-C; approximately $10-20 \%$ of worldwide attributable deaths and DALYs occurred in these subregions. The relative impact that attributable deaths and DALYs have in different subregions may also be calculated as the proportion of all deaths and DALYs attributable to non-optimal cholesterol (i.e. mean $>3.8 \mathrm{mmol} / \mathrm{l}$ ) within each specific subregion. Overall, between $2 \%$ and $25 \%$ of all deaths and 0.5 and $12 \%$ of all DALYs across subregions were attributable to non-optimal cholesterol. This excess burden was highest in European subregions where mean cholesterol levels were highest.

## 6. Trends in mean cholesterol levels over time by age, sex and Subregion

The CRA project includes estimates of avoidable burden, that is, the reduction in future burden that would be observed if current levels of exposure to a risk factor were reduced to those specified by a counterfactual distribution of exposure. To perform these calculations an estimate must be made of cholesterol distributions in the future under a "business-as-usual" scenario. This requires estimates of changes in current cholesterol levels 10,20 and 30 years into the future by age, sex and subregion.

Several basic steps were necessary to produce these estimates and each one to some extent dictated the next. First, the literature was reviewed to assess what we already know about trends in cholesterol in different populations. Second, subregions were categorized into broad groupings based on likely direction of changes in cholesterol levels over time. Third, the future trends in cholesterol distribution were estimated.

### 6.1 Literature review of cholesterol trends over time

There is evidence in the literature that risk factors such as cholesterol change over time, and a variety of studies from different world regions and populations have demonstrated increases and decreases in cholesterol levels (Capewell et al. 2000; Dobson et al. 1999; Evans et al. 2001; Law and Wald 1994; McGovern et al. 1996; Sakata and Labarthe 1996; Sigfusson et al. 1991; Suh 2001; Tuomilehto et al. 1991; Vartiainen et al. 1994, 2000). These data were reviewed by subregion. In addition, data on fat intake from the Food and Agriculture Organization of the United Nations (FAO) were reviewed as fat intake correlates with mean cholesterol levels. Changes in mean cholesterol levels in recent literature relate predominantly to AMR-A, EUR-A and WPR-A.

Figure 7.15 Percentage of total calories from fat per capita per day


The best available FAO data (www.fao.org) cover the whole world, but the country groupings differ from those of WHO. FAO data may be presented as the total number of calories supplied as fat per day, or percentage of daily calories supplied as fat (Figure 7.15). (These data only relate to total fat, and have not been subdivided further, e.g. into saturated or animal fat vs unsaturated fat.)

## AMR-A AND EUR-A

Studies in the United Kingdom of Great Britain and Northern Ireland (Capewell et al. 1999; Law and Wald 1994), the United States (Law and Wald 1994; McGovern et al. 1996), Finland (Tuomilehto et al. 1991; Vartiainen et al. 1994, 2000), Iceland (Sigfusson et al. 1991) and other parts of Europe (Law and Wald 1994) have documented reductions in mean cholesterol levels. In the more recent studies from the 1980s onwards, these decreases have been approximately $0.02-0.03 \mathrm{mmol} / \mathrm{l}$ of cholesterol per year in males and females (Capewell et al. 2000; Dobson et al. 1999; Evans et al. 2001; McGovern et al. 1996; Vartiainen et al. 2000). FAO data also show a levelling or decline of mean cholesterol levels recently in North America (AMR-A) and Europe (EUR-A). The direction of these trends correlates with those demonstrated for blood pressure in the same studies.

WPR-A
Reduction in mean cholesterol levels has been documented in New Zealand and Australia (Capewell et al. 2000; Dobson et al. 1999; Law
and Wald 1994). In contrast, there appears to have been an increase in mean cholesterol levels in Japan since the 1980s (Law and Wald 1994; Okayama et al. 1993; Sakata and Labarthe 1996) in the order of about $0.03-0.04 \mathrm{mmol} / \mathrm{l}$ per year (Sakata and Labarthe 1996). This would result in an overall increase in cholesterol levels in WPR-A, owing to the larger population size of Japan. WPR-A covers countries that would be included by FAO in Asia (Japan) and Oceania (Australia, New Zealand and the Pacific), where trends in fat consumption indicate that cholesterol levels are increasing.

## EUR-B AND EUR-C

Some decreases in cholesterol have been recorded recently in parts of eastern Europe (Bobak et al. 1997), but the data are limited. The FAO region that covers the former Soviet Union and the Russian Federation most closely maps to EUR-B and C, and also suggests a levelling off and possible decline in fat consumption.

WPR-B
Mean cholesterol levels appear to be increasing in China and in the Republic of Korea (Evans et al. 2001; Suh 2001), which correlates with increases in fat consumption in China.

## Other subregions

Very few data were available on several developing subregions (AFR-D, E, AMR-B, D, EMR-B, D, SEAR-B and D); however, certain data demonstrated that as regions become more industrialized or "acculturated", risk factor profiles such as blood pressure and cholesterol change. This is particularly evident in migration studies conducted in a variety of settings (Poulter and Sever 1994), such as Africa (Poulter et al. 1988, 1990), China (He et al. 1991a, 1991b), and the Pacific (Joseph et al. 1983; Salmond et al. 1985, 1989), which suggest that blood pressure and cholesterol levels rise after people migrate to more urbanized "acculturated" settings. The trends are as follows.

AMR-B and D: FAO data from Central and South America suggest that overall, fat consumption is increasing.

EMR-B and D; SEAR-B and D: These subregions are all included in the FAO grouping of Asia where levels of fat consumption are increasing.

AFR-D and E: FAO data suggest small increases in fat consumption.

### 6.2 Potential categories of cholesterol change

Data presented above indicate that mean cholesterol levels appear to be decreasing in many of the most developed subregions, while they are increasing or plateauing in less developed subregions. However, estimating how long the current trends will continue and quantifying the changes is difficult.

A published meta-analysis assessed the quantitative importance of dietary fat to blood concentrations of total cholesterol (Clarke et al. 1997). However, these results are difficult to extrapolate to FAO data as they relate to isocaloric replacement of fats by carbohydrates. The FAO data demonstrated not only increasing caloric fat intake, but also increasing caloric intake from other sources and overall. Therefore, the same associations do not apply. The following potential scenario was therefore based on current evidence of time trends, and previously documented changes in mean cholesterol levels.

Owing to the limited data available, these estimates are susceptible to a high degree of uncertainty. A wide range of factors could influence future risk factor levels, and it is difficult to capture these factors even with sophisticated modelling. A decision was thus made to use a relatively simple, but transparent method for the purposes of these analyses.

## Estimates of mean cholesterol levels over time

Table 7.18 presents an overview of the cholesterol trends over time, by subregion. The following estimates summarize mean cholesterol levels by subregion, sex and age, currently and in 2010, 2020 and 2030 (Tables 7.19-7.22).

Table 7.18 Scenario for changes in mean cholesterol levels (mmol/l) by subregion over the next 10,20 and 30 years $^{2}$

| Subregion | 2000 to 2010 | 2010 to 2020 | 2020 to 2030 | Comment |
| :---: | :---: | :---: | :---: | :---: |
| AMR-A | $\downarrow 0.3 \mathrm{mmol} / \mathrm{l}$ | $\downarrow 0.3 \mathrm{mmol} / \mathrm{l}$ | $\downarrow 0.2 \mathrm{~mol} / \mathrm{l}$ | $\downarrow$ attenuating over time |
| EUR-A | $\downarrow 0.3 \mathrm{mmol} / \mathrm{l}$ | $\downarrow 0.3 \mathrm{mmol} / \mathrm{l}$ | $\downarrow 0.2 \mathrm{~mol} / \mathrm{l}$ | $\downarrow$ attenuating over time |
| WPR-A | $\uparrow 0.2 \mathrm{mmol} / \mathrm{l}$ | $\rightarrow$ plateau | $\downarrow 0.1 \mathrm{~mol} / \mathrm{l}$ | $\uparrow$ due to influence of Japan over next 10 years, then plateau and start to decrease |
| AMR-B, D | $\uparrow 0.2 \mathrm{mmol} / \mathrm{l}$ | $\uparrow 0.1 \mathrm{mmol} / \mathrm{l}$ | $\rightarrow$ plateau | $\uparrow$ attenuating over time, then plateau |
| EUR-B, C | $\rightarrow$ plateau | $\downarrow 0.1 \mathrm{~mol} / \mathrm{l}$ | $\downarrow 0.1 \mathrm{~mol} / \mathrm{l}$ | $\rightarrow$ plateau initially then start to decrease as some evidence on this already happening in some regions |
| WPR-B | $\uparrow 0.3 \mathrm{mmol} / \mathrm{l}$ | $\uparrow 0.2 \mathrm{mmol} / \mathrm{l}$ | $\uparrow 0.1 \mathrm{mmol} / \mathrm{l}$ | $\uparrow$ attenuating over time |
| SEAR | $\uparrow 0.2 \mathrm{mmol} / \mathrm{l}$ | $\uparrow 0.1 \mathrm{mmol} / \mathrm{l}$ | $\rightarrow$ plateau | $\uparrow$ attenuating over time, then plateau |
| EMR | $\uparrow 0.2 \mathrm{mmol} / \mathrm{l}$ | $\uparrow 0.1 \mathrm{mmol} / \mathrm{l}$ | $\rightarrow$ plateau | $\uparrow$ attenuating over time, then plateau |
| AFR | $\uparrow 0.2 \mathrm{mmol} / \mathrm{l}$ | $\uparrow 0.1 \mathrm{mmol} / \mathrm{l}$ | $\rightarrow$ plateau | $\uparrow$ attenuating over time, then plateau |

[^22]Table 7.19 Current CRA estimates of mean cholesterol (mmol/l) by subregion, sex and age

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  | Males |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ |
| AFR-D | 4.6 | 5.1 | 5.2 | 5.2 | 4.5 | 4.7 | 4.7 | 4.7 |
| AFR-E | 4.6 | 5.1 | 5.2 | 5.2 | 4.5 | 4.7 | 4.7 | 4.7 |
| AMR-A | 4.8 | 5.6 | 6.0 | 5.9 | 5.2 | 5.5 | 5.5 | 5.2 |
| AMR-B | 4.8 | 5.4 | 5.4 | 5.2 | 5.3 | 5.0 | 4.9 | 4.8 |
| AMR-D | 4.8 | 5.4 | 5.4 | 5.2 | 5.3 | 5.0 | 4.9 | 4.8 |
| EMR-B | 4.6 | 5.1 | 5.4 | 5.6 | 4.8 | 4.9 | 5.1 | 5.4 |
| EMR-D | 4.6 | 5.1 | 5.4 | 5.6 | 4.8 | 4.9 | 5.1 | 5.4 |
| EUR-A | 5.4 | 6.3 | 6.7 | 6.5 | 5.9 | 6.1 | 6.1 | 5.9 |
| EUR-B | 4.6 | 5.3 | 5.5 | 5.4 | 5.0 | 5.1 | 5.2 | 5.2 |
| EUR-C | 5.4 | 6.2 | 6.6 | 6.5 | 5.5 | 5.8 | 5.8 | 5.7 |
| SEAR-B | 4.0 | 4.5 | 4.6 | 4.6 | 5.0 | 5.1 | 5.2 | 5.2 |
| SEAR-D | 4.9 | 5.7 | 5.8 | 5.7 | 4.8 | 5.0 | 5.0 | 5.0 |
| WPR-A | 4.9 | 5.5 | 5.6 | 5.4 | 5.2 | 5.2 | 5.1 | 5.0 |
| WPR-B | 4.3 | 4.8 | 5.1 | 5.1 | 4.5 | 4.6 | 4.7 | 4.8 |

a The same cholesterol levels apply to those aged $\geq 80$ years.

Table 7.20 Estimates of mean cholesterol (mmol/l) by subregion, sex and age in 2010

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  | Males |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ |
| AFR-D | 4.8 | 5.3 | 5.4 | 5.4 | 4.7 | 4.9 | 4.9 | 4.9 |
| AFR-E | 4.8 | 5.3 | 5.4 | 5.4 | 4.7 | 4.9 | 4.9 | 4.9 |
| AMR-A | 4.5 | 5.3 | 5.7 | 5.6 | 4.9 | 5.2 | 5.2 | 4.9 |
| AMR-B | 5.0 | 5.6 | 5.6 | 5.4 | 5.5 | 5.2 | 5.1 | 5.0 |
| AMR-D | 5.0 | 5.6 | 5.6 | 5.4 | 5.5 | 5.2 | 5.1 | 5.0 |
| EMR-B | 4.8 | 5.3 | 5.6 | 5.8 | 5.0 | 5.1 | 5.3 | 5.6 |
| EMR-D | 4.8 | 5.3 | 5.6 | 5.8 | 5.0 | 5.1 | 5.3 | 5.6 |
| EUR-A | 5.1 | 6.0 | 6.4 | 6.2 | 5.6 | 5.8 | 5.8 | 5.6 |
| EUR-B | 4.6 | 5.3 | 5.5 | 5.4 | 5.0 | 5.1 | 5.2 | 5.2 |
| EUR-C | 5.4 | 6.2 | 6.6 | 6.5 | 5.5 | 5.8 | 5.8 | 5.7 |
| SEAR-B | 4.2 | 4.7 | 4.8 | 4.8 | 5.2 | 5.3 | 5.4 | 5.4 |
| SEAR-D | 5.1 | 5.9 | 6.0 | 5.9 | 5.0 | 5.2 | 5.2 | 5.2 |
| WPR-A | 5.1 | 5.7 | 5.8 | 5.6 | 5.4 | 5.4 | 5.3 | 5.2 |
| WPR-B | 4.6 | 5.1 | 5.4 | 5.4 | 4.8 | 4.9 | 5.0 | 5.1 |

[^23]Table 7.2I Estimates of mean cholesterol (mmol/l) by subregion, sex and age in 2020

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  | Males |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ |
| AFR-D | 4.9 | 5.4 | 5.5 | 5.5 | 4.8 | 5.0 | 5.0 | 5.0 |
| AFR-E | 4.9 | 5.4 | 5.5 | 5.5 | 4.8 | 5.0 | 5.0 | 5.0 |
| AMR-A | 4.2 | 5.0 | 5.4 | 5.3 | 4.6 | 4.9 | 4.9 | 4.6 |
| AMR-B | 5.1 | 5.7 | 5.7 | 5.5 | 5.6 | 5.3 | 5.2 | 5.1 |
| AMR-D | 5.1 | 5.7 | 5.7 | 5.5 | 5.6 | 5.3 | 5.2 | 5.1 |
| EMR-B | 4.9 | 5.4 | 5.7 | 5.9 | 5.1 | 5.2 | 5.4 | 5.7 |
| EMR-D | 4.9 | 5.4 | 5.7 | 5.9 | 5.1 | 5.2 | 5.4 | 5.7 |
| EUR-A | 4.8 | 5.7 | 6.1 | 5.9 | 5.3 | 5.5 | 5.5 | 5.3 |
| EUR-B | 4.5 | 5.2 | 5.4 | 5.3 | 4.9 | 5.0 | 5.1 | 5.1 |
| EUR-C | 5.3 | 6.1 | 6.5 | 6.4 | 5.4 | 5.7 | 5.7 | 5.6 |
| SEAR-B | 4.3 | 4.8 | 4.9 | 4.9 | 5.3 | 5.4 | 5.5 | 5.5 |
| SEAR-D | 5.2 | 6.0 | 6.1 | 6.0 | 5.1 | 5.3 | 5.3 | 5.3 |
| WPR-A | 5.1 | 5.7 | 5.8 | 5.6 | 5.4 | 5.4 | 5.3 | 5.2 |
| WPR-B | 4.8 | 5.3 | 5.6 | 5.6 | 5.0 | 5.1 | 5.2 | 5.3 |

a The same cholesterol levels apply to those aged $\geq 80$ years.

Table 7.22 Estimates of mean cholesterol ( $\mathrm{mmol} / \mathrm{l}$ ) by subregion, sex and age in 2030

| Subregion | Sex and age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  |  | Males |  |  |  |
|  | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ | 30-44 | 45-59 | 60-69 | 70-79 ${ }^{\text {a }}$ |
| AFR-D | 4.9 | 5.4 | 5.5 | 5.5 | 4.8 | 5.0 | 5.0 | 5.0 |
| AFR-E | 4.9 | 5.4 | 5.5 | 5.5 | 4.8 | 5.0 | 5.0 | 5.0 |
| AMR-A | 4.0 | 4.8 | 5.2 | 5.1 | 4.4 | 4.7 | 4.7 | 4.4 |
| AMR-B | 5.1 | 5.7 | 5.7 | 5.5 | 5.6 | 5.3 | 5.2 | 5.1 |
| AMR-D | 5.1 | 5.7 | 5.7 | 5.5 | 5.6 | 5.3 | 5.2 | 5.1 |
| EMR-B | 4.9 | 5.4 | 5.7 | 5.9 | 5.1 | 5.2 | 5.4 | 5.7 |
| EMR-D | 4.9 | 5.4 | 5.7 | 5.9 | 5.1 | 5.2 | 5.4 | 5.7 |
| EUR-A | 4.6 | 5.5 | 5.9 | 5.7 | 5.1 | 5.3 | 5.3 | 5.1 |
| EUR-B | 4.4 | 5.1 | 5.3 | 5.2 | 4.8 | 4.9 | 5.0 | 5.0 |
| EUR-C | 5.2 | 6.0 | 6.4 | 6.3 | 5.3 | 5.6 | 5.6 | 5.5 |
| SEAR-B | 4.3 | 4.8 | 4.9 | 4.9 | 5.3 | 5.4 | 5.5 | 5.5 |
| SEAR-D | 5.2 | 6.0 | 6.1 | 6.0 | 5.1 | 5.3 | 5.3 | 5.3 |
| WPR-A | 4.9 | 5.5 | 5.6 | 5.4 | 5.2 | 5.2 | 5.1 | 5.0 |
| WPR-B | 4.9 | 5.4 | 5.7 | 5.7 | 5.1 | 5.2 | 5.3 | 5.4 |

[^24]
## Acknowledgements

Particular thanks are due to the Asia-Pacific Cohort Studies Collaboration secretariat, and all study collaborators for allowing us to access and analyse their data for estimates of relative risk. The Prospective Studies Collaboration assisted with analyses and provided unpublished data on relative risk estimates that could be compared with those presented here.

The authors are grateful to Varsha Parag and Derrick Bennett for biostatistical support. Angela Hurley provided valuable research assistance and Clarissa Gould-Thorpe provided secretarial support.

## Note

1 See preface for an explanation of this term.

## References

Abbott RD, Sharp DS, Burchfiel CM et al. (1997) Cross-sectional and longitudinal changes in total and high-density-lipoprotein cholesterol levels over a 20-year period in elderly men: the Honolulu Heart Program. Annals of Epidemiology, 7:417-424.
Acheson J, Hutchinson EC (1972) Controlled trial of clofibrate in cerebral vascular disease. Atherosclerosis, 15:177-183.
al-Mahroos F, al-Roomi K, McKeigue PM (2000) Relation of high blood pressure to glucose intolerance, plasma lipids and educational status in an Arabian Gulf population. International Journal of Epidemiology, 29:71-76.
al-Nuaim AR (1997) Serum total and fractionated cholesterol distribution and prevalence of hypercholesterolemia in urban and rural communities in Saudi Arabia. International Journal of Cardiology, 58:141-149.
al-Nuaim AR, al-Rubeaan K, al-Mazrou Y, al-Attas O, al-Daghari N (1996) Prevalence of hypercholesterolemia in Saudi Arabia, epidemiological study. International Journal of Cardiology, 54:41-49.
al-Nuaim AR, al-Rubeaan K, al-Mazrou Y, al-Attas O, al-Daghari N (1997) Serum total, fractionated cholesterol concentration distribution and prevalence of hypercholesterolemia in Saudi Arabia, regional variation. Annals of Saudi Medicine, 17:179-184.
al-Shammari SA, Ali M, al-Shammari A, al-Maatouq M, Tennier A, Armstrong K (1994) Blood lipid concentrations and other cardiovascular risk factors among Saudis. Family Practice, 11:153-158.
Anderson CS, Jamrozik KD, Burvill PW, Chakera TM, Johnson GA, StewartWynne EG (1993) Determining the incidence of different subtypes of stroke: results from the Perth Community Stroke Study, 1989-1990. Medical Journal of Australia, 158:85-89.
Anonymous (1968) Controlled trial of soya-bean oil in myocardial infarction. The Lancet, 2:693-699.

Anonymous (1971a) Ischaemic heart disease: a secondary prevention trial using clofibrate. Report by a research committee of the Scottish Society of Physicians. British Medical Journal, 4:775-784.

Anonymous (1971b) Trial of clofibrate in the treatment of ischaemic heart disease. Five-year study by a group of physicians of the Newcastle upon Tyne region. British Medical Journal, 4:767-775.
Anonymous (1975) Clofibrate and niacin in coronary heart disease. Journal of the American Medical Association, 231:360-381.
Anonymous (1978) A co-operative trial in the primary prevention of ischaemic heart disease using clofibrate. Report from the Committee of Principal Investigators. British Heart Journal, 40:1069-1118.

Anonymous (1980) WHO cooperative trial on primary prevention of ischaemic heart disease using clofibrate to lower serum cholesterol: mortality follow-up. Report of the Committee of Principal Investigators. The Lancet, 2:379-385.

Anonymous (1981) Distribution of some risk factors for atherosclerosis in nine Italian population samples. The research group ATS-RF2 of the Italian National Research Council. American Journal of Epidemiology, 113: 338-346.

Anonymous (1984) The Lipid Research Clinics Coronary Primary Prevention Trial results. I. Reduction in incidence of coronary heart disease. Journal of the American Medical Association, 251:351-364.

Anonymous (1989) The WHO MONICA Project. A worldwide monitoring system for cardiovascular diseases: cardiovascular mortality and risk factors in selected communities. World Health Statistics, A:27-149.

Anonymous (1992a) A coronary primary prevention study of Scottish men aged 45-64 years: trial design. The West of Scotland Coronary Prevention Study Group. Journal of Clinical Epidemiology, 45:849-860.

Anonymous (1992b) Risk factors for cardiovascular disease in the developing world. A multicentre collaborative study in the International Clinical Epidemiology Network (INCLEN). INCLEN Multicentre Collaborative Group. Journal of Clinical Epidemiology, 45:841-847.
Anonymous (1993) Effects of pravastatin in patients with serum total cholesterol levels from 5.2 to $7.8 \mathrm{mmol} /$ liter ( 200 to $300 \mathrm{mg} / \mathrm{dl}$ ) plus two additional atherosclerotic risk factors. The Pravastatin Multinational Study Group for Cardiac Risk Patients. American Journal of Cardiology, 72:1031-1037.

Anonymous (1994a) Effect of simvastatin on coronary atheroma: the Multicentre Anti-Atheroma Study (MAAS). The Lancet, 344:633-638.
Anonymous (1994b) Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). The Lancet, 344:1383-1389.
Anonymous (1995) Baseline serum cholesterol and treatment effect in the Scandinavian Simvastatin Survival Study (4S). The Lancet, 345:1274-1275.
Anonymous (1996) Current state of and recent trends in serum lipid levels in the general Japanese population. Research Committee on Serum Lipid Level

Survey 1990 in Japan. Journal of Atherosclerosis and Thrombosis, 2: 122-132.

Anonymous (1998a) Blood pressure, cholesterol, and stroke in eastern Asia. Eastern Stroke and Coronary Heart Disease Collaborative Research Group. The Lancet, 352:1801-1807.
Anonymous (1998b) Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group. New England Journal of Medicine, 339: 1349-1357.

Anonymous (2000) Pravastatin use and risk of coronary events and cerebral infarction in Japanese men with moderate hypercholesterolemia: the Kyushu Lipid Intervention Study. Journal of Atherosclerosis and Thrombosis, 7: 110-121.

Aono H, Ozawa H, Bello MC, Ito M, Saito I (1997) Prevalence of risk factors for coronary heart disease among Dominicans in the Dominican republic: comparison with Japanese and Americans using existing data. Journal of Epidemiology, 7:238-243.
APCSC (1999) Determinants of cardiovascular disease in the Asia Pacific region: protocol for a collaborative overview of cohort studies. Cardiovascular Disease Prevention, 2:281-289.

Asplund-Carlson A, Carlson LA (1994) Studies in hypertriglyceridaemia. 1. Serum triglyceride distribution and its correlates in randomly selected Swedish middle-aged men. Journal of Internal Medicine, 236:57-64.

Atkins D, Psaty BM, Koepsell TD, Longstreth WT Jr., Larson EB (1993) Cholesterol reduction and the risk for stroke in men. A meta-analysis of randomized, controlled trials. Annals of Internal Medicine, 119:136-145.
Bajekal M, Boreham R, Erens B et al. (1999) Health Survey for England '98. The Department of Health, London.
Bamford J, Sandercock P, Dennis M, Burn J, Warlow C (1990) A prospective study of acute cerebrovascular disease in the community: the Oxfordshire Community Stroke Project-1981-86.2. Incidence, case fatality rates and overall outcome at one year of cerebral infarction, primary intracerebral and subarachnoid haemorrhage. Journal of Neurology, Neurosurgery and Psychiatry, 53:16-22.
Barnes R (1965) Comparisons of blood pressure and blood cholesterol levels in New Guineans and Australians. Medical Journal of Australia, 1:611617.

Benfante R, Yano K, Hwang LJ, Curb JD, Kagan A, Ross W (1994) Elevated serum cholesterol is a risk factor for both coronary heart disease and thromboembolic stroke in Hawaiian Japanese men. Implications of shared risk. Stroke, 25:814-820.

Bennett SA, Magnus P (1994) Trends in cardiovascular risk factors in Australia. Results from the National Heart Foundation's Risk Factor Prevalence Study, 1980-1989. Medical Journal of Australia, 161:519-527.

Bhuripanyo K, Tatsanavivat P, Matrakool B, Muktabhant B, Bhuripanyo P, Harnthaveesompol S (1993) A prevalence survey of lipids abnormalities of rural area in Amphoe Phon, Khon Kaen. Journal of the Medical Association of Thailand, 76:101-108.
Biro G, Antal M, Zajkas G (1996) Nutrition survey of the Hungarian population in a randomized trial between 1992-1994. European Journal of Clinical Nutrition, 50:201-208.
Blankenhorn DH, Azen SP, Kramsch DM et al. (1993) Coronary angiographic changes with lovastatin therapy. The Monitored Atherosclerosis Regression Study (MARS). The MARS Research Group. Annals of Internal Medicine, 119:969-976.

Blankenhorn DH, Nessim SA, Johnson RL, Sanmarco ME, Azen SP, Cashin-Hemphill L (1987) Beneficial effects of combined colestipol-niacin therapy on coronary atherosclerosis and coronary venous bypass grafts. Journal of the American Medical Association, 257:3233-3240.
Blauw GJ, Lagaay AM, Smelt AH, Westendorp RG (1997) Stroke, statins, and cholesterol. A meta-analysis of randomized, placebo-controlled, double-blind trials with HMG-CoA reductase inhibitors. Stroke, 28:946-950.
Bobak M, Skodova Z, Pisa Z, Poledne R, Marmot M (1997) Political changes and trends in cardiovascular risk factors in the Czech Republic, 1985-92. Journal of Epidemiology and Community Health, 51:272-277.

Bosma H, Appels A, Sturmans F, Grabauskas V, Gostautas A (1994) Differences in mortality and coronary heart disease between Lithuania and the Netherlands: results from the WHO Kaunas-Rotterdam Intervention Study (KRIS). International Journal of Epidemiology, 23:12-19.
Boulton TJ, Cockington RA, Hamilton-Craig I, Magarey AM, Mazumdar J (1995) A profile of heart disease risk factors and their relation to parents' education, fathers' occupation and family history of heart disease in 843 South Australian families: the Adelaide Children's WHO Collaborative Study. Journal of Paediatrics and Child Health, 31:200-206.

Bovet P, Shamlaye C, Kitua A, Riesen WF, Paccaud F, Darioli R (1991) High prevalence of cardiovascular risk factors in the Seychelles (Indian Ocean). Arteriosclerosis and Thrombosis, 11:1730-1736.
Bradford RH, Shear CL, Chremos AN et al. (1991) Expanded Clinical Evaluation of Lovastatin (EXCEL) study results. I. Efficacy in modifying plasma lipoproteins and adverse event profile in 8245 patients with moderate hypercholesterolemia. Archives of Internal Medicine, 151:43-49.
Bradford RH, Shear CL, Chremos AN et al. (1990) Expanded clinical evaluation of lovastatin (EXCEL) study: design and patient characteristics of a double-blind, placebo-controlled study in patients with moderate hypercholesterolemia. American Journal of Cardiology, 66:B44-55.

Brensike JF, Levy RI, Kelsey SF et al. (1984) Effects of therapy with cholestyramine on progression of coronary arteriosclerosis: results of the NHLBI Type II Coronary Intervention Study. Circulation, 69:313-324.
Bronner LL, Kanter DS, Manson JE (1995) Primary prevention of stroke. New England Journal of Medicine, 333:1392-1400.

Brown G, Albers JJ, Fisher LD, Schaefer SM et al. (1990) Regression of coronary artery disease as a result of intensive lipid-lowering therapy in men with high levels of apolipoprotein B. New England Journal of Medicine, 323: 1289-1298.

Brown SA, Hutchinson R, Morrisett J et al. (1993) Plasma lipid, lipoprotein cholesterol, and apoprotein distributions in selected US communities. The Atherosclerosis Risk in Communities (ARIC) Study. Arteriosclerosis and Thrombosis, 13:1139-1158.

Brown SL, Salive ME, Harris TB, Simonsick EM, Guralnik JM, Kohout FJ (1994) Low cholesterol concentrations and severe depressive symptoms in elderly people. British Medical Journal, 308:1328-1332.

Bucher HC, Griffith LE, Guyatt GH (1998) Effect of HMGcoA reductase inhibitors on stroke. A meta-analysis of randomized, controlled trials. Annals of Internal Medicine, 128:89-95.

Buchwald H, Campos CT (1995) Program on the surgical control of the hyperlipidemias: 1994 report. Journal of Cardiovascular Pharmacology, 25:S3-10.
Buchwald H, Varco RL, Matts JP et al. (1990) Effect of partial ileal bypass surgery on mortality and morbidity from coronary heart disease in patients with hypercholesterolemia. Report of the Program on the Surgical Control of the Hyperlipidemias (POSCH). New England Journal of Medicine, 323: 946-955.

Bullen C, Simmons G, Trye P, Lay-Yee R, Bonita R, Jackson R (1998) Cardiovascular disease risk factors in 65-84 year old men and women: results from the Auckland University Heart and Health Study 1993-4. New Zealand Medical Journal, 111:4-7.
Burke GL, Sprafka JM, Folsom AR, Hahn LP, Luepker RV, Blackburn H (1991) Trends in serum cholesterol levels from 1980 to 1987. The Minnesota Heart Survey. New England Journal of Medicine, 324:941-946.
Burr ML, Fehily AM, Gilbert JF et al. (1989) Effects of changes in fat, fish and fibre intakes on death and myocardial reinfarction: diet and reinfarction trial (DART). The Lancet, 2:757-761.

Byington RP, Davis BR, Plehn JF et al. (2001) Reduction of stroke events with Pravastatin: the Prospective Pravastatin Pooling (PPP) Project. Circulation, 103:387-392.

Byington RP, Jukema JW, Salonen JT et al. (1995) Reduction in cardiovascular events during pravastatin therapy. Pooled analysis of clinical events of the Pravastatin Atherosclerosis Intervention Program. Circulation, 92:2419-2425.

Capewell S, Beaglehole R, Seddon M, McMurray J (2000) Explanation for the decline in coronary heart disease mortality rates in Auckland, New Zealand, between 1982 and 1993. Circulation, 102:1511-1516.

Capewell S, Morrison CE, McMurray JJ (1999) Contribution of modern cardiovascular treatment and risk factor changes to the decline in coronary heart disease mortality in Scotland between 1975 and 1994. Heart, 81: 380-386.

Carlson LA, Rosenhamer G (1988) Reduction of mortality in the Stockholm Ischaemic Heart Disease Secondary Prevention Study by combined treatment with clofibrate and nicotinic acid. Acta Medica Scandinavica, 223:405-418.

Carvalho JJ, Baruzzi RG, Howard PF et al. (1989) Blood pressure in four remote populations in the INTERSALT Study. Hypertension, 14:238-246.
Casdorph HR (1972) Normal limits for serum-cholesterol. The Lancet, 1:1076.
Cesana G, Ferrario M, Sega R, Duzioni F, Zanettini R, Grieco A (1989) [New problems in occupational medicine: findings on the risk of coronary heart disease in the services industry]. Medicina del Lavoro, 80:192-200.
Chadha S, Gopinath N, Shekhawat S (1997) Urban-rural differences in the prevalence of coronary heart disease and its risk factors in Delhi. Bulletin of the World Health Organization, 75:31-38.
Chen D, Roman G, Wu G et al. (1992) Stroke in China (Sino-MONICA-Beijing study) 1984-1986. Neuroepidemiology, 11:15-23.
Chen Z, Peto R, Collins R, MacMahon S, Lu J, Li W (1991) Serum cholesterol concentration and coronary heart disease in population with low cholesterol concentrations. British Medical Journal, 303:276-282.
Choudhury SR, Ueshima H, Kita Y et al. (1994) Alcohol intake and serum lipids in a Japanese population. International Journal of Epidemiology, 23: 940-947.
Clarke R, Frost C, Collins R, Appleby P, Peto R (1997) Dietary lipids and blood cholesterol: quantitative meta-analysis of metabolic ward studies. British Medical Journal, 314:112-117.
Collins R, Peto R (1994) Antihypertensive drug therapy: effects on stroke and coronary heart disease. In: Textbook of hypertension. Swales J, ed. Blackwell Scientific Publications, Oxford.

Collins R, Peto R, MacMahon S et al. (1990) Blood pressure, stroke, and coronary heart disease. Part 2, short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. The Lancet, 335:827-838.
Connelly PW, MacLean DR, Horlick L, O’Connor B, Petrasovits A, Little JA (1992) Plasma lipids and lipoproteins and the prevalence of risk for coronary heart disease in Canadian adults. Canadian Heart Health Surveys Research Group. Canadian Medical Association Journal, 146:1977-1987.
Connor WE, Cerqueira MT, Connor RW, Wallace RB, Malinow MR, Casdorph HR (1978) The plasma lipids, lipoproteins, and diet of the Tarahumara Indians of Mexico. American Journal of Clinical Nutrition, 31:1131-1142.

Crouse JR 3rd, Byington RP, Bond MG et al. (1992) Pravastatin, lipids, and atherosclerosis in the carotid arteries: design features of a clinical trial with carotid atherosclerosis outcome. Controlled Clinical Trials, 13:495-506.
Crouse JR 3rd, Byington RP, Bond MG et al. (1995) Pravastatin, Lipids, and Atherosclerosis in the Carotid Arteries (PLAC-II). American Journal of Cardiology, 75:455-459.

Crouse JR 3rd, Byington RP, Furberg CD (1998) HMG-CoA reductase inhibitor therapy and stroke risk reduction: an analysis of clinical trials data. Atherosclerosis, 138:11-24.

Crouse JR 3rd, Byington RP, Hoen HM, Furberg CD (1997) Reductase inhibitor monotherapy and stroke prevention. Archives of Internal Medicine, 157:1305-1310.

Dayton S, Pearce ML, Goldman H et al. (1968) Controlled trial of a diet high in unsaturated fat for prevention of atherosclerotic complications. The Lancet, 2:1060-1062.
Detre KM, Shaw L (1974) Long-term changes of serum cholesterol with choles-terol-altering drugs in patients with coronary heart disease. Veterans Administration Drug-Lipid Cooperative Study. Circulation, 50:998-1005.

Dobson AJ, McElduff P, Heller R, Alexander H, Colley P, D'Este K (1999) Changing patterns of coronary heart disease in the hunter region of New South Wales, Australia. Journal of Clinical Epidemiology, 52:761-771.

Donker GA, Labarthe DR, Harrist RB et al. (1997) Low birth weight and serum lipid concentrations at age 7-11 years in a biracial sample. American Journal of Epidemiology, 145:398-407.
Dorr AE, Gundersen K, Schneider JC Jr, Spencer TW, Martin WB (1978) Colestipol hydrochloride in hypercholesterolemic patients-effect on serum cholesterol and mortality. Journal of Chronic Diseases, 31:5-14.

Downs JR, Clearfield M, Weis S et al. (1998) Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. Air Force/Texas Coronary Atherosclerosis Prevention Study. Journal of the American Medical Association, 279:1615-1622.

Dumont JM (1993) Effect of cholesterol reduction by simvastatin on progression of coronary atherosclerosis: design, baseline characteristics, and progress of the Multicenter Anti-Atheroma Study (MAAS). Controlled Clinical Trials, 14:209-228.

Eisenberg S, Heiss G, Friedlander Y et al. (1986) Comparison of plasma lipids, lipoproteins and dyslipoproteinemia in Israel and the United States. The Lipid Research Clinics Program Prevalence Study. Atherosclerosis, 59:63-74.

Erasmus RT, Uyot C, Pakeye T (1994) Plasma cholesterol distribution in a rural Nigerian population-relationship to age, sex and body mass. Central African Journal of Medicine, 40:299-302.
Ettinger WH, Wahl PW, Kuller LH et al. (1992) Lipoprotein lipids in older people. Results from the Cardiovascular Health Study. The CHS Collaborative Research Group. Circulation, 86:858-869.
Evans A, Tolonen H, Hense HW et al. (2001) Trends in coronary risk factors in the WHO MONICA project. International Journal of Epidemiology, 30:S35-40.
Fan WX, Parker R, Parpia B et al. (1990) Erythrocyte fatty acids, plasma lipids, and cardiovascular disease in rural China. American Journal of Clinical Nutrition, 52:1027-1036.

Ferrara A, Barrett-Connor E, Shan J (1997) Total, LDL, and HDL cholesterol decrease with age in older men and women. The Rancho Bernardo Study 1984-1994. Circulation, 96:37-43.

Flight RJ, McKenzie-Pollock M, Hamilton MA, Salmond CE, Stokes YM (1984) The health status of fourth form students in Northland. New Zealand Medical Journal, 97:1-6.
Fong PC, Tam SC, Tai YT, Lau CP, Lee J, Sha YY (1994) Serum lipid and apolipoprotein distributions in Hong Kong Chinese. Journal of Epidemiology and Community Health, 48:355-359.
Frantz ID Jr, Dawson EA, Ashman PL et al. (1989) Test of effect of lipid lowering by diet on cardiovascular risk. The Minnesota Coronary Survey. Arteriosclerosis, 9:129-135.
Frantz ID Jr, Dawson EA, Kuba K, Brewer ER, Gatewood LC, Bartch GE (1975) The Minnesota Coronary Survey: effct of diet on cardiovascular events and deaths. Circulation, 51-2:I-4.

Frick MH, Elo O, Haapa K, Heinonen OP et al. (1987) Helsinki Heart Study: primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia. Safety of treatment, changes in risk factors, and incidence of coronary heart disease. New England Journal of Medicine, 317:1237-1245.
Furberg CD, Adams HP Jr, Applegate WB et al. (1994a) Effect of lovastatin on early carotid atherosclerosis and cardiovascular events. Asymptomatic Carotid Artery Progression Study (ACAPS) Research Group. Circulation, 90:1679-1687.

Furberg CD, Byington RP, Crouse JR, Espeland MA (1994b) Pravastatin, lipids, and major coronary events. American Journal of Cardiology, 73:1133-1134.

Gatchev O, Rastam L, Lindberg G, Gullberg B, Eklund GA, Isacsson SO (1993) Subarachnoid hemorrhage, cerebral hemorrhage, and serum cholesterol concentration in men and women. Annals of Epidemiology, 3:403-409.
Ghannem H, Darioli R, Limam K et al. (2001) Epidemiology of cardiovascular risk factors among schoolchildren in Sousse, Tunisia. Journal of Cardiovascular Risk, 8:87-91.
Giroud M, Milan C, Beuriat P et al. (1991) Incidence and survival rates during a two-year period of intracerebral and subarachnoid haemorrhages, cortical infarcts, lacunes and transient ischaemic attacks. The Stroke Registry of Dijon: 1985-1989. International Journal of Epidemiology, 20:892-899.

Gliksman MD, Dwyer T, Wlodarczyk J (1990) Differences in modifiable cardiovascular disease risk factors in Australian schoolchildren: the results of a nationwide survey. Preventive Medicine, 19:291-304.
Gonzalez-Villalpando C, Stern MP, Haffner SM, Gonzalez Villapando ME, Gaskill S, Rivera Martinez D (1999) Prevalence of hypertension in a Mexican population according to the sixth report of the joint national committee on prevention, detection, evaluation and treatment of high blood pressure. Journal of Cardiovascular Risk, 6:177-181.

Gorelick PB, Schneck M, Berglund LF, Feinberg W, Goldstone J (1997) Status of lipids as a risk factor for stroke. Neuroepidemiology, 16:107-115.

Graff-Iversen S, Stensvold I, Lund-Larsen PG, Nodarse LO, Tverdal A, Urdal P (1998) Serum lipids in postmenopausal or perimenopausal women using estrogen alone, estrogen with levonorgestrel, or estrogen with norethisterone, compared with nonusers: results from a cross-sectional study in two Norwegian counties 1985-1988. Journal of Clinical Epidemiology, 51: 1311-1316.
Greenland P, Goldbourt U, Yaari S et al. (1993) Divergent age-associated patterns of high density lipoprotein cholesterol and its percentage in Jewish and Moslem Arab Israeli children and adolescents: the Petach Tikva Project. American Journal of Epidemiology, 137:549-558.

Gross L, Figueredo R (1973) Long-term cholesterol-lowering effect of colestipol resin in humans. Journal of the American Geriatrics Society, 21:552-556.

He J, Klag MJ, Whelton PK et al. (1991a) Migration, blood pressure pattern, and hypertension: the Yi Migrant Study. American Journal of Epidemiology, 134:1085-1101.
He J, Tell GS, Tang YC, Mo PS, He GQ (1991b) Effect of migration on blood pressure: the Yi People Study. Epidemiology, 2:88-97.
He J, Whelton PK (2000) Selection of initial antihypertensive drug therapy. The Lancet, 356:1942-1943.
Heart Protection Study Collaborative Group (2002a) MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20536 high-risk individuals: a randomised placebo-controlled trial. The Lancet, 360:23-33.
Heart Protection Study Collaborative Group (2002b) MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20536 high-risk individuals: a randomised placebo-controlled trial. The Lancet, 360:7-22.
Hebert PR, Gaziano JM, Chan KS, Hennekens CH (1997) Cholesterol lowering with statin drugs, risk of stroke, and total mortality. An overview of randomized trials. Journal of the American Medical Association, 278:313-321.
Hebert PR, Gaziano JM, Chan KS, Hennekens CH (1997) Cholesterol lowering with statin drugs, risk of stroke, and total mortality. An overview of randomized trials. Journal of the American Medical Association, 278: 313-321.
Hebert PR, Gaziano JM, Hennekens CH (1995) An overview of trials of cholesterol lowering and risk of stroke. Archives of Internal Medicine, 155:50-55.

Heinemann L, Barth W, Hoffmeister H (1995) Trend of cardiovascular risk factors in the East German population 1968-1992. Journal of Clinical Epidemiology, 48:787-795.
Herman B, Helmert U, Greiser E (1988) Factors related to blood lipid levelsthe Bremen baseline health survey of the German Cardiovascular Prevention Study. Public Health, 102:565-575.
Hjermann I, Velve Byre K, Holme I, Leren P (1981) Effect of diet and smoking intervention on the incidence of coronary heart disease. Report from the Oslo Study Group of a randomised trial in healthy men. The Lancet, 2:1303-1310.

Hoffmeister H, Mensink GB, Stolzenberg H (1994) National trends in risk factors for cardiovascular disease in Germany. Preventive Medicine, 23: 197-205.
Holme I (1990) An analysis of randomized trials evaluating the effect of cholesterol reduction on total mortality and coronary heart disease incidence. Circulation, 82:1916-1924.
Hughes K, Yeo PP, Lun KC et al. (1990) Cardiovascular diseases in Chinese, Malays, and Indians in Singapore. II. Differences in risk factor levels. Journal of Epidemiology and Community Health, 44:29-35.
Hung TP (1993) Changes in mortality from cerebrovascular disease and clinical pattern of stroke in Taiwan. Journal of the Formosan Medical Association, 92:687-696.

Hutchinson RG, Watson RL, Davis CE et al. (1997) Racial differences in risk factors for atherosclerosis. The ARIC study. Atherosclerosis risk in communities. Angiology, 48:279-290.
Iribarren C, Jacobs DR, Sadler M, Claxton AJ, Sidney S (1996) Low total serum cholesterol and intracerebral hemorrhagic stroke: is the association confined to elderly men? The Kaiser Permanente Medical Care Program. Stroke, 27:1993-1998.
Iso H, Jacobs DR, Jr, Wentworth D, Neaton JD, Cohen JD (1989) Serum cholesterol levels and six-year mortality from stroke in 350977 men screened for the multiple risk factor intervention trial. New England Journal of Medicine, 320:904-910.

Jadue L, Vega J, Escobar MC et al. (1999) Risk factors for non communicable diseases: methods and global results of the CARMEN program basal survey. Revista Medica de Chile, 127:1004-1013.
Jeng JS, Lee TK, Chang YC et al. (1998) Subtypes and case-fatality rates of stroke: a hospital-based stroke registry in Taiwan (SCAN-IV). Journal of the Neurological Sciences, 156:220-226.
Johnson CL, Rifkind BM, Sempos CT et al. (1993) Declining serum total cholesterol levels among US adults. The National Health and Nutrition Examination Surveys. Journal of the American Medical Association, 269: 3002-3008.

Jorgensen HS, Nakayama H, Raaschou HO, Olsen TS (1995) Intracerebral hemorrhage versus infarction: stroke severity, risk factors, and prognosis. Annals of Neurology, 38:45-50.
Joseph JG, Prior IA, Salmond CE, Stanley D (1983) Elevation of systolic and diastolic blood pressure associated with migration: the Tokelau Island Migrant Study. Journal of Chronic Diseases, 36:507-516.
Jukema JW, Bruschke AV, van Boven AJ et al. (1995) Effects of lipid lowering by pravastatin on progression and regression of coronary artery disease in symptomatic men with normal to moderately elevated serum cholesterol levels. The Regression Growth Evaluation Statin Study (REGRESS). Circulation, 91:2528-2540.

Kafatos A, Schlienger JL, Deslypere JP, Ferro-Luzzi A, Cruz JA (1991) Nutritional status: serum lipids. Euronut SENECA investigators. European Journal of Clinical Nutrition, 45:S53-61.
Kesteloot H, Geboers J, Pietinen P (1987) On the within-population relationship between dietary habits and serum lipid levels in Belgium. European Heart Journal, 8:821-831.
Kitange HM, Swai AB, Masuki G, Kilima PM, Alberti KG, McLarty DG (1993) Coronary heart disease risk factors in sub-Saharan Africa: studies in Tanzanian adolescents. Journal of Epidemiology and Community Health, 47:303-307.

Lakka TA, Salonen JT (1992) Physical activity and serum lipids: a cross-sectional population study in eastern Finnish men. American Journal of Epidemiology, 136:806-818.

LaRosa JC, He J, Vupputuri S (1999) Effect of statins on risk of coronary disease: a meta-analysis of randomized controlled trials. Journal of the American Medical Association, 282:2340-2346.
Law MR (1999) Lowering heart disease risk with cholesterol reduction: evidence from observational studies and clinical trials. European Heart Journal Supplement, 1:S3-S8.
Law MR, Thompson SG, Wald NJ (1994a) Assessing possible hazards of reducing serum cholesterol. British Medical Journal, 308:373-379.

Law MR, Wald NJ (1994) An ecological study of serum cholesterol and ischaemic heart disease between 1950 and 1990. European Journal of Clinical Nutrition, 48:305-325.

Law MR, Wald NJ, Rudnicka AR (2003) Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. British Medical Journal, 326:1423.

Law MR, Wald NJ, Thompson SG (1994b) By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? British Medical Journal, 308:367-372.
Law MR, Wald NJ, Wu T, Hackshaw A, Bailey A (1994c) Systematic underestimation of association between serum cholesterol concentration and ischaemic heart disease in observational studies: data from the BUPA study. British Medical Journal, 308:363-366.

Leppala JM, Virtamo J, Fogelholm R, Albanes D, Heinonen OP (1999) Different risk factors for different stroke subtypes: association of blood pressure, cholesterol, and antioxidants. Stroke, 30:2535-2540.

Leren P (1970) The Oslo diet-heart study. Eleven-year report. Circulation, 42:935-942.

Lewington S, MacMahon S (1999) Blood pressure, cholesterol, and common causes of death: a review. Prospective studies collaboration. American Journal of Hypertension, 12:S96-98.
Li SC, Schoenberg BS, Wang CC, Cheng XM, Bolis CL, Wang KJ (1985) Cerebrovascular disease in the People's Republic of China: epidemiologic and clinical features. Neurology, 35:1708-1713.

Lindeberg S, Nilsson-Ehle P, Terent A, Vessby B, Schersten B (1994) Cardiovascular risk factors in a Melanesian population apparently free from stroke and ischaemic heart disease: the Kitava study. Journal of Internal Medicine, 236:331-340.

Lupien PJ, Moorjani S, Rochon J et al. (1985) Lipid profile of a FrenchCanadian population: 1. Association of plasma lipid and lipoprotein levels with age, relative body weight and education. Canadian Medical Association Journal, 133:1127-1133.
MacMahon S (1994) Blood pressure and risks of cardiovascular disease. In: Textbook of hypertension. Swales J, ed. Blackwell Scientific Publications, Oxford.

MacMahon S, Peto R, Cutler J et al. (1990) Blood pressure, stroke, and coronary heart disease. Part I, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. The Lancet, 335:765-774.
Mahley RW, Palaoglu KE, Atak Z et al. (1995) Turkish Heart Study: lipids, lipoproteins, and apolipoproteins. Journal of Lipid Research, 36: 839-859.

Mann GV, Shaffer RD, Anderson RS, Sandstead HH (1964) Cardiovascular disease in the Masai. Journal of Atherosclerosis, 4:289-312.
Mann JI, Duncan A, Ball MJ et al. (1991) Blood lipid levels in New Zealand. New Zealand Medical Journal, 104:371-374.

Mann JI, Lewis B, Shepherd J et al. (1988) Blood lipid concentrations and other cardiovascular risk factors: distribution, prevalence, and detection in Britain. British Medical Journal, 296:1702-1706.

Masia R, Pena A, Marrugat J et al. (1998) High prevalence of cardiovascular risk factors in Gerona, Spain, a province with low myocardial infarction incidence. REGICOR investigators. Journal of Epidemiology and Community Health, 52:707-715.
McCaughan D (1981) The long-term effects of probucol on serum lipid levels. Archives of Internal Medicine, 141:1428-1432.
McGovern PG, Pankow JS, Shahar E et al. (1996) Recent trends in acute coronary heart disease-mortality, morbidity, medical care, and risk factors. The Minnesota Heart Survey Investigators. New England Journal of Medicine, 334:884-890.

Miettinen TA, Huttunen JK, Naukkarinen V et al. (1985) Multifactorial primary prevention of cardiovascular diseases in middle-aged men. Risk factor changes, incidence, and mortality. Journal of the American Medical Association, 254:2097-2102.
Ministry of Health (1999) NZ food: NZ people. Key results of the 1997 National Nutrition Survey. Ministry of Health, Wellington.
Misra A, Sharma R, Pandey RM, Khanna N (2001) Adverse profile of dietary nutrients, anthropometry and lipids in urban slum dwellers of northern India. European Journal of Clinical Nutrition, 55:727-734.

Mitwalli AH, al-Maatouq A, al-Wakeel J, Alam AA (1994) Hypercholesterolemia in healthy adult males-a public survey in central Saudi Arabia. Annals of Saudi Medicine, 14:499-502.

Molla A, Manser WW, Lalani R, Badruddin SH, Mohammad Z, Khurshid M (1990) Blood lipids in a healthy Karachi population. Journal of Tropical Medicine and Hygiene, 93:295-299.
Muldoon MF, Manuck SB, Mendelsohn AB, Kaplan JR, Belle SH (2001) Cholesterol reduction and non-illness mortality: meta-analysis of randomised clinical trials. British Medical Journal, 322:11-15.
Myllykangas M, Pekkanen J, Rasi V, Haukkala A, Vahtera E, Salomaa V (1995) Haemostatic and other cardiovascular risk factors, and socioeconomic status among middle-aged Finnish men and women. International Journal of Epidemiology, 24:1110-1116.

Nakayama T, Date C, Yokoyama T, Yoshiike N, Yamaguchi M, Tanaka H (1997) A 15.5-year follow-up study of stroke in a Japanese provincial city. The Shibata Study. Stroke, 28:45-52.
Neaton JD, Blackburn H, Jacobs D et al. (1992) Serum cholesterol level and mortality findings for men screened in the multiple risk factor intervention trial. Multiple Risk Factor Intervention Trial Research Group. Archives of Internal Medicine, 152:1490-1500.

Neaton JD, Wentworth D (1992) Serum cholesterol, blood pressure, cigarette smoking, and death from coronary heart disease. Overall findings and differences by age for 316099 white men. Multiple Risk Factor Intervention Trial Research Group. Archives of Internal Medicine, 152: 56-64.

Neaton JD, Wentworth DN, Cutler J, Stamler J, Kuller L (1993) Risk factors for death from different types of stroke. Multiple risk factor intervention trial research group. Annals of Epidemiology, 3:493-499.
Nikkila M, Heikkinen J (1990) Serum cholesterol, high-density lipoprotein cholesterol and five-year survival in elderly people. Age and Ageing, 19:403-408.

Nissinen A, Piha T, Tuomilehto J, Romo M, Puska P (1987) Comparison of the levels of cardiovascular risk factors between eastern and south-western Finland in 1982. Acta Medica Scandinavica, 222(5):389-400.

Nyarko NK, Adubofour KO, Ofei F, Pobee JO, Owusu SK (1994) Serum lipids and lipoprotein in adult Ghanaians. Journal of Internal Medicine, 236: 251-253.

Oelofse A, Jooste PL, Steyn K et al. (1996) The lipid and lipoprotein profile of the urban black South African population of the Cape Peninsula-the BRISK study. South African Medical Journal, 86:162-166.
Okayama A, Ueshima H, Marmot MG, Nakamura M, Kita Y, Yamakawa M (1993) Changes in total serum cholesterol and other risk factors for cardiovascular disease in Japan 1980-1989. International Journal of Epidemiology, 22:1038-1047.

Okesina AB, Oparinde DP, Akindoyin KA, Erasmus RT (1999) Prevalence of some risk factors of coronary heart disease in a rural Nigerian population. East African Medical Journal, 76:212-216.

Olferev AM, Volozh OI, Polesskii VA et al. (1990) Characteristics of the plasma lipid profile in the male population of Tallinn aged 20-54 years. Cor et Vasa, 32:441-447.

Olferev AM, Volozh OI, Sokolova MA et al. (1991) Characteristics of the blood serum lipoprotein pattern in female Tallinn residents aged 35-54 years. Cor et Vasa, 33:472-479.
Olusi SO, Prabha K, Sulaiman A (1997) Prevalence of hypercholesterolemia in a Kuwati hospital outpatient population. Annals of Saudi Medicine, 17:108-111.

Onat A, Surdum-Avci G, Senocak M, Ornek E, Gozukara Y (1992) Plasma lipids and their interrelationship in Turkish adults. Journal of Epidemiology and Community Health, 46:470-476.
Page LB, Damon A, Moellering RC, Jr. (1974) Antecedents of cardiovascular disease in six Solomon Islands societies. Circulation, 49:1132-1146.
Pignone M, Phillips C, Mulrow C (2000) Use of lipid lowering drugs for primary prevention of coronary heart disease: meta-analysis of randomised trials. British Medical Journal, 321:983-986.
Pitt B, Mancini GB, Ellis SG, Rosman HS, Park JS, McGovern ME (1995) Pravastatin limitation of atherosclerosis in the coronary arteries (PLAC I): reduction in atherosclerosis progression and clinical events. PLAC I investigation. Journal of the American College of Cardiology, 26:1133-1139.

Posadas-Romero C, Tapia-Conyer R, Lerman-Garber I et al. (1995) Cholesterol levels and prevalence of hypercholesterolemia in a Mexican adult population. Atherosclerosis, 118:275-284.

Poulter N, Sever P (1994) Blood pressure in other populations. A. Low blood pressure populations and the impact of rural-urban migration. In: Textbook of hypertension. Swales J, ed. Blackwell Scientific Publications, Oxford.
Poulter NR, Khaw KT, Hopwood BE et al. (1990) The Kenyan Luo migration study: observations on the initiation of a rise in blood pressure. British Medical Journal, 300:967-972.
Poulter NR, Khaw KT, Sever PS (1988) Higher blood pressures of urban migrants from an African low-blood pressure population are not due to selective migration. American Journal of Hypertension, 1:S143-145.

PSC (1995) Cholesterol, diastolic blood pressure, and stroke: 13000 strokes in 45000 people in 45 prospective cohorts. The Lancet, 346:1647-1653.

Puska P, Matilainen T, Jousilahti P et al. (1993) Cardiovascular risk factors in the Republic of Karelia, Russia, and in North Karelia, Finland. International Journal of Epidemiology, 22:1048-1055.
Razay G, Heaton KW, Bolton CH (1992) Coronary heart disease risk factors in relation to the menopause. Quarterly Journal of Medicine, 85: 889-896.

Reddy KK, Ramachandraiah T, Reddanna P, Thyagaraju K (1994) Serum lipid peroxides and lipids in urban and rural Indian men. Archives of Environmental Health, 49:123-127.
Research Committee to the Medical Council (1965) Low-fat diet in myocardial infarction: a controlled trial. The Lancet, 2:501-504.
Robinson D, Bevan EA, Hinohara S, Takahashi T (1992) Seasonal variation in serum cholesterol levels-evidence from the UK and Japan. Atherosclerosis, 95:15-24.

Rose GA, Thomson WB, Williams RT (1965) Corn oil in treatment of ischaemic heart disease. British Medical Journal, 1:1531-1533.
Rosengren A, Eriksson H, Larsson B et al. (2000) Secular changes in cardiovascular risk factors over 30 years in Swedish men aged 50: the study of men born in 1913, 1923, 1933 and 1943. Journal of Internal Medicine, 247: 111-118.

Ross SD, Allen IE, Connelly JE et al. (1999) Clinical outcomes in statin treatment trials: a meta-analysis. Archives of Internal Medicine, 159: 1793-1802.

Sacks FM, Pfeffer MA, Moye LA et al. (1996) The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events Trial investigators. New England Journal of Medicine, 335:1001-1009.

Sahni R, Maniet AR, Voci G, Banka VS (1991) Prevention of restenosis by lovastatin after successful coronary angioplasty. American Heart Journal, 121:1600-1608.

Sakata K, Labarthe DR (1996) Changes in cardiovascular disease risk factors in three Japanese national surveys 1971-1990. Journal of Epidemiology, 6:93-107.

Salmond CE, Joseph JG, Prior IA, Stanley DG, Wessen AF (1985) Longitudinal analysis of the relationship between blood pressure and migration: the Tokelau Island Migrant Study. American Journal of Epidemiology, 122:291-301.

Salmond CE, Prior IA, Wessen AF (1989) Blood pressure patterns and migration: a 14-year cohort study of adult Tokelauans. American Journal of Epidemiology, 130:37-52.

Salonen R, Nyyssonen K, Porkkala E et al. (1995) Kuopio Atherosclerosis Prevention Study (KAPS). A population-based primary preventive trial of the effect of LDL lowering on atherosclerotic progression in carotid and femoral arteries. Circulation, 92:1758-1764.

Salvaggio A, Periti M, Miano L, Quaglia G, Marzorati D (1991) Coffee and cholesterol, an Italian study. American Journal of Epidemiology, 134: 149-156.

Scrimgeour EM, McCall MG, Smith DE, Masarei JR (1989) Levels of serum cholesterol, triglyceride, HDL-cholesterol, apoproteins A-I and B, and plasma glucose, and prevalence of diastolic hypertension and cigarette smoking in Papua New Guinea highlanders. Pathology, 21:46-50.

Sever PS, Gordon D, Peart WS, Beighton P (1980) Blood-pressure and its correlates in urban and tribal Africa. The Lancet, 2:60-64.

Shear CL, Franklin FA, Stinnett $S$ et al. (1992) Expanded Clinical Evaluation of Lovastatin (EXCEL) study results. Effect of patient characteristics on lovas-tatin-induced changes in plasma concentrations of lipids and lipoproteins. Circulation, 85:1293-1303.
Shepherd J, Cobbe SM, Ford I et al. (1995) Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. New England Journal of Medicine, 333:1301-1307.
Sigfusson N, Sigvaldason H, Steingrimsdottir L et al. (1991) Decline in ischaemic heart disease in Iceland and change in risk factor levels. British Medical Journal, 302:1371-1375.
Simons LA, Friedlander Y, McCallum J et al. (1991) The Dubbo study of the health of elderly: correlates of coronary heart disease at study entry. Journal of the American Geriatrics Society, 39:584-590.
Smith WC, Tunstall-Pedoe H, Crombie IK, Tavendale R (1989) Concomitants of excess coronary deaths-major risk factor and lifestyle findings from 10 359 men and women in the Scottish Heart Health Study. Scottish Medical Journal, 34:550-555.
Sprafka JM, Burke GL, Folsom AR, Luepker RV, Blackburn H (1990) Continued decline in cardiovascular disease risk factors: results of the Minnesota Heart Survey, 1980-1982 and 1985-1987. American Journal of Epidemiology, 132:489-500.
Srinivasan SR, Wattigney W, Webber LS, Berenson GS (1991) Race and gender differences in serum lipoproteins of children, adolescents, and young adultsemergence of an adverse lipoprotein pattern in white males: the Bogalusa Heart Study. Preventive Medicine, 20:671-84.
Steyn K, Benade AJ, Langenhoven ML, Joubert G, Rossouw JE (1987) Hypercholesterolaemia in the coloured population of the Cape Peninsula (CRISIC study). South African Medical Journal, 71:483-486.
Steyn K, Jooste PL, Langenhoven ML et al. (1985) Coronary risk factors in the coloured population of the Cape Peninsula. South African Medical Journal, 67:619-625.

Strachan D, Rose G (1991) Strategies of prevention revisited: effects of imprecise measurement of risk factors on the evaluation of "high-risk" and "population-based" approaches to prevention of cardiovascular disease. Journal of Clinical Epidemiology, 44:1187-1196.
Sudlow CL, Warlow CP (1997) Comparable studies of the incidence of stroke and its pathological types: results from an international collaboration. International stroke incidence collaboration. Stroke, 28:491-499.
Suh I (2001) Cardiovascular mortality in Korea: a country experiencing epidemiologic transition. Acta Cardiologica, 56:75-81.

Suh I, Jee SH, Kim HC, Nam CM, Kim IS, Appel LJ (2001) Low serum cholesterol and haemorrhagic stroke in men: Korea medical insurance corporation study. The Lancet, 357:922-925.

Suzuki K, Kutsuzawa T, Takita K et al. (1987) Clinico-epidemiologic study of stroke in Akita, Japan. Stroke, 18:402-406.

Swai AB, McLarty DG, Kitange HM et al. (1993) Low prevalence of risk factors for coronary heart disease in rural Tanzania. International Journal of Epidemiology, 22:651-659.
Szatrowski TP, Peterson AV, Jr., Shimizu Y et al. (1984) Serum cholesterol, other risk factors, and cardiovascular disease in a Japanese cohort. Journal of Chronic Diseases, 37:569-584.

Tanaka H, Ueda Y, Hayashi M et al. (1982) Risk factors for cerebral hemorrhage and cerebral infarction in a Japanese rural community. Stroke, 13:62-73.

Tao S, Li Y, Xiao Z et al. (1992) Serum lipids and their correlates in Chinese urban and rural populations of Beijing and Guangzhou. PRC-USA Cardiovascular and Cardiopulmonary Epidemiology Research Group. International Journal of Epidemiology, 21:893-903.
Thrift AG, Dewey HM, Macdonell RA, McNeil JJ, Donnan GA (2001) Incidence of the major stroke subtypes: initial findings from the North East Melbourne Stroke Incidence Study (NEMESIS). Stroke, 32:1732-1738.
Thune I, Njolstad I, Lochen ML, Forde OH (1998) Physical activity improves the metabolic risk profiles in men and women: the Tromso Study. Archives of Internal Medicine, 158:1633-1640.

Tian HG, Nan Y, Liang XQ et al. (1995) Relationship between serum lipids and dietary and non-dietary factors in a Chinese population. European Journal of Clinical Nutrition, 49:871-882.

Tuomilehto J, Bonita R, Stewart A, Nissinen A, Salonen JT (1991) Hypertension, cigarette smoking, and the decline in stroke incidence in eastern Finland. Stroke, 22:7-11.
Vaccarino V, Borgatta A, Gallus G, Sirtori CR (1995) Prevalence of coronary heart disease risk factors in northern-Italian male and female employees. European Heart Journal, 16:761-769.
van Beurden EK, James R, Henrikson D, Tyler C, Christian J (1991) The North Coast Cholesterol Check Campaign. Results of the first three years of a large-scale public screening programme. Medical Journal of Australia, 154:385-391.

Vartiainen E, Jousilahti P, Alfthan G, Sundvall J, Pietinen P, Puska P (2000) Cardiovascular risk factor changes in Finland, 1972-1997. International Journal of Epidemiology, 29:49-56.
Vartiainen E, Puska P, Pekkanen J, Tuomilehto J, Jousilahti P (1994) Changes in risk factors explain changes in mortality from ischaemic heart disease in Finland. British Medical Journal, 309:23-27.

Verschuren WM, Boerma GJ, Kromhout D (1994) Total and HDL-cholesterol in the Netherlands: 1987-1992. Levels and changes over time in relation to age, gender and educational level. International Journal of Epidemiology, 23:948-956.

Viikari J, Akerblom HK, Nikkari T et al. (1985) Atherosclerosis precursors in Finnish children and adolescents. IV. Serum lipids in newborns, children and adolescents. Acta Paediatrica Scandinavica, Supplement, 318:103-109.
Wallace RB, Colsher PL (1992) Blood lipid distributions in older persons. Prevalence and correlates of hyperlipidemia. Annals of Epidemiology, 2:15-21.

Warlow C, Dennis M, VanGijn J et al. (1996) Stroke. A practical guide to management. Blackwell Science Ltd, Oxford.

Waters D, Higginson L, Gladstone P et al. (1994) Effects of monotherapy with an HMG-CoA reductase inhibitor on the progression of coronary atherosclerosis as assessed by serial quantitative arteriography. The Canadian Coronary Atherosclerosis Intervention Trial. Circulation, 89:959968.

Waters D, Higginson L, Gladstone P, Kimball B, LeMay M, Lesperance J (1993) Design features of a controlled clinical trial to assess the effect of an HMG CoA reductase inhibitor on the progression of coronary artery disease. Canadian Coronary Atherosclerosis Intervention Trial Investigators Montreal, Ottawa, and Toronto, Canada. Controlled Clinical Trials, 14:45-74.

Watts GF, Lewis B, Brunt JN et al. (1992) Effects on coronary artery disease of lipid-lowering diet, or diet plus cholestyramine, in the St Thomas' Atherosclerosis Regression Study (STARS). The Lancet, 339:563-569.
Weintraub WS, Boccuzzi SJ, Klein JL et al. (1994) Lack of effect of lovastatin on restenosis after coronary angioplasty. Lovastatin Restenosis Trial Study Group. New England Journal of Medicine, 331:1331-1337.

White HD, Simes RJ, Anderson NE et al. (2000) Pravastatin therapy and the risk of stroke. New England Journal of Medicine, 343:317-326.

Woo J, Leung SS, Ho SC, Sham A, Lam TH, Janus ED (1997) Dietary practices and lipid intake in relation to plasma lipid profile in Hong Kong Chinese. European Journal of Clinical Nutrition, 51:467-471.
Woodhill JM, Palmer AJ, Leelarthaepin B, McGilchrist C, Blacket RB (1978) Low fat, low cholesterol diet in secondary prevention of coronary heart disease. Advances in Experimental Medicine and Biology, 109:317330.

Wu Y, Yao C, Wu Z et al. (1992) Interim report of Sino-MONICA-Beijing for the years 1985-1989. Chinese Medical Sciences Journal, 7:125-129.
Yang XS, Kesteloot H, Huang DX (1986) Serum cholesterol in China and the West. Chinese Medical Journal, 99:183-186.

Yano K, Popper JS, Kagan A, Chyou PH, Grove JS (1994) Epidemiology of stroke among Japanese men in Hawaii during 24 years of follow-up: the Honolulu Heart Program. Health Reports, 6:28-38.

Yano K, Reed DM, Curb JD, Hankin JH, Albers JJ (1986) Biological and dietary correlates of plasma lipids and lipoproteins among elderly Japanese men in Hawaii. Arteriosclerosis, 6:422-433.

Yano K, Reed DM, MacLean CJ (1989) Serum cholesterol and hemorrhagic stroke in the Honolulu Heart Program. Stroke, 20:1460-1465.
Zhuang HZ, Hana QQ, Chen HZ (1986) Study of serum lipids and lipoproteins of healthy subjects in Shanghai. Chinese Medical Journal, 99:657-659.
Appendix A: Sampling methods, response rate and cholesterol measuring techniques of studies included in CHOLESTEROL DATA REVIEW
Africa

| Subregion | Country or area | Study reference | Sampling methods | Response rate | Cholesterol measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | Ghana | Nyarko et al. (1994) | Employees from the University of Ghana were randomly selected | * | Trained staff \& certified lab <br> All fasting samples <br> Storage of sample unnecessary, enzymatic analysis |
|  | Nigeria | Erasmus et al. (1994) | Recruitment from a rural village | 82\% | Trained staff \& certified lab Non-fasting samples Storage of sample unnecessary, extraction analysis |
|  | Nigeria | Okesina et al. (1999) | Houses in villages selected randomly, then individuals randomly selected | * | All fasting samples ?where sample stored, enzymatic-colorimetric analysis |
|  | Seychelles | Bovet et al. (1991) | Age and sex stratified random sample from national census | 84-89\% | Certified lab <br> All fasting samples <br> Sample stored in freezer, enzymatic analysis |


| AFR-E | South Africa | Oelofse et al. (1996) | A stratified proportional sample from target population using census data | * | Trained staff <br> Non-fasting samples <br> Sample stored in fridge, enzymatic analysis |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | South Africa | Steyn et al. (1985, 1987) | A stratified sample by age and sex from census data | * | Trained staff \& certified lab Non-fasting samples Sample stored in freezer, enzymatic analysis |
|  | United Republic of Tanzania | Kitange et al. (1993) | Whole population in four villages, and a random sample in another four villages | 60-94\% | Certified lab <br> All fasting samples <br> Sample stored in freezer, enzymatic analysis |
|  | United Republic of Tanzania | Swai et al. (1993) | Community-based survey of eight villages in three regions in rural United Republic of Tanzania | 90.9-96\% | Trained staff \& certified lab All fasting samples Sample stored in freezer, enzymatic analysis |
|  | Zimbabwe | Allain and Matenga <br> (T. Allain et al., personal communication, 2001) | Stratified sampling | * | Trained staff \& certified lab Some fasting samples Sample analysed immediately ? method |
| * Data unavailable. |  |  |  |  |  |
| ? Specific details not given. |  |  |  |  |  |

Americas

| Subregion | Country or area | Study reference | Sampling methods | Response rate | Cholesterol measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AMR-A | Canada | Connelly et al. (1992) | Probability sample from health insurance register | 60-70\% | Trained staff \& certified lab All fasting samples Sample stored in fridge, enzymatic analysis |
|  | Canada | Lupien et al. (1985) | Random sample from provincial electoral register | 60\% | Trained staff \& certified lab All fasting samples ?where sample stored, ?method of analysis (Technicon Autoanalyzer II) |
|  | Halifax | MONICA study <br> (Anonymous 1989) | Sample from public health service register | 4I-63\% | Trained staff \& certified lab All fasting samples Sample stored in fridge, enzymatic analysis |
|  | USA | Abbott et al. (1997) | Sample from long-term study | * | Trained staff \& certified lab All fasting samples Sample stored in fridge, ?method of analysis (LRC program) |
|  | USA | Brown et al. (1993) | Sample from four communities | 42-68\% | Trained staff \& certified lab All fasting samples Sample stored in freezer, enzymatic analysis |
|  | USA | Burke et al. (1991) | Clusters of 40 households in the metropolitan area were randomly selected | * | Trained staff \& certified lab Non-fasting samples ?where sample stored, ?method of analysis (Technicon Autoanalyzer II: LRC program) |
|  | USA | Donker et al. (1997) | Sample by age and race from survey of school-age children | * | Trained staff \& certified lab ?fasting sample ?where sample stored, ?method of analysis (Technicon Autoanalyzer: LRC program) |

USA
USA
USA
USA
USA
USA
USA
USA
Americas (continued)

| Subregion | Country or area | Study reference | Sampling methods | Response rate | Cholesterol measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | USA | Yano et al. (I986) | Random sample from cohort studies | 86\% | Trained staff \& certified lab <br> All fasting samples <br> Sample stored in freezer, ?method of analysis <br> (LRC program) |
|  | Stanford | MONICA study (Anonymous 1989) | Sample from commercial household directory | 66\% | Trained staff \& certified lab Non-fasting samples Sample stored in fridge, extraction analysis |
| AMR-B | Brazil | INCLEN (Anonymous 1992b) | Random sample of males from 12 centres in 7 countries | 82-92\% | Trained staff \& certified lab ?fasting sample ?where sample stored, ?method of analysis |
|  | Chile | INCLEN (Anonymous 1992b) | Random sample of males from 12 centres in 7 countries | 67-76\% | Trained staff \& certified lab ?fasting sample <br> ?where sample stored, ?method of analysis |
|  | Chile | Jadue et al. (1999), <br> L. Jadue, personal communication, 2001 | Stratified sample | * | Trained staff <br> All fasting samples <br> Immediate analysis enzymatic analysis |
|  | Colombia | INCLEN (Anonymous 1992b) | Random sample of males from 12 centres in 7 countries | 83\% | Trained staff \& certified lab ?fasting sample <br> ?where sample stored, ?method of analysis |

Trained staff \& certified lab
?fasting sample
?asting sample stored in freezer, ?method of analysis
Trained staff
Trained staff
All fasting samples
?where sample stor
? where sample stored, ?method of analysis
Trained staff \& certified lab
All fasting samples
Sample stored in freezer, extraction analysis
Trained staff \& certified lab
Non-fasting samples
Sample stored in freezer, enzymatic analysis
Trained staff \& certified lab
All fasting samples
Sample stored in fridge, enzymatic analysis

| AMR-D |
| :--- | :--- |
| * Data unavailable. |
| ? Specific details not given. |

Aono et al. (1997)
Clusters of 2000 individuals from selected
ysnueds mous su!!durs pay!?..15s Town


Random sample of population from a selected area-a complete
household survey was performed
Random sample from population register

## 80.2\%

58\%

## \% 8

 Random stratified sample inMexico City
 communication, 2001
‘(6661) (
C. Gonzalez, personal communication, 2001 Posadas-Romero et al. (1995)
L. Yamamoto et al.,
L. Yamamoto et al., personal communication,
2001
72-77\%
72-77\%
64\%

Mexico

## Mexico

Mexico
Data unavailable.
Eastern Mediterranean

| Subregion | Country or area | Study reference | Sampling | Response rate | Cholesterol measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
| EMR-B | Bahrain | al-Mahroos et al. (2000) | Stratified sample by age from national population register | 62\% | Trained staff \& ?certified lab Some fasting samples ?where sample stored, enzymatic analysis |
|  | Jordan | H. Jaddou, personal communication, 2001 | Stratified sample from the population | 62\% | Trained staff \& certified lab <br> All fasting samples <br> Sample stored in freezer, enzymatic analysis |
|  | Kuwait | Olusi et al. (1997) | Sample taken from outpatient requests | * | Trained staff \& certified lab <br> ?fasting sample <br> ?where sample stored, enzymatic analysis |
|  | Saudi Arabia | al-Nuaim et al. (1996, 1997) | Stratified cluster random sample from target population using census data | 92\% | Trained staff \& certified lab <br> ?fasting sample <br> Sample stored in freezer, enzymatic calorimetric analysis |
|  | Saudi Arabia | al-Nuaim (1997) | Stratified random sample from target population using census data | 92\% | Trained staff \& certified lab <br> ?fasting sample <br> Sample stored in freezer, enzymatic calorimetric analysis |


|  | Saudi Arabia | al-Shammari et al. (1994) | Random sample from family practice <br> health clinics | $*$ |
| :--- | :--- | :--- | :--- | :--- |

Europe

| Subregion | Country or area | Study reference | Sampling | Response rate | Cholesterol measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
| EUR-A | Belgium | Kesteloot et al. (1987) | Sample from a survey performed in the Belgian army | * | Trained staff \& certified lab ?fasting sample ?where sample stored, enzymatic analysis |
|  | Charleroi | MONICA study <br> (Anonymous 1989) | Sample from population register | 59\% | Trained staff \& certified lab Non-fasting samples Sample stored in freezer, enzymatic analysis |
|  | Ghent | MONICA study <br> (Anonymous 1989) | Sample from public health service register | 54-57\% | Trained staff \& certified lab <br> Non-fasting samples <br> Sample stored in freezer, enzymatic analysis |
|  | Denmark, Glostrup | MONICA study <br> (Anonymous 1989) | Sample from population register | 79-80\% | Trained staff \& certified lab All fasting samples Sample stored in fridge, enzymatic analysis |
|  | Finland | Lakka and Salonen (1992) | Random sample from selected region | 83\% | Trained staff \& certified lab All fasting samples Storage not necessary, enzymatic analysis |
|  | Finland | Myllkangas et al. (1995) | Stratified random sample from population register | 79\% | Trained staff \& certified lab All fasting samples Sample stored in freezer, enzymatic analysis |
|  | Finland | Nikkila and Heikkinen (1990) | Sample from health survey | * | Trained staff \& certified lab ?fasting sample ?where sample stored, enzymatic analysis |
|  | Finland | Nissinen et al. (1987) | Random sample from population of two areas using the national population register | 76-80\% | Trained staff \& certified lab ?fasting sample Storage not necessary, enzymatic analysis |

\(\left.\left.$$
\begin{array}{llll}\text { Finland } & \text { Puska et al. (1993) } & \begin{array}{l}\text { Stratified random sample from } \\
\text { population register }\end{array} & 68-81 \%\end{array}
$$ $$
\begin{array}{l}\text { Trained staff \& certified lab } \\
\text { All fasting samples } \\
\text { Storage not neessary, enzymatic analysis }\end{array}
$$\right] \begin{array}{l}Trained staff \& certified lab <br>
?fasting sample <br>

Storage not necessary, enzymatic analysis\end{array}\right]\)| Trained staff \& certified lab |
| :--- |
| All fasting samples |
| Sample stored in freezer, enzymatic analysis |

Europe (continued)

| Subregion | Country or area | Study reference | Sampling | Response rate | Cholesterol measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Germany | Herman et al. (1988) | Random sample from health survey | 71\% | Trained staff \& certified lab ?fasting sample <br> ?where sample stored, ?method of analysis (Boehringer Mannheim CHOD-PAP) |
|  | Germany | Hoffmeister et al. (1994) | Randomly selected districts on resident registries | 69\% | Trained staff \& certified lab Non-fasting samples ?where sample stored, enzymatic analysis |
|  | Germany | MONICA study (Anonymous 1989) | Sample from national X-ray screening register | 72\% | Trained staff \& certified lab <br> Non-fasting samples <br> Sample stored in freezer, direct analysis |
|  | Augsburg (Rural) | MONICA study (Anonymous 1989) | Sample from population register | 82-84\% | Trained staff \& certified lab <br> Non-fasting samples <br> Sample stored in fridge, enzymatic analysis |
|  | Augsburg (Urban) | MONICA study (Anonymous 1989) | Sample from population register | 76-80\% | Trained staff \& certified lab <br> Non-fasting samples <br> Sample stored in fridge, enzymatic analysis |
|  | Bremen | MONICA study (Anonymous 1989) | Sample from resident register | 71\% | Trained staff \& certified lab Non-fasting samples Sample stored in freezer, enzymatic analysis |
|  | Rhein-Neckar | MONICA study (Anonymous 1989) | Sample from population register | 74\% | ?Trained staff \& ?certified lab ?fasting sample <br> ?where sample stored, enzymatic analysis |

Former German Democratic
Republic, Sample from national $X$-ray
$70-80 \%$
$77 \%$
$76 \%$
$72-81 \%$ ?fasting sample
?where sample stored, ?method of analysis
Trained staff \& certified lab
All fasting sample
Trained staff \& certified lab
Non-fasting samples
Sample stored in freezer, direct analysis
Trained staff \& certified lab
Non-fasting samples
Sample stored in freezer, direct analysis
Trained staff \& certified lab
All fasting samples
Storage not necessary, extraction analysis
Trained staff \& certified lab
?fasting sample
?where sample stored, ?method of analysis
(LRC program)
Trained staff \& certified lab
All fasting samples
Storage not necessary, ?method of analysis
(Gilford automated instrument)
?Trained staff \& certified lab
?fasting sample
?where sample stored, ?method of analysis
Trained staff \& certified lab
All fasting sample
?where sample stored, ?method of analysis
?where sample stored, ?method of analysis
Europe (continued)

| Subregion | Country or area | Study reference | Sampling | Response rate | Cholesterol measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Italy | Salvaggio et al. (Law and Wald 1994; Salvaggio et al. 1991) | Sample from a study on preventative medicine | * | Trained staff \& certified lab ?fasting sample <br> ?where sample stored, ?method of analysis |
|  | Italy | Vaccarino et al. (1995) | All employees of IBM asked to participate | 45\% | Trained staff \& certified lab All fasting samples ?where sample stored, enzymatic analysis |
|  | Brianza | MONICA study (Anonymous 1989) | Sample from population register | 70-71\% | Trained staff \& certified lab All fasting samples Storage not necessary, enzymatic analysis |
|  | Friuli | MONICA study (Anonymous 1989) | Sample from regional health roll | 80-82\% | Trained staff \& certified lab Non-fasting samples Sample stored in freezer, enzymatic analysis |
|  | Latina | MONICA study (Anonymous 1989) | Sample from electoral rolls | 76\% | Trained staff \& certified lab <br> All fasting samples <br> Sample stored in fridge, enzymatic analysis |
|  | Netherlands | Vershuren et al. (1994) | Random sample of population selected from the municipal registry of each town | 50-57\% | Trained staff \& certified lab Non-fasting samples ?where sample stored, enzymatic analysis |
|  | Netherlands | Bosma et al. (1994) | Sample taken from a 10 -year follow-up to the Kaunus-Rotterdam Intervention Study (KRIS) | * | Trained staff \& certified lab <br> ?fasting sample <br> ?where sample stored, ?method of analysis |
|  | Norway | Graff-Iverson et al. (1998) | Sample from health survey | 89\% | Trained staff \& certified lab Non-fasting samples ?where sample stored, enzymatic analysis |


| Population-based cohort study | $81 \%$ |
| :--- | :--- |
| Two-stage population random <br> sample stratified by age from census <br> data <br> Sample from population register | $73 \%$ |
| Selected at random from population <br> register | $63 \%$ |
| Selected at random | $55-76 \%$ |
| Sample from population register | $84-86 \%$ |
| Sample from population register | $79-83 \%$ |
| Sample from population register | $61-69 \%$ |
| Sample from health survey | $55 \%$ |

Thune et al. (1998)
Masia et al. (1998)
MONICA study
(Anonymous 1989)
Asplund-Carlson and
Carlson (1994)
Rosengren et al. (2000)
MONICA study
(Anonymous 1989)
MONICA study
(Anonymous 1989)
MONICA study
(Anonymous 1989)
Brown et al. (1994)
Norway
Spain
Catalonia
Sweden
Sweden
Gothenburg
Switzerland,
Ticino
Vaud;
Fribourg
United Kingdom

$$
\begin{aligned}
& \text { Trained staff \& certified lab } \\
& \text { Non-fasting samples } \\
& \text { ?where sample stored, enzymatic analysis } \\
& \text { Trained staff \& certified lab } \\
& \text { All fasting samples } \\
& \text { Sample stored in freezer, enzymatic analysis } \\
& \text { Trained staff \& certified lab } \\
& \text { All fasting samples } \\
& \text { Sample stored in fridge, enzymatic analysis } \\
& \text { Trained staff \& certified lab } \\
& \text { All fasting samples } \\
& \text { Sample stored in freezer, ?method of analysis } \\
& \text { (Boehringer Mannheim) } \\
& \text { Trained staff \& certified lab } \\
& \text { All fasting samples } \\
& \text { ?where sample stored, enzymatic analysis } \\
& \text { Trained staff \& certified lab } \\
& \text { All fasting samples } \\
& \text { Sample stored in freezer, enzymatic analysis } \\
& \\
& \text { ?Trained staff \& certified lab } \\
& \text { Non-fasting samples } \\
& \text { Sample stored in fridge, enzymatic analysis } \\
& \text { Trained staff \& certified lab } \\
& \text { Non-fasting samples } \\
& \text { Sample stored in fridge, enzymatic analysis } \\
& \text { Trained staff \& certified lab } \\
& \text { Non-fasting samples } \\
& \text { ?where sample stored, enzymatic analysis } \\
& \hline
\end{aligned}
$$

Europe (continued)

| Subregion | Country or area | Study reference | Sampling | Response rate | Cholesterol measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | United Kingdom | Mann et al. (1988) | Opportunistic case finding from GP lists and random sampling from age-sex registers | 73\% | Trained staff \& certified lab <br> Some fasting samples <br> ?where sample stored, enzymatic analysis |
|  | England | Bajekal et al. (1999) | Community sample | Not stated | Trained \& certified staff Standard sphg one cuff 2 measures taken sitting after 5 mins rest, R arm |
|  | England | Razay et al. (1992) | Stratified random sample from regional health survey | 73\% | Trained staff \& certified lab <br> All fasting samples <br> Sample stored in freezer, enzymatic analysis |
|  | Northern Ireland, Belfast | MONICA study (Anonymous 1989) | Sample from GP lists | 57-70\% | Trained staff \& certified lab <br> Non-fasting samples <br> Sample stored in fridge, enzymatic analysis |
|  | Scotland | Smith et al. (1989) | Sample taken from selected districts, then random sampling from GP lists | * | Trained staff \& certified lab <br> Non-fasting samples <br> ?where sample stored, enzymatic analysis |
|  | Scotland, Glasgow | MONICA study (Anonymous 1989) | Sample from GP lists | 50-64\% | Trained staff \& certified lab <br> Non-fasting samples <br> Sample stored in fridge, enzymatic analysis |
|  | Multiple sites in EUR-A | Kafatos et al. (1991) | Stratified random sample by age and sex of target population from II countries | * | Trained staff \& certified lab <br> All fasting samples <br> Sample stored in freezer, enzymatic-colorimetric analysis |

Poland,
Tarnobrzeg
Voivodship
(Anonymous 1989)
MONICA study
(Anonymous 1989) (Anonymous 1989)
Mahley et al. (I995),
R. Mahley, personal communication, 2001
Onat et al. (1992)

70-80\% | Trained staff \& certified lab |
| :--- |
| All fasting samples |
| Sample stored in freezer, direct analysis |
| Trained staff \& certified lab |
| All fasting samples |
| Sample stored in fridge, direct analysis |
| Trained staff \& certified lab |
| All fasting sample |
| Sample stored in freezer, enzymatic analysis |
| Trained staff \& certified lab |
| Some fasting samples |
| ?where sample stored, enzymatic analysis |

| EUR-C | Former Czechoslovakia | MONICA study <br> (Anonymous 1989) | Random sample from population register | * | Trained staff \& certified lab All fasting samples Sample stored in fridge, direct analysis |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Estonia | Olferev et al. (1990, 1991) | Random sample from population register | 70-72\% | ?Trained staff \& certified lab All fasting samples ?where sample stored, ?method of analysis |
|  | Hungary | Biro et al. (1996) | Random sample from Budapest and seven counties in Hungary | * | Trained staff \& certified lab ?fasting sample ?where sample stored, ?method of analysis |
|  | Hungary | Kafatos et al. (1991) | Stratified random sample by age and sex of target population from II countries | * | Trained staff \& certified lab All fasting samples Sample stored in freezer, enzymatic-colorimetric analysis |

Europe (continued)

| Subregion | Country or area | Study reference | Sampling | Response rate | Cholesterol measurement |
| :---: | :---: | :--- | :--- | :--- | :--- |
|  | Budapest | MONICA study | Sample from population register <br> (Anonymous 1989) | $75-80 \%$ | Trained staff \& certified lab <br> Non-fasting samples |
|  | Pecs | MONICA study | Sample from population register <br> (Anonymous 1989) <br> (Anonymous 1989) | $70-80 \%$ | Trained staff \& certified lab |

[^25]South-East Asia

| Subregion | Country or area | Study reference | Sampling | Response rate | Cholesterol measurement |
| :--- | :--- | :--- | :--- | :--- | :--- |
| SEAR-B | Indonesia | INCLEN (Anonymous <br> I992b) | Random sample of males from 12 <br> centres in 7 countries | $70 \%$ | Trained staff \& certified lab <br> ?fasting sample |
|  | Thailand | Bhuripanyo et al. (1993) | Stratified random sample from <br> community survey | ?where sample stored, ?method of analysis |  |

Western Pacific

| Subregion | Country or area | Study reference | Sampling | Response rate | Cholesterol measurement |
| :--- | :--- | :--- | :--- | :--- | :--- |
| WPR-A | Australia | APCSC-Busselton <br> (APCSC secretariat, <br> personal communication, | A stratified random sample <br> representative of population <br> 200I) |  | $*$ |

\(\left.$$
\begin{array}{lllll}\text { Newcastle } & \begin{array}{l}\text { MONICA study } \\
\text { (Anonymous 1989) }\end{array} & \text { Sample from electoral register } & \text { 68-82\% } & \begin{array}{l}\text { Trained staff \& certified lab } \\
\text { All fasting samples } \\
\text { Sample stored in fridge, extraction \& enzymatic } \\
\text { analysis }\end{array} \\
\text { Perth } & \begin{array}{l}\text { MONICA study } \\
\text { (Anonymous 1989) }\end{array} & \text { Sample from electoral register } & \text { 81-84\% }\end{array}
$$ \begin{array}{l}Trained staff \& certified lab <br>
All fasting samples <br>
Sample stored in fridge, extraction \& enzymatic <br>

analysis\end{array}\right]\)| Trained staff \& certified lab |
| :--- |
| ?fasting sample |
| ?where sample stored, ?method of analysis |

Western Pacific (continued)

| Subregion | Country or area | Study reference | Sampling | Response rate | Cholesterol measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Japan | Robinson et al. (Law and Wald 1994; Robinson et al. 1992) | Sample from medical centre | * | Trained staff \& certified lab All fasting samples ?where sample stored, enzymatic analysis |
|  | Japan | Serum lipid Survey <br> (Anonymous 1996) | Sample from 39 institutes from various districts | * | Trained staff \& certified lab All fasting samples Sample stored in fridge, enzymatic analysis |
|  | Japan | 1990 National Survey (Sakata and Labarthe 1996) | Randomly selected from National Health Survey districts | 81.5\% | Trained staff \& certified lab <br> All fasting samples <br> Sample stored in fridge, enzymatic analysis |
|  | New Zealand | APCSC-Fletcher challenge (APCSC secretariat, personal communication, 2001) | Employees of one centre, and random sample from electoral roll | 74\% | Trained staff \& certified lab Non-fasting samples Storage not necessary, enzymatic analysis |
|  | New Zealand | Bullen et al. (1998) | Age stratified random sampling from electoral rolls | 66-68\% | Trained staff \& certified lab Non-fasting samples Sample not necessary, enzymatic analysis |
|  | New Zealand | Flight et al. (1984) | Randomly chose schools then students | 81\% | Trained staff \& certified lab ?fasting sample ?where sample stored, ?method of analysis |
|  | New Zealand | Mann et al. (1991) | Random sample from electoral roll | 94\% | Trained staff \& certified lab <br> Fasting samples <br> Sample stored in freezer, enzymatic analysis |


|  | New Zealand | National Survey (Ministry of Health 1999) | Random sample of individuals from random households constructed from census data | * | Trained staff \& certified lab <br> Non-fasting samples <br> Sample stored in freezer, ?method of analysis |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Auckland | MONICA study (Anonymous 1989) | Sample from electoral register | 81\% | Trained staff \& certified lab <br> Non-fasting samples <br> Sample stored in freezer, extraction analysis |
|  | Singapore | APCSC-Singapore Heart Survey (APCSC secretariat, personal communication, 2001) | Random sampling of population of Singapore | * | Trained staff \& certified lab <br> ?fasting sample <br> ?where sample stored, ?method of analysis |
|  | Singapore | Hughes et al. (1990) | Random sample of census districts, units, houses, then individuals (weighted) | 52-66\% | Trained staff \& certified lab <br> All fasting samples <br> Sample stored in fridge, enzymatic analysis |
| WPR-B | China | APCSC-Anzhen 02 (APCSC secretariat, personal communication, 2001) | Cluster sampling | * | Trained staff \& certified lab <br> ?fasting sample <br> ?where sample stored, ?method of analysis |
|  | China | APCSC-Hong Kong SAR (APCSC secretariat personal communication, 2001) | Stratified random sampling in Hong Kong SAR | 85\% | Trained staff \& certified lab <br> ?fasting sample <br> ?where sample stored, ?method of analysis |
|  | China | APCSC-Seven Cities cohort (APCSC secretariat, personal communication, 2001) | Study samples selected from Chinese populations of farmers, workers and fishermen | * | Trained staff \& certified lab <br> ?fasting sample <br> ?where sample stored, ?method of analysis |

Western Pacific (continued)

| Subregion | Country or area | Study reference | Sampling | Response rate | Cholesterol measurement |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | China | APCSC-Six Cohorts (APCSC secretariat, personal communication, 2001) | Cluster sampling of six cohorts in villages, including farmers of Shanxi, Shaanxi, Guangxi, Jiangsu province, minors of Hebei province and fishermen of Zhejiang | $\geq 85 \%$ | Trained staff \& certified lab ?fasting sample <br> ?where sample stored, ?method of analysis |
|  | China | INCLEN (Anonymous 1992b) | Random sample of males from 12 centres in 7 countries | 76-99\% | Trained staff \& certified lab ?fasting sample <br> ?where sample stored, ?method of analysis |
|  | China | Tao et al. (1992) | Cluster sampling from four regions in China | 87-88\% | Trained staff \& certified lab All fasting samples Storage not necessary, enzymatic analysis |
|  | China | Tian et al. (1995) | Random survey of population in Tianjin | 96\% | Trained staff \& certified lab <br> All fasting samples <br> Storage not necessary, enzymatic analysis |
|  | China | Yang et al. (1986) | Sample from Beijing region | * | Trained staff \& certified lab <br> Non-fasting samples <br> Sample stored in freezer, ?method of analysis |
|  | China | Zhuang et al. (1986) | Sample from specific region using urban and rural population | * | Trained staff \& certified lab All fasting samples ?where sample stored, ?method of analysis |
|  | Beijing | MONICA study <br> (Anonymous 1989) | Sample from register of households | 89-90\% | Trained staff \& certified lab ?fasting sample |

?where sample stored, ?method of analysis
Trained staff \& certified lab
Sample stored in freezer, ?method of analysis
Trained staff \& certified lab
?fasting sample
?where sample stored, enzymatic analysis
Trained staff \& certified lab
?fasting sample
?where sample stored, ?method of analysis
Sample stored in freezer, ?method of analysis
Trained \& certified lab
Sample stored in freezer, enzymatic analysis
Trained staff \& certified lab
?fasting sample
?where sample stored, ?method of analysis
?
Sample from population register
Two townships in Taiwan, China

age range during a period of seven weeks-some non-randomized
subjects included due to low participation
Randomly selected from three rural 58-64\% (F) communities and one urban $\quad 52-60 \%$ (M) community
Random sample of males from 12
centres in 7 countries
Taiwan, China APCSC-CVDFACTS/Two

Scrimgeour et al. (1989)
INCLEN (Anonymous
1992b)
Hong Kong
SAR
Hong Kong
SAR
Taiwan, China
Papua New
Guinea
Philippines
Data unavailable.
Specific details not given.

Appendix B: Methodology for estimating stroke subtypes by age, sex, subregion and fatal and non-fatal events
Subregional- and age-specific proportions of ischaemic and haemorrhagic fatal and non-fatal strokes were estimated so that weighted RRs could be applied. Based on published data, an assumption was made that these are the same for males and females. Several steps were undertaken to estimate stroke subtype proportions, and these are outlined below.

## Step 1. Assess age patterns of haEmorrhagic and <br> ISCHAEMIC STROKE

The most reliable data on age patterns of stroke subtypes are available from a small selection of "gold standard" incidence studies, most of which were included in a review (Sudlow and Warlow 1997). These studies include the Melbourne stroke study (Thrift et al. 2001), Oxfordshire stroke project (Bamford et al. 1990), Perth community stroke study (Anderson et al. 1993) and the Dijon study (Giroud et al. 1991). Broadly, similar age patterns have also been seen in other studies.

Utilizing the combined age-specific data from these four studies, the percentage of ischaemic strokes in 10-year age categories is presented in Figure B.1. (The number of strokes in the youngest age group is small compared to the other age groups and is therefore less reliable.)

Figure B.I Percentage of ischaemic strokes by age using data from four "gold standard" studies with linear regression line


Table B.I Estimates of the proportion of stroke subtypes by age

|  | Age group (years) |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Stroke subtypes | $30-44$ | $45-59$ | $60-69$ | $70-79$ | $\geq 80$ |
| $\%$ ischaemic | 77 | 81 | 84 | 87 | 89 |
| $\%$ haemorrhagic | 23 | 19 | 16 | 13 | 11 |

The linear regression was then used to estimate the proportions of ischaemic stroke by GBD age groups using the midpoint in each GBD age category. The remaining strokes were classified as haemorrhagic (Table B.1).

## Step 2. Estimate the age-specific subtype proportions for fatal and NON-FATAL EVENTS

Table B. 1 relates to total strokes, but these proportions will differ for fatal and non-fatal strokes as the case fatality rates for the ischaemic and haemorrhagic strokes are different. The "gold standard" incidence studies (Sudlow and Warlow 1997) suggest that the one-month case fatality rates range from $10 \%$ to $23 \%$ (crude average $=14 \%$ ) for ischaemic stroke, and $35 \%$ to $54 \%$ (crude average $=45 \%$ ) for haemorrhagic stroke. If it is assumed that these case fatality rates apply for all age groups, these crude percentages can be applied to each age group to provide estimates of the overall proportion of stroke subtypes for fatal and non-fatal events (Table B.2).

These percentages may then be converted in proportions of ischaemic and haemorrhagic stroke within the fatal and non-fatal categories for each age group (Table B.3).

Table B. 2 Estimates of subtype proportions for fatal and non-fatal events by age

| Age group (years) | Total stroke (\%) |  | Fatal stroke (\%) |  | Non-fatal stroke (\%) |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Ischaemic | Haemorrhagic | Ischaemic | Haemorrhagic | Ischaemic | Haemorrhagic |
| 30-44 | 77 | 23 | 11 | 10 | 66 | 13 |
| 45-59 | 81 | 19 | 11 | 9 | 70 | 10 |
| 60-69 | 84 | 16 | 12 | 7 | 72 | 9 |
| 70-79 | 87 | 13 | 12 | 6 | 75 | 7 |
| $\geq 80$ | 89 | 11 | 12 | 5 | 77 | 6 |

Table B. 3 Percentage of stroke subtypes within fatal and non-fatal categories

|  | Fatal stroke (\%) |  |  | Non-fatal stroke (\%) |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Age group (years) | Ischaemic | Haemorrhagic |  | Ischaemic | Haemorrhagic |
| $30-44$ | 51 | 49 |  | 84 | 16 |
| $45-59$ | 57 | 43 |  | 87 | 13 |
| $60-69$ | 62 | 38 | 89 | 11 |  |
| $70-79$ | 68 | 32 | 91 | 9 |  |
| $\geq 80$ | 72 | 28 | 93 | 7 |  |

Step 3. Estimate how these age-specific subtype proportions can be APPLIED TO DIFFERENT REGIONS
The age-specific subtype proportions in Table B. 3 would only apply to subregions such as AMR-A and EUR-A, from where the data on which the estimates were based came. Estimates were also required for other subregions. The literature suggests that haemorrhagic stroke is relatively more common (about 30-40\% of all strokes) in Japan (Nakayama et al. 1997; Suzuki et al. 1987), China (Chen et al. 1992; Wu et al. 1992) and Taiwan, China (Hung 1993; Jeng et al. 1998) and lower in countries in North America and Europe (about 10-15\% of all strokes) (Sudlow and Warlow 1997). Unfortunately, reliable data on the remainder of the world are extremely limited or non-existent, and there are no internally consistent WHO estimates available.

The level of cholesterol is important in determining the overall ratio of ischaemic to haemorrhagic strokes in any world region (APCSC 1999). Subregions may therefore be crudely ranked based on their mean cholesterol levels, grouped into three basic subregional categories, and average percentages of haemorrhagic stroke applied. An example is given in Table B.4.

Table B. 4 Subregional grouping by mean cholesterol and percentage haemorrhagic stroke ${ }^{\text {a }}$

| Subregional grouping | Cholesterol level | Overall \% haemorrhagic stroke |
| :--- | :--- | :---: |
| AMR-A, EUR-A, EUR-C | $\geq 5.8 \mathrm{mmol} / \mathrm{l}$ | 15 |
| AFR-D, AFR-E, AMR-B, AMR-D, | $5.1-5.7 \mathrm{mmol} / \mathrm{I}$ | 20 |
| EMR-B, EMR-D, EUR-B, SEAR-D, WPR-A |  |  |
| SEAR-B, WPR-B | $\leq 5.0 \mathrm{mmol} / \mathrm{l}$ | 30 |

[^26]The case fatality rates used previously ( $14 \%$ ischaemic, $45 \%$ haemorrhagic) are applicable to WPR-A, AMR-A and EUR-A, but case fatality would be higher in other subregions. Limited data are available, but when case fatality rates of $20 \%$ for ischaemic stroke and $60 \%$ for haemorrhagic stroke (Chen et al. 1992; Li et al. 1985) are applied to other subregions, five potential scenarios result (Table B.5).

These proportions may then be translated into age-specific proportions for fatal and non-fatal ischaemic and haemorrhagic strokes (Table B.6).

Table B. 5 Subregional grouping by mean cholesterol and percentage haemorrhagic stroke

| Subregional grouping | \% haemorrhagic <br> stroke | Ischaemic stroke <br> case fatality (\%) | Haemorrhagic <br> case fatality (\%) |
| :--- | :---: | :---: | :---: |
| EUR-A, AMR-A | 15 | 14 | 45 |
| EUR-C | 15 | 20 | 60 |
| WPR-A | 20 | 14 | 45 |
| EMR-B, EMR-D, EUR-B, AFR-D, | 20 | 20 | 60 |
| AFR-E, AMR-B, AMR-D, SEAR-D | 30 | 20 | 60 |
| SEAR-B, WPR-B |  |  |  |

Table B. 6 Proportion of fatal and non-fatal stroke subtypes by age and subregion

| Age group (years) | Fatal stroke (\%) |  | Non-fatal stroke (\%) |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Ischaemic | Haemorrhagic | Ischaemic | Haemorrhagic |
| EUR-A, AMR-A |  |  |  |  |
| 30-44 | 51 | 49 | 84 | 16 |
| 45-59 | 57 | 43 | 87 | 13 |
| 60-69 | 62 | 38 | 89 | 11 |
| 70-79 | 68 | 32 | 91 | 9 |
| $\geq 80$ | 72 | 28 | 93 | 7 |
| EUR-C |  |  |  |  |
| 30-44 | 53 | 47 | 87 | 13 |
| 45-59 | 59 | 41 | 90 | 10 |
| 60-69 | 64 | 36 | 91 | 9 |
| 70-79 | 69 | 31 | 93 | 7 |
| $\geq 80$ | 73 | 27 | 94 | 6 |

Table B. 6 Proportion of fatal and non-fatal stroke subtypes by age and subregion (continued)

| Age group (years) | Fatal stroke (\%) |  | Non-fatal stroke (\%) |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Ischaemic | Haemorrhagic | Ischaemic | Haemorrhagic |
| WPR-A |  |  |  |  |
| 30-44 | 41 | 59 | 78 | 22 |
| 45-59 | 48 | 52 | 82 | 18 |
| 60-69 | 53 | 47 | 85 | 15 |
| 70-79 | 60 | 40 | 88 | 12 |
| $\geq 80$ | 64 | 36 | 90 | 10 |
| EMR-B, EMR-D, EUR-B, AFR-D, AFR-E, AMR-B, AMR-D, SEAR-D |  |  |  |  |
| 30-44 | 43 | 57 | 82 | 18 |
| 45-59 | 50 | 50 | 85 | 15 |
| 60-69 | 55 | 45 | 88 | 12 |
| 70-79 | 61 | 39 | 91 | 9 |
| $\geq 80$ | 66 | 34 | 92 | 8 |
| SEAR-B, WPR-B |  |  |  |  |
| 30-44 | 28 | 72 | 70 | 30 |
| 45-59 | 35 | 65 | 77 | 23 |
| 60-69 | 41 | 59 | 81 | 19 |
| 70-79 | 49 | 51 | 85 | 15 |
| $\geq 80$ | 54 | 46 | 88 | 12 |

## Chapter 8

# Overweight and obesity (High body mass index) 

W. Philip T. James, Rachel Jackson-Leach, Cliona Ni Mhurchu, Eleni Kalamara, Maryam Shayeghi, Neville J. Rigby, Chizuru Nishida and Anthony Rodgers

## Summary

It is widely acknowledged that being overweight is associated with an amplified risk of disease, particularly if body fat is deposited within the abdomen, as suggested by a high waist-circumference measurement. This chapter aims to estimate the burden of disease attributable to overweight and obesity as indicated by a high body mass index (BMI), by age, sex and subregion. ${ }^{1}$

BMI, which is calculated as weight ( kg ) divided by height squared $\left(\mathrm{m}^{2}\right)$, was chosen as a simple measurement of body weight in relation to height. While increases in both body fat and lean tissue cause increments in BMI, relationships between body weight and health are conventionally expressed in terms of BMI rather than body fat. Data on population weight and height, often collected as part of general medical or economic surveys, were obtained, typically from specially-commissioned analyses from ministries of health. Where these data sets or published representative information were lacking, earlier data published for each country were used. All information based on studies of select groups within a population were excluded. In addition, only data obtained by actual measurement of heights and weights by trained observers were included. As data were not available for some countries, it was necessary to extrapolate from data for other countries or subregions when deriving estimates of BMIs for the different age groups in each subregion.

Analyses of the relationship between BMI and both mortality and morbidity suggested that the theoretical optimum mean population BMI was approximately $21 \mathrm{~kg} / \mathrm{m}^{2}$. This value is far removed from those now found in many parts of the world. The analyses based on this continuous relationship therefore replaced the usual categorical analyses based on rates of overweight and obesity in the different subregions.

The disease outcomes assessed in relation to excess weight were type II diabetes (diabetes mellitus), ischaemic heart disease, stroke, hypertensive heart disease, osteoarthritis, and cancers of the postmenopausal breast, colon, endometrium and kidney. As it was evident that adult BMIs of $>21 \mathrm{~kg} / \mathrm{m}^{2}$ were associated with the development of disease, the burden of disease attributable to high BMI was calculated from this baseline. New analyses based on 33 cohort studies carried out within the Asia-Pacific region were used to estimate the incremental risk of cardiovascular disease associated with each unit increase in BMI above 21 $\mathrm{kg} / \mathrm{m}^{2}$. The relationship between BMI and the risk of type II diabetes was derived from both unpublished and published data comprising measured anthropometry and fasting blood sugar measurements, extracted from nationally representative studies. Equivalent increments in the risks of co-morbidities associated with body-weight gain were assumed for all parts of the world.

High mean BMIs and elevated rates of overweight and obesity were found in the Americas, Europe, the Middle East and in the Western Pacific. It is estimated that rates of obesity vary geographically from $2-3 \%$ in some Asian countries to $75 \%$ in several Pacific Island nations. Currently, there are more than 300 million obese and more than 750 million overweight individuals in the world.

The proportions of the global burden of disease attributable to increases in BMI were $58 \%$ for type II diabetes, $21 \%$ for ischaemic heart disease, $39 \%$ for hypertensive disease, $23 \%$ for ischaemic stroke, $12 \%$ for colon cancer, $8 \%$ for postmenopausal breast cancer and $32 \%$ for endometrial cancer in women, and $13 \%$ for osteoarthritis. This means that the global burden of disease attributable to excess BMI in adults amounted to more than 30 million disability-adjusted life years (DALYs) in 2000, mostly incurred from ischaemic heart disease and type II diabetes. There were two and a half million deaths associated with this exposure. These are average global figures and there are remarkable variations by subregion and by disease. Thus EUR-C has the greatest burden of DALYs, this being dominated by the impact of high BMI on ischaemic heart disease, whereas the two African subregions have the lowest burden of DALYs. The burden of diabetes attributable to high BMI is greatest in WPR-B and AMR-B, with AMR-A also having a substantial burden. DALYs attributable to stroke were also dominated by the impact of high BMIs in both EUR-C and WPR-B, while the burden of DALYs caused by cancer was substantial in the European subregions, AMR-A, AMR-B and WPR-B.

Current trends were used to predict the increases in BMI and disease burden that are likely to occur by 2030, assuming that no new measures are taken to counteract the rapid recent increases in body weight in all parts of the world. On this basis, it is predicted that the burden of disease will increase substantially in most parts of the world, but there will probably be remarkable variations by subregion.

## 1. Introduction

Although the measurement and analysis of body weights and heights have been recognized as general indices of health for many years, it is only comparatively recently that the World Health Organization (WHO) has set out criteria for assessing underweight and overweight in both children and adults (WHO 1995). These new analyses of the impact of excess body weight came from insurance data generated in the first half of the 20th century which were used to identify optimum weights-forheight above which life expectancy was reduced, for both men and women. In the second half of the 20th century, it became clear that abnormalities in blood lipids relating to the risks of ischaemic heart disease were amplified by excessive body-weight gain, as was the risk of high blood pressure, type II diabetes, gallbladder disease and some cancers. It also became clear that the mechanical impact of excess body weight induced breathlessness and promoted arthritis in the weightbearing joints. In developed countries, overweight women were stigmatized, with marked consequences on their sense of well-being, social interactions and even their employment and marriage prospects.

The traditional concerns of governments and policy-makers have focused on undernutrition, with greater emphasis being placed on the continuing problem of childhood protein-energy malnutrition, which is found especially in children aged $0-4$ years. This condition is still prevalent in many countries despite economic progress (James et al. 2000), as described in chapter 2. Many nations now have reasonable systems for monitoring children's growth and can provide estimates of the prevalence of stunting, wasting and overweight in children aged $<5$ years (de Onis and Blössner 2000). Unfortunately, the value of monitoring the weights and heights of older children and adults has not been appreciated until fairly recently. Since the 1997 WHO Expert Consultation on Obesity (WHO 2000), there has been a substantial increase in the number of publications presenting newly-analysed data from past studies in different parts of the world. Thus, the regular national NHANES surveys in the United States of America (Stevens et al. 1999) allowed the magnitude of the problem of overweight to be recognized, and many cardiovascular surveys, for example the WHO MONICA surveys (Dobson et al. 1998), also documented high prevalences of overweight and obesity in Europe and Australasia. Other surveys such as the INTERSALT study (Dyer and Elliott 1999) revealed high prevalences of excess weight in some developing countries, including Brazil (James and Francois 1988). The data presented by many of these studies are not representative and do not include validation of the measurements of height and body weight. Nevertheless, it is apparent that many new national surveys are now being undertaken and a much more extensive database is expected to become available within the next few years.

This chapter was based on an extensive search of the literature to identify appropriate data sets and also specifically-commissioned analyses provided by a number of individuals, organizations and governments.

## 2. Choice of exposure variable

### 2.1 Definitions of body weight and of risk factors

## The use of the BMI

The present analysis is based exclusively on the use of the BMI, which is calculated as weight $(\mathrm{kg})$ divided by height squared $\left(\mathrm{m}^{2}\right)$. The height and weight of both children and adults are crude indices of the impact of many environmental factors, (including diet and infections) on the genetic growth potential of the individual over short and long periods of time, and affect many health outcomes.

BMI is the most appropriate simple indicator by which weight-forheight can be related to health outcome. WHO (1995) therefore proposed the use of BMI to monitor both undernutrition and overweight. The power of height is taken as 2.0 although it has been shown in many analyses that 1.5 might be more appropriate for women on the basis that this index in population studies proves to be approximately heightindependent (Micozzi et al. 1986). Nevertheless, international convention, as represented by two major WHO Technical Consultations (WHO 1995,2000 ), endorsed the use of a common BMI scheme for adults irrespective of sex or age.

Preliminary analyses of the global burden of disease associated with higher BMI, based on the current data sets, suggested that the population distribution of BMI values for men and women in each age group provided more valuable information than simply the proportions of the population who are classified as overweight and obese. These categories of overweight and obesity are used extensively by clinicians for patient management decisions, by the public and by policy-makers. Therefore, the proportions of overweight and obese people in the population are included in this chapter despite the fact that this information was not used in the calculation of the contribution of different values of BMI to the disease burden.

In these subsidiary analyses, the standard WHO BMI categories were used, except that the term "overweight" was taken as referring to BMI values of $25.0-29.9 \mathrm{~kg} / \mathrm{m}^{2}$ only and did not include the "obese" category, i.e. BMI of $\geq 30 \mathrm{~kg} / \mathrm{m}^{2}$, since these two groups, overweight and obesity, are often referred to independently. More extreme categories of obesity have been specified (WHO 2000), but were not included in the current global analyses.

Recently, it has been proposed that a lower BMI range of "healthy", "normal" or "acceptable" weights should be applied to groups of Asian
people (see below), but in the current analysis an assessment was made of the impact of increments in BMI on disease risk in different parts of the world, which was therefore not dependent on different schemes for categorizing overweight and obesity.

## BODY WEIGHT IN CHILDREN

It has become increasingly common for epidemiologists to express heights and weights of children in terms of the same BMI as used in adults, despite detailed analyses showing that BMI varies by age and sex during growth. Criteria have therefore now been developed for specifying the normal weight-for-height of children in terms of BMI for each age group, by sex, until adult height is achieved at approximately 18 years of age. There are three approaches to the categorical analysis of BMI in children: the traditional approach whereby an "abnormal" group is taken to be more than two standard deviations from the mean, the new International Obesity Task Force (IOTF) approach that relates BMI categories in childhood to the accepted classification in adults (Cole et al. 2000), and a new Centers for Disease Control and Prevention (CDC) set of standards whereby obesity is specified as $>95$ th BMI percentile of carefully selected representative data from the United States (Ogden et al. 2002).

It is recognized that children of similar body proportions but of different heights at the same age will have different BMI values and that to obtain height-independent indices would require a sequential adjustment in the power value of height from about age 5 years upwards (Franklin 1999). Nevertheless, given that population comparisons are being made here, rather than the monitoring of the growth of individuals, weights and heights for children have been expressed in terms of BMI and these calculations have been applied only to children aged 5-18 years. The large body of nationally representative data for children aged $<5$ years collated by WHO and presented in chapter 2 provides the relevant information for this age group. However, the burden of disease estimates presented here are only for adults aged $\geq 30$ years. This age limit was chosen because there are as yet insufficient prospective studies of an appropriate magnitude in children and young adults to allow quantitative analyses of the impact of excess weight gain on the incidence of noncommunicable diseases in individuals aged $<30$ years.

### 2.2 Other exposure determinants

## Body fat

It is often assumed that health-related data would ideally be related to good measures of body fatness, and that the combination of weight and height in the form of BMI provides a crude index of body fatness. In practice, however, too few studies have measured body fatness and health outcomes at different ages and in different societies to allow an analysis
of whether a more specific measure of body fat than BMI would give greater predictive power for health outcomes. Given the many prospective studies that use BMI, the current convention has been maintained while recognizing that recent data show that different ethnic groups have substantially different proportions of body fat at the same value of BMI. For example, the ratio of fat: lean tissue is highest in Indian people, while values for Chinese people are intermediate between those for Indians and Caucasian peoples (Deurenberg et al. 2002), and Polynesians are increasingly recognized as having a relatively high proportion of lean tissue (Swinburn et al. 1999). On this basis, Deurenberg and colleagues have suggested that different BMI values should be chosen if the intention is to standardize international comparisons on the basis of body fat (Deurenberg et al. 1998, 2002). There are, however, no international studies as yet which would allow all population groups to be set a particular BMI value based on their fat:lean tissue ratios and the relationship of these indices to health outcomes.

Therefore, this chapter maintains the current convention of using BMI as an indicator of body fatness in adults. It is recognized that not only do women have substantially more fat tissue than men at equivalent BMIs (Shetty and James 1994), but also that both men and women lose lean tissue during the course of their adult lives such that, at an equivalent BMI, a 75 -year-old man or woman has substantially greater proportion of fat than a 25 year old (James et al. 1988). This is not a cohort effect since the same changes have been shown, at least in men, in the Baltimore study of ageing, which evaluated the changing body composition of the same men over a 50-year period (James et al. 1989).

## CORrECTIONS FOR UNUSUAL BODY PROPORTIONS

The proportions of the major body parts which contribute to height may be different in different ethnic groups. For example, some African tribes are considered to have exceptionally long legs, whereas the indigenous populations of Central America are often cited as being small with very short legs (Norgan 1994a). It can readily be shown that even if the proportions of both the trunk and legs are equal in very short and tall peoples, the actual BMIs of these peoples will be very different. This has led Norgan (1994b) to develop a simple correction for BMI measurement based on the ratio of sitting height to total height. Although this is well-recognized (WHO 1995) and is valuable when looking at particular groups, the limited availability of good data on sitting height meant that it was not possible to incorporate this correction for BMI into the current analyses.

## WAIST CIRCUMFERENCE

Originally it was hoped that sufficient data would become available on waist circumference to allow an assessment of the usefulness of this measure in predicting the health of different communities. There are
many analyses that demonstrate that waist circumference provides a reasonable indicator of the quantity of abdominal fat, which correlates with the amount of intra-abdominal or visceral fat (Despres et al. 2001). This fat is considered to be metabolically rather different from subcutaneous fat in its responsiveness to dietary change and in its array of metabolite and hormonal outputs. An excess of abdominal fat has been associated with a range of metabolic abnormalities and diseases (Despres et al. 2001). The measurement of waist circumference is often found to be more valuable than BMI itself, for example, in predicting the likelihood of ischaemic heart disease (Lapidus et al. 1984; Larsson et al. 1984) or diabetes (Chan et al. 1994). The National Institutes of Health (NIH) report from the United States (NIH 1998) used waist circumference measurements as a suitable indicator of additional risk within a given range of BMI. In some studies, there seems to be additional predictive power when the waist:hip ratio rather than just waist circumference is used. The hip measurement indicates the degree of fat accumulation around the hips and this deposition may help in some way to limit the health impact of abdominal fat accumulation (Seidell et al. 2001b). Nevertheless, there seems to be increasing acceptance that, for general use, a single measure of waist circumference provides a simple index of fat distribution and additional risk (Seidell et al. 2001a).

Proposals have been also made for lower cut-off points for measurements of waist circumference for use in Asian communities (WHO/IASO/IOTF 2000) and new Chinese analyses have also proposed different values (Zhou 2002). There is also now increasing evidence that many communities, e.g. African and Hispanic Americans in the United States, Indians in India and elsewhere, the Chinese and Latin Americans have a greater propensity as adults to accumulate excess adipose tissue in the abdominal area than Caucasians in Europe or the United States (Ford et al. 2002; Sánchez-Castillo et al. 2003; Sargeant et al. 2002; Singh et al. 1995; Zhou et al. 2002). Although the selective accumulation of abdominal fat is indicative of a much greater risk of diabetes, hypertension, ischaemic heart disease, strokes and gall bladder disease, cross-sectional studies of the African diaspora in West Africa, the Caribbean and the United States show that the relationship of waist circumference to disease seemed to vary by region, perhaps because of concomitant regional dietary differences (Okosun et al. 1998, 2000). Nationally representative data and long-term cohort studies of the health impact of different indices of abdominal obesity in different communities are also currently insufficient to allow the use of some measures of waist circumference to estimate the BMI-disease relationship in different parts of the world.

The propensity to abdominal obesity within a community seems to be markedly influenced by stunting or a small size in childhood (Schroeder et al. 1999) and also by size at birth (Barker 1998). Changes in the hypo-thalamic-pituitary-adrenal axis controlling pituitary hormone and
corticosteroid metabolism in response to fetal nutritional deprivation and early postnatal events are also evident experimentally (Seckl et al. 2001) and abdominal obesity is associated with abnormal control of corticosteroid metabolism (Björntorp and Rosmond 1999). Evidence from India shows that children aged 4 and 8 years who were born small and later showed accelerated growth had a propensity to abdominal obesity with greater insulin resistance and higher blood pressure (Yajnik 2000). The current data available on a global basis do not, however, allow a systematic adjustment of health risk based on birth weights in different parts of the world, or the prediction of childhood BMIs from infant birth weights. A substantial proportion of the world's population that has been existing on marginal diets for centuries may have been sensitized to excess body-weight gain, this being reflected in the greater propensity to accumulate abdominal fat and in the higher prevalence of the metabolic syndrome of multiple risk factors for chronic diseases of adults such as diabetes, hypertension and ischaemic heart disease in Hispanic and non-Caucasian ethnic groups (Ford et al. 2002).

## 3. Methods of identifying sources AND STUDIES

### 3.1 Studies of interest

Studies of interest were identified using the following methods:

- Searches of the Medline and Embase databases were conducted systematically for all 191 countries of the world. Medline searches were performed with the keywords "BMI" and "obesity", each paired with "cardiovascular disease", "hyperlipidaemia", "cholesterol", "stroke", "ischaemic heart disease", "osteoarthritis", "diabetes mellitus type II", "cerebrovascular disease", and in combination with each country name, i.e. $2 \times 8 \times 191$ searches. Example: BMI AND cardiovascular disease AND country X. Both United Kingdom and American English spellings were used in the searches. Embase searches were performed with the keywords "BMI", "obesity", "body mass", "body height", "weight", "children", and "adults". Countries were not specified in these searches.
- IOTF contacted each WHO Regional Nutrition Officer to request help with the analyses. The precise format for these was specified and each region was asked to help identify appropriate contacts from whom reliable national data on both BMI and diet could be obtained.
- Numerous direct contacts were made with governments and individuals to determine whether unpublished data were available. It is significant that with obesity now becoming a high profile issue throughout the world, many investigators, on learning of this
project, stated that they now wished to publish in their own right information on prevalence rates that had remained unpublished for several years.
- Relevant data sets were retrieved from online databases, or purchased and re-analysed. These included a United States Agency for International Development (USAID)-sponsored series of Demographic and Health Surveys (DHS) conducted by Macro International, the United States National Health and Nutrition Examination Survey (NHANES) III, and the 1998 Health Survey for England.
- Analyses of 33 cardiovascular cohort studies being conducted in the Asia-Pacific region and participating in the Asia-Pacific Cohort Studies Collaboration (APCSC) were also used to derive the relative risks of cardiovascular disease associated with increases in BMI.

For data from the identified studies, the following inclusion criteria were used.

- Nationally representative data were preferred.
- Clinical data were excluded whenever possible because they reflect a subgroup of the population with particular medical problems and could not be considered nationally representative.
- Only measured anthropometric data were used to assess the national information on BMI. A good correlation between measured and reported weights and heights can be found (Flegal and Troiano 2000), but many international analyses of reported vs measured heights and weights, including those from the United States, reveal discrepancies which underestimate weight and overestimate height, these discrepancies being particularly apparent in the groups of overweight people (see Niedhammer et al. 2000). Australian analyses have also shown that there can be very substantial differences in the prevalences of overweight and obesity as judged by the two approaches (Anonymous 1999). Nevertheless, for prospective analyses, the overall associations between reported weights and heights and health outcomes are unlikely to be seriously affected, even if the magnitude of the association becomes more uncertain. Therefore, some results of major studies employing self-reported weights and heights were used to illustrate disease relationships (although the quantitative associations used in estimates of impact on the health of the population are all derived from studies employing actual measurement of BMI). Given the levels of inaccuracy and bias associated with self-reported BMI, the use of such data was considered inappropriate for the purpose of estimating exposure. A small bias could have a substantial impact on the estimated prevalences of overweight and obesity in the tail of the BMI distribution. This criterion excludes many high profile publications that rely on self-reported weights and heights, particularly from

European Union surveys, and a large number of studies from the United States.

- A sufficiently large sample size was required, with preference being given to studies investigating $\geq 1000$ individuals. For countries with no data, studies with smaller samples were not excluded.
- The earliest cut-off date for data collection was 1990 , whenever possible. Where no suitable studies were available, studies dating from 1980 onwards were considered.
- For diabetes, only representative population measures of fasting plasma glucose were used. Ideally, representative data from children and adults tested with a standard glucose load are desirable for a full assessment of the prevalence of diabetes in relation to BMI, but there are very few studies with nationally representative data available from developing countries. No studies which involved the self-reporting of the presence of diabetes were considered appropriate for estimating national prevalences of diabetes, since surveys have repeatedly shown that representative assessments of population groups find a substantial proportion of unrecognized cases of diabetes within the community. In general, as age and BMI increase, the number of individuals with unrecognized diabetes also increases and there can therefore be substantial systematic biases. On this basis, only nationally representative data on fasting blood glucose levels, in combination with measured weights and heights, were used to assess the risk of diabetes associated with body-weight gain. The reason for relying on prevalence rather than incidence data for diabetes to estimate risk in relation to excess body weight is set out in a later section.


### 3.2 Measurements in children

The quantitative assessments of the disease burden caused by high BMI reported here only apply to adults aged $\geq 30$ years. However estimates were also made of BMI values among children, by age, sex and subregion. This was for two reasons. First, these estimates are relevant to estimates of avoidable burden, since BMI levels track over time and so they will guide projections of the distribution of adult BMI in the next few decades. Second, high BMI is responsible for a disease burden in children and there is increasing evidence that high BMI in childhood markedly enhance the risks of disease once these children become adults; these relationships need to be included in hazard size estimates in the future. Therefore, these analyses attempted to estimate the distribution of BMIs for each 1 -year age group for ages $5-18$ years (i.e. until the typical end of child and adolescent normal growth [see the WHO Expert Technical Consultation on Anthropometry, WHO 1995]) and for ages 18-29 years combined.

In analysing the basis of overweight in children (Dietz and Bellizzi 1999), a concept was developed by IOTF which allowed a coherent set of nationally representative data on BMI percentiles at age 18 years to be obtained from both developed countries (prior to the recent emergence of many children with clear clinical obesity) and developing countries. The percentile values corresponding to BMIs of 25 and $30 \mathrm{~kg} / \mathrm{m}^{2}$ at age 18 years were used to derive sex- and age-specific cut-off points for the categorical analysis of overweight (see Cole et al. 2000). In the United States, CDC have also produced reference curves, but these are based arbitrarily on the 85 th and 95 th percentiles of carefully selected nationally representative data (Ogden et al. 2002). The data for this chapter are presented according to the IOTF system, Roberts and Dallal (2001) having concluded that the IOTF reference levels were more suitable for international comparisons.

## 4. Estimating mean BMI And prevalences of OVERWEIGHT AND OBESITY

Given that representative data were not available for all 191 countries, it proved necessary to undertake a number of extrapolations. Some of the principal approaches are set out below.

### 4.1 Description of subregional availability of data

AFR-D
Childhood mean BMI data were from Mali and Senegal. In adults, subregional estimates were based on data from Cameroon, the Gambia, Ghana, Nigeria, Mali and Senegal. Data were also obtained for the Seychelles but were not included in the estimate. Mean BMI data are outlined in Table 8.1.

## AFR-E

In children, subregional mean BMI data were from Ethiopia, South Africa and Zimbabwe. In adults, estimates were based on data from Ethiopia, Kenya, Malawi, South Africa, the United Republic of Tanzania and Zimbabwe. Data available from Kenya and the United Republic of Tanzania were limited to females only. Mean BMI data are outlined in Table 8.2.

## AMR-A

For children, subregional mean BMI data were derived from both Canada and the United States. Although the Canadian data were not nationally representative, it was felt that these data should be used until nationally representative data become available in the required format. In adults, data were available from all countries in the subregion. Data from Cuba were provided in different age categories and were adjusted

Table 8.I Mean BMI in AFR-D

| Country (reference) | Sex | Mean BMI (kg/m ${ }^{2}$ ) Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| Cameroon ${ }^{2}$ (Rotimi et al. 1995) | Male | - | 23.7 | 24.4 | 24.0 | - | - | - |
|  | Female | - | 24.6 | 24.8 | 25.0 | - | - | - |
| Gambia (Van der Sande et al. 1997) | Male | - | 19.6 | 20.5 | 20.9 | 21.0 | 20.0 | - |
|  | Female | - | 21.0 | 21.9 | 21.8 | 21.3 | 20.9 | - |
| Ghana (DHS data provided by Macro International 1998) | Male | - | - | - | - | - | - | - |
|  | Female | - | 21.8 | 22.4 | 21.4 | - | - | - |
| Mali (Re-analysed by | Male | 14.8 | 18.9 | 20.5 | 20.8 | 20.3 | 19.6 | 20.2 |
| Ferro-Luzzi, personal communication) | Female | 14.9 | 19.9 | 21.1 | 20.6 | 20 | 19.5 | 20.8 |
| Nigeria (Okesina et al. 1999) | Male | - | 19.8 | 20.9 | 21.5 | - | - | - |
|  | Female | - | 21.0 | 21.8 | 20.3 | - | - | - |
| Senegal (Re-analysed by Ferro-Luzzi, personal communication) | Male | 14.2 | 18.2 | 19.9 | 21.0 | 20.7 | 19.8 | 19.2 |
|  | Female | 14.3 | 19.6 | 21.4 | 22.1 | 22.2 | 21.3 | 20.7 |
| Seychelles (Bovet et al. 1991) | Male | - | 22.9 | 23.5 | 23.1 | 23.2 | - | - |
|  | Female | - | 23.2 | 25.7 | 27.2 | 27.5 | - | - |

- No data.
a Data provided for Cameroon were estimated from graphs as actual figures were not available.
according to the methodology outlined in section 4.4. Mean BMI data are outlined in Table 8.3.

AMR-B
Data on mean BMI in childhood were derived using findings from Argentina, Brazil and Mexico. The standard deviation for mean BMI in children was not available. To estimate standard deviation, the methodology outlined in section 4.6 was applied to data from EMR-B. For adults, subregional data were taken from Argentina, Barbados, Brazil, Mexico and Paraguay. Data were also obtained for Saint Lucia but were not included in the subregional analysis. Mean BMI data are outlined in Table 8.4.

AMR-D
Limited data were available for this subregion and neither population sample presented for children was considered to be nationally representative. However, it was considered inappropriate to exclude these data and to extrapolate from other subregions. For Guatemala, only data from children living in high altitude areas were considered in the calculations. Data for adults were only available for females. Subregional estimates for females were used to derive estimates for males, using

Table 8.2 Mean BMI in AFR-E

| Country (reference) | Sex | Mean BMI (kg/m ${ }^{2}$ ) <br> Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| Ethiopia (Re-analysed by | Male | 14.2 | 17.5 | 18.3 | 18.0 | 18.0 | 17.9 | 19.8 |
| Ferro-Luzzi, personal communication) | Female | 14.5 | 18.9 | 18.6 | 17.3 | 16.7 | 17.6 | 18.6 |
| Kenya (DHS data provided | Male | - | - | - | - | - | - | - |
| by Macro International, 1998) | Female | - | 21.7 | 22.3 | 22.0 | - | - | - |
| Malawi (Chilima and Ismail | Male | - | - | - | 19.8 | 19.8 | 19.7 | - |
| 1998) | Female | - | - | - | 20.5 | 20.5 | 19.6 | - |
| South Africa (T. Puoane | Male | 13.8 | 21.5 | 24.2 | 25.3 | 24.8 | 24.4 | - |
| et al. I998, unpublished document) ${ }^{\text {a }}$ | Female | 14.0 | 24.4 | 28.5 | 29.9 | 28.8 | 27.7 | - |
| United Republic of Tanzania | Male | - | - | - | - | - | - | - |
| (DHS data provided by Macro International, 1996) | Female | - | 21.8 | 22.3 | 21.6 | - | - | - |
| Zimbabwe (Re-analysed by | Male | 15.3 | 19.5 | 20.8 | 21.0 | 21.0 | 20.1 | 20.0 |
| Ferro-Luzzi, personal communication) | Female | 15.4 | 21.3 | 23.0 | 23.5 | 21.8 | 20.5 | 20.3 |

- No data.
a Anthropometric patterns in South Africa: results from the National Demographic and Adult Health Survey 1998.


## Table 8.3 Mean BMI in AMR-A

| Country (reference) | Sex | Mean BMI (kg/m ${ }^{2}$ ) Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| Canada (Hanley et al. 2000; ${ }^{\text {a }}$ | Male | 19.1 | 23.7 | 25.6 | 26.8 | 26.6 | 26.3 | - |
| Macdonald et al. $1997{ }^{\text {b }}$ ) | Female | 20.2 | 23.2 | 24.1 | 26.3 | 26.7 | 26.4 | - |
| Cuba (Provided by C. | Male | - | 22.2 | 23.5 | 23.5 | - | - | - |
| Nishida, personal communication, 1992) ${ }^{\text {b }}$ | Female | - | 22.4 | 24.3 | 25.4 | - | - | - |
| USA (NHANES) ${ }^{\text {b }}$ | Male | 18.5 | 24.2 | 26.6 | 27.8 | 27.5 | 26.8 | 25.1 |
|  | Female | 18.6 | 24.0 | 26.4 | 28.0 | 27.6 | 27.0 | 25.0 |

- No data.
a Childhood data.
b Adult data.

Table 8.4 Mean BMI in AMR-B

| Country (reference) | Sex | Mean BMI (kg/m ${ }^{2}$ ) <br> Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| Argentina (Hernandez et al. 1987) | Male | 17.1 | 23.6 | 25.4 | 27.4 | 27.8 | 26.6 | - |
|  | Female | 18.0 | 22.3 | 24.0 | 26.3 | 26.7 | 26.5 | - |
| Barbados ${ }^{\text {a }}$ (Rotimi et al. 1995) | Male | - | 25.5 | 25.5 | 26.4 | - | - | - |
|  | Female | - | 28.0 | 28.3 | 29.2 | - | - | - |
| Brazil (Monteiro and Conde 1999) | Male | 16.9 | 22.1 | 23.8 | 24.2 | 24.1 | - | - |
|  | Female | 17.4 | 23.0 | 25.4 | 26.3 | 26.3 | - | - |
| Mexico (Arroyo et al. 2000; C.P. Sánchez Castillo, personal communication, 2002 ${ }^{\text {b }}$ ) | Male | 19.2 | 24.6 | 27.2 | 27.6 | 27.0 | 25.7 | 24.7 |
|  | Female | 19.8 | 25.3 | 28.3 | 29.6 | 28.8 | 27.3 | 25.5 |
| Paraguay ${ }^{\text {a }}$ (Jimenez et al. 1998) | Male | - | 22.3 | 25.2 | 25.1 | 23.1 | - | - |
|  | Female | - | 21.9 | 27.0 | 29.4 | 27.6 | - | - |
| Saint Kitts and Nevis, and Saint Lucia ${ }^{\mathrm{a}, \mathrm{b}}$ (Rotimi et al. 1995) | Male | - | 23.5 | 23.8 | 24.2 | - | - | - |
|  | Female | - | 26.0 | 26.6 | 27.2 | - | - | - |
| - No data. |  |  |  |  |  |  |  |  |
| a Data for Barbados, Paraguay, Saint Kitts and Nevis, and Saint Lucia were estimated from graphs as actual figures were not available. |  |  |  |  |  |  |  |  |
| b Childhood data. |  |  |  |  |  |  |  |  |

Table 8.5 Mean BMI in AMR-D

| Country (reference) | Sex | Mean BMI (kg/m ${ }^{2}$ ) <br> Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| Guatemala (DHS 1998; | Male | 14.6 | - | - | - | - | - | - |
| Martorell et al. 1995 ${ }^{\text {b }}$ | Female | 14.9 | 24.4 | 25.8 | 26.8 | - | - | - |
| Peru (DHS data provided by | Male | 15.2 | - | - | - | - | - | - |
| Macro International, 1996; ${ }^{\text {a }}$ Gonzales et al. 1994) | Female | 15.3 | 24.4 | 25.8 | 26.5 | - | - | - |
| - No data. |  |  |  |  |  |  |  |  |
| a Adult data. |  |  |  |  |  |  |  |  |
| b Childhood data. |  |  |  |  |  |  |  |  |

methodology outlined in section 4.4. The methodology outlined in section 4.4 was also used to obtain estimates for upper age categories. The original mean BMI data for females are outlined in Table 8.5.

EMR-B
In children, subregional data were from Bahrain, Lebanon and Saudi Arabia. In adults, regional estimates were derived from Bahrain, Cyprus,

Table 8.6 Mean BMI in EMR-B

| Country (reference) | Sex | Mean BMI (kg/m ${ }^{2}$ ) <br> Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| Bahrain (al-Mannai et al. 1996;, ${ }^{\text {a,b }}$ Musaiger and al-Mannai 200I a,c Musaiger and Gregory $2000^{\text {d }}$ ) | Male | 16.0 | 22.9 | 27.2 | 26.7 | 25.1 | - | - |
|  | Female | 16.9 | 24.6 | 30.4 | 30.0 | 29.9 | - | - |
| Iran (Islamic Republic of) (Pishad 1996) ${ }^{\text {a }}$ | Male | - | 21.0 | 23.8 | 24.3 | 23.1 | 22.7 | - |
|  | Female | - | 21.8 | 24.7 | 25.0 | 24.1 | 22.5 | - |
| Jordan (Ajlouni et al. 1998) ${ }^{\text {a }}$ | Male | - | 24.9 | 27.0 | 28.5 | 28.0 | 26.2 | - |
|  | Female | - | 26.3 | 30.8 | 32.5 | 31.9 | 30.1 | - |
| Kuwait (al-Isa 1995) ${ }^{\text {a }}$ | Male | - | 26.7 | 28.3 | 28.6 | 25.0 | - | - |
|  | Female | - | 26.7 | 30.3 | 31.5 | 29.8 | - | - |
| Lebanon (Data re-analysed by N. Hwalla and N. Adra, 1996) ${ }^{\text {a,b }}$ | Male | 17.8 | 23.5 | 25.8 | 26.7 | 26.1 | 25.5 | 23.5 |
|  | Female | 17.8 | 22.3 | 25.4 | 28.1 | 29.2 | 27.2 | 26.0 |
| Saudi Arabia (al-Nuaim et al. 1996) ${ }^{\text {a }}$ | Male | - | 23.5 | 26.1 | 27.0 | 26.0 | - | - |
|  | Female | 21.0 | 24.5 | 27.6 | 28.7 | 27.0 | - | - |
| United Arab Emirates <br> (el Mugamer et al. 1995) ${ }^{\text {a }}$ | Male | - | 24.7 | 25.6 | 26.5 | 24.6 | - | - |
|  | Female | - | 26.8 | 27.8 | 28.8 | 25.4 | - | - |
| - No data. |  |  |  |  |  |  |  |  |
| a Adult data. |  |  |  |  |  |  |  |  |
| b 18-29 years. |  |  |  |  |  |  |  |  |
| c >30 years. |  |  |  |  |  |  |  |  |
| ${ }^{\text {d }}$ Childhood data. |  |  |  |  |  |  |  |  |

the Islamic Republic of Iran, Jordan, Kuwait, Lebanon, Saudi Arabia and the United Arab Emirates. Only data for Lebanon were provided in the appropriate age categories; all other data were subject to the methodology outlined in section 4.4. Mean BMI data are outlined in Table 8.6.

## EMR-D

Childhood data were not available in this subregion in the required format. Using the methodology outlined in section 4.5 , it was concluded that it would be most appropriate to use data from the AFR-E subregion in order to determine the estimates for children. Mean BMI data for adults were limited to Pakistan and Egypt (females only). No data were available for the $\geq 80$ years age category, therefore the methodology outlined in section 4.4 was applied. Mean BMI data are outlined in Table 8.7.

## EUR-A

Availability of mean BMI data is described in Table 8.8.

Table 8.7 Mean BMI for adults in EMR-D

|  |  | Mean BMI (kg/m²) <br> Age group (years) |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
|  | Country (reference) | Sex | $15-29$ | $30-44$ | $45-59$ | $60-69$ | $70-79$ | $\geq 80$ |
| Egypt (DHS, data provided by | Male | - | - | - | - | - | - |  |
| Macro International, I992-1995) | Female | 25.3 | 27.2 | 27.1 | - | - | - |  |
| Pakistan (Data provided by Dr | Male | 20.7 | 21.8 | 21.9 | 21.6 | 21.0 | - |  |
| Habibullah, 1998) | Female | 21.1 | 22.5 | 22.7 | 22.3 | 21.3 | - |  |
| No data. |  |  |  |  |  |  |  |  |

EUR-B
In children, data on mean BMI were available from Bulgaria, Poland, Slovakia and Turkey. In adults, data on mean BMI were available from Romania, Slovakia, Tajikistan, Turkey and Uzbekistan, and are outlined in Table 8.9.

EUR-C
Limited data were available for children in EUR-C, from the Russian Federation only. In adults, data were available for Hungary, Latvia, Lithuania and the Russian Federation and are outlined in Table 8.10.

SEAR-B
There were no data available for children in this subregion, thus data from AFR-E were used, as specified in section 4.5. As no data for adults were available for Indonesia or Sri Lanka, subregional figures were based on data from Thailand, as outlined in Table 8.11. Normally these data would have been excluded as they came from attendees at a dental clinic. However, as no other data were available in the required format at the time, it was decided that they should be included in the absence of any more appropriate data.

SEAR-D
Data for children were available from Nepal. Data for adults were available for Bangladesh, India (females only) and Nepal. Mean BMI values available are shown in Table 8.12.

WPR-A
Data were available for both children and adults for Australia and Japan, as outlined in Table 8.13.

WPR-B
As no data were available for children in WPR-B, data from AFR-E were used as specified in section 4.5. Data for adults were available for China,

Table 8.8 Mean BMI in EUR-A


Table 8.9 Mean BMI in EUR-B

| Country (reference) | Sex | Mean BMI (kg/m ${ }^{2}$ ) <br> Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| Bulgaria (Data re-analysed by S. Petrova, 1998) ${ }^{\text {b }}$ | Male <br> Female | $\begin{aligned} & 18.0 \\ & 17.9 \end{aligned}$ | - | - | - | - | - | - |
| Poland (Data re-analysed by I. Palczewska, 1999) ${ }^{\text {b }}$ | Male Female | $\begin{aligned} & 17.6 \\ & 17.3 \end{aligned}$ | — | - | - | - | - | - |
| Romania (N. Hâncu, personal communication, 1999) ${ }^{\text {a }}$ | Male <br> Female | - | $\begin{aligned} & 23.9 \\ & 22.5 \end{aligned}$ | $\begin{aligned} & 26.4 \\ & 26.9 \end{aligned}$ | $\begin{aligned} & 27.1 \\ & 28.0 \end{aligned}$ | - | - | - |
| Slovakia (K. Babinska, personal communication, 1995-1999); ${ }^{\text {a }}$ (Data collated by A. Bederova and re-analysed by K. Babinska, 1995-1999) ${ }^{\text {b }}$ | Male <br> Female | $\begin{aligned} & 18.0 \\ & 18.2 \end{aligned}$ | $\begin{aligned} & 21.7 \\ & 20.8 \end{aligned}$ | $\begin{aligned} & 26.7 \\ & 24.5 \end{aligned}$ | $\begin{aligned} & 27.7 \\ & 27.0 \end{aligned}$ | $\begin{aligned} & 28.4 \\ & 28.9 \end{aligned}$ | $\begin{aligned} & 26.8 \\ & 28.2 \end{aligned}$ | - |
| Tajikistan (A. Robertson, personal communication, 1998) ${ }^{\text {a }}$ | Male <br> Female | — | $\begin{aligned} & 17.8 \\ & 18.0 \end{aligned}$ | $\begin{aligned} & 20.8 \\ & 20.4 \end{aligned}$ | $\begin{aligned} & 25.4 \\ & 25.5 \end{aligned}$ | $\begin{aligned} & 26.8 \\ & 27.3 \end{aligned}$ | - | - |
| Turkey (Data re-analysed by G. Pekcan and N. Rak, 1993-1999) | Male <br> Female | $\begin{aligned} & 16.7 \\ & 16.8 \end{aligned}$ | 21.9 24.0 | 25.6 27.7 | 26.3 29.6 | 26.3 | 25.6 29.2 | 24.8 27.5 |
| Uzbekistan (A. Robertson, personal communication, 1999) ${ }^{2}$ | Male Female | - | $\begin{aligned} & 20.7 \\ & 20.2 \end{aligned}$ | $\begin{aligned} & 21.2 \\ & 20.1 \end{aligned}$ | $\begin{aligned} & 22.5 \\ & 22.0 \end{aligned}$ | — | - | - |
| - No data. <br> a Adult data. <br> b Childhood data. |  |  |  |  |  |  |  |  |

Table 8.IO Mean BMI in EUR-C

| Country (reference) | Sex | Mean BMI (kg/m ${ }^{2}$ ) <br> Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| Hungary (Zajkas and Biro | Male | - | 24.0 | 26.8 | 28.1 | 28.4 | 25.4 | 25.0 |
| 1998) | Female | - | 22.4 | 25.4 | 28.2 | 29.8 | 27.0 | 27.7 |
| Latvia (A. Robertson, personal communication, 1997) | Male | - | 25.6 | 27.0 | 29.0 | 29.2 | - | - |
|  | Female | - | 22.1 | 24.2 | 27.6 | 28.8 | - | - |
| Lithuania (A. Robertson, personal communication, 1999) | Male | - | 24.8 | 25.7 | 26.5 | 26.8 | - | - |
|  | Female | - | 23.2 | 24.9 | 27.6 | 28.7 | - | - |
| Russian Federation (Data re-analysed by AD Deev, Russian Longitudinal Monitoring Survey-RLMS 1992) | Male | 18.1 | 22.8 | 25.1 | 25.9 | 25.6 | 25.2 | 24.8 |
|  | Female | 17.7 | 22.8 | 26.6 | 28.5 | 28.6 | 27.2 | 25.2 |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

[^27]Table 8.II Mean BMI for adults in SEAR-B

|  |  | Mean BMI $\left(\mathrm{kg} / \mathrm{m}^{2}\right)$ <br> Age group (years) |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Country (reference) | Sex | $15-29$ | $30-44$ | $45-59$ | $60-69$ | $70-79$ | $\geq 80$ |
| Thailand (Chaichareon et al. | Male | 20.8 | 22.6 | 23.4 | 23.0 | 22.6 | 22.6 |
| 1992) | Female | 20.8 | 22.7 | 23.9 | 24.3 | 22.5 | 22.5 |

Table 8.I2 Mean BMI in SEAR-D

| Country (reference) | Sex | Mean BMI (kg/m ${ }^{2}$ ) <br> Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| Bangladesh (M. Q-K and K. | Male | - | 19.0 | 19.5 | 18.9 | 19.1 | 18.0 | - |
| Talukder, personal communication, 2000) | Female | - | 19.7 | 19.9 | 19.2 | 18.7 | 19.7 | - |
| India (DHS data provided | Male | - | - | - | - | - | - | - |
| by Macro International, 1998) | Female | - | 19.5 | 20.9 | 21.6 | - | - | - |
| Nepal (Data re-analysed by | Male | 14.4 | 19.0 | 20.1 | 19.8 | 19.6 | 18.2 | 19.4 |
| A. Ferro-Luzzi personal communication, (997) ${ }^{\text {a,b }}$ | Female | 18.5 | 20.4 | 20.7 | 20.2 | 19.6 | 19.1 | 16.0 |

- No data.
a Adult data.
b Childhood data.

Table 8.13 Mean BMI in WPR-A

|  |  | Mean BMI (kg/m <br> Age <br> group (years) |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
|  | Country (reference) | Sex | $5-14$ | $15-29$ | $30-44$ | $45-59$ | $60-69$ | $70-79$ | $\geq 80$ |
| Australia (National | Male | 18.1 | 24.6 | 26.8 | 26.8 | 27.6 | 27.1 | 24.7 |  |
| Nutrition Survey 1995, data | Female | 18.6 | 22.5 | 24.6 | 27.1 | 27.2 | 26.4 | 25.4 |  |
|  <br> re-analysed by T. Gill, 2000) |  |  |  |  |  |  |  |  |  |
| Japan (Yoshiike et al. 1998) | Male | 17.6 | 21.8 | 23.0 | 23.4 | 22.7 | 22.3 | 21.5 |  |
|  | Female | 16.9 | 20.5 | 22.1 | 23.3 | 23.5 | 23.0 | 22.3 |  |

Table 8.14 Mean BMI in WPR-B

|  |  | Mean BMI (kg/m $\left.{ }^{2}\right)$ <br> Age group (years) |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Country (reference) | Sex | $15-29$ | $30-44$ | $45-59$ | $60-69$ | $70-79$ | $\geq 80$ |
| China | Male | 21.6 | 22.9 | 23.2 | 23.0 | 22.3 | - |
|  | Female | 22.7 | 23.0 | 23.7 | 23.9 | 22.6 | - |
| Malaysia (Khor et al. 1999) | Male | 21.6 | 23.6 | 23.0 | 21.4 | - | - |
|  | Female | 22.3 | 24.8 | 24.5 | 22.6 | - | - |
| Republic of Korea (Jones | Male | 22.5 | 22.8 | 23.2 | 21.8 | 21.6 | 20.8 |
| et al. I994) | Female | 21.1 | 21.8 | 23.0 | 22.4 | 22.2 | 21.1 |
| Samoa (McGarvey I99I) | Male | 24.9 | 25.8 | 28.0 | 27.7 | 26.5 | - |
|  | Female | 26.0 | 27.9 | 30.3 | 29.8 | 28.6 | - |
| Solomon Islands (Eason | Male | 22.9 | 23.5 | 23.6 | - | - | - |
| et al. I987) | Female | 24.4 | 24.9 | 24.2 | - | - | - |
| Viet Nam (Giay and Khoi | Male | 19.3 | 19.5 | 19.0 | 18.2 | - | - |
| I994) | Female | 19.8 | 19.4 | 18.6 | 17.8 | - | - |

- No data.

Malaysia, the Republic of Korea, Samoa, the Solomon Islands and Viet Nam, as outlined in Table 8.14.

### 4.2 Obtaining subregional estimates from COUNTRY-SPECIFIC ESTIMATES

In order to obtain subregional estimates of mean BMI, prevalences of overweight and obesity and their associated standard errors for sex- and age-specific categories, a meta-analysis was initially considered. In this approach, estimates from different studies would be combined into a single weighted estimate, using the variance of the estimate as the weight. The combined estimate of the variances would then be the inverse of the sum of the study-specific variances. There were two major drawbacks which made this approach unsuitable.

- Countries with unknown variances would be excluded from the subregional analysis for the estimate of mean BMI or prevalences of overweight and obesity.
- The method of weighting by the variance of the study (which is highly dependent on the sample size and the design of the study, with "better", larger studies having the smallest variances) assumes an equal population for each sample. The approach does not take into account differences in population sizes.

A second approach was to obtain a single estimate of mean BMI and the prevalences of overweight and obesity by using a populationweighted average. Standard deviations were estimated using standard statistical relationships. ${ }^{2}$

For simplicity, a simple average of the standard errors of the prevalences of overweight or obesity was used to obtain subregional estimates. In cases where no country data were available and the prevalences of overweight or obesity were based on predictions, the subregional standard error was that of the prediction.

### 4.3 CONVERTING MEAN BMI TO PROPORTIONS OF OVERWEIGHT and obesity and vice versa

BMI distributions are skewed in almost all age and sex categories throughout the world. Thus mean BMI with standard deviation does not accurately describe the whole BMI distribution and usually no further details of the distribution are available. However, Rose and Shipley (1990) showed that the prevalence of obesity ( $\mathrm{BMI} \geq 30 \mathrm{~kg} / \mathrm{m}^{2}$ ) was highly correlated with mean BMI in the selected group of adults, the prevalence of obesity increasing by $4.22 \%$ per unit $\left(1 \mathrm{~kg} / \mathrm{m}^{2}\right)$ increase in mean BMI. This approach was repeated in the present analysis, which was based on adult data (all ages $\geq 18$ years combined) from 36 countries with continuous and categorical data for females and 26 countries with continuous and categorical data for males (Table 8.15). The mean BMI values for countries used in this estimation varied widely from $18.1 \mathrm{~kg} / \mathrm{m}^{2}$ to $29.2 \mathrm{~kg} / \mathrm{m}^{2}$. Mali was excluded from the current analyses in males because it was an outlier. Mean sex-specific BMI vs percentages of overweight and obesity are shown in Figures 8.1 and 8.2.

There is a clear positive linear association for both sexes. The regression equations that describe the graphs are:

Figure 8.I Mean BMI and percentage of the adult population that is overweight


[^28]Table 8.15 Countries from which mean BMI data were used to derive equations for calculating the percentages of overweight and obese adults in the population

| For female overweight and obesity | For male overweight and obesity |
| :---: | :---: |
| Australia | Australia |
| Bangladesh | Bangladesh |
| Brazil | Brazil |
| China | China |
| Denmark | Denmark |
| Egypt | Ethiopia |
| Ethiopia | Germany |
| Germany | Hungary |
| Ghana | Iceland |
| Guatemala | Japan |
| Hungary | Kuwait |
| Iceland | Lebanon |
| Japan | Nepal |
| Kenya | Norway |
| Kuwait | Republic of Korea |
| Lebanon | Russian Federation |
| Malawi ${ }^{\text {a }}$ | Senegal |
| Mali | Slovakia |
| Nepal | Switzerland |
| Norway | Thailand |
| Peru | Turkey |
| Republic of Korea | United Kingdom |
| Russian Federation | USA |
| Senegal | Uzbekistan |
| Slovakia | Zimbabwe |
| South Africa ${ }^{\text {a }}$ |  |
| Switzerland |  |
| Thailand |  |
| Turkey |  |
| United Kingdom |  |
| United Republic of Tanzania |  |
| USA |  |
| Uzbekistan |  |
| Zimbabwe |  |

[^29]Figure 8.2 Mean BMI and percentage of the adult population that is obese


Note: The predicted relationship is shown, together with the $95 \%$ confidence intervals of the prediction. The quadratic equations are given below.

Males $(\%$ overweight $)=-110.4+5.7 \times($ mean BMI)
Females $(\%$ overweight $)=-74.2+3.97 \mathrm{x}($ mean BMI)
In both cases, the $\beta$ coefficient was highly statistically significant and the models accounted for $>90 \%$ of the total variation. There was also a strong correlation between mean BMI and percentages of obese adults in the population, as seen in Figure 8.2.

Males $(\%$ obese $)=205.1-20.4 \times($ mean BMI $)+0.5 \times(\text { mean BMI })^{2}$
Females $(\%$ obese $)=168.5-17.4 \times($ mean BMI $)+0.4 \times(\text { mean BMI })^{2}$
The following conditions applied:

- the predictions had to be positive (since they are percentages);
- the predictions must be $<100$; and
- the sum of the predicted percentages of people in the overweight and obese categories must be $\leq 100$ (the sum is equal to 100 in the extreme case whereby no individuals belong to the underweight or normal BMI categories).

These conditions hold simultaneously for mean BMIs of $21.3-29.7 \mathrm{~kg} / \mathrm{m}^{2}$ for males and $20.1-33.9 \mathrm{~kg} / \mathrm{m}^{2}$ for females. Predictions which fell outside these ranges were therefore not considered. These four models (Equations 1-4) for predicting the prevalences of overweight and obesity from mean BMIs were then applied to other populations when necessary.

Finally, the assumption was made that the equations held true for each age group and every country, so that the estimates could be
derived uniformly. This assumption seemed justified because wherever data were available the patterns of mean BMI and percentages of overweight and obesity were similar across age groups in most of the countries assessed.

### 4.4 Age and sex extrapolation

## Obtaining estimates for the required age groups

Apart from data personally donated to or re-analysed by IOTF, all other data were reported in different age categories from those required for this work. To obtain data in these age categories, it was assumed that:

- the numbers of persons in each year within an age group were the same and equal to the total number of people in the age group divided by the number of years in the age group.
- the mean BMI and the standard deviation for each year within an age group were the same and equal to the mean BMI and standard deviation in the age group as a whole.
Single years or convenient groups of years were treated as different strata which were then combined to obtain the desired estimates in any age categorization.


## EXtrapolation to adult age groups for which no data WERE AVAILABLE

It was not always possible to obtain data for all age groups, particularly for the oldest (70-79 and $\geq 80$ years) at a subregional level. WPR-B, where this problem was initially encountered for both sexes in people aged $>70$ years, is used as an example, although data subsequently became available for this subregion.

In the majority of the available data worldwide, the mean BMI increased with age and then started falling with rising age. Figure 8.3 shows the relationship between mean BMI and age in WPR-B. It was assumed that the mean BMI remained constant within each age group and changed only when moving from one age group to another.

The following regression equations describe the graphs:

$$
\begin{aligned}
& \text { Females }(\text { Mean BMI })=16.71+0.27 \times(\text { age })-0.0024 \times\left(\text { age }^{2}\right) \\
& \text { Males }(\text { Mean } B M I)=16.62+0.14 \times(\text { age })-0.0011 \times\left(\text { age }^{2}\right)
\end{aligned}
$$

Using these equations, the sex-specific mean BMI for each single year from age 70 years and above could be predicted. The overall mean BMI in the age group 70-79 years was set equal to the average of the mean BMIs for the individual years. The same procedure was used to estimate the mean BMI in the age group $\geq 80$ years. The results are shown in Table 8.16. The illustrated approach was used for EMR-B.

Table 8.16 Predicted mean BMI values in the oldest age groups in WPR-B

|  | Mean BMI $\left(\mathrm{kg} / \mathrm{m}^{2}\right)$ |  |
| :--- | :---: | :---: |
| Age group (years) | Females | Males |
| $70-79$ | 23.9 | 24.2 |
| $\geq 80$ | 22.8 | 23.8 |

Figure 8.3 The relationship between mean BMI and age in WPR-B


## Extrapolating data from one sex to the other

Many countries reported results either for both sexes combined or for one sex only. For example, on a subregional basis, no suitable data were available for males in AMR-D. The crude mean BMI for females in all subregions (calculated as the mean of the subregion-specific estimates for all ages combined) was $23.7 \mathrm{~kg} / \mathrm{m}^{2}$ and was 0.4 units greater than the crude mean BMI for men in all subregions ( $P=0.7$ ). To determine whether this was the case in each age group, the age-specific differences in mean BMI between females and males were estimated for all subregions. For all age groups apart from the oldest (i.e. $\geq 80$ years), the mean BMI for females was greater than that for males, as shown in Table 8.17.

These values served as correction factors when using data from males to estimate mean BMI for females and vice versa (without considering whether the differences between mean BMI for males and for females were significant). Thus, for each age group the respective correction factor was subtracted from mean BMI values for females to obtain the values for males. In most studies, it is found that women tend to have

# Table 8.17 Age-dependent differences between mean BMI for females and for males 

|  | Age group (years) |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $18-29$ | $30-44$ | $45-59$ | $60-69$ | $70-79$ | $\geq 80$ |
| Differences in mean BMI <br> (females-males) | 0.3 | 0.5 | 0.7 | 0.8 | 0.3 | -0.1 |

higher BMIs than men and this seems to be related to the deposition of fat rather than metabolically active lean tissue with body-weight gain. Women lay down a lower proportion of lean tissue and thus have to put on more weight before the slower increase in the mass of lean tissue raises the basal metabolic rate sufficiently to add to the exercise costs of their greater weight and achieve an energy output which finally matches their energy intake (James and Reeds 1997).

### 4.5 Estimating Childhood data

When no data were available for children for any of the countries within a subregion, an estimate was made of the distribution of BMI in children for that subregion by extrapolating from subregions with data and having equivalent economies, as judged by their gross national products (GNP). In the absence of any other data, WPR-A was extrapolated for AMR-A.

In general, plotting the mean age-standardized BMI for adults for each subregion (calculated as a simple average of population-weighted means for both sexes and for all ages $\geq 18$ years) vs GNP (on a subregional basis) shows a broad relationship between increasing GNP and mean BMI in subregions with GNPs of $<$ US $\$ 10000$ per year (Figure 8.4). For countries with the lowest GNPs, the assumption was made that BMIs for children would be similar in subregions with low GNP and low mean BMI for adults. This is subject to large uncertainty, given the complex underlying factors that determine body weight and height.

At the time of writing, no data were available for children in EMRD and WPR-B. The GNPs of countries in EMR-D were similar to those in AFR-E; the average GNP in EMR-D was US\$ 652 (with a range of US $\$ 110-1290$ ), whereas the average GNP in AFR-E was US $\$ 647$, (with a range of US\$80-3520). Mean BMI values were also similar in both populations, with values being slightly higher in EMR-D. The distribution of BMIs in children in AFR-E was therefore applied to EMR-D. The distributions of GNP and BMI for adults in SEAR-B and WPR-B were also similar, but there were no data for children, so the AFR-E values were used for these subregions, AFR-E being the closest to these subregions economically and also in terms of BMI.

Further problems arose when the mean BMIs for children were available, but not the standard deviations, as in AMR-B. The standard

Figure 8.4 The relationship between mean BMI for adults and GNP in nine subregions

deviations of BMIs for children in AMR-B were therefore also assumed to have the same variability as in EMR-B.

Very little in the way of categorical data was available for children and it was therefore necessary to extrapolate extensively, pending appropriate data becoming available.

### 4.6 The estimated mean BMI and standard deviation for EACH SUBREGION, BY AGE AND SEX

These data are presented in the standard format proposed for the CRA project. The mean BMIs for each year for ages 5-17 years inclusive for males and females separately were obtained and are given in Tables 8.18 and 8.19.

The initial BMI analyses were made as previously, considering children in 1-year age groups, firstly for the countries with data. From these values, the mean BMIs of the 1-year age groups within the subregion were estimated. Assuming equivalent numbers of people in each year of each specified age group, it was possible to provide mean BMIs, standard deviations and confidence intervals for the age groups 514 years and 15-29 years used in the CRA analyses. It was recognized that these values could not be used in the usual way to predict different categories of excess weight because of the normal changes in body weight
Table 8.18 The mean BMIs for children and adults in all subregions, by sex and age ${ }^{\text {a }}$

|  |  | Mean BMI (kg/m ${ }^{2}$ ) <br> Age group (years) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Subregion | Sex | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 14.2 | 14.0 | 14.1 | 14.1 | 14.5 | 14.6 | 14.7 | 14.9 | 15.1 | 15.7 | 16.1 | 16.3 | 16.9 | 20.1 | 21.2 | 21.7 | 20.5 | 20.5 | 20.0 |
|  | Female | 13.9 | 13.6 | 13.8 | 14.0 | 14.0 | 14.5 | 15.1 | 15.1 | 15.9 | 17.0 | 17.2 | 19.0 | 18.9 | 21.3 | 22.1 | 21.0 | 21.0 | 20.2 | 20.8 |
| AFR-E | Male | 14.1 | 13.8 | 13.8 | 13.9 | 13.9 | 14.1 | 14.5 | 16.2 | 16.1 | 16.7 | 17.5 | 17.8 | 18.1 | 19.6 | 21.0 | 21.2 | 20.8 | 22.0 | 19.8 |
|  | Female | 13.9 | 13.6 | 13.6 | 13.8 | 14.1 | 14.1 | 14.6 | 16.9 | 17.0 | 17.8 | 18.7 | 19.4 | 20.2 | 21.7 | 22.9 | 22.8 | 22.3 | 20.4 | 18.9 |
| AMR-A | Male | 15.9 | 16.4 | 16.9 | 17.2 | 18.2 | 18.5 | 19.4 | 20.0 | 20.5 | 22.5 | 22.2 | 22.5 | 23.4 | 24.5 | 26.4 | 27.6 | 27.4 | 26.7 | 25.1 |
|  | Female | 16.0 | 16.0 | 17.3 | 17.2 | 18.3 | 18.6 | 19.8 | 20.1 | 22.1 | 22.3 | 22.2 | 22.9 | 23.0 | 24.1 | 26.1 | 27.7 | 27.5 | 26.9 | 25.0 |
| AMR-B | Male | 15.4 | 15.6 | 15.7 | 16.0 | 16.3 | 16.7 | 17.0 | 17.5 | 18.2 | 18.8 | 19.4 | 20.1 | 20.4 | 23.7 | 25.0 | 25.6 | 25.5 | 26.1 | 24.7 |
|  | Female | 15.2 | 15.6 | 16.0 | 16.2 | 16.6 | 17.0 | 17.5 | 18.6 | 19.8 | 20.4 | 20.8 | 21.5 | 21.7 | 24.1 | 26.2 | 27.3 | 27.1 | 26.9 | 25.5 |
| AMR-D | Male | 15.8 | 15.5 | 14.9 | 15.3 | 15.2 | 15.1 | 15.9 | 16.0 | 17.1 | 17.6 | 18.8 | 19.4 | 20.1 | 24.1 | 25.3 | 25.9 | 26.0 | 26.3 | 26.3 |
|  | Female | 15.6 | 15.3 | 15.2 | 15.1 | 14.9 | 14.7 | 16.0 | 16.7 | 18.9 | 19.2 | 20.5 | 21.7 | 22.1 | 24.4 | 25.8 | 26.6 | 26.8 | 26.6 | 26.2 |
| EMR-B | Male | 15.4 | 15.7 | 16.9 | 16.7 | 17.9 | 17.4 | 17.6 | 19.7 | 19.5 | 21.3 | 20.8 | 22.3 | 23.8 | 21.9 | 24.6 | 25.3 | 24.3 | 23.1 | 23.5 |
|  | Female | 15.2 | 15.4 | 16.4 | 17.2 | 16.3 | 18.0 | 18.9 | 19.1 | 20.5 | 21.7 | 22.0 | 21.6 | 21.4 | 22.8 | 25.8 | 26.5 | 25.5 | 23.3 | 26.0 |
| EMR-D | Male | 14.1 | 13.8 | 13.8 | 13.9 | 13.9 | 14.1 | 14.5 | 16.2 | 16.1 | 16.7 | 17.5 | 17.8 | 18.1 | 20.7 | 21.8 | 21.9 | 21.6 | 21.0 | 20.1 |
|  | Female | 13.9 | 13.6 | 13.6 | 13.8 | 14.1 | 14.1 | 14.6 | 16.9 | 17.0 | 17.8 | 18.7 | 19.4 | 20.2 | 22.3 | 23.8 | 22.8 | 22.3 | 21.3 | 18. |


| EUR-A | Male | 16.2 | 16.6 | 16.5 | 16.7 | 16.9 | 17.3 | 18.1 | 18.6 | 19.6 | 19.8 | 21.0 | 22.2 | 22.5 | 24.7 | 26.3 | 27.2 | 27.8 | 27.5 | 26.1 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Female | 16.3 | 16.2 | 16.5 | 16.4 | 17.2 | 17.6 | 18.7 | 18.9 | 20.1 | 20.5 | 22.1 | 22.3 | 22.9 | 23.5 | 25.1 | 27.5 | 28.4 | 27.8 | 25.8 |
| EUR-B | Male | 15.1 | 15.5 | 15.5 | 16.0 | 16.7 | 17.2 | 18.6 | 18.4 | 18.6 | 19.3 | 19.7 | 20.4 | 20.6 | 22.1 | 25.0 | 26.5 | 27.3 | 25.8 | 24.8 |
|  | Female | 15.0 | 15.2 | 15.5 | 15.8 | 16.4 | 16.9 | 17.6 | 18.2 | 18.9 | 19.9 | 20.0 | 20.5 | 21.4 | 23.2 | 25.6 | 27.9 | 28.8 | 29.0 | 27.5 |
| EUR-C | Male | 16.4 | 15.7 | 17.0 | 17.4 | 17.1 | 17.8 | 18.0 | 18.8 | 19.6 | 20.5 | 20.6 | 21.4 | 21.6 | 23.3 | 25.2 | 26.1 | 25.9 | 25.2 | 24.8 |
|  | Female | 15.7 | 15.6 | 16.4 | 16.9 | 17.1 | 17.1 | 17.9 | 18.2 | 19.4 | 20.2 | 21.3 | 21.3 | 21.3 | 23.1 | 26.5 | 28.4 | 28.7 | 27.2 | 25.4 |
| SEAR-B | Male | 14.1 | 13.8 | 13.8 | 13.9 | 13.9 | 14.1 | 14.5 | 16.2 | 16.1 | 16.7 | 17.5 | 17.8 | 18.1 | 20.8 | 22.6 | 23.4 | 23.0 | 22.6 | 22.6 |
|  | Female | 13.9 | 13.6 | 13.6 | 13.8 | 14.1 | 14.1 | 14.6 | 16.9 | 17.0 | 17.8 | 18.7 | 19.4 | 20.2 | 20.8 | 22.7 | 23.9 | 24.3 | 22.5 | 22.5 |
| SEAR-D | Male | 13.8 | 14.3 | 13.9 | 13.6 | 14.1 | 14.2 | 14.4 | 15.2 | 14.8 | 15.8 | 16.3 | 17.2 | 17.7 | 19.0 | 19.6 | 19.0 | 19.2 | 18.0 | 19.4 |
|  | Female | 14.1 | 14.2 | 13.8 | 14.2 | 14.5 | 14.2 | 14.8 | 15.2 | 17.1 | 18.0 | 18.5 | 19.1 | 20.3 | 19.5 | 20.8 | 21.4 | 18.9 | 19.6 | 16.0 |
| WPR-A | Male | 16.3 | 15.6 | 15.9 | 16.5 | 16.8 | 17.6 | 18.1 | 19.2 | 19.9 | 20.4 | 20.9 | 21.4 | 21.8 | 22.4 | 23.7 | 23.9 | 23.3 | 22.8 | 21.9 |
|  | Female | 16.2 | 15.3 | 15.6 | 16.0 | 16.4 | 17.0 | 17.5 | 18.4 | 19.5 | 19.8 | 20.1 | 20.5 | 20.5 | 20.9 | 22.5 | 23.8 | 23.9 | 23.3 | 22.6 |
| WPR-B | Male | 14.1 | 13.8 | 13.8 | 13.9 | 13.9 | 14.1 | 14.5 | 16.2 | 16.1 | 16.7 | 17.5 | 17.8 | 18.1 | 21.5 | 22.8 | 23.1 | 22.8 | 22.3 | 20.8 |
|  | Female | 13.9 | 13.6 | 13.6 | 13.8 | 14.1 | 14.1 | 14.6 | 16.9 | 17.0 | 17.8 | 18.7 | 19.4 | 20.2 | 22.5 | 22.8 | 23.5 | 23.6 | 22.6 | 21.1 |
|  | analysed | 1-yea | age gr | ups in | hildhoo |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Note: Fip | in shad | cells | es | ated | outin | in th | etho | logy | tion. |  |  |  |  |  |  |  |  |  |  |  |

Table 8.19 The standard deviations of the BMIs for children and adults in all subregions, by sex and agea

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 1.5 | 1.3 | 1.9 | 1.5 | 1.1 | 1.2 | 1.7 | 1.3 | 1.5 | 1.6 | 1.7 | 2.0 | 2.0 | 2.0 | 3.4 | 5.4 | 2.9 | 2.9 | 2.7 |
|  | Female | 1.3 | 1.3 | 1.3 | 1.2 | 1.3 | 1.6 | 1.6 | 1.9 | 2.1 | 3.0 | 2.4 | 2.6 | 2.7 | 3.8 | 3.9 | 4.6 | 3.8 | 3.6 | 2.3 |
| AFR-E | Male | 1.4 | 1.2 | 1.2 | 1.2 | 1.1 | 1.4 | 1.0 | 3.0 | 2.1 | 2.3 | 2.4 | 2.4 | 2.4 | 3.1 | 5.2 | 4.6 | 4.2 | 3.6 | 1.8 |
|  | Female | 1.3 | 1.3 | 1.3 | 1.3 | 1.2 | 1.8 | 1.4 | 2.6 | 2.5 | 3.0 | 2.9 | 2.9 | 2.9 | 4.2 | 4.8 | 5.2 | 5.1 | 5.1 | 2.4 |
| AMR-A | Male | 1.7 | 2.5 | 2.9 | 4.2 | 4.3 | 4.3 | 4.4 | 4.5 | 4.5 | 4.7 | 4.7 | 4.7 | 4.8 | 4.8 | 5.0 | 5.0 | 4.4 | 4.3 | 4.1 |
|  | Female | 4.0 | 4.0 | 4.1 | 4.1 | 4.3 | 4.3 | 4.4 | 4.5 | 4.7 | 4.7 | 4.7 | 4.8 | 4.8 | 5.9 | 7.1 | 6.5 | 6.1 | 5.8 | 4.9 |
| AMR-B | Male | 1.5 | 1.4 | 2.3 | 1.9 | 2.8 | 4.1 | 3.8 | 3.7 | 3.7 | 3.9 | 3.6 | 3.5 | 3.2 | 4.1 | 4.1 | 4.3 | 4.0 | 3.9 | 3.8 |
|  | Female | 1.4 | 2.1 | 2.8 | 3.8 | 3.3 | 3.6 | 3.9 | 4.1 | 4.1 | 4.7 | 4.0 | 4.2 | 5.1 | 4.8 | 5.1 | 5.3 | 5.2 | 5.2 | 4.3 |
| AMR-D | Male | 1.1 | 1.0 | 1.0 | 1.3 | 1.3 | 1.3 | 1.4 | 1.3 | 1.2 | 1.4 | 1.5 | 1.5 | 1.6 | 2.8 | 3.8 | 4.0 | 3.5 | 3.5 | 3.5 |
|  | Female | 1.2 | 1.0 | 1.0 | 1.8 | 1.8 | 1.8 | 1.8 | 1.8 | 2.3 | 2.5 | 1.7 | 2.4 | 2.6 | 3.4 | 4.4 | 4.6 | 4.1 | 4.1 | 4.1 |
| EMR-B | Male | 1.5 | 2.0 | 2.9 | 2.1 | 3.0 | 2.4 | 2.9 | 4.5 | 4.0 | 3.4 | 3.8 | 3.2 | 3.7 | 3.8 | 4.1 | 3.9 | 4.0 | 3.7 | 3.6 |
|  | Female | 1.4 | 1.9 | 2.2 | 5.2 | 2.6 | 2.9 | 3.4 | 4.1 | 4.3 | 4.8 | 5.3 | 5.7 | 4.9 | 4.1 | 5.1 | 5.3 | 4.8 | 3.7 | 5.8 |
| EMR-D | Male | 1.4 | 1.2 | 1.2 | 1.2 | 1.1 | 1.4 | 1.0 | 3.0 | 2.1 | 2.3 | 2.4 | 2.4 | 2.4 | 5.6 | 6.9 | 4.8 | 5.8 | 5.8 | 5.8 |
|  | Female | 1.3 | 1.3 | 1.3 | 1.3 | 1.2 | 1.8 | 1.4 | 2.6 | 2.5 | 3.0 | 2.9 | 2.9 | 2.9 | 7.3 | 8.5 | 6.5 | 7.4 | 7.4 | 7.4 |

















| $\bar{\sim}$ | $\underset{\sim}{N}$ | $\stackrel{\alpha}{i} \stackrel{\infty}{\mathrm{i}}$ | さ̣ | $\stackrel{0}{0} \text { m }$ | $\underline{0}$ | 끈 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\underset{m}{\omega} \underset{\sim}{N}$ | $\bar{\sim}-$ | $\begin{array}{cc} \circ \\ \underset{\sim}{\sim} \end{array}$ | $\underset{~}{~!}$ | $\underset{O}{\infty} \underset{\sim}{-}$ | m = | ก̣ |
| 끈 | $\stackrel{O}{-}$ | $\underset{\sim}{n}$ | ㅍㅡㅡㅡㄹ | $\pm \underset{\circ}{\circ}$ | $\stackrel{\star}{\sim} \xrightarrow{\infty}$ | 프́ |


EUR-A
EUR-B
EUR-C
SEAR-B
SEAR-D
WPR-A
WPR-B

[^30]Note: Figures in shaded cells were estimated as outlined in the methodology section.

Table 8.20 Mean BMIs by subregion, sex and age

| Subregion | Sex | Mean BMI ( $\mathrm{kg} / \mathrm{m}^{2}$ ) Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 14.6 | 19.2 | 21.2 | 21.7 | 20.5 | 20.5 | 20.0 |
|  | Female | 14.6 | 20.6 | 22.1 | 21.0 | 21.0 | 20.2 | 20.8 |
| AFR-E | Male | 14.7 | 19.2 | 21.0 | 21.2 | 20.8 | 22.0 | 19.8 |
|  | Female | 14.9 | 21.2 | 22.9 | 22.8 | 22.3 | 20.4 | 18.9 |
| AMR-A | Male | 18.5 | 24.1 | 26.4 | 27.6 | 27.4 | 26.7 | 25.1 |
|  | Female | 18.8 | 23.8 | 26.1 | 27.7 | 27.5 | 26.9 | 25.0 |
| AMR-B | Male | 16.7 | 22.9 | 25.0 | 25.6 | 25.5 | 26.1 | 24.7 |
|  | Female | 17.3 | 23.5 | 26.2 | 27.3 | 27.1 | 26.9 | 25.5 |
| AMR-D | Male | 15.8 | 23.0 | 25.3 | 25.9 | 26.0 | 26.3 | 26.3 |
|  | Female | 16.1 | 23.7 | 25.8 | 26.6 | 26.8 | 26.6 | 26.2 |
| EMR-B | Male | 17.8 | 22.0 | 24.6 | 25.3 | 24.3 | 23.1 | 23.5 |
|  | Female | 17.9 | 22.5 | 25.8 | 26.5 | 25.5 | 23.3 | 26.0 |
| EMR-D | Male | 14.7 | 20.0 | 21.8 | 21.9 | 21.6 | 21.0 | 20.1 |
|  | Female | 14.9 | 21.6 | 23.8 | 22.8 | 22.3 | 21.3 | 18.9 |
| EUR-A | Male | 17.6 | 24.2 | 26.3 | 27.2 | 27.8 | 27.5 | 26.1 |
|  | Female | 17.9 | 23.3 | 25.1 | 27.5 | 28.4 | 27.8 | 25.8 |
| EUR-B | Male | 17.1 | 21.7 | 25.0 | 26.5 | 27.3 | 25.8 | 24.8 |
|  | Female | 17.0 | 22.7 | 25.6 | 27.9 | 28.8 | 29.0 | 27.5 |
| EUR-C | Male | 18.0 | 22.9 | 25.2 | 26.1 | 25.9 | 25.2 | 24.8 |
|  | Female | 17.6 | 22.7 | 26.5 | 28.4 | 28.7 | 27.2 | 25.4 |
| SEAR-B | Male | 14.7 | 20.2 | 22.6 | 23.4 | 23.0 | 22.6 | 22.6 |
|  | Female | 15.0 | 20.5 | 22.7 | 23.9 | 24.3 | 22.5 | 22.5 |
| SEAR-D | Male | 14.4 | 18.6 | 19.6 | 19.0 | 19.2 | 18.0 | 19.4 |
|  | Female | 15.0 | 19.5 | 20.8 | 21.4 | 18.9 | 19.6 | 16.0 |
| WPR-A | Male | 17.7 | 22.2 | 23.7 | 23.9 | 23.3 | 22.8 | 21.9 |
|  | Female | 17.2 | 20.8 | 22.5 | 23.8 | 23.9 | 23.3 | 22.6 |
| WPR-B | Male | 14.7 | 20.8 | 22.8 | 23.1 | 22.8 | 22.3 | 20.8 |
|  | Female | 15.0 | 21.9 | 22.8 | 23.5 | 23.6 | 22.6 | 21.1 |

Note: Figures in shaded cells were estimated as outlined in the methodology section.
and BMI during childhood development. The data on mean BMIs are presented in Table 8.20. Table 8.21 contains the standard deviations for these estimates, while Table 8.22 lists the number of subjects measured and used in this analysis, in order to provide a preliminary perspective on the validity of these estimates.

### 4.7 Final estimates of the prevalence of overweight and obesity, by age, sex and subregion

Table 8.23 shows the subregional prevalences of overweight, as defined by a BMI of between 25.0 and $29.9 \mathrm{~kg} / \mathrm{m}^{2}$. For the age groups $5-14$ years

Table 8.2I The standard deviations of the mean BMI estimates by subregion, sex and age

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 1.5 | 2.0 | 3.4 | 5.4 | 2.9 | 2.9 | 2.7 |
|  | Female | 1.6 | 3.5 | 3.9 | 4.6 | 3.8 | 3.6 | 2.3 |
| AFR-E | Male | 1.6 | 2.9 | 5.2 | 4.6 | 4.2 | 3.6 | 1.8 |
|  | Female | 1.7 | 3.9 | 4.8 | 5.2 | 5.1 | 5.1 | 2.4 |
| AMR-A | Male | 3.8 | 4.8 | 5.0 | 5.0 | 4.4 | 4.3 | 4.1 |
|  | Female | 4.3 | 5.7 | 7.1 | 6.5 | 6.1 | 5.8 | 4.9 |
| AMR-B | Male | 2.9 | 4.0 | 4.1 | 4.3 | 4.0 | 3.9 | 3.8 |
|  | Female | 3.4 | 4.7 | 5.1 | 5.3 | 5.2 | 5.2 | 4.3 |
| AMR-D | Male | 1.2 | 2.5 | 3.8 | 4.0 | 3.5 | 3.5 | 3.5 |
|  | Female | 1.7 | 3.1 | 4.4 | 4.6 | 4.1 | 4.1 | 4.1 |
| EMR-B | Male | 2.9 | 3.7 | 4.1 | 3.9 | 4.0 | 3.7 | 3.6 |
|  | Female | 3.3 | 4.4 | 5.1 | 5.3 | 4.8 | 3.7 | 5.8 |
| EMR-D | Male | 1.6 | 4.9 | 6.9 | 4.8 | 5.8 | 5.8 | 5.8 |
|  | Female | 1.8 | 6.3 | 8.5 | 6.5 | 7.4 | 7.4 | 7.4 |
| EUR-A | Male | 2.6 | 3.7 | 3.7 | 3.9 | 3.8 | 3.8 | 3.7 |
|  | Female | 2.6 | 4.1 | 4.7 | 4.7 | 5.2 | 4.9 | 4.3 |
| EUR-B | Male | 2.6 | 3.1 | 3.9 | 3.8 | 3.7 | 3.6 | 4.8 |
|  | Female | 2.5 | 3.9 | 5.2 | 6.0 | 5.1 | 4.7 | 4.8 |
| EUR-C | Male | 2.9 | 2.8 | 3.5 | 3.9 | 3.8 | 3.9 | 3.9 |
|  | Female | 2.8 | 3.9 | 5.0 | 5.1 | 5.2 | 5.1 | 5.0 |
| SEAR-B | Male | 1.6 | 2.2 | 2.8 | 2.3 | 3.9 | 2.5 | 2.5 |
|  | Female | 1.8 | 3.5 | 2.3 | 2.1 | 4.1 | 4.5 | 4.5 |
| SEAR-D | Male | 1.1 | 2.0 | 3.2 | 2.8 | 2.9 | 2.7 | 3.1 |
|  | Female | 1.6 | 3.0 | 4.0 | 4.6 | 3.7 | 5.9 | 2.0 |
| WPR-A | Male | 2.2 | 3.4 | 3.4 | 3.2 | 3.2 | 3.2 | 3.6 |
|  | Female | 1.7 | 3.5 | 4.1 | 3.7 | 3.7 | 3.8 | 3.7 |
| WPR-B | Male | 1.6 | 2.6 | 3.3 | 3.1 | 3.3 | 3.6 | 1.8 |
|  | Female | 1.8 | 4.2 | 4.1 | 3.6 | 5.0 | 3.7 | 1.9 |

Note: Figures in shaded cells were estimated as outlined in the methodology section.
and 15-29 years, these prevalence figures can be used in their condensed form because allowances have already been made for the age- and sexspecific cut-off points in the BMI percentiles corresponding to a BMI of $25.0 \mathrm{~kg} / \mathrm{m}^{2}$ at age 18 years. The rather crude nature of the calculation of BMI does not take into account the different ages at which pubertal changes occur in different subregions of the world; but the figures for prevalence are more robust and usable than the data on mean BMI described earlier, given the implications of definitions and cut-offs for overweight and obesity among adolescents.

Table 8.22 The number of subjects used when estimating the mean and standard deviation in each age category for Tables 8.20 and 8.21

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 1254 | 2394 | 2008 | 1625 | 474 | 236 | 32 |
|  | Female | 1180 | 4333 | 3695 | 1893 | 450 | 196 | 24 |
| AFR-E | Male | 3120 | 3899 | 1969 | 1295 | 879 | 333 | 18 |
|  | Female | 3340 | 9099 | 5687 | 2265 | 1514 | 519 | 4 |
| AMR-A | Male | 2822 | 8585 | 11133 | 5286 | 2530 | 1773 | 695 |
|  | Female | 2944 | 13641 | 14664 | 6405 | 2455 | 1825 | 781 |
| AMR-B | Male | 59127 | 39400 | 5590 | 3829 | 798 | 13 | - |
|  | Female | 57726 | 38066 | 7290 | 5603 | 865 | 17 | - |
| AMR-D | Male | 1378 | 204 | - | - | - | - | - |
|  | Female | 1325 | 7120 | 5652 | 337 | - | - | - |
| EMR-B | Male | 1000 | 4761 | 3717 | 2060 | 1170 | 227 | 102 |
|  | Female | 1271 | 5541 | 3617 | 2193 | 1099 | 267 | 70 |
| EMR-D | Male | - | 8535 | 7626 | 4641 | 2482 | 1870 | - |
|  | Female | - | 12129 | 11256 | 5103 | 2354 | 1418 | - |
| EUR-A | Male | 8549 | 8935 | 13341 | 14339 | 4471 | 2581 | 359 |
|  | Female | 9215 | 10132 | 14966 | 16798 | 5615 | 3413 | 566 |
| EUR-B | Male | 4063 | 3472 | 1402 | 2381 | 499 | 298 | 52 |
|  | Female | 4240 | 6138 | 3184 | 3411 | 1012 | 353 | 63 |
| EUR-C | Male | 1202 | 1966 | 2776 | 1949 | 860 | 210 | 79 |
|  | Female | 1150 | 2171 | 3349 | 2415 | 1442 | 561 | 233 |
| SEAR-B | Male | - | 503 | 395 | 215 | 93 | 27 | 13 |
|  | Female | - | 1273 | 988 | 423 | 105 | 20 | 12 |
| SEAR-D | Male | 392 | 919 | 765 | 412 | 159 | 36 | 5 |
|  | Female | 392 | 2430 | 2224 | 790 | 134 | 38 | 3 |
| WPR-A | Male | 949 | 5074 | 7784 | 7676 | 4303 | 2191 | 482 |
|  | Female | 895 | 5557 | 9390 | 9418 | 5179 | 3023 | 767 |
| WPR-B | Male | - | 3014 | 2636 | 2406 | 1397 | - | - |
|  | Female | - | 4307 | 4026 | 3499 | 1882 | - | - |

- No data.

The prevalences of obesity corresponding to the subregional groupings by sex and age are listed in Table 8.24. Again, the accepted WHO criterion is taken, i.e. $\mathrm{BMI} \geq 30 \mathrm{~kg} / \mathrm{m}^{2}$, and children have been assessed in relation to the cut-off percentile for BMI proposed by IOTF. This percentile varies by age and sex during childhood as growth and pubertal changes occur, but these collated values provide an overall estimate of the prevalences of substantial excess weight in both children and adults.

The estimated standard errors for the prevalences of overweight and obesity are given in Tables 8.25 and 8.26 .

Table 8.23 The prevalence of overweight in children and adults, by subregion, sex and age

| Subregion | Sex | Prevalence of overweight (\%) Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 0.0 | 1.3 | 3.7 | 11.0 | 4.9 | 6.7 | 0.7 |
|  | Female | 0.2 | 5.0 | 11.7 | 8.6 | 13.7 | 7.4 | 0.8 |
| AFR-E | Male | 0.4 | 0.1 | 0.5 | 1.3 | 1.9 | 0.7 | 0.0 |
|  | Female | 0.7 | 5.9 | 8.4 | 7.1 | 1.2 | 1.8 | 0.0 |
| AMR-A | Male | 17.0 | 26.9 | 41.8 | 43.9 | 45.9 | 43.6 | 42.7 |
|  | Female | 19.0 | 19.3 | 22.6 | 28.5 | 34.6 | 33.4 | 35.0 |
| AMR-B | Male | 19.5 | 30.1 | 45.3 | 47.2 | 43.9 | 42.3 | 42.8 |
|  | Female | 19.8 | 28.0 | 39.0 | 39.5 | 40.2 | 31.1 | 29.9 |
| AMR-D | Male | 16.8 | 24.0 | 33.8 | 37.2 | 38.8 | 40.6 | 40.6 |
|  | Female | 15.0 | 26.3 | 38.0 | 41.2 | 32.3 | 31.5 | 30.0 |
| EMR-B | Male | 15.1 | 18.1 | 33.1 | 37.2 | 23.8 | 25.5 | 36.4 |
|  | Female | 17.7 | 17.9 | 39.5 | 37.8 | 25.1 | 24.0 | 25.0 |
| EMR-D | Male | 0.3 | 4.8 | 11.4 | 11.9 | 10.4 | 7.9 | 5.0 |
|  | Female | 0.7 | 12.2 | 22.6 | 23.9 | 13.9 | 9.1 | 0.0 |
| EUR-A | Male | 15.3 | 29.7 | 47.4 | 53.2 | 52.8 | 57.2 | 46.9 |
|  | Female | 18.0 | 17.3 | 28.1 | 39.4 | 41.9 | 50.0 | 37.7 |
| EUR-B | Male | 16.9 | 14.3 | 37.8 | 42.7 | 40.8 | 43.9 | 28.8 |
|  | Female | 15.0 | 17.6 | 30.8 | 32.1 | 32.8 | 42.9 | 36.5 |
| EUR-C | Male | 20.0 | 20.5 | 37.1 | 41.6 | 39.3 | 42.2 | 35.8 |
|  | Female | 16.9 | 17.1 | 33.6 | 37.8 | 38.1 | 36.3 | 31.0 |
| SEAR-B | Male | 0.3 | 2.7 | 19.0 | 27.0 | 24.7 | 17.5 | 17.5 |
|  | Female | 0.7 | 5.5 | 20.3 | 34.6 | 37.1 | 21.9 | 21.9 |
| SEAR-D | Male | 0.0 | 0.0 | 6.6 | 9.4 | 5.4 | 4.9 | 0.0 |
|  | Female | 0.0 | 3.9 | 12.9 | 18.8 | 8.7 | 6.8 | 0.0 |
| WPR-A | Male | 17.2 | 17.2 | 27.3 | 28.8 | 24.0 | 20.9 | 14.4 |
|  | Female | 19.3 | 9.4 | 15.8 | 25.0 | 28.5 | 26.3 | 22.3 |
| WPR-B | Male | 1.5 | 13.8 | 24.3 | 27.0 | 28.3 | 13.6 | 9.0 |
|  | Female | 1.4 | 10.5 | 18.7 | 21.8 | 21.6 | 14.0 | 9.6 |

Note: Figures in shaded cells were estimated as outlined in the methodology section.

## 5. RISK FACTOR-DISEASE RELATIONSHIPS

This section reviews evidence for causality relating BMI to different disease and injury outcomes; provides a rationale for choosing $21.0 \pm 1.0$ $\mathrm{kg} / \mathrm{m}^{2}$ (mean $\pm \mathrm{SD}$ ) as the BMI value of theoretical population minimumrisk of adverse health effects; and summarizes the sources of data for the hazard estimates required for estimates of the attributable fraction for the population. To date, there have been no systematic reviews of cohort studies that present age- and sex-specific associations of adverse health outcomes with BMI as a continuous variable (rather than for

Table 8.24 The estimated prevalences of obesity, by subregion, sex and age

| Subregion | Sex | Prevalence of obesity (\%) Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 0.2 | 0.0 | 0.0 | 0.3 | 1.2 | 0.3 | 0.0 |
|  | Female | 0.1 | 0.9 | 2.7 | 0.9 | 2.8 | 0.4 | 0.0 |
| AFR-E | Male | 0.2 | 0.0 | 0.1 | 0.0 | 0.0 | 0.4 | 0.0 |
|  | Female | 0.1 | 0.5 | 2.5 | 2.1 | 0.9 | 0.4 | 0.0 |
| AMR-A | Male | 8.4 | 11.5 | 19.0 | 24.2 | 23.9 | 19.5 | 8.0 |
|  | Female | 9.4 | 12.7 | 23.7 | 32.2 | 28.9 | 24.3 | 15.0 |
| AMR-B | Male | 7.1 | 9.2 | 18.5 | 22.2 | 19.7 | 20.0 | 20.7 |
|  | Female | 5.7 | 11.8 | 28.5 | 38.5 | 32.8 | 22.6 | 20.8 |
| AMR-D | Male | 3.2 | 3.3 | 9.0 | 12.1 | 16.3 | 18.1 | 18.1 |
|  | Female | 2.6 | 5.6 | 14.6 | 19.4 | 24.6 | 23.3 | 20.8 |
| EMR-B | Male | 6.1 | 3.7 | 6.8 | 8.9 | 6.2 | 3.0 | 9.1 |
|  | Female | 3.3 | 6.4 | 14.5 | 17.1 | 18.0 | 15.8 | 25.0 |
| EMR-D | Male | 0.2 | 0.9 | 1.7 | 1.7 | 2.0 | 2.5 | 0.0 |
|  | Female | 0.1 | 4.4 | 12.4 | 13.9 | 7.3 | 4.6 | 0.0 |
| EUR-A | Male | 9.4 | 6.9 | 14.4 | 18.7 | 22.7 | 19.2 | 11.6 |
|  | Female | 13.2 | 9.5 | 13.5 | 22.5 | 31.6 | 34.3 | 15.7 |
| EUR-B | Male | 3.2 | 2.4 | 11.0 | 15.1 | 15.0 | 12.6 | 13.5 |
|  | Female | 2.6 | 7.2 | 22.8 | 31.6 | 32.3 | 39.5 | 33.3 |
| EUR-C | Male | 5.7 | 1.9 | 9.2 | 14.7 | 14.7 | 8.9 | 9.0 |
|  | Female | 3.7 | 5.7 | 22.2 | 35.4 | 36.4 | 26.8 | 18.2 |
| SEAR-B | Male | 0.2 | 0.5 | 2.3 | 2.3 | 2.2 | 7.5 | 7.5 |
|  | Female | 0.1 | 0.6 | 3.1 | 5.5 | 5.7 | 9.4 | 9.4 |
| SEAR-D | Male | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
|  | Female | 0.0 | 0.8 | 2.9 | 4.9 | 0.8 | 0.2 | 0.0 |
| WPR-A | Male | 8.4 | 7.2 | 4.7 | 4.5 | 3.7 | 3.3 | 1.9 |
|  | Female | 9.4 | 3.5 | 4.5 | 5.7 | 6.5 | 5.3 | 3.5 |
| WPR-B | Male | 0.4 | 1.2 | 4.8 | 6.5 | 7.6 | 0.4 | 0.0 |
|  | Female | 0.3 | 1.7 | 6.7 | 9.7 | 10.0 | 3.4 | 1.1 |

Note: Figures in shaded cells were estimated as outlined in the methodology section. The obesity rates in children were particularly variable in some subregions, e.g. EUR-B, because they were based on relatively small samples and the obesity cut-off point reflects an extreme percentile distribution of BMIs.
cut-off points, so that the full impact can be captured), and all affected outcomes and for different subregions. Therefore some trade-offs had to be made to obtain the best estimates of hazard size for this project. The following analysis is based on a series of systematic reviews, the two principal ones being of BMI and vascular disease, and BMI and cancer.

Table 8.25 The estimated standard errors of the measured or predicted prevalence of overweight, by subregion, sex and age

|  |  | Standard errors of the prevalence of overweight |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Age group (years) |  |  |  |  |  |  |  |  |

- No data.

Note: Figures in shaded cells were estimated as outlined in the methodology section.

### 5.1 Evidence for causality

Traditionally, excess body weight and obesity have been considered as risk factors in epidemiological terms, despite the fact that the International Statistical Classification of Diseases (ICD) has specified obesity as a disease in its own right since 1948. The effects of excess weight as a risk factor are at least partly mediated through changes in other risk

Table 8.26 The estimated standard errors of the measured or predicted prevalences of obesity, by subregion, sex and age

| Subregion | Sex | Standard errors of the prevalence of obesity Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 0.11 | 0.53 | 0.67 | 0.89 | 0.85 | 1.10 | - |
|  | Female | 0.11 | 0.43 | 0.73 | 1.21 | 1.00 | 1.03 | - |
| AFR-E | Male | 0.27 | 0.35 | 0.66 | 0.35 | 0.38 | 1.51 | - |
|  | Female | 0.16 | 0.34 | 0.90 | 1.89 | 2.06 | 1.48 | - |
| AMR-A | Male | 1.80 | 1.19 | 0.90 | 1.14 | 1.57 | 1.62 | 1.20 |
|  | Female | 1.90 | 1.09 | 0.99 | 0.92 | 1.08 | 1.05 | 1.40 |
| AMR-B | Male | 3.40 | 1.52 | 1.09 | 1.61 | 1.32 | 0.95 | - |
|  | Female | 2.10 | 0.92 | 0.74 | 0.94 | 0.79 | 0.63 | - |
| AMR-D | Male | 1.01 | 0.11 | - | - | - | - | - |
|  | Female | 1.31 | 0.73 | 1.45 | 5.50 | - | - | - |
| EMR-B | Male | 3.40 | 1.99 | 2.35 | 2.40 | 2.35 | 2.68 | 8.70 |
|  | Female | 2.10 | 1.58 | 2.31 | 3.14 | 3.42 | 3.46 | 25.0 |
| EMR-D | Male | 0.26 | 0.00 | - | - | - | - | - |
|  | Female | 0.16 | 0.77 | 1.20 | 4.10 | - | - | - |
| EUR-A | Male | 1.71 | 1.86 | 1.46 | 2.29 | 2.36 | 1.42 | 2.25 |
|  | Female | 1.98 | 1.84 | 0.99 | 1.21 | 2.21 | 1.39 | 2.25 |
| EUR-B | Male | 1.00 | 0.78 | 2.76 | 2.40 | 4.40 | 4.30 | 4.70 |
|  | Female | 1.29 | 0.49 | 1.04 | 2.37 | 4.15 | 5.55 | 5.90 |
| EUR-C | Male | 1.94 | 0.86 | 1.25 | 1.65 | 4.10 | 2.10 | 7.50 |
|  | Female | 1.37 | 0.97 | 1.15 | 1.75 | 4.25 | 6.05 | 9.00 |
| SEAR-B | Male | 0.25 | 0.24 | 0.70 | 1.00 | 1.50 | 5.10 | 7.30 |
|  | Female | 0.15 | 0.16 | 0.60 | 1.10 | 2.30 | 6.50 | 8.40 |
| SEAR-D | Male | 0.00 | 0.00 | - | - | - | - | - |
|  | Female | 0.00 | 0.23 | 0.50 | 0.80 | - | - | - |
| WPR-A | Male | 1.80 | 1.15 | 0.91 | 1.04 | 1.26 | 1.46 | 2.05 |
|  | Female | 1.90 | 1.02 | 0.74 | 0.89 | 1.09 | 1.24 | 1.74 |
| WPR-B | Male | 0.24 | 0.70 | 0.85 | 1.17 | 1.11 | 0.90 | - |
|  | Female | 0.14 | 0.46 | 0.70 | 1.08 | 1.14 | 1.00 | 0.63 |

- No data.

Note: Figures in shaded cells were estimated as outlined in the methodology section.
factors, such as blood pressure and abnormal blood lipids. More recently, it has become clear that excess weight is not only of value in predicting the risk of suffering from particular diseases, but that intentional weight loss reduces these intermediate risk factors (such as high blood pressure); and experimental evidence is also emerging for reduced disease outcomes (Sjöström et al. 1999). This experimental evidence provides the strongest evidence of causality and is summarized in the following sections. The observational evidence from cohort studies linking

Table 8.27 Health outcomes considered in relation to excess bodyweight gain

| Disease | ICD revision |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | ICD-6/7 ${ }^{\text {a }}$ | ICD-8 ${ }^{\text {b }}$ | ICD-9 BTL | ICD-9 | ICD-IO |
| Ischaemic heart disease | A08I | A083 | B27 | 410-414, | 120-I25 |
|  |  |  |  | Proportion of: 427.I, 427.4, 427.5, 428, 429.0-429.2, 429.9, 440.9 |  |
| Cerebrovascular disease | A070 | A085 | B29 | 430-438 | 160-169 |
| Hypertensive disease | $\begin{aligned} & \text { A083, } \\ & \text { A084 } \end{aligned}$ | A082 | B26 | 40I-405 | 110-113 |
| Type II diabetes | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | EII |
| Osteoarthritis | ... | ... | ... | 715 | MI5-M19 |
| Colon and rectum cancers | $\begin{aligned} & \text { A047, } \\ & \text { A048 } \end{aligned}$ | $\begin{aligned} & \text { A048, } \\ & \text { A049 } \end{aligned}$ | $\begin{aligned} & \text { B093, } \\ & \text { B094 } \end{aligned}$ | 153, 154 | C18-C2I |
| Breast cancer (females only) | A05 I | A054 | BII3 | 174 | C50 |
| Endometrial cancer | A053 | A056 | B122 | 179, 182 | C54-C55 |

BTL Basic tabulation list.
... Not available.
a Intermediate list of 150 causes.
. List A: list of I50 causes.

BMI to adverse health outcomes is also described in more detail in the subsequent sections, as these studies provided estimates of exposuredisease hazard size. Together, these sources of data show that the associations found between BMI and many disease outcomes satisfy the widely-accepted criteria for causal relationships: they are strong, consistent, have a dose-response relationship and are biologically plausible (Hill 1965).

### 5.2 Health outcomes caused by excess body weight

The health outcomes that were considered are presented in Table 8.27; the causal relationship between excess body weight and these conditions is dealt with later.

### 5.3 The optimum population BMI (THEORETICAL-MINIMUM-RISK)

The theoretical-minimum-risk distribution of BMI in the population is that which is associated with the lowest health risks related to BMI. This choice of the theoretical minimum needs to take into account the fact that there are hazards associated with low as well as high BMIs (Shetty
and James 1994). The optimum trade-off is based on a balance between the level down to which the risk of developing diseases associated with high BMI persists (described in subsequent sections, but generally $20 \mathrm{~kg} / \mathrm{m}^{2}$, once the confounding effects of smoking and co-morbid prevalent diseases are accounted for) and the health hazards and reduced physical capacity found at lower BMIs. The handicaps related to underweight stem from a chronic dietary energy deficit and other phenomena related to undernutrition. These are summarized in Figures 8.5 and 8.6.

Analyses have been made (James and Francois 1994) of the relationship between the proportion of adults in the population who are underweight (i.e. BMI of $<18.5 \mathrm{~kg} / \mathrm{m}^{2}$ ) and the median BMI of the population. This definition of underweight was accepted by WHO in both the 1995

Figure 8.5 Morbidity and mortality at lower BMIs


Note: Data for Figure 8.5 were collated by Shetty and James (1994).
Source: (a) de Vasconcellos, based on data presented at a meeting "Functional Significance of Low Body Mass Index (BMI)", Rome 4-6 November, 1992. Fig. 6.3: BMI and the probability of illness among Brazilian women (PNSN Survey 1989, de Vasconcellos, personal communication); (b) P. Francois, unpublished data, 1990. BMI and "equivalent days" of illness from collated times of impaired activity among women in Rwanda and Burundi, 1982; (c) Adapted from J. Pryer, 1990. BMI and loss of labour days due to illness in Bangladesh; (d) Satyanarayana et al. I991. Mortality rates for men according to BMI categories (Hyderabad).

Figure 8.6 Median or mean BMI and the prevalence of (a) underweight and (b) overweight in different populations


Source: (a) James and Francois (1994); (b) Regression lines taken from Figure 8.I.
report on the uses of anthropometry (WHO 1995) and in the more recent report on obesity (WHO 2000). Figure 8.6(a) presents unpublished representative data obtained from 27 large national surveys of developing countries. By specifying underweight adults as those with a BMI of $<18.5$ $\mathrm{kg} / \mathrm{m}^{2}$ it becomes clear that an optimum median BMI needs to be about 22.0 or more to ensure that fewer than about $10 \%$ of women are underweight (see below).

In order to compare the proportions of people with low and high BMIs in the population at different mean BMIs, data relating underweight (i.e. BMIs of $<18.5 \mathrm{~kg} / \mathrm{m}^{2}$ ), to overweight, (i.e. BMIs of $\geq 25.0 \mathrm{~kg} / \mathrm{m}^{2}$ ), are required. To simplify this comparison, the two figures are presented side by side in Figure 8.6, but it should be noted that they are drawn from different data sets and that the data relating to underweight are presented in relation to the median rather than the mean BMI. However, at lower BMIs the mean and median values are very similar. It can be seen that when Figures 8.6(a) and (b) are compared, a BMI of $21.0-22.0 \mathrm{~kg} / \mathrm{m}^{2}$ emerges as an optimum BMI at which the chances of there being substantial proportions of either underweight or overweight people in the population are minimized. At a BMI of about $21 \mathrm{~kg} / \mathrm{m}^{2}$, the minimum proportion of underweight and overweight people in the population is about $10 \%$ for males, according to Figure 8.6(a), and for both sexes, according to Figure $8.6(\mathrm{~b})$. To achieve a proportion of only $10 \%$ of women being underweight requires that the mean BMI of the population be about $22.5 \mathrm{~kg} / \mathrm{m}^{2}$. However, at this mean BMI already
about $15 \%$ of the female population is overweight, according to Figure 8.6(a). Given the clear evidence for health hazards in women, even at BMIs of $23 \mathrm{~kg} / \mathrm{m}^{2}$ and above (Willett et al. 1995), and the modest handicaps currently evident for women within the first grade of underweight, i.e. BMIs of $17.0-18.4 \mathrm{~kg} / \mathrm{m}^{2}$ (James et al. 1988), we concluded that at present a universal mean BMI of $21.0 \mathrm{~kg} / \mathrm{m}^{2}$ should be chosen as the optimum for both sexes in populations throughout the world.

Overall, the theoretical-minimum-risk was estimated to occur at a BMI of $21.0 \pm 1.0 \mathrm{~kg} / \mathrm{m}^{2}$ (mean $\pm \mathrm{SD}$ ). It should be noted that this is below the level which some cohort studies have estimated to be the nadir of associations with mortality. This is because of the corrections for effects of disease on BMI that have been employed here and the focus on nonfatal as well as fatal events. This optimum BMI is similar to the lower limit of the range proposed by the WHO Technical Consultation on Obesity (WHO 2000), i.e. $21.0-23.0 \mathrm{~kg} / \mathrm{m}^{2}$. The inclusion of the upper limit of $23.0 \mathrm{~kg} / \mathrm{m}^{2}$ in the original WHO Technical Consultation stemmed from a concern that in some developing countries there might be a need for higher reserves of body energy to cope with potential natural disasters and crop failures leading to food deprivation. In practice, as will become evident, the increase in the rates of diseases associated with increases in BMI is evident from BMIs of about $20 \mathrm{~kg} / \mathrm{m}^{2}$.

### 5.4 Body weight and cardiovascular disease

## CAUSAL RELATIONSHIPS BETWEEN EXCESS WEIGHT AND CARDIOVASCULAR DISEASE

Non-optimum levels of blood pressure, cholesterol and glucose are leading causes of cardiovascular disease. There is strong evidence that excess body weight is associated with adverse levels of these risk factors. A recent Cochrane systematic review focusing on randomized controlled trials (Mulrow et al. 2001) identified 18 trials totalling 2611 hypertensive participants with an average body weight of 84 kg . The data suggested that weight loss in the range of $4-8 \%$ of body weight produced an average reduction in systolic blood pressure of 3.0 mmHg , consistent with earlier reviews (Goldstein 1992; MacMahon et al. 1987). It has also been clearly shown in meta-analyses of intervention trials that losing excess weight improves blood lipid profiles (Dattilo and Kris-Etherton 1992), with a fall in total serum cholesterol and triglyceride levels, and an increase in high density lipoprotein (HDL) concentrations. It is reasonable to expect that such reductions in major intermediate risk factors would translate into reduced incidence of cardiovascular disease. Direct evidence from randomized trials on clinical outcomes is limited, owing to the challenges in the modern environment to achieve and maintain weight loss. Nonetheless, evidence is emerging for substantial reductions in diabetes incidence in trials which have weight loss as a major feature of the intervention in individuals at high risk. A detailed discussion can
be found in, for example, Williamson and Pamuk (1993) and in the latest publications from Sweden, where very large numbers of individuals are currently in the middle of long-term trials of the health impact of weight loss induced by surgical reconstruction of the intestine (Sjöström et al. 1999, 2000).

There is also evidence from prospective observational cohort studies for positive associations between BMI and a range of cardiovascular disease outcomes. These data provide the main source of estimates of hazard size in this analysis and are summarized in the relevant following sections. Some further insight into causality from selected crosssectional and prospective studies is summarized in Figure 8.7, which

Figure 8.7 The relationship between BMI, high blood pressure (systolic pressure $\geq 160 \mathrm{mmHg}$ ) and concentrations of blood lipid in men aged 40-59 years in the United Kingdom


[^31]shows a progressive rise in total cholesterol with increasing BMI, from a BMI of $20 \mathrm{~kg} / \mathrm{m}^{2}$, and a sustained fall in HDL from BMI $20-30 \mathrm{~kg} / \mathrm{m}^{2}$. In addition, there is a clear association between high BMI and increases in serum triacylglyceride (triglyceride) concentration. Many studies have shown that there is a potential independent additional risk of ischaemic heart disease with increases in triglyceride concentrations.

The mechanisms by which increased body weight leads to the induction of cardiovascular diseases and excess mortality are not always clear. The effect is partly related to the frequent concomitant lack of physical fitness and physical activity in the overweight, but it is generally accepted that body-weight gain per se enhances insulin resistance, and thus physical inactivity is not the sole explanation. The development of insulin resistance is a powerful predictor of excess levels of triglycerides in the blood and of the propensity to develop type II diabetes.

In the INTERSALT study (Dyer et al. 1989), a significant association was found between BMI and systolic and diastolic blood pressures, which was independent of age, alcohol intake, smoking habits and urinary sodium and potassium excretion (this excretion rate being taken as an index of intake). Other cross-sectional analyses with measurements of BMI, blood pressure and lipids in West Africa, the Caribbean and the United States have also clearly shown increases in BMI associated with rising blood pressure (Wilks et al. 1996). Several mechanisms may explain why changes in body weight lead to alterations in blood pressure. Physical activity is one contributor, with inactivity tending to promote both weight gain and blood pressure increases. Weight change induced by diet without altering physical activity, however, also leads to changes in blood pressure. The mechanisms by which weight gain promotes a rise in blood pressure may involve the accentuation of insulin resistance, increases in the tone of the sympathetic nervous system control of the arterioles and the production by the adipose tissue itself of a variety of vasoactive cytokines and hormones, such as angiotensinogen, which increase blood pressure (Gorzelniak et al. 2002). These vasoactive compounds act in part by reducing sodium excretion by the kidney, thereby increasing the blood volume and therefore blood pressure.

It is well recognized that physical inactivity is an important contributor to body-weight gain and also increases the risk of diabetes, cardiovascular disease and some cancers. Data on the risks of excess BMI stratified by levels of physical inactivity (e.g. Figure 8.8 and also Paffenbarger et al. 1970) show that the hazards of high BMI are present at all levels of physical activity. This indicates that physical inactivity does not account for the full relationship between BMI and disease.

Figure 8.8 BMI, vigorous exercise and incidence of fatal heart attacks among male British civil servants


Note: This study was based on observations of 17944 male British civil servants aged 40-65 years who self-reported at survey between 1968 and I970. BMI values were taken at age $40-59$ years.
Average follow-up was 8.5 years. Death certificates were supplied by the National Health Service Central Register. Rates were standardized for age. Note that the group with the highest BMI and reporting vigorous exercise was "fewer than five people".
Source: Adapted from Morris et al. (1980) (Table III p. I209). Morris defined vigorous exercise as $>6$ METS (metabolic equivalents, ml oxygen $/ \mathrm{kg} \mathrm{min}^{-1}$ ) or $7.5 \mathrm{kcal}_{\mathrm{min}}{ }^{-1}$, e.g. swimming, hill climbing, gardening, brisk walking, for longer than 30 minutes per day.

## Systematic review of BMI and cardiovascular disease outcomes: the Asia-Pacific Cohort Studies Collaboration (APCSC)

To date, the only available systematic review providing age-, sex- and outcome-specific hazard size as a continuous variable is the APCSC, which included data from 33 cohorts (12 studies from Japan, 11 from mainland China, two from Singapore, two from Taiwan [China], one from Hong Kong Special Administrative Region of China [Hong Kong SAR], one from the Republic of Korea, one from New Zealand and three from Australia) (Table 8.28). The heights and body weights of individual participants were measured in all studies and BMI was calculated. Data from participants recorded as having a BMI of $<12 \mathrm{~kg} / \mathrm{m}^{2}$ or $>60 \mathrm{~kg} / \mathrm{m}^{2}$ were excluded from the analysis.

There was evidence of confounding due to disease at baseline and therefore health events that occurred within the first 3 years of followup were excluded from all analyses. Smokers were not excluded, but smoking was included as a covariate in all analyses. Participants were categorized as "smokers" if they classed themselves as current smokers, former smokers or ever-smokers (people who have smoked at any time), and as "non-smokers" if they indicated that they had never smoked.

Some cohorts included a smoking category entitled "not current"; participants in this category were excluded from the analyses since it was not possible to determine if they were former smokers or never-smokers.

In total, 310283 participants contributed 2148354 person-years of follow-up with a mean duration of 6.9 years. Data on baseline BMI, smoking habits and $>3$ years of follow-up were available for these participants. The mean age of the participants at baseline was 47 years and $41 \%$ were female. Ten per cent of participants were from Japan, $15 \%$ were from mainland China, $55 \%$ were from other parts of Asia (Singapore, Taiwan [China], Hong Kong SAR and the Republic of Korea), and $20 \%$ were from Australia and New Zealand (ANZ). The contributing data sets are set out in Table 8.28.

The overall mean baseline BMI was $23.6 \mathrm{~kg} / \mathrm{m}^{2}$. The mean BMI for the Asian populations was $22.9 \mathrm{~kg} / \mathrm{m}^{2}$ while that for the ANZ populations was $26.4 \mathrm{~kg} / \mathrm{m}^{2}$. Table 8.29 presents the calculated increments in risk of cardiovascular disease associated with a one unit $\left(1 \mathrm{~kg} / \mathrm{m}^{2}\right)$ decrease in BMI and Figure 8.9 summarizes the relationship of BMI with all ischaemic heart disease events (adjusted for age, sex, cohort and smoking).

Figure 8.9 The relationship between BMI and all ischaemic heart disease events in the Asia-Pacific Cohort Studies Collaboration

Table 8.28 Asia Pacific Cohort Studies Collaboration: cohorts contributing to the analyses of BMI

| Country or area | Study name | $n$ | Start year | Mean follow-up (years) | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | $\begin{gathered} \text { Mean } \\ \text { age } \\ \text { (years) } \end{gathered}$ | BMI (kg/m²) |  | No. of stroke events | No. of haemorrhagic stroke events | No. of ischaemic stroke events | No. of IHD events |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  | Mean | SD |  |  |  |  |
| Australia | Busselton | 7166 | 1966-1981 | 21.2 | 52 | 44 | 24.7 | 3.8 | 619 | 57 | 118 | 915 |
|  | Melbourne | 41051 | 1990-1994 | 5.7 | 59 | 55 | 26.9 | 4.4 | 21 | 8 | 5 | 84 |
|  | Perth | 10118 | 1979-1994 | 12.7 | 48 | 45 | 25.2 | 3.9 | 55 | 9 | 4 | 164 |
| Mainland China | Anzhen | 8165 | 1991 | 4.3 | 55 | 53 | 23.9 | 3.7 | 130 | 32 | 91 | 24 |
|  | Anzhen 02 | 735 | 1992 | 3.0 | 1 | 46 | 24.1 | 3.1 | 3 | 0 | 3 | 0 |
|  | Capital Iron \& Steel Company | 3664 | 1974 | 13.3 | 0 | 45 | 23.0 | 2.7 | 116 | 37 | 75 | 52 |
|  | CISCH | 2132 | 1992 | 3.3 | 51 | 44 | 24.7 | 3.5 | 9 | 0 | 0 | 13 |
|  | East Beijing | 321 | 1979 | 14.8 | 20 | 41 | 23.5 | 3.5 | 8 | 3 | 4 | 3 |
|  | Fangshan | 1407 | 1991 | 3.7 | 67 | 47 | 24.0 | 3.4 | 15 | 2 | 9 | 3 |
|  | Huashan | 44 | 1992 | 3.1 | 27 | 55 | 22.9 | 2.9 | 1 | 1 | 0 | 0 |
|  | Seven Cities Cohorts | 10282 | 1987 | 9.8 | 53 | 52 | 22.1 | 3.2 | 164 | 112 | 52 | 37 |
|  | Six Cohorts | 8824 | 1982 | 8.3 | 12 | 45 | 21.1 | 2.4 | 109 | 33 | 71 | 43 |
|  | Tianjin | 4487 | 1984 | 6.0 | 39 | 53 | 22.9 | 3.8 | 116 | 43 | 10 | 36 |
|  | Yunnan | 6238 | 1992 | 4.5 | 3 | 55 | 21.6 | 2.9 | 59 | 46 | 12 | 3 |
| Hong Kong SAR | Hong Kong | 842 | 1985 | 3.4 | 67 | 77 | 22.0 | 3.9 | 12 | 1 | 0 | 23 |
| Taiwan, China | CVDFACTS | 5387 | 1989 | 6.6 | 56 | 47 | 23.5 | 3.4 | 22 | 6 | 4 | 7 |
|  | Kinmen | 178 | 1993 | 3.5 | 51 | 67 | 23.1 | 3.8 | 4 | 0 | 0 | 3 |

Table 8.28 Asia Pacific Cohort Studies Collaboration: cohorts contributing to the analyses of BMI (continued)

| Country or area | Study name | $n$ | Start year | Mean follow-up (years) | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | $\begin{gathered} \text { Mean } \\ \text { age } \\ \text { (years) } \end{gathered}$ | BMI ( $\mathrm{kg} / \mathrm{m}^{2}$ ) |  | No. of stroke events | No. of haemorrhagic stroke events | No. of ischaemic stroke events | No. of IHD events |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  | Mean | SD |  |  |  |  |
| Japan | Aito Town | 1124 | 1980 | 15.4 | 59 | 51 | 22.9 | 3.1 | 20 | 4 | 0 | 8 |
|  | Akabane | 1804 | 1985 | 11.3 | 56 | 54 | 22.5 | 3.0 | 35 | 5 | 14 | 27 |
|  | Civil Service Workers | 9020 | 1991 | 6.6 | 33 | 47 | 22.5 | 2.7 | 1 | 0 | 0 | 1 |
|  | Hisayama | 1427 | 1961 | 19.9 | 56 | 55 | 21.6 | 2.7 | 285 | 55 | 209 | 78 |
|  | Konan | 1033 | 1987 | 7.0 | 56 | 51 | 21.9 | 3.0 | 10 | 2 | 6 | 3 |
|  | Miyama | 986 | 1988 | 6.5 | 56 | 60 | 22.2 | 3.0 | 4 | 0 | 2 | 1 |
|  | Ohasama | 2155 | 1992 | 4.2 | 64 | 59 | 23.3 | 3.1 | 47 | 9 | 33 | 2 |
|  | Saitama | 3534 | 1986 | 9.7 | 62 | 54 | 22.4 | 2.9 | 47 | 11 | 24 | 20 |
|  | Shibata | 2207 | 1977 | 16.8 | 58 | 56 | 22.5 | 3.0 | 173 | 30 | 67 | 62 |
|  | Shigaraki Town | 2457 | 1991 | 4.8 | 59 | 57 | 22.5 | 3.0 | 3 | 1 | 2 | 1 |
|  | Shirakawa | 4590 | 1974 | 16.8 | 54 | 48 | 21.5 | 2.8 | 74 | 24 | 36 | 60 |
|  | Tanno/Soubetsu | 737 | 1977 | 15.3 | 14 | 51 | 22.9 | 2.9 | 42 | 12 | 15 | 31 |
| New Zealand | Fletcher Challenge | 2383 | 1992 | 4.8 | 22 | 40 | 26.5 | 4.3 | 21 | 2 | 5 | 29 |
| Republic of Korea | KMIC | 160159 | 1990 | 5.5 | 33 | 44 | 23.0 | 2.5 | 999 | 295 | 429 | 256 |
| Singapore | Singapore Heart | 2326 | 1982 | 12.5 | 49 | 40 | 23.4 | 4.3 | 69 | 8 | 21 | 63 |
|  | Singapore NHS92 | 3300 | 1992 | 6.2 | 52 | 39 | 23.2 | 4.2 | 39 | 3 | 13 | 21 |
| Total or average ${ }^{\text {a }}$ |  | 310283 |  | 6.9 | 41 | 47 | 23.0 | 3.1 | 3332 | 851 | 1334 | 2073 |

[^32]Table 8.29 Asia Pacific Cohort Studies Collaboration: summary of associations of cardiovascular disease with a one-unit ( $1 \mathrm{~kg} / \mathrm{m}^{2}$ ) decrease in BMI, by age

|  | No. of studies | No. of participants | No. of events | Hazard ratio ${ }^{\text {a }}$ | 95\% Cl |
| :---: | :---: | :---: | :---: | :---: | :---: |
| All ischaemic heart disease events |  |  |  |  |  |
| 30-44 years | 31 | 137916 | 73 | 0.89 | (0.84-0.95) |
| 45-59 years | 32 | 205109 | 504 | 0.91 | (0.88-0.93) |
| 60-69 years | 33 | 76301 | 511 | 0.95 | (0.93-0.97) |
| 70-79 years | 28 | 28366 | 576 | 0.96 | (0.94-0.98) |
| $\geq 80$ years | 26 | 5869 | 414 | 0.97 | (0.95-1.00) |
| Deaths from hypertensive disease |  |  |  |  |  |
| 30-44 years | 31 | 136265 | 0 | 1.00 | (1.00-1.00) |
| 45-59 years | 32 | 205007 | 5 | 0.92 | (0.74-1.13) |
| 60-69 years | 33 | 76272 | 16 | 0.86 | (0.76-0.96) |
| 70-79 years | 28 | 28446 | 29 | 0.89 | (0.82-0.98) |
| $\geq 80$ years | 26 | 5979 | 39 | 0.94 | (0.86-I.03) |
| All ischaemic stroke events |  |  |  |  |  |
| 30-44 years | 31 | 137917 | 41 | 0.85 | (0.77-0.94) |
| 45-59 years | 32 | 205109 | 411 | 0.92 | (0.88-0.95) |
| 60-69 years | 33 | 76345 | 345 | 0.94 | (0.91-0.97) |
| 70-79 years | 28 | 28449 | 316 | 0.94 | (0.91-0.98) |
| $\geq 80$ years | 26 | 5967 | 222 | 0.98 | (0.94-1.02) |

Note: The first 3 years of follow-up were excluded and results were adjusted for age, sex, cohort and smoking habits.
a The age pattern of the hazard ratios presented was smoothed and the resulting age-specific estimates were used to derive the estimates of global burden of disease attributable to BMI:
Ischaemic heart disease: $0.88,0.92,0.93,0.95,0.98$.
Hypertensive disease: $0.88,0.91,0.93,0.95,0.97$.
Ischaemic stroke: $0.82,0.85,0.88,0.90,0.93$.

The overall reduction in risk of ischaemic heart disease associated with a reduction in BMI of one unit in the age group 45-59 years in the APCSC data amounted to $9 \%$, a very similar figure to that provided by the large North American and European prospective studies (reviewed in the next section).

The findings for stroke from the APCSC meta-analysis are displayed in Figure 8.10. A continuous relationship between increasing BMI and risk of non-fatal stroke is evident, but little association was seen between BMI and the risk of fatal stroke. In examining stroke subtypes, a continuous relationship between increasing BMI and risk of ischaemic stroke was apparent, but a weaker association was seen between BMI and the

Figure 8.10 The relationship between stroke events and BMI in adults in the cohort studies of the Asia-Pacific Cohort Study Collaboration

risk of haemorrhagic stroke. This suggests that the lack of association seen between BMI and fatal stroke may reflect in part the higher proportion of fatal strokes that are of the haemorrhagic subtype. Therefore estimates of global burden of disease attributable to stroke in this work were based solely on ischaemic stroke for which a continuous association with BMI is evident.

## Other systematic reviews of BMI and cardiovascular outcomes

A systematic review of large cohort studies investigating ischaemic heart disease has recently been completed (Whitlock et al. 2002). These studies were almost all from North America and Europe and are described in Table 8.30. Most of the studies included measured BMI and dealt with
Table 8.30 The relationship between risk of ischaemic heart disease and BMI: large North American and European prospective cohort studies ${ }^{\text {a }}$

| Study (reference) | Maximum no. of cases analysed | Type of IHD outcome | BMI | Age at recruitment (years) | Shape of association | Risk reduction per kg/m ${ }^{2}$ BMI reduction (\%) ${ }^{\text {b }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Nurses' Health study (Willett et al. 1995) | 1292 | Fatal and non-fatal | Self-reported | 30-55 | Positive | 11 |
| Adventist Mortality study (Lindsted et al. 1991) | 1004 | Fatal | Self-reported | 30-89 | Positive | 8 |
| British civil servants' study (Morris et al. 1980) | 977 | Fatal and non-fatal | Not reported | 40-59 | Positive? (no significance tests) | 4 |
| British Regional Heart study (Shaper et al. 1997) | 974 | Fatal and non-fatal | Measured | 40-59 | Positive | 7 |
| Manitoba Follow-Up study (Tate et al. 1998) | 915 | Fatal and non-fatal | Measured | 18-62 | Positive or J-shaped | 4 |
| Social Insurance Institution study (Rissanen et al. 1990) | 868 | Non-fatal | Measured | 25-64 | Positive | 4 |
| Whitehall study (Jarrett et al. 1982) | 727 | Fatal | Measured | 40-64 | Positive | 6 |
| NHANES I Epidemiologic Follow-up study (Cooper and Ford 1992) | $\begin{array}{r} \text { (Male) } 563 \\ \text { (Female) } 696 \end{array}$ | Fatal and non-fatal | Measured | $\begin{aligned} & 25-75 \\ & 25-75 \end{aligned}$ | Not significant Positive or J-shaped | $\begin{array}{r} \text { (Male) } 6 \\ \text { (Female) } 2 \end{array}$ |
| Gothenburg men's study (Rosengren 1999) | 686 | Fatal | Measured | 47-55 | $J$-shaped | 8 |
| Framingham study (Higgins et al. 1988) | $\begin{array}{r} \text { (Male) } 659 \\ \text { (Female) } 540 \end{array}$ | Fatal and non-fatal | Measured | $\begin{aligned} & 35-69 \\ & 35-69 \end{aligned}$ | Positive or J-shaped Positive or J-shaped | $\begin{array}{r} \text { (Male) } 5 \\ \text { (Female) } 4 \end{array}$ |
| Seven Countries study (Keys et al. 1972) | 632 | Fatal and non-fatal | Measured | 40-59 | Positive or J-shaped | 7 |
| Physicians' Health study (Rexrode 2001) | 548 | Fatal and non-fatal | Self-reported | 40-84 | Positive | 7 |
| Eastern and Southwestern Finland study (Jousilahti et al. 1999) | 520 | Fatal and non-fatal | Measured | 25-64 | Positive or J-shaped | 4 |
| Honolulu Heart Program (Rhoads and Kargan 1983) | 511 | Non-fatal | Measured | 45-68 | Positive | 10 |

[^33]Adapted from Whitlock et al. (2002). The large American Cancer Society (1960-1972) cohort (Calle et al. 1999; Stevens et al. 1998) recorded self-reported height and weight. However, cause-specific associations have not been published. A positive association between risk of all cardiovascular death, steeper in younger age groups, has been shown.
non-fatal as well as fatal outcomes. All but one showed a positive or J-shaped relationship between BMI and the risk of either fatal or nonfatal ischaemic heart disease. When the potential reduction in risk was calculated for each unit decrease in BMI, results were found to be reasonably similar, with most studies indicating a $5-10 \%$ reduction in rates of ischaemic heart disease. The overall unweighted average reduction in risk per unit BMI difference was $6 \%$, with studies using self-reported BMI giving a weighted average of $9 \%$, whereas the weighted average for the studies using measured BMI amounted to a $5 \%$ reduction. The mean age at death for these cohorts was estimated to be age 50-60 years, indicating that these data are highly consistent with those of the APCSC outlined earlier.

The conclusions of other systematic reviews are also consistent with the APCSC results, although comparisons are limited by the use of BMI categories; the main results are summarized in Table 8.31.

### 5.5 Type II diabetes

## Causal relationships between excess weight and type II diabetes

The relationship between excess body-weight gain and type II diabetes is now considered so strong that there is increasing use of the term "diabesity" as a unifying concept. Not only is there a close association between higher BMIs and the risk of developing type II diabetes, but weight gain itself has also been identified as a particularly important risk factor. The impact of weight gain is markedly enhanced if it occurs in young adults who were already overweight or obese when they entered adult life (Colditz et al. 1995). More direct evidence for the importance of increases in weight in the development of diabetes comes from intervention studies. Over $80 \%$ of very obese diabetic adults treated by gastric bypass surgery to induce marked weight loss, for example 30 kg , become non-diabetic and over an 8 -year post-surgery follow-up period, the incidence of new cases of diabetes in these patients is minimal (Sjöström et al. 2000). Four prospective studies, three of which were randomly controlled intervention studies, have also shown that changes in diet and exercise that induce a modest loss of weight in overweight or obese subjects with glucose intolerance can markedly reduce the subsequent development of type II diabetes over periods of 3-6 years (Diabetes Prevention Program Research Group 2002; Eriksson and Lindgarde 1991; Tuomilehto et al. 2001; Xiao-Ren Pan 1997). Weight-loss trials among obese patients with type II diabetes have also shown marked improvements in diabetic states or even a return to normal glucose tolerance. Lean et al. (1990) have also shown that the degree of weight loss achieved in newlydiagnosed patients with type II diabetes predicts their future life span. Williamson and Pamuk (1993) also observed that in the United States overweight and obese women with co-morbidities have not only a reduced overall mortality but also a selective reduction in death from

Table 8.3I The relationship between BMI and the development of cardiovascular disease: analyses from two systematic reviews

Table 8.3I(a) Australia

|  | Relative risks associated with overweight and obesity |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Overweight (BMI 25-29.9 $\mathrm{kg} / \mathrm{m}^{2}$ ) |  |  |  | Obese ( BMI $\geq 30 \mathrm{~kg} / \mathrm{m}^{2}$ ) |  |  |  |
|  | Males |  | Females |  | Males |  | Females |  |
|  | <65 | $\geq 65$ | <65 | $\geq 65$ | <65 | $\geq 65$ | <65 | $\geq 65$ |
| Ischaemic heart disease based on Harris et al. (1993, 1997); Manson et al. (1990); Rimm et al. (1995) | 1.35 | 1.00 | 1.40 | 1.00 | 1.80 | 1.20 | 2.00 | 1.25 |
| Ischaemic stroke based on Rexrode et al. (1997) | 1.35 | 1.00 | 1.25 | 1.00 | 1.50 | 1.15 | 1.60 | 1.20 |
| Hypertension based on Ascherio et al. (1992); Sjöström et al. (1992); Witteman et al. (1989) | 1.40 | 1.40 | 1.40 | 1.40 | 2.35 | 2.35 | 2.35 | 2.35 |

Source: Mathers et al. (1999).

Table 8.3I(b) United Kingdom

|  | Relative risks associated with obesity$\left(\mathrm{BMI} \geq 30 \mathrm{~kg} / \mathrm{m}^{2}\right)$ |  |
| :---: | :---: | :---: |
|  | Males | Females |
| Myocardial infarction | 3.2 | 1.5 |
| Stroke | 1.3 | 1.3 |
| Hypertension | 2.6 | 4.2 |
| Angina | 1.8 | 1.8 |
| Relative risks specified only in relation to obesity (BMI $\geq 30 \mathrm{~kg} / \mathrm{m}^{2}$ ). Derived from 48 unspecified studies after a systematic review of 3537 studies. |  |  |
| Source: National Audit |  |  |

diabetes-related conditions if they intentionally lose modest amounts of weight of up to 9 kg (Williamson and Pamuk 1993).

The mechanisms whereby weight gain leads to the development of type II diabetes are the subject of intense investigation, with the development of insulin resistance seen as dominant. Type II diabetes develops when the pancreatic capacity to generate insulin cannot maintain the markedly increased demand induced by insulin resistance. Insulin resis-
tance itself is affected not only by increases in weight, particularly if the extra energy is stored in abdominal, i.e. visceral, fat, but also by dietary composition. Dietary fat induces insulin resistance (Marshall et al. 1994; Sarkkinen et al. 1996; Vessby et al. 2001) and there is increasing interest in the possibility that rapidly absorbed carbohydrates, which cause sudden increases in concentrations of blood glucose, place extra demands on the pancreas (Willett et al. 2002). Adipose tissue itself, particularly visceral adipose tissue, secretes cytokines such as interleukin-6 (IL-6) and tumour necrosis factor (TNF $\alpha$ ) which are recognized to be important inducers of insulin resistance. Circulating adiponectin, an adipocytederived hormone which markedly improves insulin sensitivity is reduced as the fat cells expand with body-weight gain, and is also modulated by sex hormones (Nishizawa et al. 2002). Physical inactivity also contributes to insulin resistance, with vigorous exercise leading to a rapid restoration of insulin sensitivity.

The quantitative contributions of each of these components are still uncertain, but the carefully controlled Chinese Prevention study on preventing type II diabetes (Xiao-Ren Pan et al. 1997) provides some information. In this study, three groups of overweight adults with glucose intolerance were assigned to either a low fat, low saturated fat and high vegetable and fruit diet, or to a modest increase in physical activity equivalent to 30 minutes of brisk walking daily, or were advised on both dietary and physical activity interventions. There was a marked and equivalent reduction in the incidence of type II diabetes in all three groups over the 6 years of follow-up. This implies that there is little interaction between diet and physical activity but that both can contribute to the development of insulin sensitivity.

Physical activity is potentially a major confounding effect. Physical activity is recognized as beneficial in enhancing the sensitivity of tissues to insulin, thereby enhancing the body's capacity to handle glucose. As noted above, increasing physical activity alone without dietary change reduced the incidence of diabetes in modestly overweight Chinese adults with glucose intolerance (Xiao-Ren Pan et al. 1997). This finding was consistent with those from a further two major controlled trials from Finland and the United States, which suggested that the combination of changes in diet and activity together with only modest weight losses reduced the incidence of type II diabetes in susceptible adults by $56 \%$, this benefit increasing to $76 \%$ in those aged $>60$ years (Diabetes Prevention Program 2002; Tuomilehto et al. 2001). In these trials, the modest weight loss, for example $5 \%$, was induced by a low-fat diet combined with physical activity. Given the important effect of physical activity, the quantification of any interaction between BMI and physical activity when determining the risk of onset of type II diabetes is necessary. As with coronary disease, the increase in risk of diabetes with increasing BMI is seen at all levels of physical activity (Figure 8.11).

Figure 8.1I The interaction of higher BMIs with physical activity in determining the age-adjusted incidence of type II diabetes per 10000 person-years of follow-up, in males


Note: "Incidence" relates to the age-adjusted incidence of type II diabetes per 10000 person-years of follow-up in male graduates of the University of Pennsylvania.

Source: Adapted from Helmrich et al. (1991).

Sources of hazard size estimates for BMI and type II diabetes

## Incidence studies

The conclusions of systematic reviews from Australia (Anonymous 1999; Mathers et al. 1999) and the United Kingdom of Great Britain and Northern Ireland (National Audit Office 2001) are given in Table 8.32. In one of the Australian reports (Mathers et al. 1999), the risks of developing diabetes were arbitrarily halved in an attempt to take into account the impact of physical activity on diseases relating to obesity (see below). The first three of the four studies (Carey et al. 1997; Colditz et al. 1990, 1995; Njolstad et al. 1998) noted by this report involved very large groups of professionals in the United States (e.g. $>100000$ subjects with individual prospective follow-ups of 12-18 years). Unfortunately, these data are based on self-reported heights, weights and disease. Although weights and heights reported in these study populations are found to have a good correlation with observed measurements, they are still likely to systematically underestimate the actual prevalences of overweight and obesity. The Victoria report (Anonymous 1999) looked at this issue and

Table 8.32 The relative risks of developing type II diabetes associated with overweight and obesity, as assessed by two governmental reviews

| Study | Age (years) | Overweight <br> (BMI 25.0-29.9 kg/m ${ }^{2}$ ) |  | $\begin{gathered} \text { Obese } \\ \left(\mathrm{BMI} \geq 30 \mathrm{~kg} / \mathrm{m}^{2}\right) \end{gathered}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Males | Females | Males | Females |
| Australia ${ }^{\text {a }}$ (Mathers | <65 | 1.8 | 1.8 | 3.2 | 3.2 |
| et al. 1999) | $\geq 65$ | 1.8 | 1.8 | 3.2 | 3.2 |
| United Kingdom (National Audit Office 200I) | All ages | - | - | 5.2 | 12.7 |

- No data.
a Relative risk values arbitrarily halved to help account for any independent impact of physical activity or residual confounding.
found from its own analysis that "people who are obese selectively underestimated their weight and/or overestimated their height more than others. As a result, the proportion of people who are obese by measurement is $60 \%$ higher than estimates based on self-reported height and weight. The greatest discrepancies are found in adolescent men and older people". How this systematic underestimation of obesity impacts on the estimates of relative risk per BMI unit is unclear, but it may well result in overestimation, i.e. a bias away from the null. This is because the distribution of self-reported BMI is narrower than the distribution of actual BMI, and so the slope of associations between BMI and risk of disease is artificially steep. The use of self-reported disease status is another source of uncertainty in hazard estimates because in many such studies diabetes has been found to be underreported or under-diagnosed (Harris et al. 1998). If undrrecording of diabetes is associated with BMI level, this would also result in bias in hazard estimates (away from the null, if relatively more cases of diabetes failed to be identified among people with low BMI).

The prospective studies on male and female health professionals in the United States (Chan et al. 1994; Colditz et al. 1995) suggest age-adjusted relative risks of self-reported diabetes of between 4 and 14 in men aged $40-75$ years and reporting a BMI of about $30 \mathrm{vs}<23 \mathrm{~kg} / \mathrm{m}^{2}$; whereas the relative risk for self-reported diabetes in the nurses of similar age and with reported BMIs of about 30 was about 28 compared with the rates for nurses with reported BMIs of $<22 \mathrm{~kg} / \mathrm{m}^{2}$. A more robust estimate of the relative risk of developing type II diabetes comes from a detailed Norwegian study based on a sampling system which is representative of their most northern county and in which objective measurements of weight and height as well as measurements of fasting glucose concentrations were made (Table 8.33).

Table 8.33 Relative risk of developing type II diabetes in Norwegian men and women aged 35-52 years, by sex-specific quartiles of baseline BMIs

| Quartiles of baseline <br> $\mathrm{BMI}\left(\mathrm{kg} / \mathrm{m}^{2}\right)$ |  | Relative risk of type II <br> diabetes over I2 years |  |
| :--- | :--- | :--- | ---: |
| Males | Females | Males | Females |
| $<23.2$ | $<22.0$ | 1.0 | 1.0 |
| $23.2-25.0$ | $22.0-24.1$ | 1.8 | 1.8 |
| $25.1-27.0$ | $24.2-27.1$ | 2.5 | 2.1 |
| $\geq 27.1$ | $\geq 27.2$ | 13.0 | 30.0 |

Note: The above relative risk of developing diabetes for both males and females are interpolated from a graph in Njolstad et al. (1998).

By chance, the quartiles of BMI for men again allow an approximate distinction between those with BMIs of $<23.0 \mathrm{~kg} / \mathrm{m}^{2}$ and $<25.0 \mathrm{~kg} / \mathrm{m}^{2}$, i.e. their second quartile, the upper limit of which coincides with the WHO distinction between normal and overweight. The values for women need to be adjusted. These data relate to men and women aged 35-64 years, but were age-standardized. Nevertheless, it is clear that the relative risks of diabetes as derived from these data in general agree with data from the United States prospective studies, and that the values given for the upper quartile will be low estimates of the true risk of diabetes in those Norwegian men and women who have a BMI of $\geq 30.0 \mathrm{~kg} / \mathrm{m}^{2}$.

## Prevalence studies

The ideal source of estimates of the hazard size would be large long-term trials or, in their absence, large, long-term cohort studies with direct measurements of BMI and blood glucose. The principal cause of type II diabetes appears to be excess weight gain in childhood and in adult life, and so relatively short-term prospective studies of the incidence of diabetes may not capture the full impact of persistent excess BMI. Large, nationally representative cross-sectional studies are available that use objective measurements of BMI and diabetes (in contrast to the few large, long-term prospective studies outlined earlier). Although the diagnosis of diabetes may itself lead to a loss in weight, this is likely to result in only a moderate bias to the null, since current BMI correlates so closely with BMI levels over many previous decades. The estimates of hazard size for this analysis of the burden of diabetes attributable to BMI were therefore derived from the age- and sex-specific associations of type II diabetes (based on fasting glucose values) with BMI from the Japanese National survey (Yoshike, personal communication; see Figure 8.12). The incremental risks for the required age categories were estimated on a linear basis (Table 8.34).

Table 8.34 The relative risk of developing type II diabetes, per unit ( $1 \mathrm{~kg} / \mathrm{m}^{2}$ ) increase in BMI

| Disease | Sex | Country source | Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-17 | 18-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| Type II | Males | Japan | - | - | 1.36 | 1.24 | 1.18 | 1.27 | 1.27 |
| diabetes | Females | Japan | - | - | 1.47 | 1.34 | 1.21 | 1.20 | 1.20 |

- No data.

Source: These estimates were taken from a diagram of results (from the Japanese National survey) and are likely to be subject to some inaccuracy.

Figure 8.12 The prevalence of type II diabetes (based on fasting blood glucose concentrations) by age, sex, BMI and decade of adult life, in a representative sample of the Japanese population


Males

Females

Note: Based on rates of type II diabetes determined from measurements of fasting blood glucose $\geq 126$ $\mathrm{mg} / \mathrm{dl}$ or where treatment is already being given for type II diabetes (Yoshike, personal communication).

The proportional increases in the prevalence of diabetes associated with a given increase in BMI were consistent in the Japanese national survey with associations observed in a Danish representative survey (Drivsholm et al. 2001; see Figure 8.13). They are also broadly consistent with the results of prospective studies that measured weight and height, that is, the Norwegian study that assessed the risk of diabetes,

Figure 8.13 A comparison of the prevalence of type II diabetes (based on fasting blood glucose concentrations and measured BMI) in Danish and Japanese adults of comparable age

with measurements of blood glucose (Table 8.33) and the United States NHANES III analyses of prevalence rates of diagnosed diabetes (Table 8.35).

A perspective on the overall burden of ill-health associated with higher BMI can be gauged from nationally representative studies of measured BMI and the prevalence of type II diabetes. Thus Must et al. (1999) used the NHANES III study and highlighted the marked 20 -fold increase in prevalence rates of reported type II diabetes in men and women aged $>55$ years compared with those aged $<55$ years, even in people with a measured "normal weight", that is, with BMIs of $<25.0 \mathrm{~kg} / \mathrm{m}^{2}$. Superimposed on this age-related increase was the impact of high BMIs: the prevalence ratios for people aged $<55$ years with high BMI compared with those with a normal weight varied from 3.3 in men and 3.8 in women with BMIs of $25-29.9 \mathrm{~kg} / \mathrm{m}^{2}$, up to $8-11$ in men and women with BMIs of $35-39.9 \mathrm{~kg} / \mathrm{m}^{2}$ (but with very broad confidence limits). Prevalence ratios were lower in the older age groups (1.8 in the overweight men and women, and 4.2 in men and 3.2 in women with BMIs of $35-39.9 \mathrm{~kg} / \mathrm{m}^{2}$ ).

Recently, there has been an interest in the issue of ethnic differences in the incidence of co-morbidities that result from increases in weight, with reports of an increased propensity to diabetes in Pima Indians, in those of Hispanic origin and in Asians, compared to the American Cau-

Table 8.35 Estimated prevalence ratios of type II diabetes by weight status category in adults in the United States-representative NHANES III study

| Age group (years) | Prevalence ratio of type II diabetes |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Weight status category |  |  |  |  |
|  | Adjusted prevalence in individuals of normal weight ${ }^{2}$ | Overweight <br> (BMI 25.0- <br> $29.9 \mathrm{~kg} / \mathrm{m}^{2}$ ) | Obesity class I (BMI 30.0$34.9 \mathrm{~kg} / \mathrm{m}^{2}$ ) | Obesity class 2 <br> (BMI 35.0- <br> $39.9 \mathrm{~kg} / \mathrm{m}^{2}$ ) | Obesity class 3 <br> (BMI $\geq 40.0 \mathrm{~kg} / \mathrm{m}^{2}$ ) |
| Males |  |  |  |  |  |
| <55 | 0.2 | 3.27 (1.17-9.05) | 10.14 (4.03-25.08) | 7.95 (2.44-25.23) | 18.08 (6.71-46.84) |
| $\geq 55$ | 5.3 | 1.77 (1.26-2.47) | 2.56 (1.71-3.74) | 4.23 (2.09-7.59) | 3.44 (1.11-8.32) |
| Females |  |  |  |  |  |
| <55 | 0.4 | 3.82 (1.75-8.21) | 2.49 (1.01-6.12) | 10.67 (4.02-27.11) | 12.87 (5.69-28.05) |
| $\geq 55$ | 7.9 | 1.81 (1.41-2.31) | 2.19 (1.56-3.01) | 3.24 (2.13-4.67) | 5.76 (4.17-7.42) |

a For people with BMI of $18.5-24.9 \mathrm{~kg} / \mathrm{m}$, prevalence data were specified only for white adults, with current smokers included if aged $<55$ years, but also former smokers in the age group $\geq 55$ years. Data for both age groups are adjusted for age, and prevalence ratios are adjusted for race and ethnicity, as well as smoking status. The ratios are set out in relation to the group of individuals of normal weight. The $95 \%$ confidence limits are shown in parentheses.

Note: The prevalence of type II diabetes depended on self-reporting because fasting glucose levels were available for only $44 \%$ of the sample.
Source: Must et al. (1999).
casian population (Edelstein et al. 1997; Seidell et al. 2001a). Figure 8.13 shows that the prevalence of diabetes in the Japanese population markedly exceeds that in the Danish population, for both sexes and for each BMI category, but the gradient of risk appears to be roughly the same in the two communities. Similar enhanced risks of diabetes are evident in Mexicans assessed in the latest national survey of the prevalence of diabetes and other diseases in relation to excess weight in Mexico (Sánchez-Castillo et al. 2003). These data were compared with the recalculated data relating only to those with fasting blood glucose measurements, from the NHANES III study in the United States. Prevalence rates for type II diabetes are $2-3$-fold greater in Mexicans than in non-Hispanic Caucasians in the United States, on an age-standardized basis and, more importantly, prevalences in Mexicans exceed those in people of the United States, when BMI is taken into account. The differences were somewhat smaller, but still statistically significant when the age-dependent prevalences of diabetes were related to the waist circumferences of the two national groups. In the year 2000, the prevalence of abdominal obesity was greater in Mexicans than in non-Hispanic Caucasians in the United States, as measured in the NHANES III study (Sánchez-Castillo et al. 2003). Similar data are now emerging from mainland China Hong Kong SAR and India. In the near future, these comparisons may be extended so that the increment in risk for each pop-
ulation at each BMI level, taking age into account, can be obtained with some assurance on the basis of measured body weights, heights and fasting glucose levels. It will also be necessary to take into account the relationship between BMI and diabetes in Pacific Islanders, which may well prove to be different from that in other peoples, given their greater lean tissue : fat ratios (Bell et al. 2001). Nevertheless, it seems reasonable on present evidence to conclude that although in Asians and some other populations the absolute risk of type II diabetes is amplified, the incremental gradient of risk of diabetes in relation to increasing BMI is approximately the same as that found in Caucasians. On this basis, similar values for relative risk can be applied to all subregional groups.

### 5.6 Osteoarthritis

There is a well-documented association between high BMI and the development of osteoarthritis in both men and women (Cicuttini and Spector 1998). Osteoarthritis is an abnormality which involves damage to and eventually the destruction of the articular cartilage of the joint. New bone is formed at the joint surfaces, probably in response to the cartilaginous damage. The relationship between excess weight and the development of osteoarthritis has been studied in a number of population-based prospective, cross-sectional and retrospective trials, showing excess weight as the most important preventable risk factor for osteoarthritis. Cross-sectional studies have consistently reported increased risks of osteoarthritis in association with body-weight gain, with early studies suggesting a $2-7$-fold greater risk in those individuals in the top compared with the bottom tertile of BMI (Table 8.36) and, in some studies, a greater risk in women than men (Cicuttini and Spector 1998). As the data were adjusted for race, ethnicity and smoking as well as age, no age-related data were available, but these are the best nation-

Table 8.36 Prevalence of osteoarthritis by weight status category, in adults

| Sex | Prevalence of osteoarthritis (\%) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Weight status category |  |  |  |  |  |
|  | Underweight $B M I<18.5$ | Normal weight BMI I8.5-24.9 | Overweight BMI 25.0-29.9 | Obesity class 1 <br> BMI 30.0-34.9 | Obesity class 2 <br> BMI 35.0-39.9 | Obesity class 3 $B M I \geq 40.0$ |
| Males $(n=6987)$ | 0.39 | 2.59 | 4.55 | 4.66 | 5.46 | 10.04 |
| Females $(n=7689)$ | 7.79 | 5.22 | 8.51 | 9.94 | 10.39 | 17.19 |

Source: Table 3 in Must et al. (1999).
ally representative data available for assessing the quantitative impact of BMI increases on the prevalence of osteoarthritis.

A number of mechanisms have been proposed to explain the association of adult body-weight gain with osteoarthritis. The physical burden associated with an increased load on the joints seems straightforward, but changes in movement and gravitational stresses as weight gain occurs are also a factor. Other mechanisms have, however, been invoked, including systemic changes in metabolism associated with hypertension, raised blood glucose and cholesterol concentrations, insulin resistance and elevated concentrations of blood uric acid, as well as hormonal changes induced by the metabolic effects of additional adipose tissue (Cicuttini and Spector 1998). Several of these factors could be acting on the metabolic integrity of the articular cartilage, as could other dietary factors, such as high fat intake, which have also been linked to this disease. The associations with hypertension tend to disappear once concomitant increased body weight is taken into account, and the link with hypercholesterolaemia is not sufficiently robust to warrant special consideration. Abnormal glucose metabolism is a more plausible mechanism, with the possible involvement of growth hormone (Cicuttini and Spector 1998), but epidemiological studies have not shown a consistent link between type II diabetes and osteoarthritis. Raised uric acid concentrations have been associated with osteoarthritis, but again data supporting the relationships are inconsistent.

It is clear that excess weight gain precedes the development of osteoarthritis rather than the reverse. This was initially reported in retrospective recall studies of reported former weights (Anderson and Felson 1998; Hart and Spector 1993), but three studies (Anderson and Felson 1998; Cicuttini et al. 1996; Hart and Spector 1993) have shown a strong association of excess weight gain with asymptomatic radiological evidence of osteoarthritis, such radiological evidence having been clearly shown to be a predictor of future disability (Acheson et al. 1974; Hochberg et al. 1989). Twin studies have shown that the heavier twin has a greater risk of developing osteoarthritis (Cicuttini et al. 1996), and the incidence of osteoarthritis is markedly enhanced in overweight women (Schouten et al. 1992). In population studies, the incidence of disability once osteoarthritis has developed is also particularly marked if the subjects are obese (Verbrugge et al. 1991). In addition to predicting the greater risk of developing osteoarthritis, weight gain has also been shown to enhance the progression of the disease. Among middle-aged women with early stage unilateral knee damage, those who were overweight had a $34 \%$ chance of developing osteoarthritis in the contralateral knee within 2 years and $22 \%$ also showed radiological progression of the disease in the initially affected joint (Cicuttini and Spector 1998).

Detailed population studies regarding the incremental risk of developing osteoarthritis of any joint over a wide range of BMIs are rare,
however, with Must et al. (1999) being the principal source of data. Must et al. used the NHANES III study with measurements of BMI and with reported rates of disability from osteoarthritis to derive risk values. Given the need to consider the linear relationship between BMI and disabilities rather than the type of categorical analyses shown in Table 8.36, the Must et al. (1999) study was used here. This study, as well as several others, shows that the absolute risk of osteoarthritis is greater in women than in men, but that the incremental risk of osteoarthritis per unit increase in BMI is approximately the same in both sexes. On the basis of this survey in the United States, the relative risks of osteoarthritis were derived without regard to adult ages or sex, as shown in Table 8.37. These analyses were not adjusted either for age or smoking status, but Must et al. (1999) found that there were no interactions between weight class and race or ethnicity for either sex.

## Gout

Gout is a metabolic disease of uric acid metabolism, which has a genetic component and which involves acute episodes of arthritis leading to chronic arthritis and disability. Although much less common than osteoarthritis, gout has also been linked with excess weight gain in both cross-sectional and longitudinal studies (Cicuttini and Spector 1998). The abdominal distribution of body fat seems to be a particularly important risk factor, in keeping with mechanistic studies which suggest that hormonal or other agents which alter uric acid metabolism, such as diet and alcohol, are involved in precipitating an attack of gout. Although multivariate analyses have shown the importance of BMI from as early as age 35 years and excess weight gain as distinct risk factors (Roubenoff et al. 1991), with other large studies suggesting that excess weight as early as in adolescence is predictive of gout, there is no clear picture of the prevalence of or associated disability linked to gout on an international basis. Thus, although weight reduction is clearly recognized as potentially important in avoiding and limiting the risk of gout and its associated handicaps, it was decided not to include estimates of the

Table 8.37 The relative risk of developing osteoarthritis per unit ( $1 \mathrm{~kg} / \mathrm{m}^{2}$ ) increase in BMI

|  |  | Age group (years) |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
|  |  | Country |  |  |  |  |  |  |  |  |
| Disease | Sex | source | $15-17$ | $18-29$ | $30-44$ | $45-69$ | $60-69$ | $70-79$ | $\geq 80$ |  |
| Osteoarthritis | Males | USA | - | - | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 |  |
|  | Females | USA | - | - | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 |  |

- No data.
effects of excess BMI on gout and only to use data on weight gain in association with osteoarthritis.


### 5.7 Cancer

There have recently been two major re-analyses, incorporating systematic reviews, of the associations between excess weight gain and the development of different forms of cancer (Bergström et al. 2001; IARC 2002), which supersede those of Australia (Mathers et al. 1999) and the United Kingdom (National Audit office 2001). The latest review by the International Agency for Research on Cancer (IARC) lists the risk of different cancers in relation to both body weight and physical activity (IARC 2002). These analyses concluded that cancers of the colon, breast (postmenopausal), endometrium and kidney were statistically related to weight gain, each analysis being usually based on a systematic review of a large number of both case-control and prospective studies. The present analysis also drew heavily on the recent series of meta-analyses conducted by IARC staff and colleagues, which provided estimates of the coefficient of risk per unit BMI increase (Bergström et al. 2001). Since this chapter used continuous rather than categorical data, the conclusions of the main IARC report on statistically significant findings are used to select the cancers to be considered, but the coefficients of risk are taken from Bergström et al. (2001). Table 8.38 sets out these estimates. In view of the much lower incidence of and the uncertainty of the global statistics for kidney cancer, the estimates of the burden of cancer were confined to cancers of the breast in postmenopausal women, colon and endometrium.

## Breast cancer

A distinction needs to be made between the risks of developing breast cancer before and after the menopause. Premenopausal breast cancer is less likely to develop in women with high BMIs, but this effect is only seen at BMIs of about $>28 \mathrm{~kg} / \mathrm{m}^{2}$ and rates of mortality are not lower among women with higher BMIs. For the present analyses, the role of excess weight gain in the development and the burden of disease from breast cancer was confined to cases arising in postmenopausal women. Over 100 studies conducted in many populations have found that women with higher BMIs are at greater risk of postmenopausal breast cancer (IARC 2002). Bergström et al. (2001) used 27 of these studies to quantitatively evaluate the impact of excess weight. Age, age at menarche, parity, alcohol intake and diet are recognized confounders, but when all of these were taken into account the coefficient of risk per unit BMI was still statistically significant (Table 8.38). Furthermore, there are studies which overall suggest that those women who have limited their weight gain or have lost weight in early adult life tend to have a reduced risk of postmenopausal breast cancer (IARC 2002). Over 50 studies (e.g. Huang et al. 1997; Tretli 1989; Törnberg and Carstensen 1994) also

Table 8.38 The relative risks of developing cancer associated with increases in BMI, as calculated for European populations

| Cancer | Coefficient ( $\pm 95 \%$ Cl) ${ }^{\text {a }}$ (Increase in incidence rates per unit [ $1 \mathrm{~kg} / \mathrm{m}^{2}$ ] increase in BMI ) | Relative risk of developing cancer |  | No. of studies |
| :---: | :---: | :---: | :---: | :---: |
|  |  | Overweight ${ }^{\text {b }}$ | Obesity ${ }^{\text {c }}$ |  |
| Postmenopausal breast | 1.03 (1.02-1.04) | 1.12 | 1.25 | 13 |
| Colon | 1.03 (1.01-1.05) | 1.15 | 1.33 | 6 |
| Endometrial | 1.10 (1.07-1.14) | 1.59 | 2.52 | 4 |
| Kidney | 1.06 (1.03-1.08) | 1.36 | 1.84 | 2 |

a Coefficient describes the increase in incidence rates per unit increase in BMI irrespective of age (and includes the $95 \%$ confidence intervals) based on covariance analysis.
b Defined as BMI $25.0-29.0 \mathrm{~kg} / \mathrm{m}^{2}$; the relative risk is calculated from each study's provision of the distribution of BMIs within the studied population.
c Defined as BMI $\geq 30.0 \mathrm{~kg} / \mathrm{m}^{2}$; the relative risk is calculated from each study's provision of the distribution of BMIs within the studied population.
Notes: 27 eligible studies but analysis restricted to cohort studies with incident cases only; adjustment for age, reproductive factors, alcohol and diet did not materially affect the relationship.
No effect was seen when restricting the data to incident cases; all studies accounted for age, diet (where data were available), alcohol and/or physical activity.
Data provided for men but the data were almost identical for women (I.07, CI I.05-I.09). Studies were restricted to incident cases, with age adjustment and allowance for smoking but these restrictions did little to influence the results from the seven studies considered.

Source: Bergström et al. (2001).
suggest that women with a higher BMI at the time of diagnosis have poorer survival and an increased likelihood of recurrent breast cancer, irrespective of their menopausal status and after adjusting for the stage of cancer development and the type of treatment used.

Mechanistically, it seems clear that the risk of developing postmenopausal breast cancer is increased in women with raised plasma and tissue concentrations of estrogens. The activity of these hormones is greater when there are lower circulating concentrations of the sex hormone-binding globulin (SHBG). Obesity, with its associated insulin resistance, lowers SHBG levels; overweight women are also found to have higher circulating concentrations of total and bioavailable androgens and estrogens. Confirmation of the importance of these hormonal changes comes from the observation that women exposed to combined estrogens and progesterones as part of postmenopausal hormone replacement therapy subsequently have increased rates of breast cancer, the risk being greater in those on combined compared with estrogen-alone treatment (Weiderpass et al. 1999). The reduced risk associated with a late menarche and in women with anovulatory cycles is considered to relate to the lower exposure of the breast to bioactive estrogens (IARC 2002).

## Colon cancer

Bergström et al. (2001) considered 19 studies, 12 of which were prospective, which generally tended to show a stronger relationship between excess weight and incidence of cancer in men than in women. The studies also allowed an assessment to be made of the confounding effects of physical activity, age, family history of colon cancer, ethnicity, social class and diet. For the full quantitative analyses, only six studies could be included (Gerhardsson et al. 1990; Giovannucci et al. 1995; Kune et al. 1990; Lee and Paffenbarger 1992; Martinez et al. 1997; Thun et al. 1992); and in these studies no sex-specific differences could be found, nor did restricting the analysis to incident cases alter the estimate. The same general relationships of weight to the development of large colonic adenomas were found in the IARC analyses (IARC 2002), the development of adenoma being seen as part of the progression of cellular changes leading to the development of colon cancer. Although fewer studies have considered rectal cancers separately, no relationship was found between BMI and rectal cancer.

The mechanisms by which weight gain might accentuate the risk of developing large adenomas and colon cancer are unclear, but the stronger association of high BMIs with large rather than small adenomas suggests that excess weight operates at a relatively late stage in the promotion of tumour formation. Excess weight is associated with a wide range of hormonal and metabolic effects that may be involved in the promotion of cancer. Dietary factors could, in theory, be confounders with high meat intake, especially processed meat, and a low intake of fibre-rich vegetables and fruit being particularly linked to colon cancer and also being part of a weight-gain-inducing, energy-dense diet. However, several of the studies also assessed diet and the impact of higher BMIs seemed to be independent of the direct dietary effects.

## ENDOMETRIAL CANCER

Both case-control and cohort studies have shown a relationship between higher BMI and increased risk of developing endometrial cancer, even after adjusting for other risk factors relating to the reproductive system, such as age at birth of first child and parity (Le Marchand et al. 1991). There was a remarkably consistent relationship between high BMI and risk of endometrial cancer in 22 of the 25 studies assessed by IARC (2002), studies which considered only cancer incidence and adjusted for all suggested confounders showing similar relationships. Bergström et al. (2001) used four of the seven cohort and 17 case-control studies reviewed by IARC (2002) to calculate by meta-analysis the incremental risk shown in Table 8.38. The overall risks of endometrial cancer appeared to be equivalent at different ages, but adult weight gain seems to be particularly important whatever the early adult weight status. Upper body fatness may be particularly conducive to the process of car-
cinogenesis, but the standard measures of abdominal fatness have given inconsistent results.

The dominant mechanistic theory relates to the unopposed estrogen hypothesis, according to which estrogenic contraceptives or hormone replacement therapy enhance the risk of endometrial cancer, whereas progesterone-containing preparations confer protection. Estrogens are known to induce endometrial proliferation via local production of insulin growth factor (IGF-1), whereas progesterone induces the production of an endometrial IGF-1-binding protein. Women with low levels of plasma SHBG, high levels of androgens and, after the menopause, elevated levels of total and bioavailable estrogens have an increased risk of endometrial cancer, as have younger women with the polycystic ovarian disease, which is associated with chronic anovulation and therefore low rates of production of progesterone. All these findings, therefore, fit the concept of excess available bioactive estrogen, which induces endometrial cell proliferation. Insulin resistance and higher concentrations of circulating IGF-1 induced by the lower concentrations of IGF-binding proteins in women who gain weight may also be involved. The IARC report notes that, given the substantial changes in insulin resistance, IGF-1 and estrogen status which accompany weight loss, it is possible that weight reduction quite late in life could reduce the risk of the estrogen-promoted cancers of the postmenopausal breast and endometrium.

### 5.8 Body weight and total mortality

Clearly, the net associations of BMI with total mortality will depend crucially on the component causes of death, which vary substantially by age, sex and population. Effects on total mortality were not estimated directly in these analyses for this reason, only cause-specific estimates being made. However, for completeness, data relating BMI to total mortality (usually derived from middle-aged North American populations) are reviewed here.

An extensive review of the relationship between BMI and total allcause mortality, based on a detailed systematic review, was published by Troiano et al. (1996). About 1000 citations dating from 1861 to 1991 were identified; of these, 22 suitable studies were selected, with 56 substudy groups (e.g. in relation to age, ethnicity, sex and smoking status) and 354 BMI groups. Most of the studies dealt with Caucasian men, but 14 substudies assessed Caucasian women. Only two substudies in Asian men (not women) were available and one substudy in Samoan men and women combined was assessed. The authors distinguished data derived from insurance companies from data derived from other populations because they demonstrated that insurance data, particularly those from the United States, were associated with lower mortality in relation to BMI than that found in the general population. It was inferred that these data, which formed the basis of many earlier official reviews on the
impact of obesity for governments, for example, in the United States and the United Kingdom, related to groups in society who were relatively affluent and therefore, for a variety of reasons, able to sustain better health.

Troiano et al. (1996) demonstrated evidence of a U-shaped curve of mortality in relation to BMI, but this curve varied depending on whether the data were derived from the United States or elsewhere (predominantly northern Europe and Scandinavia). When all data were included, as shown in Table 8.39, there was clear evidence of a minimum mortality at BMIs of $24-27 \mathrm{~kg} / \mathrm{m}^{2}$, the lower value being that obtained with longer, i.e. 30 -year follow-ups of adults. Most of these data relate to individuals who were aged about 50 years and were followed for 30 years. The table dealing with all the studies shows a statistically significant increase of $z$-scores when the $z$-scores exceed 1.65 . Clearly, even BMIs of $\leq 23$ and $\geq 28 \mathrm{~kg} / \mathrm{m}^{2}$ are associated with higher mortality in these overall groups. It is noteworthy, however, that these analyses included

Table 8.39 The probability of death associated with BMI level, for either 10 or 30 years of follow-up (smokers and nonsmokers included)

| BMI ( $\mathrm{kg} / \mathrm{m}^{2}$ ) | Odds ratio |  | $z$-score differences |  |
| :---: | :---: | :---: | :---: | :---: |
|  | 10 years | 30 years | 10 years | 30 years |
| 19 | 2.08 | 2.89 | 1.16 | 2.29 |
| 20 | 1.86 | 1.97 | 1.64 | 2.43 |
| 21 | 1.65 | 1.48 | 2.00 | 2.50 |
| 22 | 1.46 | 1.21 | 1.74 | 2.32 |
| 23 | 1.30 | 1.06 | 1.25 | 1.68 |
| 24 | 1.16 | - | 0.84 | - |
| 25 | 1.07 | 1.00 | 0.50 | 0.11 |
| 26 | 1.01 | 1.07 | 0.18 | 0.82 |
| 27 | - | 1.20 | - | 1.45 |
| 28 | 1.04 | 1.39 | 0.63 | 2.10 |
| 29 | 1.15 | 1.68 | 1.27 | 2.87 |
| 30 | 1.36 | 2.12 | 2.18 | 3.84 |
| 31 | 1.73 | 2.69 | 3.04 | 4.84 |
| 32 | 2.23 | 3.49 | 3.13 | 5.03 |

[^34]both smokers and non-smokers. Non-smokers, considered over the 30 -year period, had systematically lower risks at any BMI than smokers and Troiano et al. (1996) used these data for their baseline mortality curves.

There has been extensive discussion over the last 30 years regarding the repeated finding of higher mortality rates associated with lower BMIs. It was recognized that the original inclusion of data from smokers in such calculations had a marked effect because smokers are at greater risk of mortality, but tend to be thinner because of their reduction in appetite and their increased metabolic rate, that is, increased total energy expenditure, which leads to lower body weights when these effects are in energy balance. Thus the excess of smokers in the group of "thin" adults imposes higher mortality rates on the group overall, despite the lower BMIs.

Further, the mortality rates of the groups with low BMIs may be enhanced by the presence of individuals with as yet undiagnosed diseases, for example, cancer, who may have lost weight before symptoms emerged or a diagnosis was made. A convention has therefore developed whereby the early deaths are excluded and only those deaths that occur $2-5$ years after the initiation of any study are considered. By doing this, it is frequently found that the U-shaped curve converts to a J-shaped curve or $\log$-linear relationship.

The data previously considered were standardized to men and women aged about 50 years. Stevens et al. (1998) have recently presented a detailed analysis based on about 325000 men and women taking part in the American Cancer Society's Cancer Prevention Study. As already noted, this analysis in the United States relates to a relatively affluent fraction of the population and unfortunately used reported, not measured, weights and heights. Nevertheless, the sample is valuable in indicating the likely effect of age when assessing total mortality rates over a 12 -year follow-up period. For this purpose, data were expressly chosen which related only to life-long non-smokers. Figure 8.14 shows the agerelated risk of death from all causes. Note that men and women with a BMI of $19.0-21.9 \mathrm{~kg} / \mathrm{m}^{2}$ had the lowest total mortality. Small increases in risk were apparent at BMIs of $22.0-27.0 \mathrm{~kg} / \mathrm{m}^{2}$, but the increase in relative risk was more obvious in men than in women and in those aged $<75$ years. The risk of death from cardiovascular disease related to BMI was more clear-cut than the relative risk for all causes of death, as found in many studies (see below).

Figure 8.15 shows the decline in relative risk per unit increase in BMI with age. A very clear trend is seen, with the incremental risks being statistically significantly above those for the reference group up to the age of 75 years.

From Figure 8.14, it seems reasonable to conclude that the ideal BMI should be between 19.0 and $21.9 \mathrm{~kg} / \mathrm{m}^{2}$. These values relate to the individual risk of death in groups within a single population (in this

Figure 8.14 The relative risk of death from all causes, according to age and BMI: American Cancer Society's studies using selfreported weights and heights for BMI calculations


[^35]Figure 8.15 The change with age in the relative risk of death from all causes associated with a one-unit ( $1 \mathrm{~kg} / \mathrm{m}^{2}$ ) increase in BMI, in never-smokers


Note: The bars represent $95 \%$ confidence intervals and the trends are significant in all cases.
Source: Taken from Stevens et al. (1998) with the reference group having a BMI of $19.0-21.9 \mathrm{~kg} / \mathrm{m}^{2}$ in each age group.
case, the United States) and seem to apply to all age groups considered in the CRA analysis, from age 30-79 years. These findings are in keeping with the earlier analyses of the relationship between mean BMI and the minimizing of the prevalence of overweight, but would allow a higher proportion of underweight, which in a developing country would be a disadvantage. Given that the circumstances encountered by thin adults in the United States are probably different to those in developing countries, it seems reasonable to conclude that the data reinforce the choice of a low value for the minimum population mean BMI of $21.0-22.0 \mathrm{~kg} / \mathrm{m}^{2}$.

### 5.9 Excluded health outcomes

A number of other conditions have been widely quoted as concomitants of body-weight gain in affluent societies, including breathlessness, sleep apnoea, back pain, dermatitis, reactive depression and social isolation, menstrual disorders, infertility and gall bladder disease. The reasons for excluding these conditions are as follows.

Breathlessness: this is hard to quantify on an international basis and comparable data to those obtained in affluent societies in relation to excess weight are difficult to find in many parts of the world. It is also not clear to what extent different degrees of physical fitness in different parts of the world might affect an assessment of perceived breathlessness in those with excess weight.

Sleep apnoea is a well-described clinical complication of excess weight, which affects very obese children and adults. There are few data on its prevalence however, and it is generally considered to occur in those with more extreme forms of obesity.

Back pain: there are many causes of back pain, which is a very prevalent condition with substantial economic implications. Few studies however provide reliable estimates of hazard due to any specific risk.

Dermatitis: skin problems occur commonly in people who are obese, but few data are available other than from the developed world. Again, the description of skin problems associated with obesity is largely clinical and therefore this condition is not included.

Reactive depression and social isolation: it is well recognized that the societal response to obese individuals varies widely across the world. In some developing countries, the overweight and obese have traditionally been seen as successful individuals who are sufficiently wealthy or resourceful to have acquired enough food. In these circumstances, there is little indication of any social isolation or ensuing depression resulting from the overweight and obese being excluded from social interactions. This is quite different from accounts widely reported in North America and Europe where being obese, particularly for children and young women, is a social stigma which has clearly been related to poorer access to employment opportunities, lower earning power, a tendency to marry less affluent partners and a tendency to become personally distressed and socially isolated. Given the diversity in cultural perceptions of the benefits or handicaps of being overweight or obese, no attempt has been made at this stage to use representative data on body weight in adults in relation to mental health outcome.

Menstrual disorders and infertility: although severe underweight and anorexia nervosa have classically been associated with amenorrhoea and infertility, obesity is now increasingly recognized as a major feature of, for example, polycystic ovarian disease, which is associated with menstrual abnormalities, hirsutism and infertility. Weight loss markedly improves the condition of patients with this disorder and restores fertility. Little is known about the prevalence of this disease in different societies. The issue of infertility does not seem as yet to be of great societal concern if considered simply in terms of the probability of maintaining the population size. There is also no suggestion that the marked decline
in fertility seen, for example, in Europe (France, Italy and Spain), relates to the increasing prevalence of adiposity. Societal and social issues appear to be of far greater importance, so this outcome is also excluded from the current analysis.
Gallbladder disease: it has been recognized for several decades that excess weight gain is associated with a greater propensity to the development of gallstones and gall bladder disease. This relationship is clearcut in developed societies, but as data on gall bladder disease are not collected systematically in many countries it is not possible to undertake an appropriate international analysis of the risk associated with weight gain.

### 5.10 Risk reversibility

The evidence on the reversibility of health hazards after reducing excessive BMIs has been given in each section setting out the relationship between increases in BMI and the development of individual diseases such as type II diabetes, ischaemic heart disease, hypertension and stroke, as well as cancers. The speed of reversibility depends on the condition. Thus an elevated blood pressure associated with weight gain can start to reverse within days of the beginning of weight loss; in association with other dietary measures and increased physical activity, lower body weights can also then limit the incidence of hypertension for 10 years (Stamler et al. 1980). Changes in blood lipids also begin within days of weight loss, although the restitution of low concentrations of HDL cholesterol to normal requires a period of several weeks at a stable lower body weight (Dattilo and Kris-Etherton 1992). The impact of weight loss on the development of ischaemic heart disease is more difficult to distinguish from other dietary changes accompanying the intended weight loss. Thus when individuals at a high risk of suffering a myocardial infarction are trained to markedly reduce their intake of fat, together with their intake of salt and sugars, as well as undertaking exercise training, they lose substantial amounts of weight and show a marked decrease not only in the principal risk factors for ischaemic heart disease (i.e. hypertension, dyslipidaemias and glucose intolerance) but also show a reduction within weeks in the incidence of angina and then in rates of myocardial infarction (Ornish et al. 1990); this is clearly evident within a 5 -year period. The time needed to alter insulin resistance is also short (i.e. within a few days to weeks) and the resulting reduced incidence of type II diabetes becomes evident within 1 or 2 years, although to obtain statistically robust data in most studies about 3 years is needed (Tuomilehto et al. 2001).

## 6. Results

Tables $8.40-8.42$ show the proportion of the included diseases, the number of deaths and disease burden attributable to increases in BMI
above $21.0 \mathrm{~kg} / \mathrm{m}^{2}$ in the different subregions of the world in 2000. Figure 8.16 summarizes the contributions of the diseases considered in this analysis to the total global burden of ill-health attributable to the effect of high BMI in 2000.

No lives are lost because of arthritis, but the global total mortality for cancers is appreciable, amounting to 74000 for colon cancer, 47000 for breast cancer and 32000 for endometrial cancer, i.e. a total of 153000 cancer deaths in 2000. However, these figures are dwarfed by the 491000 deaths from diabetes, 489000 from ischaemic stroke, 1168000 from ischaemic heart disease and 290000 from hypertensive disease. Thus by considering only the major diseases affected by high BMI, a total of 2592000 deaths in 2000 were attributable to this risk factor. There are marked regional differences in this mortality burden, some of which reflect major differences in population sizes.

It is evident from the impact of weight gain on risk of arthritis that those subregions with a large proportion of adults with high BMIs, for example, AMR-A, AMR-B, EUR-A and EUR-B, are likely to have more of a burden linked directly to the physical demands of weight bearing, but much depends on the quality of the documentation of degrees of handicap associated with arthritis globally. The proportions of diabetes attributable to BMI increases range from $38-88 \%$ according to subregion. For diabetes and cardiovascular diseases, while the gradient of

Figure 8.16 The contribution of high BMI to the global burden of illhealth in the year 2000

Table 8.40 Percentage of different diseases which is attributable to high BMI in adults aged $\geq 30$ years, by subregion

| Subregion | Osteoarthritis |  | Colon cancer |  | Postmenopausal breast cancer |  | Endometrial cancer |  | Type II diabetes |  | Stroke ${ }^{\text {a }}$ |  | Ischaemic heart disease |  | Hypertensive disease |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female |
| AFR-D | 6 | 6 | 4 | 5 | NA | 3 | NA | 15 | 38 | 46 | 14 | 13 | 13 | 12 | 24 | 22 |
| AFR-E | 7 | 9 | 5 | 7 | NA | 4 | NA | 25 | 44 | 58 | 17 | 20 | 15 | 19 | 29 | 32 |
| AMR-A | 22 | 24 | 17 | 18 | NA | 12 | NA | 52 | 83 | 88 | 40 | 37 | 37 | 32 | 63 | 58 |
| AMR-B | 17 | 23 | 13 | 16 | NA | 10 | NA | 46 | 71 | 83 | 34 | 40 | 32 | 37 | 52 | 60 |
| AMR-D | 18 | 21 | 13 | 14 | NA | 8 | NA | 41 | 69 | 77 | 31 | 34 | 31 | 32 | 52 | 54 |
| EMR-B | 15 | 15 | 10 | 13 | NA | 7 | NA | 43 | 62 | 72 | 26 | 32 | 30 | 32 | 45 | 51 |
| EMR-D | 10 | 15 | 7 | 11 | NA | 5 | NA | 34 | 55 | 72 | 18 | 25 | 21 | 26 | 36 | 44 |
| EUR-A | 22 | 24 | 18 | 18 | NA | 13 | NA | 50 | 78 | 84 | 35 | 31 | 35 | 29 | 56 | 50 |
| EUR-B | 15 | 22 | 15 | 19 | NA | 12 | NA | 48 | 71 | 84 | 35 | 41 | 34 | 38 | 56 | 65 |
| EUR-C | 17 | 25 | 14 | 19 | NA | 13 | NA | 52 | 68 | 83 | 33 | 37 | 33 | 33 | 56 | 64 |
| SEAR-B | 9 | 10 | 7 | 8 | NA | 4 | NA | 24 | 44 | 56 | 18 | 21 | 17 | 19 | 32 | 36 |
| SEAR-D | 2 | 5 | 1 | 4 | NA | 2 | NA | 14 | 13 | 43 | 3 | 10 | 4 | 10 | 7 | 21 |
| WPR-A | 11 | 12 | 8 | 9 | NA | 5 | NA | 28 | 55 | 65 | 20 | 18 | 19 | 15 | 31 | 27 |
| WPR-B | 9 | 10 | 7 | 8 | NA | 5 | NA | 26 | 47 | 62 | 18 | 19 | 17 | 18 | 30 | 34 |
| World | 11 | 14 | 11 | 13 | NA | 8 | NA | 42 | 50 | 66 | 21 | 25 | 21 | 22 | 36 | 41 |
|  | plicable age of ccomp | ke attrib ng this b | ble to ). | BMI wa | sed up | estimates | r ischae | c stroke | fractio | for total | oke are | ven in the | nex tab | for the | r (se | he CD- |

Table 8.4I Mortality (000s) from diseases caused by high BMI in adults aged $\geq 30$ years, by subregion

| Subregion | Osteoarthritis | Colon cancer | Postmenopausal breast cancer | Endometrial cancer | Type II diabetes | Stroke | Ischaemic heart disease | Hypertensive disease | Total |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | 0 | 0 | 0 | 0 | 7 | 7 | 14 | 4 | 32 |
| AFR-E | 0 | I | 1 | 0 | 16 | 11 | 20 | 7 | 56 |
| AMR-A | 0 | 12 | 8 | 4 | 62 | 28 | 135 | 22 | 271 |
| AMR-B | 0 | 4 | 4 | 5 | 103 | 35 | 78 | 33 | 262 |
| AMR-D | 0 | 1 | 0 | 2 | 11 | 4 | 9 | 6 | 33 |
| EMR-B | 0 | 1 | 0 | 0 | 10 | 5 | 33 | 14 | 63 |
| EMR-D | 0 | 1 | 2 | 0 | 22 | 16 | 65 | 19 | 125 |
| EUR-A | 0 | 25 | 14 | 8 | 70 | 66 | 167 | 29 | 379 |
| EUR-B | 0 | 5 | 3 | 3 | 23 | 51 | 137 | 36 | 258 |
| EUR-C | 0 | 10 | 6 | 6 | 16 | 124 | 284 | 21 | 467 |
| SEAR-B | 0 | 2 | 1 | 1 | 32 | 13 | 34 | 19 | 102 |
| SEAR-D | 0 | 1 | 2 | 0 | 35 | 24 | 82 | 9 | 153 |
| WPR-A | 0 | 3 | I | 1 | 9 | 10 | 15 | 2 | 41 |
| WPR-B | 0 | 9 | 4 | 2 | 74 | 95 | 95 | 68 | 347 |
| World | 0 | 74 | 47 | 32 | 491 | 489 | 1168 | 290 | 2592 |

Table 8.42 Burden of disease in DALYs (000s) from diseases caused by high BMI in adults aged $\geq 30$ years, by subregion

| Subregion | Osteoarthritis | Colon cancer | Postmenopausal breast cancer | Endometrial cancer | Type II diabetes | Stroke | Ischaemic heart disease | Hypertensive disease | Total |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | 37 | 6 | 5 | 2 | 148 | 110 | 199 | 57 | 564 |
| AFR-E | 55 | 9 | 14 | 5 | 240 | 188 | 282 | 94 | 887 |
| AMR-A | 243 | 104 | 77 | 46 | 1171 | 400 | 1243 | 195 | 3479 |
| AMR-B | 193 | 41 | 40 | 77 | 1367 | 504 | 892 | 309 | 3423 |
| AMR-D | 22 | 5 | 5 | 22 | 164 | 51 | 91 | 62 | 422 |
| EMR-B | 33 | 7 | 6 | 4 | 273 | 84 | 447 | 136 | 990 |
| EMR-D | 70 | 12 | 17 | 6 | 515 | 217 | 865 | 208 | 1910 |
| EUR-A | 348 | 185 | 127 | 66 | 876 | 619 | 1271 | 166 | 3658 |
| EUR-B | 176 | 46 | 35 | 37 | 401 | 559 | 1304 | 306 | 2864 |
| EUR-C | 266 | 95 | 66 | 69 | 526 | 1291 | 2741 | 208 | 5262 |
| SEAR-B | 85 | 21 | 18 | 7 | 555 | 175 | 406 | 201 | 1468 |
| SEAR-D | 95 | 8 | 20 | 5 | 916 | 309 | 1156 | 117 | 2626 |
| WPR-A | 74 | 31 | 11 | 9 | 224 | 135 | 134 | 11 | 629 |
| WPR-B | 421 | 104 | 45 | 33 | 1503 | 1280 | 1139 | 709 | 5234 |
| World | 2118 | 676 | 486 | 386 | 8877 | 5921 | 12170 | 2780 | 33414 |

Figure 8.I7 Selected subregional differences in the age-related proportion of type II diabetes attributable to increases in BMI in women

relative risk was considered the same in the different subregions, the absolute risk varied substantially because of other determinants of these diseases such as dietary variation, which causes additional changes in both serum cholesterol levels and blood pressure.

Figure 8.17 presents subregional differences in the age-related proportion of type II diabetes attributable to high BMI in women. The picture of differences is magnified and the clear downward gradient in the attributable fraction with increasing age is evident. Similar relationships are observed for all three cardiovascular end-points, but not for the cancers or osteoarthritis, where relative risks were independent of age.

## 7. Discussion

These analyses highlight the very substantial burden of ill-health incurred by increases in adult BMIs with the greatest impact in disease specific terms being the burden associated with the development of type II diabetes. The attributable fraction for high BMIs does vary markedly by subregion, which implies that other factors contribute to or interact with increases in BMI. The overall burden by subregion, however, depends on
the prevalences of both high BMIs and of the particular disease (e.g. ischaemic heart disease) as well as the overall size of the population. The current analyses do not take account of the interactions with other risk factors, for example, physical inactivity, which could confound the estimated burden attributable to high BMIs per se. However, increases in body weight also amplify other major risk factors such as blood pressure and blood cholesterol levels, and so discriminating a selective effect of high BMIs from dietary factors and physical activity is not straightforward. Nevertheless, on the basis of the current data, the impact of increases in BMI on the development of type II diabetes and on cardiovascular diseases and cancers in most parts of the world is substantial.

This is the first global analysis of the risks attributable to high BMIs and the results reinforce the recent recognition by WHO (2000) that this is one of the largest unrecognized public health problems that now need to be addressed. With the seemingly inexorable rise in mean BMIs in various populations, the projected impact on the global health burden will be very substantial by 2030 unless effective public health measures can be introduced soon.

## 8. Future exposure

### 8.1 Regional time trends in adult BMIs

There are as yet only a modest number of studies dealing with BMI changes in different populations, the most comprehensive being that conducted by Pelletier and Rahn (1998). In this study, a series of small data sets were found for several countries within different regions of the world. The data sets differed by sex, age and setting, i.e. whether they were studied in a clinical context, were related to urban or rural groups or differed in terms of their economic status. By specifying these variables, Pelletier and Rahn (1998) were able to identify the magnitude and significance of these variables and still develop generic equations for the overall BMI changes with time in different regions. These estimates of time trends per decade are presented in Table 8.43.

Given the fact that these data relate to the 1980s and that the equations are often based on data from clinics or other small groups, more recent evidence from larger and more representative samples was sought, again with the specification that the BMIs should be measured and that mean BMI values were available.

Popkin et al. (2002) have recently presented valuable new data on annual increments in the prevalence of overweight and obesity. Unfortunately this approach does not use the continuous approach to BMI-risk relationships and extrapolating from the equations for the curvilinear relationship between the mean BMI and the prevalence of

Table 8.43 Predicted regional rates of change in BMI per decade for adults

| Country or area | Mean BMI change per decade |
| :--- | :---: |
| Sub-Saharan Africa | 0.207 |
| South and South-East Asia | 0.169 |
| India | 0.048 |
| Australasia | -0.295 |
| Polynesia and Micronesia | 0.957 |
| Latin America and the Caribbean | 0.112 |
| China | 0.106 |

obesity at BMIs of $\geq 30 \mathrm{~kg} / \mathrm{m}^{2}$ (see Figure 8.2 ) proved too inaccurate, given the data provided on annual increment in obesity and the need to extrapolate to 2030.

There is, however, a series of recent analyses that attempt to look at changes in body weight over the last 10-40 years. Some of these analyses, for example from Brazil (Monteiro et al. 1995), several European countries (Molarius et al. 2000), India (Shetty 2002), Japan (Yoshiike 2002), and the United States (Flegal and Troiano 2000), are based on repeated national representative data and allow some preliminary subregional estimates to be developed. To these data it was possible to add background data sets while recognizing that many were not representative of a country, let alone a subregion and that sex-specific rates may well be very different in different societal settings.

The uncertainty associated with assuming that both sexes and all age groups are likely to respond in the same way over the next 30 years is illustrated by Figure 8.18, which sets out the detailed age- and sex-based time trends in mean BMI for the Japanese population (Yoshiike 2002). It is evident that whereas the mean BMI for men has increased in a linear fashion between 1977 and 1995 in all age groups, the mean BMI for women aged $<50$ years has shown a progressive reduction amounting to about 0.3 BMI units per decade. Given that the mean BMI is now $20.4 \mathrm{~kg} / \mathrm{m}^{2}$ for young women aged $20-29$ years and $22.2 \mathrm{~kg} / \mathrm{m}^{2}$ for women aged $30-49$ years, it is clear that predicting a continuation of this trend for the subsequent 35 years, that is, to 2030, would mean that the average BMI of the younger women would fall to $19.1 \mathrm{~kg} / \mathrm{m}^{2}$; this in turn would imply an appreciable increase in the proportion of underweight women with potentially increased morbidity (Shetty and James 1994). Yet we know that Japanese children, both girls and boys, are becoming heavier. Therefore modelling the changes in BMI by age and sex for the next 30 years presents difficulties. What is clear, however, is that the age-related changes in BMI were

Figure 8.18 Annual changes in mean BMI, by age and sex


Note: Adjusted for age distribution by use of the new world population (WHO 1993).
Source: The National Nutrition Survey, Japan, 1976-1995.
maintained from 1976 to 1995 so these are unlikely to change, unless the weights of Japanese children have shown a particularly acute recent increase.

Bearing in mind the uncertainty of these predictions, the assumption was made that the current trends in any subregion would persist in unremitting manner for the next 30 years. It was assumed that data obtained from a country within a subregion would apply to the whole adult population within that subregion. Where more than one data set was available, the values were adjusted not only for the intervals of study and the date on which measurements were made, but also for the differential population numbers in the countries being measured, with appropriate adjustments to the overall population numbers in the subregion. Only adult data were considered; no allowances were made for
children's BMIs since this would have additionally required the prediction of the probability that the current BMIs of children, expressed in percentiles of BMI for age and sex, would be maintained on the same percentiles into adult life on the basis of the BMI percentile charts produced by the IOTF (Cole et al. 2000). Where the data only covered adults aged $\leq 60$ years, it was assumed that the same delta changes were also occurring in the older age groups, thus maintaining the current agerelated differences in mean BMI within the subregion.

## Different subregional approaches

Given the paucity of data, the following extrapolations proved necessary for certain subregions.
$A F R-D$ : The data relied predominantly on the observed changes in BMI in Ghanaian women, with additional data being available from Mauritius and the Seychelles. The latter two island data sets for both sexes were then used in conjunction with data from South Africa to assess the differential sex-specific trends. These differentials were then used in conjunction with the Ghanaian female data to derive male values for the subregion.
$A F R-E$ : Here it proved necessary to use data on the secular increases found in rural South African men and women (Temple et al. 2001). It was expected that urban trends might have been greater, but given categorical analyses suggesting more modest increases in the prevalence of overweight in the United Republic of Tanzania and Zambia, the rural South African data were applied to the whole subregion.
$A M R-A$ : This was relatively straightforward given the availability of data sets from both Canada (Tremblay et al. 2002) and the United States (Flegal and Troiano 2000). The changing prevalence of overweight and obesity in Cuba was noted (Rodriguez-Ojea et al. 2002), but it was not possible to obtain suitable data on mean BMI changes. Therefore the North American data sets were applied to the whole subregion.
$A M R-B$ : There are ample Brazilian data on secular trends and a series of unpublished analyses from other countries, but the only published and available Brazilian data were those giving overall adult values (Monteiro et al. 1995). However, new analyses from a national survey carried out in Mexico in 2000 were made available (Sánchez-Castillo et al. 2003), together with earlier representative Mexican data provided by Arroyo and colleagues (Arroyo et al. 2000).
$A M R-D$ : In the absence of satisfactory data, the equations of Pelletier and Rahn (1998) were applied.
$E M R-B$ : Kuwaiti data (al-Isa 1997) were applied to the whole subregion, it being recognized that there were numerous published data sets from small clinical and other studies available, as well as several recent unpub-
lished national data sets which highlight the marked secular trends in the subregion.
EMR-D: Moroccan categorical analyses (Benjelloun 2002) having suggested marked BMI increases, the Kuwaiti data were also applied to this subregion.

EUR-A: There were ample data sets from Belgium, Denmark, Finland, Germany, the Netherlands and the United Kingdom that could be used to derive an overall set of population-adjusted, sex-specific and agerelated predictions.

EUR-B: Reliance was placed on data on secular trends in Poland.
EUR-C: Although some categorical Hungarian trends data were available, it was concluded from subregional cross-sectional data that the Polish trends data should be applied to this subregion.

SEAR-B: In the absence of data on mean BMI trends, Pelletier and Rahn's equation (1998) for south and east Asia was used.

SEAR-D: Indian data from the repeated nationally-representative surveys were used for this subregion.

WPR-A: Both Australian and the extensive Japanese representative data could be used for this subregion.

WPR-B: Although new data sets are becoming available for the numerically dominant country (China), comparable data sets with mean BMIs from a variety of adult men and women were only available from Zhou (2002). To these data sets were added information from Malaysia, the Republic of Korea and Samoa.

The predicted mean BMIs for the different subregions in 2030 are set out in Table 8.44. As already indicated, these predicted values encompass considerable uncertainties but do suggest that if current trends continue then in some subregions (e.g. AMR-A, AMR-B, EUR-A and EUR-B) half of the adults in each of several age groups will have estimated BMIs of $>30 \mathrm{~kg} / \mathrm{m}^{2}$, that is, the WHO classification of obesity. Given that there is now intense concern about escalating rates of obesity, it is very likely that new public health measures will be adopted to limit this rise, but so far the efforts of millions of people in affluent societies to either slim or limit weight gain seems to have been of only modest success. The challenge is to arrest the current trends towards increases in BMI and, if possible, to reverse the public health burden associated with weight gain. Although traditionally the prevalence rates of overweight and obesity are used as an index of the health burden, the current analyses show that the full range of BMIs should be considered. This in turn emphasizes the need to take population-based approaches to preventive strategies for minimizing the hazards of excess weight gain.

Table 8.44 Predicted mean BMIs in each subregion in 2030, by sex and age
$\left.\begin{array}{lllllll}\hline & & & & \left.\text { BMI (kg/m}{ }^{2}\right) \\ & & & & \\ \text { Age (years) }\end{array}\right]$
a Figures in parentheses refer to the percentage increase over the 2000 estimate.

### 8.2 The implications of excessive weight gain among children

The foregoing analyses of the disease burden associated with increases in BMI relate only to adults aged $\geq 30$ years. However, there is now rapidly mounting concern regarding the increasing prevalence of overweight and obesity in children (Ebbeling et al. 2002). The impact of psychosocial, physical and metabolic problems has not been incorporated in the current analyses, but may well have to be considered within the next $5-10$ years. Overweight children in many countries are handicapped by the social stigma of being considered overtly fat by their peers. In
addition, these children have a propensity to bone and joint deformation during their growing phase, as well as breathlessness and, in more extreme cases, sleep apnoea arising from the mechanical handicap of being heavy. It is now clear that overweight children also have higher blood pressure, serum lipid abnormalities and increasing insulin resistance, all of which are hallmarks of early metabolic disease and susceptibility to atherosclerosis and other cardiovascular problems (DiPietro et al. 1994; Must et al. 1992; Unger et al. 1990). Within the last 5 years, paediatricians have observed that clinics for children with type I diabetes now include a remarkably increasing number of very overweight children with type II diabetes. In many parts of the world, type II diabetes occurring in adolescence is now a more prevalent condition than type I diabetes (Rosenbloom et al. 1999). As with adult type II diabetes, these children show a remarkable improvement in their glucosehandling capacity and in other risk factors if they lose weight, but many soon develop a need for insulin therapy and their condition becomes difficult to control. Children with poorly controlled diabetes are known to have accelerated atherosclerosis with microvascular disease leading to blindness and renal failure in their early 30s. Concern regarding obesity and type II diabetes in children is now being accentuated by the recognition that mothers who develop gestational diabetes produce larger babies who are then very susceptible to pre-adolescent obesity and to the development of type II diabetes in adolescence (Silverman et al. 1998).

People who are overweight in childhood are more prone to be obese when they enter adult life and these individuals are also likely to continue to gain weight in adulthood. Such individuals then have up to a 100 -fold increased risk of developing type II diabetes compared with normal weight children who do not gain excessive weight once they are adults (Colditz et al. 1995). Early data from insurance companies also showed that heavy young adults had a much greater likelihood of suffering premature deaths (Blair and Haines 1966) and this is now recognized to be related to their greater propensity to hypertension, dyslipidaemia, glucose intolerance with insulin resistance and accelerated cardiovascular disease. These observations are relevant to any assessment of the future burden of disease arising from high BMIs in adults because the rapidly rising prevalence of overweight and obesity in children is likely to amplify the currently predicted risks of excess weight in young adults.

## Acknowledgements

We are most grateful for the support of numerous individuals and institutions in gathering data for this analysis. In addition we wish to acknowledge the assistance of Mrs Jean James in preparing the manuscript for this chapter.

The following contributed invaluably by providing re-analysed data to assist in the IOTF's analyses: N. Adra, P. Arroyo, K. Babinska, G. Biro, A. Deev, T.P. Gill, N. Hwalla, M. Lahti-Koski, B. Lajos, A. Ferro-Luzzi, A. Kaic-Rak, G. Mensink, J.C. Montero, A. Nielsen, I. Palczewska, G. Pekcan, N. Rak, C.P. Sánchez-Castillo, G. Whitlock and the Asia-Pacific Cohort Studies Collaboration.

In addition, the following assisted by allowing access to data or suggesting sources:
T. Athanassouli, W. Becker, Z. Brazdova, W. Chen, W. Freire, V. Gudnason, K. Habibullah, U. Hagman, B. Heitman, N. Katsilambros, J. Luthy, V. Lambert, A.O. Musaiger, A. Neilson, L. Ovesen, R. Prattala, W. Rand, K.S. Reddy, A. Robertson, J. Seidell, L. Steingrimsdottir, M.Q.-K. and K. Talukder, R. Uauy, and M. Vasconcellos.

## Notes

1 See preface for an explanation of this term.
2 The standard deviation (SD) was sometimes missing for different sexes and age groups. This problem was dealt with as follows:
SD missing for one sex; for example, there were some subregions where SDs were available only for females

For all subregions with SDs available for both men and women, SDs were in general higher in females than males. The mean difference between the SDs of females and males varied from $0.24-2.25$ units, with the largest differences observed between higher SD values. For example, subregions with SDs of $>5-6$ units in women tended to have SDs in these women which were $>1$ unit higher than those in men, whereas with female SD values of $3-4.5$, the sex differences in SDs were $<1$.

The mean of the differences when the SD values were $<1$ unit was 0.6 units, and for those $>1$ unit was 1.63 . As a result, to estimate male SD from female SD in subregions where SD values were $<5$ in most age groups, a simple subtraction of 0.6 from the SD for females was used to estimate the male SD. In other subregions where SDs were higher in females, a value of 1 was subtracted from the female SD to estimate the corresponding male SD. The same approach was applied to all age groups. These differences in SDs by sex are then reflected in the differences in prevalences of overweight and obesity in men and women with the same mean BMI.

For example, in AMR-D an SD was available for females only and the values varied from 3.8-4.6 units. Subtracting 0.6 from these gave the corresponding values for men of 3.2, 3.8 and 4.0. In EMR-D, again only SDs for women were obtained with values varying from 6.4 to 8.5 . In this case, the correction factor of 1.6 was applied to obtain the estimates for men.
SD missing for some age groups
There was no clear pattern of SD by age group. For the age groups where no SDs were available, it was assumed that the SD was equal to an average
of the observed SD for the subregion. For example, in AMR-D, the mean SD for females aged $18-59$ years was $(3.4+4.4+4.6) / 3=4.13$.

This figure was then applied to the remaining age groups where no SDs for females were available.

## References

Acheson RM, Kelsy JL, Ginsburg GN (1974) The New Haven survey of joint diseases. XVI. Impairment, disability and arthritis. British Journal of Preventive Medicine, 27:168-176.
Ajlouni K, Jaddou H, Batieha A (1998) Obesity in Jordan. International Journal of Obesity and Related Metabolic Disorders, 22:624-628.
al-Isa AN (1995) Prevalence of obesity among adult Kuwaitis. International Journal of Obesity, 19:431-433.
al-Isa AN (1997) Temporal changes in body mass index and prevalence of obesity among Kuwaiti men. Annals of Nutrition and Metabolism, 41:307-314.
Allison DB, Fontaine KR, Manson JE, Stevens J, VanItallie TB (1999) Annual deaths attributable to obesity in the United States. Journal of the American Medical Association, 282:1530-1538.
al-Mannai A, Dickerson JW, Morgan JB, Khalfan H (1996) Obesity in Bahraini adults. Journal of the Royal Society of Health, 116:30-32, 37-40.
al-Nuaim AR, al-Rubeaan K, al-Mazrou Y, al-Attas O, al-Daghari N, Khoja T (1996) High prevalence of overweight and obesity in Saudi Arabia. International Journal of Obesity, 20:547-552.
Anderson JJ, Felson DT (1998) Factors associated with osteoarthritis of the knee in the first national health and nutritional examination survey (NHANES I). American Journal of Epidemiology, 128:179-189.
Anonymous (1999) Victorian burden of disease study: morbidity. Public Health Division, Department of Human Services, Melbourne.
Arroyo P, Loria A, Fernández V et al. (2000) Prevalence of pre-obesity and obesity in urban adult Mexicans in comparison with other large surveys. Obesity Research, 8:179-185.
Ascherio A, Rimm EB, Giovannucci EL et al. (1992) A prospective study of nutritional factors and hypertension among US men. Circulation, 86:1475-1484.
Barker DJP (1998) Mothers, babies and health in later life. Churchill Livingstone, London.
Bell AC, Swinburn BA, Simmons D, Wang W, Amosa H, Gatland B (2001) Heart disease and diabetes risk factors in Pacific Islands communities and associations with measures of body fat. New Zealand Medical Journal, 114:364-365.
Benjelloun S (2002) Nutrition transition in Morocco. Public Health Nutrition, 5:135-140.

Bergmann KE, Mensink GBM (1999) Korpermasse und Ubergewicht. Gesundheitswesen, 62:S115-120.

Bergström A, Pisani P, Tenet V, Wolk A, Adami HO (2001) Overweight as an avoidable cause of cancer in Europe. International Journal of Cancer, 91:421-430.

Berkham LF, Breslow L (1983) Health and ways of living: the Alameda county studies. Oxford University Press. New York.
Björntorp P, Rosmond R (1999) The hypothalamic origin of the metabolic Syndrome X. Annals of the New York Academy of Sciences, 892:297307.

Blair BF, Haines LW (1966) Mortality experience according to build at the higher durations. Transactions of the Actuarial Society, 18:35-41.

Bovet P, Shamlaye C, Kitua A et al. (1991) High prevalence of cardiovascular risk factors in the Seychelles (Indian Ocean). Arteriosclerosis Thrombosis, 11:1730-1736.

Calle EE, Thun MJ, Petrelli JM, Rodriguez C, Heath CW Jr (1999) Body-mass index and mortality in a prospective cohort of US adults. New England Journal of Medicine, 341:1097-1105.
Carey VJ, Walters EE, Colditz GA et al. (1997) Body fat distribution and risk on non-insulin dependent diabetes mellitus in women. The nurses' health study. American Journal of Epidemiology, 145:614-619.
Chaichareon P, Leeiahagul P, Tanphaichitr V (1992) Body mass index in adults with dental diseases. Internal Medicine, 8:113-117.
Chan JM, Rimm EB, Colditz GA, Stampfer MJ, Willett WC (1994) Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. Diabetes Care, 17:961-969.

Chilima DM, Ismail SJ (1998) Anthropometric characteristics of older people in rural Malawi. European Journal of Clinical Nutrition, 52:643649.

Cicuttini FM, Baker JR, Spector TD (1996) The association of obesity with osteoarthritis of the hand and knee in women: a twin study. Journal of Rheumatology, 23:1221-1226.
Cicuttini FM, Spector TD (1998) Obesity, arthritis and gout. In: Handbook of obesity. Bray GA, Bouchard C, James WPT, eds. Marcel Dekker Inc, New York.

Colditz GA, Willett WC, Rotnitzky A, Manson JE (1995) Weight gain as a risk factor for clinical diabetes mellitus in women. Annals of Internal Medicine, 122:481-486.
Colditz GA, Willett WC, Stampfer MJ et al. (1990) Weight gain as a risk factor for clinical diabetes mellitus in women. American Journal of Epidemiology, 132:501-513.

Cole TJ, Bellizzi MC, Flegal KM, Dietz WH (2000) Establishing a standard definition for child overweight and obesity worldwide: international survey. British Medical Journal, 320:1240-1243.

Cooper RS, Ford E (1992) Comparability of risk factors for coronary heart disease among blacks and whites in the NHANES-I Epidemiologic Follow-up Study. Annals of Epidemiology, 2:637-645.

Dattilo AM, Kris-Etherton PM (1992) Effects of weight reduction on blood lipids and lipoproteins: a meta-analysis. American Journal of Clinical Nutrition, 56:320-328.

Dawber TR, Meadors GF, Moore FE (1951) Epidemiological approaches to heart disease: the Framingham study. American Journal of Public Health, 41:279-286.
de Onis M, Blössner M (2000) Prevalence and trends of overweight among preschool children in developing countries. American Journal of Clinical Nutrition, 72:1032-1039.

Despres JP, Lemieux I, Prud'homme D (2001) Treatment of obesity: need to focus on high risk abdominally obese patients. British Medical Journal, 322:716-720.

Deurenberg P, Deurenberg-Yap M, Guricci S (2002) Asians are different from Caucasians and from each other in their body mass index/body fat percent relationship. Obesity Reviews, 3:141-146.
Deurenberg P, Yap M, van Staveren WA (1998) Body mass index and percent body fat: a meta-analysis among different ethnic groups. International Journal of Obesity, 22:1164-1171.

Diabetes Prevention Program Research Group (2002) Reduction in the incidence of Type 2 diabetes with lifestyle intervention or Metformin. New England Journal of Medicine, 346:393-403.

Dietz WH, Bellizzi MC (1999) Assessment of childhood and adolescent obesity. American Journal of Clinical Nutrition, 70:S117-175.

DiPietro L, Mossberg HO, Stunkard AJ (1994) A 40-year history of overweight children in Stockholm: life time overweight, morbidity and mortality. International Journal of Obesity and Related Metabolic Disorders, 18:585-590.
Dobson AJ, Evans A, Ferrario M et al. (1998) Changes in estimated coronary risk in the 1980s: data from 38 populations in the WHO MONICA project. World Health Organization. Monitoring trends and determinants in cardiovascular diseases. Annals of Medicine, 30:199-205.

Drivsholm T, Ibsen H, Schroll M, Davidsen M, Borch-Johnsen K (2001) Increasing prevalence of diabetes mellitus and impaired glucose tolerance among 60-year-old Danes. Diabetic Medicine, 18:126-132.

Dyer, AR, Elliot P (1989) The INTERSALT study: relations of body mass index to blood pressure. INTERSALT Co-operative Research Group. Journal of Human Hypertension, 3:299-308.

Eason RJ, Pada J, Wallace R, Henry A, Thornton R (1987) Changing patterns of hypertension, diabetes, obesity and diet among Melanesians and Micronesians in the Solomon Islands. Medical Journal of Australia, 146:465-469, 473.

Ebbeling CB, Pawlak DB, Ludwig DS (2002) Childhood obesity: public-health crisis, common sense cure. The Lancet, 360:473-482.

Edelstein SL, Knowler WC, Bain RP et al. (1997) Predictors of progression from impaired glucose tolerance to NIDDM: an analysis of six prospective studies. Diabetes, 46:701-710.

Eichholzer M, Luthy J, Gutzwiller F (1999) Epidemiology of overweight in Switzerland: results of the Swiss National Health Survey 1992-3. Schweizerische Medizinische Wochenschrift, 129:353-361.
el Mugamer IT, Ali Zayat AS, Hossain MM, Pugh RN (1995) Diabetes, obesity and hypertension in urban and rural people of Bedouin origin in the United Arab Emirates. Journal of Tropical Medical Hygiene, 98:407415.

Epstein FH, Napier JA, Block WD et al. (1970) The Tecumseh study: design, progress, and perspectives. Archives of Environmental Health, 21:402-407.
Eriksson KF, Lindgarde F (1991) Prevention of type 2 (non-insulin-dependent) diabetes mellitus by diet and physical exercise. The 6 -year Malmo feasibility study. Diabetologia, 34:891-898.
Flegal KM, Troiano RP (2000) Changes in the distribution of body mass index of adults and children in the US population. International Journal of Obesity, 24:807-818.
Ford ES, Giles WH, Dietz WH (2002) Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. Journal of the American Medical Association, 287: 356-359.

Franklin MF (1999) Comparison of weight and height relations in boys from 4 countries. American Journal of Clinical Nutrition, 70:S157-162.

Gerhardsson de Verdier M, Hagman U, Steineck G, Rieger A, Norell SE (1990) Diet, body mass and colorectal cancer: a case-referent study in Stockholm. International Journal of Cancer, 46: 832-838.
Giay T, Khoi HH (1994) Use of body mass index in the assessment of adult nutritional status in Vietnam. European Journal of Clinical Nutrition, 48:S124-130.
Giovannucci E, Ascherio A, Rimm EB, Colditz GA, Stampfer MJ, Willett WC (1995) Physical activity, obesity, and risk for colon cancer and adenoma in men. Annals of Internal Medicine, 122:327-334.
Goldstein DJ (1992) Beneficial health effects of modest weight loss. International Journal of Obesity, 16:397-415.
Gonzales GF, Villena A, Gonez C, Zevallos M (1994) Relationship between body mass index, age, and serum adrenal androgen levels in Peruvian children living at high altitude and at sea level. Human Biology, 66:145-153.
Gorzelniak K, Engeli S, Janke J, Luft FC, Sharma AM (2002) Hormonal regulation of the human adipose-tissue renin-angiotensin system: relationship to obesity and hypertension. Journal of Hypertension, 20:965-973.
Hanley AJ, Harris SB, Gittelsohn J, Wolever TM, Saksvig B, Zinman B (2000) Overweight among children and adolescents in a Native Canadian community: prevalence and associated factors. American Journal of Clinical Nutrition, 71:693-700.

Harris ML, Goldstein DA, Flegal KM et al. (1998) Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults. Diabetes Care, 21:518-524.

Harris TB, Ballard-Barbasch R, Madans J, Makuc DM, Feldman JJ (1993) Overweight, weight loss, and risk of coronary heart disease in older women. The NHANES I Epidemiologic Follow-up Study. American Journal of Epidemiology, 137:1318-1327.
Harris TB, Launer LJ, Madans J, Feldman JJ (1997) Cohort study of effect of being overweight and change in weight on risk of coronary heart disease in old age. British Medical Journal, 314:1791-1794.

Hart DJ, Spector TD (1993) The relationship of obesity, fat distribution and osteoarthritis in women in the general population: the Chingford study. Journal of Rheumatology, 20:331-335.

Helmrich SP, Ragland DR, Leung RW, Paffenbarger RS Jr (1991) Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. New England Journal of Medicine, 325:147-152.
Hernandez RE, Cardonnet LJ, Libman C, Gagliardino JJ (1987) Prevalence of diabetes and obesity in an urban population of Argentina. Diabetes Research and Clinical Practice, 3:277-283.
Higgins M, Kannel W, Garrison R et al. (1988) Hazards of obesity-the Framingham experience. Acta Medica Scandinavica, Supplementum, 723:23-36.

Hill AB (1965) The environment and disease: association or causation? Proceedings of the Royal Society of Medicine, 58:295-300.

Hochberg MC, Lawrence RC, Everett DF, Coroni-Huntley J (1989) Epidemiological associations of pain in osteoarthritis of the knee: data from the National Health and Nutrition Examination Survey and Nutrition Examination-I Epidemiological Follow-up Survey. Seminars in Arthritis and Rheumatism, 18:4-9.
Huang Z, Hankinson SE, Colditz GA et al. (1997) Dual effects of weight and weight gain on breast cancer risk. Journal of the American Medical Association, 278:1407-1411.
IARC (2002) IARC Handbook of cancer prevention, Vol. 6. Weight control and physical activity. Vainio H, Bianchini F, eds. International Agency for Research on Cancer, Lyon.

James WPT, Ferro Luzzi A, Waterlow JC (1988) Definition of chronic energy deficiency in adults. Report of a Working Party of the International Dietary Energy Consultative Group. European Journal of Clinical Nutrition, 42:969-981.

James WPT, Ralph A, Ferro-Luzzi A (1989) Energy needs of the elderly: a new approach. In: Human nutrition. Ageing and the elderly. Munro HN, Danford DE, eds. Plenum Press, New York.
James WPT, Francois P (1994) The choice of cut-off point for distinguishing normal body weight from underweight or 'chronic energy deficiency' in adults. European Journal of Clinical Nutrition, 48:S179-184.

James WPT, Norum K, Smitasiri S et al. (2000) Ending malnutrition by 2020: an agenda for change in the millennium. (Final report to the ACC/SCN by the Commission on the nutrition challenges of the 21st century. Supplement to the Food and Nutrition Bulletin, September/October 2000.) UNU International Nutrition Foundation, USA.
James WPT, Reeds PJ (1997) Nutrient partitioning. In: Handbook on obesity. Bray GA, Bouchard C and James WPT, eds. Marcel Dekker Inc, New York.
Jarrett RJ, Shipley MJ, Rose G (1982) Weight and mortality in the Whitehall study. British Medical Journal, 285:535-537.
Jimenez JT, Palacios M, Canete F et al. (1998) Prevalence of diabetes mellitus and associated cardiovascular risk factors in an adult urban population in Paraguay. Diabetic Medicine, 15:334-338.

Jones DW, Kim JS, Andrew ME, Kim SJ, Hong YP (1994) Body mass index and blood pressure in Korean men and women: the Korean National Blood Pressure Survey. Journal of Hypertension, 12:1433-1437.
Jousilahti P, Vartiainen E, Tuomilehto J et al. (1999) Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14786 middle-aged men and women in Finland. Circulation, 99:1165-1172.
Kain J, Uauy R, Vio F, Albala C (2002) Trends in overweight and obesity prevalence in Chilean children: comparison of three definitions. European Journal of Clinical Nutrition, 56:200-204.

Keys A, Aravanis C, Blackburn H et al. (1972) Coronary heart disease: overweight and obesity as risk factors. Annals of Internal Medicine, 77:15-27.

Khor GL, Yusof AM, Siong Tee E, Kandiah M, Lee Huang MS (1999) Prevalence of overweight among Malaysian adults from rural communities. Asia Pacific Journal of Clinical Nutrition, 8:272-279.
Kromeyer-Hauschild K, Zellner K, Jaeger U, Hoyer H (1999) Prevalence of overweight and obesity among school children in Jena (Germany). International Journal of Obesity, 23:1143-1150.
Kune GA, Kune S, Watson LF (1990) Body weight and physical activity as predictors of colorectal cancer risk. Nutrition and Cancer, 13:9-17.
Lahti-Koski M, Vartiainen E, Mannisto S, Pietinen P (2000) Age, education and occupation as determinants of trends in body mass index in Finland from 1982 to 1997. International Journal of Obesity, 24:1669-1676.
Lapidus L, Bengtsson C, Larsson B, Pennert K, Rybo E, Sjöström L (1984) Distribution of adipose tissue and risk of cardiovascular disease and death: a 12 year follow up of participants in the population study of women in Gothenburg, Sweden. British Medical Journal, 289:1257-1261.
Larsson B, Svärdsudd K, Welin L, Wilhelmsen L, Björntorp P, Tibblin G (1984) Abdominal adipose tissue distribution, obesity and risk of cardiovascular disease and death: 13 year follow up of participants in the study of men born in 1913. British Medical Journal, 288:1401-1404.
Lean ME, Powrie JK, Anderson AS, Garthwaite PH (1990) Obesity, weight loss and prognosis in type 2 diabetes. Diabetic Medicine, 7:228-233.

Lee I-M and Paffenbarger RS Jr (1992) Quetelet's index and risk of colon cancer in college alumni. Journal of the National Cancer Institute, 84:1326-1331.

Le Marchand L, Wilkens LR, Mi MP (1991) Early-age body size, adult weight gain and endometrial cancer risk. International Journal of Cancer, 48: 807-811.

Lew EA, Garfinkel L (1979) Variations in mortality by weight among 750000 men and women. Journal of Chronic Diseases, 32:563-576.

Lindsted K, Tonstad S, Kuzma JW (1991) Body mass index and patterns of mortality among Seventh-day Adventist men. International Journal of Obesity, 15:397-406.
Macdonald SM, Reeder BA, Chen Y, Despres JP (1997) Obesity in Canada: a descriptive analysis. Canadian Heart Health Surveys Research Group. Canadian Medical Association Journal, 157:S3-9.

MacMahon S, Cutler J, Brittain E, Higgins M (1987) Obesity and hypertension: epidemiological and clinical issues. European Heart Journal, 8:S57-70.

Manson JE, Colditz GA, Stampfer MJ et al. (1990) A prospective study of obesity and risk of coronary heart disease in women. New England Journal of Medicine, 322:882-889.

Marshall JA, Shetterly S, Hoag S et al. (1994) Dietary fat predicts conversion from impaired glucose tolerance to NIDDM. The San Luis Valley Diabetes study. Diabetes Care, 17:50-56.
Martinez ME, Giovannucci E, Spiegelman D, Hunter DJ, Willett WC, Colditz GA (1997) Leisure-time physical activity, body size and colon cancer in women. Nurses' Health Study Research Group. Journal of the National Cancer Institute, 89:948-955.

Martorell R, Schroeder DG, Rivera JA, Kaplowitz HJ (1995) Patterns of linear growth in rural Guatemalan adolescents and children. Journal of Nutrition, 125:S1060-1067.

Mathers C, Vos T, Stevenson C (1999) The Burden of disease and injury in Australia, 1999. (AIHW Catalogue No. PHE 17.) AIHW, Canberra.
McCarthy SN, Harrington KE, Kiely M et al. (2001) Analyses of the anthropometric data from the North/South Ireland Food Consumption Survey. Public Health Nutrition, 4:1099-1106.
McGarvey ST (1991) Obesity in Samoans and a perspective on its etiology in Polynesians. American Journal of Clinical Nutrition, 53:S1586-1594.
Micozzi MS, Albanes D, Jones DY, Chumlea WC (1986) Correlations of body mass indices with weight, stature, and body composition in men and women in NHANES I and II. American Journal of Clinical Nutrition, 44:725-731.

Molarius A, Seidell JC, Sans S, Tuomilehto J, Kuulasmaa K (2000) Educational level, relative body weight, and changes in their association over 10 years: an international perspective from the WHO MONICA project. American Journal of Public Health, 90:1260-1268.
Monteiro CA, Conde WL (1999) Time trends in overweight prevalence in children, adolescents and adults from less and more developed regions of

Brazil. In: Progress in obesity research. Guy Grand B, Ailhaud G, eds. John Libbey \& Company, London.
Monteiro CA, Mondini L, de Souza AL, Popkin BM (1995) The nutrition transition in Brazil. European Journal of Clinical Nutrition, 49:105-113.
Moreno LA, Sarria A, Fleta J, Rodriguez G, Bueno M (2000) Trends in body mass index and overweight prevalence among children and adolescents in the region of Aragon (Spain) from 1985 to 1995. International Journal of Obesity, 24:925-931.
Morris JN, Everitt MG, Pollard R et al. (1980) Vigorous exercise in leisure-time: protection against coronary heart disease. The Lancet, 2:1207-1210.
Mulrow DC, Chiquette E, Angel L et al. (2000) Dieting to reduce body weight for controlling hypertension in adults. Cochrane Database System Review, 2:CD000484 [online database].
Musaiger AO, al-Mannai MA (2001) Weight, height, body mass index and prevalence of obesity among adult population in Bahrain. Annals Human Biology, 10:346-350.
Musaiger AO, Gregory WB (2000) Profile of body composition of school children in Bahrain. International Journal of Obesity, 24:1093-1096.
Must A, Jacques PF, Dallal GE, Bajema CJ, Dietz WH (1992) Long-term morbidity and mortality of overweight adolescents. A follow-up of the Harvard Growth Study of 1922 to 1935. New England Journal of Medicine, 327:1350-1355.
Must A, Spadano J, Coakley EH (1999) The disease burden associated with overweight and obesity. Journal of the American Medical Association, 282:1523-1529.

National Audit Office (2001) Tackling obesity in England. A report by the Comptroller and Auditor General. National Audit Office, London.

Niedhammer I, Bugel I, Bonenfant S, Goldberg M, Leclerc A (2000) Validity of self-reported weight and height in the French GAZEL cohort. International Journal of Obesity Related Metabolic Disorders, 24:1111-1118.
NIH (1998) Clinical guidelines on the identification, evaluation and treatment of overweight and obesity in adults: the evidence report. US Department of Health and Human Services, National Institutes of Health, National Heart, Lung and Blood Institute, USA.
Nishizawa H, Shimomura I, Kishida K et al. (2002) Androgens decrease plasma adiponectin, an insulin-sensitizing adipocyte-derived protein. Diabetes, 51:2734-2741.

Njolstad I, Arnesen E, Lund-Larsen PG (1998) Sex differences in risk factors for clinical diabetes mellitus in a general population: a 12 -year follow-up of the Finnmark study. American Journal of Epidemiology, 147:49-58.
Norgan NG (1994a) Population differences in body composition in relation to the body mass index. European Journal of Clinical Nutrition, 48:S10-27.
Norgan NG (1994b) Relative sitting height and the body mass index. Annals of Human Biology, 21:78-82.

Ogden CL, Kuczmarski RJ, Flegal KM et al. (2002) Centers for Disease Control and Prevention 2000 growth charts for the United States: improvements to the 1977 National Center for Health Statistics version. Pediatrics, 109:45-60.

Okesina AB, Oparinde DP, Akindoyin KA, Erasmus RT (1999) Prevalence of some risk factors of ischaemic heart disease in a rural Nigerian population. East African Medical Journal, 76:212-217.
Okosun IS, Cooper RS, Rotimi CN, Osotimehin B, Forrester T (1998) Association of waist circumference with risk of hypertension and type 2 diabetes in Nigerians, Jamaicans, and African-Americans. Diabetes Care, 21:1836-1842.

Okosun IS, Rotimi CN, Forrester TE et al. (2000) Predictive value of abdominal obesity cut-off points for hypertension in Blacks from West African and Caribbean islands nations. International Journal of Obesity, 24:180-186.
Ornish D, Brown SE, Scherwitz LW et al. (1990) Can lifestyle changes reverse coronary heart disease? The Lancet, 336:129-133.
Paffenbarger RS Jr, Laughlin MD, Gima AS, Black RA (1970) Work activity of longshoremen as related to death from coronary heart disease and stroke. New England Journal of Medicine, 282:1109-1114.
Parsons TJ, Power C, Logan S, Summerbell CD (1999) Childhood predictors of adult obesity: a systematic review. International Journal of Obesity, 23: S1-107.

Pelletier DL and Rahn M (1998) Trends in body mass index in developing countries. Food and Nutrition Bulletin, 19:223-239.

Pishdad GR (1996) Overweight and obesity in adults aged 20-74 in southern Iran. International Journal of Obesity, 20:963-965.

Popkin BM (2002) An overview on the nutrition transition and its health implications: the Bellagio meeting. Public Health Nutrition, 5:93-103.

Reilly JJ, Dorosty AR, Emmett PM, ALSPAC Study Team (2000) Identification of the obese child: adequacy of the body mass index for clinical practice and epidemiology. International Journal of Obesity, 24:1623-1627.
Rexrode KM, Buring JE, Manson JE (2001) Abdominal and total adiposity and risk of coronary heart disease in men. International Journal of Obesity, 25:1047-1056.
Rexrode KM, Hennekens CH, Willett WC et al. (1997) A prospective study of body mass index, weight change, and risk of stroke in women. Journal of the American Medical Association, 277:1539-1545.

Rhoads GG, Kagan A (1983) The relation of coronary disease, stroke, and mortality to weight in youth and in middle age. The Lancet, 1:492-495.

Rimm EB, Stampfer MJ, Giovannucci E et al. (1995) Body size and fat distribution as predictors of coronary heart disease among middle-aged and older US men. American Journal of Epidemiology, 141:1117-1127.
Rissanen A, Heliövaara M, Knekt P et al. (1990) Risk of disability and mortality due to overweight in a Finnish population. British Medical Association, 301:835-837.

Roberts SB, Dallal GE (2001) The new childhood growth charts. Nutrition Reviews, 59:31-36.

Rodriguez-Ojea A, Jimenez S, Berdasco A, Esquivel M (2002) The nutrition transition in Cuba in the nineties: an overview. Public Health Nutrition, 5:129-133.

Rose G, Shipley M (1990) The population mean predicts the number of deviant individuals. British Medical Journal, 301:1031-1034.
Rosenbloom AL, Joe JR, Young RS, Winter WE (1999) Emerging epidemic of type 2 diabetes in youth. Diabetes Care, 22:345-354.
Rosengren A, Wedel H, Wilhelmsen L (1999) Body weight and weight gain during adult life in men in relation to coronary heart disease and mortality. A prospective population study. European Heart Journal, 20:269277.

Rotimi CN, Cooper RS, Ataman SL et al. (1995) Distribution of anthropometric variables and the prevalence of obesity in populations of West African origin. Obesity Research, 3:595-105.

Roubenoff R, Klag MJ, Mead LA, Liang KY, Seidler AJ, Hochberg MC (1991) Incidence and risk factors for gout in white men. Journal of the American Medical Association, 266:3004-3007.
Sargeant LA, Bennett FI, Forrester TE, Cooper RS, Wilks RJ (2002) Predicting incident diabetes in Jamaica: the role of anthropometry. Obesity Research, 10:792-798.

Sarkkinen E, Schwab U, Niskanen L et al. (1996) The effects of monounsatu-rated-fat enriched diet and polyunsaturated-fat enriched diet on lipid and glucose metabolism in subjects with impaired glucose tolerance. European Journal of Clinical Nutrition, 50:592-598.
Sanchez-Castillo CP, Velázquez-Monroy O, Berber A et al. (2003) Anthropometric cut-off points for predicting chronic diseases in the Mexican National Health Survey 2000. Obesity Research, 11:442-451.
Schouten JSAG, van den Ouweland FA, Valkenburg HA (1992) A twelve year follow-up study in the general population on prognostic factors of cartilage loss in osteoarthritis of the knee. Annals of the Rheumatic Diseases, 51:932-937.

Schroeder DG, Martorell R, Flores R (1999) Infant and child growth and fatness and fat distribution in Guatemalan adults. American Journal of Epidemiology, 149:177-185.

Seckl JR, Walker BR (2001) Minireview: $11 \beta$-hydroxysteroid dehydrogenase Type 1: a tissue-specific amplifier of glucocorticoid action. Endocrinology, 142:1371-1376.

Seidell JC, Kahn HS, Williamson DF, Lissner L, Valdez R (2001a) Report from a Centers for Disease Control and Prevention Workshop on use of adult anthropometry for public health and primary health care. American Journal of Clinical Nutrition, 73:123-126.
Seidell JC, Perusse L, Despres JP, Bouchard C (2001b) Waist and hip circumferences have independent and opposite effects on cardiovascular disease risk
factors: the Quebec Family study. American Journal of Clinical Nutrition, 74:315-321.

Shaper AG, Wannamethee SG, Walker M (1997) Body weight: implications for the prevention of coronary heart disease, stroke and diabetes mellitus in a cohort study of middle aged men. British Medical Journal, 314:13111317.

Shetty PS (2002) Nutrition transition in India. Public Health Nutrition, 5: 175-182.
Shetty PS, James WPT (1994) Body Mass Index. A measure of chronic energy deficiency in adults. (FAO Food and Nutrition Paper No. 56.) Food and Agriculture Organization of the United Nations, Rome.

Silverman BL, Rizzo TA, Cho NH, Metzger BE (1998) Long-term effects of the intrauterine environment. The Northwestern University Diabetes in Pregnancy Center. Diabetes Care, 21:B142-149.

Singh RB, Niaz MA, Agarwal P, Beegum R, Rastogi SS, Singh NK (1995) Epidemiologic study of central obesity, insulin resistance and associated disturbances in the urban population of North India. Acta Cardiologica, 3:215-225.

Sjöström CD, Lissner L, Wedel H, Sjöström L (1999) Reduction in incidence of diabetes, hypertension and lipid disturbances after intentional weight loss induced by bariatric surgery: the SOS intervention study. Obesity Research, 7:477-484.

Sjöström CD, Peltonen M, Wedel H, Sjöström L (2000) Differentiated long-term effects of intentional weight loss on diabetes and hypertension. Hypertension, 36:20-25.
Sjöström L, Larsson B, Backman L et al. (1992) Swedish obese subjects (SOS). Recruitment for an intervention study and a selected description of the obese state. International Journal of Obesity Related Metabolic Disorders, 16:465-479.
Stamler J, Farinaro E, Mojonnier L, Hall Y, Moss D, Stamler R (1980) Prevention and control of hypertension by nutritional-hygienic means. Long-term experience of the Chicago Coronary Prevention Evaluation Program. Journal of the American Medical Association, 243:1819-1823.

Stam-Moraga MC (1999) Sociodemographic and nutritional determinants of obesity in Belgium. International Journal of Obesity, 23:S1-9.
Stampfer MJ, Willett WC, Colditz GA, Rosner B, Speizer FE, Hennekens CH (1985) A prospective study of post-menopausal estrogen therapy and coronary heart disease. New England Journal of Medicine, 313:10441049.

Stevens J, Cai J, Juhaeri, Thun MJ, Williamson DF, Wood JL (1999) Consequences of the use of different measures of effect to determine the impact of age on the association between obesity and mortality. American Journal of Epidemiology, 150:399-407.
Stevens J, Cai J, Panuk ER, Williamson DF, Thun MJ, Wood JL (1998) The effect of age on the association between body-mass index and mortality. New England Journal of Medicine, 338:1-7.

Swinburn BA, Ley SJ, Carmichael HE, Plank LD (1999) Body size and composition in Polynesians. International Journal of Obesity, 23:1178-1183.
Tate RB, Manfreda J, Cuddy TE (1998) The effect of age on risk factors for ischemic heart disease: the Manitoba Follow-up study, 1948-1993. Annals of Epidemiology, 8:415-421.
Temple NJ, Steyn K, Hoffman M, Levitt NS, Lombard CJ (2001) The epidemic of obesity in South Africa: a study in a disadvantaged community. Ethnic Disorders, 11:431-437.
Thun MJ, Calle EE, Namboodiri MM et al. (1992) Risk factors for fatal colon cancer in a large prospective study. Journal of the National Cancer Institute, 84:1491-1500.

Tremblay MS, Katzmarzyk PT, Willms JD (2002) Temporal trends in overweight and obesity in Canada, 1981-1996. International Journal of Obesity, 26:518-543.

Tretli S (1989) Height and weight in relation to breast cancer morbidity and mortality. A prospective study of 570000 women in Norway. International Journal of Cancer, 44:23-30.
Trichopoulou A, Gnardellis C, Lagiou A, Benetou V, Trichopoulos D (2000) Body Mass Index in relation to energy expenditure among adults in Greece. Epidemiology, 11:333-336.
Troiano RP, Frongillo EA, Sobal J, Levitsky DA (1996) The relationship between body weight and mortality: a quantitative analysis of combined information from existing studies. International Journal of Obesity, 20:63-75.

Tuomilehto J, Lindstöm J, Eriksson JG et al. (2001) Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. New England Journal of Medicine, 344:1343-1350.

Törnberg SA, Carstensen JM (1994) Relationship between Quetelet's index and cancer of breast and female genital tract in 47000 women followed for 25 years. British Journal of Cancer, 69:358-361.
Unger R, Kreeger L, Christoffel KK (1990) Childhood obesity. Medical and familial correlates and age of onset. Clinical Pediatrics, 29:368-373.
van der Sande MA, Bailey R, Faal H et al. (1997) Nationwide prevalence study of hypertension and related non communicable diseases in the Gambia. Tropical Medicine and International Health, 12:1039-1048.

Verbrugge LM, Gates, DM, Ike RW (1991) Risk factors for disability among US adults with arthritis. Journal of Clinical Epidemiology, 44:167-182.

Vessby B, Uusitupa M, Hermansen K et al. (2001) Substituting dietary monounsaturated for saturated fat improves insulin sensitivity in healthy men and women-the KANWU study. Diabetologia, 44:312-319.

Visscher TL, Kromhout D, Seidell JC (2002) Long-term and recent time trends in the prevalence of obesity among Dutch men and women. International Journal of Obesity Related Metabolic Disorders, 26:1218-1224.
Weiderpass E, Adami HO, Baron JA et al. (1999) Risk of endometrial cancer following estrogen replacement with and without progestins. Journal of the National Cancer Institute, 91:1131-1137.

Whitlock G, Lewington S, Ni Mhurchu C (2002) Coronary heart disease and body mass index: a systematic review of the evidence from large prospective cohort studies. Seminars and Vascular Medicine, 2:369-381.

WHO (1985) Energy and protein requirements. Report of a joint FAO/WHO/UNU expert consultation. (WHO Technical Report Series No. 724.) World Health Organization, Geneva.
WHO (1993) World health statistics. Annual report. World Health Organization, Geneva, 1994.
WHO (1995) Physical status: the use and interpretation of anthropometry. (Technical Report Series No. 854.) World Health Organization, Geneva.
WHO (2000) Obesity: preventing and managing the global epidemic. (WHO Obesity Technical Report Series No. 894.) World Health Organization, Geneva.

WHO/IASO/IOTF (2000) The Asia-Pacific perspective: redefining obesity and its treatment. February 2000. Full document available from: http://www.idi.org.au/obesity_report.htm.

Wilks R, McFarlane-Anderson N, Bennett F et al. (1996) Obesity in peoples of the African diaspora. In: The origins and consequences of obesity. Chadwick DJ, Cardew G, eds. John Wiley \& Sons, New York.
Willett W, Manson J, Liu S (2002) Glycemic index, glycemic load, and risk of type 2 diabetes. American Journal of Clinical Nutrition, 76:S274280.

Willett WC, Manson JE, Stampfer MJ et al. (1995) Weight, weight change, and coronary heart disease in women: risk within the "normal" weight range. Journal of the American Medical Association, 273:461-465.
Williamson DF, Pamuk ER (1993) The association between weight and increased longevity. A review of the evidence. Annals of Internal Medicine, 119:731-736.

Williamson DF, Pamuk E, Thun M, Flanders D, Byers T, Heath C (1995) Prospective study of intentional weight loss and mortality in never-smoking overweight US white women aged 40-64 years. American Journal of Epidemiology, 141:1128-1141.
Witteman JC, Willett WC, Stampfer MJ et al. (1989) A prospective study of nutritional factors and hypertension among US women. Circulation, 80:1320-1327.

Xiao-Ren Pan, Guang-wei Li, Ying-Hua Hu et al. (1997) Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes study. Diabetes Care, 20:537-544.

Yajnik C (2000) Interactions of perturbations in intrauterine growth and growth during childhood on the risk of adult-onset disease. Proceedings of the Nutrition Society, 59:1-9.
Yoshiike N, Matsumura Y, Zaman MM, Yamaguchi M (1998) Descriptive epidemiology of BMI in Japanese adults in a representative sample from the National Nutrition Survey 1990-1994. International Journal of Obesity, 22:684-687.

Yoshiike N, Seino F, Tajima S et al. (2002) Twenty-year changes in the prevalence of overweight in Japanese adults: the National Nutrition Survey 1976-1995. Obesity Reviews, 3:183-190.
Zajkas G, Biro G (1998) Data on the prevalence of obesity in Hungarian adult population between 1985-88 and 1992-94. Zeitschrift für Ernärungswissenschaft, 37:S134-135.
Zhou B-F (2002) Predictive values of body mass index and waist circumference for risk factors of certain related diseases in Chinese adults-study on optimal cut-off points of body mass index and waist circumference in Chinese adults. Biomedical and Environmental Sciences, 15:83-95.

Zhou B, Wu Y, Yang J, Li Y, Zhang H, Zhao L (2002) Overweight is an independent risk factor for cardiovascular disease in Chinese populations. Obesity Reviews, 3:147-156.
Zimmermann MB, Hess SY, Hurrell RF (2000) A national study of the prevalence of overweight and obesity in 6-12 year old Swiss children; body mass index, body-weight perceptions and goals. European Journal of Clinical Nutrition, 54:568-572.

## Chapter 9

# Low fruit and vegetable CONSUMPTION 

Karen Lock, Joceline Pomerleau, Louise Causer and Martin McKee

## Summary

This analysis assessed the levels of mean dietary intakes of fruit and vegetables (excluding potatoes) measured in grams per day. The theoretical-minimum-risk distribution for fruit and vegetable intake was estimated to be $600 \mathrm{~g} /$ day in adults, $480 \mathrm{~g} /$ day in children aged $5-14$ years, and $330 \mathrm{~g} /$ day in children aged $0-4$ years. It is proposed to use set intervals of $80 \mathrm{~g} /$ day of fruit and vegetables (equivalent to one serving) to elaborate the distributional transition.

The effect of fruit and vegetable consumption in preventing ischaemic heart disease (IHD), cerebrovascular disease, lung cancer, stomach cancer, colon and rectum cancers and oesophageal cancer was estimated. The choice of outcomes was guided mainly by previous reviews of the literature (Law and Morris 1998; Ness and Powles 1997; World Cancer Research Fund and American Institute for Cancer Research 1997), which suggested a protective effect of fruit and vegetables for IHD, stroke and cancers of the lung and digestive tract.

Health outcomes that could be reconsidered for inclusion in the next revision of the Global Burden of Disease (GBD) study include cancers of the larynx, pancreas, bladder, ovary, endometrium, thyroid and prostate; type II diabetes; chronic obstructive pulmonary disease; and cataract.

Estimates were based primarily on representative population-based surveys of dietary intake identified using a comprehensive search of the literature and contacts with experts. Data were obtained for 26 countries from nine subregions ${ }^{1}$ and pooled statistically within each subregion. When no survey data were available in a subregion, per capita food supply statistics were combined with estimates of the distribution of intakes by sex and age. Systematic extrapolations were made when the original data did not meet the comparative risk assessment (CRA) categories and when part of the estimates were unavailable.

Estimates of fruit and vegetable intake were highest in EUR-A, followed by WPR-A. The lowest intakes were found in AMR-B, EURC, SEAR-B, SEAR-D and AFR-E. Intakes varied by age groups, with children and the elderly generally having lower intakes than middle-aged adults.

Standard deviations varied considerably by subregion, sex and age group, with a median of $223 \mathrm{~g} /$ day. Estimates tended to be, on average, lower in women than in men (but with variations by age group), and they were generally lower in young children. In some subregions, standard deviations were also slightly smaller in the elderly.

Risk factor-disease relationships for each selected outcome were estimated based on the results of a systematic review of the literature combined with meta-analysis. The relative risk estimates derived were applied to all age groups between the ages of 15 and 70 years. To take account of age attenuation the relative risks were reduced by a quarter for ages $70-79$ years, and by half for the age group $\geq 80$ years. For those aged $<15$ years, a relative risk of 1 was applied.

Due to the limited information available on subregional differences in risks, constant estimates were applied to all subregions. Risks were assessed relative to the chosen theoretical-minimum-risk distribution fruit and vegetable intake.

The protective effects of fruit and vegetables were expressed as relative risk estimates associated with an $80 \mathrm{~g} /$ day increase in fruit and vegetable intake. The relative risk estimates were as follows: 0.90 (95\% CI $0.82-0.99)$ for IHD, $0.94(0.89-0.99)$ for ischaemic stroke, 0.96 (0.93-0.99) for lung cancer, 0.94 ( $0.86-1.03$ ) for gastric cancer, 0.99 (0.97-1.02) for colorectal cancer and 0.94 (0.88-1.01) for oesophageal cancer.

CRA found that the lack of dietary fruit and vegetables contributes an important share of the worldwide disease burden. It was estimated that increasing individual fruit and vegetable consumption up to the theoretical-minimum-risk distribution could reduce the worldwide burden of disease for IHD and ischaemic stroke by about $31 \%$ ( $30 \%$ in men and $31 \%$ for women) and $19 \% ~(18 \%$ in men and $19 \%$ in women), respectively. For stomach and oesophageal cancer, the potential reduction in disease attributable to an increase in fruit and vegetable intake was $19 \%$ and $20 \%$, respectively. Attributable risk fractions were lower for lung and colorectal cancer ( $12 \%$ and $2 \%$ ). The total worldwide mortality attributable to inadequate fruit and vegetable consumption is estimated to be 2.726 million deaths or 26.662 million disabilityadjusted life years (DALYs) per year.

Although these results need to be interpreted in the light of the limitations of the current methods, including the potential for residual confounding and misclassification of exposure, they clearly highlight the importance of increased fruit and vegetable consumption in improving public health worldwide.

## 1. Introduction

### 1.1 Definition of risk factor

In this chapter, the selected risk factor is the mean dietary intake of fruit and vegetables. Intake is treated as a continuous variable and is measured in grams/person per day. The estimates provided exclude potatoes in order to be consistent with current international recommendations for the intake of fruit and vegetables (WHO 1990).

### 1.2 Choice of exposure variable, reasons and implications

Accumulating epidemiological evidence has suggested that fruit and vegetables in the diet can reduce the risk of major diseases such as cardiovascular diseases and certain cancers, thus reducing premature deaths (Klerk et al. 1998; World Cancer Research Fund and American Institute for Cancer Research 1997). This consistent pattern of findings has led several national and international bodies to advocate an increase in intake to at least $400 \mathrm{~g} /$ day (excluding potatoes) (WHO 1990; World Cancer Research Fund and American Institute for Cancer Research 1997).

While the first round of the GBD study did not look explicitly at the overall impact of nutrition on the global burden of disease, an attempt was later made in Sweden to estimate the burden of disease that could be attributed to different causal factors in the European Union (EU) (National Institute of Public Health 1997). It was estimated that dietrelated factors directly contributed to $8.3 \%$ of the estimated number of DALYs lost, almost half of this being attributed to low fruit and vegetable intake ( $3.5 \%$ compared with $3.7 \%$ for overweight and $1.1 \%$ for high saturated fat intake). However, these figures do not take account of potential interactions between the different factors. In comparison, tobacco smoking accounted for $9 \%$ of the burden of disease in the EU. These findings are similar to those of recent studies from Australia and New Zealand (Mathers et al. 1999; Tobias 2001). In these countries, it was estimated that about $3 \%$ ( $2.8 \%$ in Australia and $2.4 \%$ in New Zealand) of the burden of disease could be attributed to low fruit and vegetable consumption. The Australian study also reported that approximately $10 \%$ of all cancers could be due to an insufficient intake of fruit and vegetables.

### 1.3 CHOICE OF THEORETICAL-MINIMUM-RISK DISTRIBUTION

The CRA project estimates the effects of shifting risk factor distributions towards a counterfactual distribution, rather than the difference between "exposed" and "unexposed" groups. Fruit and vegetable consumption is unusual in the CRA project in that there is a suggested inverse risk factor-disease relationship (i.e. the potential protective effect of fruit and vegetables for various disease outcomes is considered). Hence, the
theoretical-minimum-risk distribution (the distribution of exposure that would yield the lowest population risk) involves estimating an upper consumption that is protective.

For this exercise, we chose limits that could be realistically attained. However, we do not know what the true upper intakes that would give the highest level of protection would be; they could reflect levels of intakes that are much higher than current consumption patterns. Thus, choosing a theoretical-minimum-risk distribution is currently difficult to do with certainty. Evidence has clearly shown that those in the highest categories of fruit and vegetable consumption have lower risk compared with those in the lowest consumption categories. It is not yet clear whether there is a threshold effect for fruit and vegetable consumption (although many studies have presented a linear dose-response relationship). Nor is it clear whether the same threshold would apply to all protective effects (e.g. there appears to be some evidence to suggest a lower threshold for colorectal cancer from research in individuals with very low intake) (Terry et al. 2001). However, due to the current uncertainty in this area, the same threshold was used for all selected health outcomes. Better information on threshold effects is a priority for future work.

In this project, the counterfactual was chosen to be a constant level using a minimum risk approach based on the following information.

- Most studies showing risk differences between the highest and lowest quartiles or quintiles of intakes have been performed in western Europe and North America. In the absence of other evidence, it was decided to base the counterfactual on the ranges of intakes observed in these populations.
- According to food balance sheet data, the amounts of fruit and vegetables available for consumption by the populations of these subregions range from about 300 to $1200 \mathrm{~g} /$ person per day (FAO 1998a). The highest availability is found in Greece where the quantity of fruit and vegetables available to consumers is $1100-1200 \mathrm{~g} /$ person per day. Assuming approximately $33 \%$ waste at the household level (Joffe and Robertson 2001), the mean daily intake of fruit and vegetables could be estimated to be approximately $700-800 \mathrm{~g} /$ person. However, dietary survey data gathered for this study showed that the mean intakes in adults of any given country rarely went above $500 \mathrm{~g} /$ day, and never above $550 \mathrm{~g} / \mathrm{day}$, even in countries known for their high intakes of fruit and vegetables (Israel and Italy). It was thus decided to limit the theoretical-minimum-risk distribution to $600 \mathrm{~g} /$ day in adults.
- As described in section 2.5 , it was assumed that children consume less fruit and vegetables than adults ( $45 \%$ less in children aged $0-4$ years and $20 \%$ less in children aged $5-14$ years).
- As data gathered for this project showed almost no difference in fruit and vegetable intake between men and women (see section 2.5), it was decided to use the same theoretical-minimum-risk distribution for both sexes.
- In order to allow for slight population variability, it was decided to apply around the theoretical-minimum-risk distribution, a margin of uncertainty of $50 \mathrm{~g} /$ day.

The final theoretical-minimum-risk distribution levels used in this project are listed in Table 9.1.

It is proposed to use set intervals of $80 \mathrm{~g} /$ day of fruit and vegetables for the elaboration of the distributional transition. This amount has been estimated in nutritional studies to be equivalent to approximately one serving of fruit and vegetables and would constitute a plausible and feasible change for individuals towards the selected counterfactual level (World Cancer Research Fund and American Institute for Cancer Research 1997).

Most studies to date that have investigated the relationship between fruit and vegetable intake and risk of diseases were performed on populations that reported only a relatively narrow range of intake which in most cases did not reach the selected theoretical-minimum-risk distribution. As a result, it is not possible to know whether the same relative risks would apply in populations where high amounts are consumed. It is expected that future epidemiological studies incorporating populations with wider range of intakes, such as the European Prospective Investigation of Cancer (EPIC) study, may shed light on the impact of high fruit and vegetable intakes (higher than those investigated in current studies) on the risk of cardiovascular disease and cancer.

Table 9.I Theoretical-minimum-risk distribution by age group

| Age group (years) | Theoretical-minimum-risk distribution and SD <br> (fruit and vegetable intake g/person per day) |
| :--- | :---: |
| $0-4$ | $330 \pm 50$ |
| $5-14$ | $480 \pm 50$ |
| $15-29$ | $600 \pm 50$ |
| $30-44$ | $600 \pm 50$ |
| $45-59$ | $600 \pm 50$ |
| $60-69$ | $600 \pm 50$ |
| $70-79$ | $600 \pm 50$ |
| $\geq 80$ | $600 \pm 50$ |

## 2. Estimating risk factor levels

### 2.1 Methods

The methods used to obtain estimates of risk factor levels are described in the following sections.

### 2.2 Criteria for considering sources and studies

Potential sources of data on intake and supply
Data on dietary intake and supply of fruit and vegetables may be available at the national, household and individual level. The following subsections briefly describe these potential sources of information and whether they were used for the CRA project.

## National level

The most commonly used source of information at the national level is food balance sheet data published by the Food and Agriculture Organization of the United Nations (FAO) for over 175 countries (1983). Food balance sheets provide standardized estimates of the average amount of food available per person on a daily basis. They are calculated by estimating the quantity of food produced in a given country added to the quantity of food imported (adjusted for changes in stocks), and subtracting the food exported, lost in storage, transport, and processing, fed to livestock or used for seeds and non-dietary purposes. The estimated national food supply is then divided by population size estimates to derive per capita figures (in $\mathrm{kg} / \mathrm{person}$ per year). The main limitation of food balance statistics is that they tend to reflect national food availability patterns rather than actual dietary intake and are thus a reflection of both intake and wastage at the retail, food service and household levels (Kantor et al. 1997). As a result, they cannot provide information on the dietary intake of different population subgroups and they tend to overestimate food consumption, particularly in developed countries. However, time trends in food availability tend to parallel those observed with household surveys (Sekula et al. 1991), so FAO food balance sheets constitute a useful tool for international comparisons and time trend analysis.

## Household level

Household-based surveys, where the unit of measurement is the household rather than the individual, are often performed to explore the diversity of food consumption patterns among communities. They can give information about dietary patterns of groups, making distinctions between geographical regions, income brackets and family structures. The most frequently used types of household surveys are the food account method, the inventory method, the household food record
method and the list-recall method (Nelson and Bingham 1997). Household surveys have several limitations: they cannot provide information on individuals; they are subject to sampling errors; they sometimes exclude foods consumed outside the home or certain food groups (e.g. sweets, alcoholic beverages, etc.); and some methods are subject to recall bias. In addition, they are available for only a limited number of countries and the diversity of the methods used make international comparisons difficult. Due to these limitations, data from household surveys were not used in this project.

## Individual level

It is generally agreed that there is a considerable lack of internationally comparable data at an individual level. This is partly due to the difficulties associated with measuring the dietary intake of individuals, including potential measurement error and bias. In spite of this, data collected at the individual level provide invaluable information on the mean dietary intakes of population subgroups (e.g. stratified by age and sex) and variability in intakes. They are thus essential for stratified intake estimates.

Data on present or recent food consumption are collected using four main techniques: (i) the 24-hour recall; (ii) food records (with or without weighing of foods); (iii) food frequency questionnaires; and (iv) food history. Details of these methods and their limitations can be found elsewhere (Nelson and Bingham 1997; Willett 1998a). The choice of a method to collect data at the individual level will normally depend on the objectives of the study and on the resources available. When the main objective is to obtain the mean consumption of a large group of individuals, it is generally sufficient to use a single 24 -hour recall or a oneday food record. This approach is often used in large national surveys of dietary intake as it represents a relatively small burden for the respondents and is associated with relatively low costs. The main caveat of using only one day of information is that it tends to artificially increase the standard deviation of the estimates due to large intra-individual variation in intakes (Cameron and van Staveren 1988). Thus, the observed distribution of intakes has extreme values that are higher and lower than any of the true long-term averages for any individual. Including several days of data collection for each respondent will normally dampen day-to-day variation, but it will also increase the burden on the respondents and the costs. If the objective of the study is to assess the distribution of food consumption in a group or the position of an individual's intake within the population, more complex methods such as repeated 24 -hour recalls or food records, food frequency questionnaires or dietary history are needed. These approaches have been used mostly in cohort studies or smaller, more focused surveys of dietary intake; they have less frequently been used in national surveys of dietary intake.

## Sources of data used in this project

Only dietary surveys with data collected at the individual level can provide information on mean intakes and variability in intakes (standard deviations) in population subgroups. Thus, we set out to identify data from at least one valid and representative population-based survey of dietary intake for each of the 191 countries covered by the CRA.

However, this was not realistically possible as currently only a few countries (mainly economically developed) have conducted representative national or regional surveys of dietary intake at the individual level, and a few others have performed surveys in selected sections of the population only. Conversely, for the majority of countries in the world, yearly estimates of available food supply exist in the form of the FAO food balance sheets. These food balance sheets were used to complement data collected at the individual level, when required. The methods used and the subregions to which they were applied are described in section 2.5.

## Criteria for including sources of individual level data

The main criteria used for including sources of individual level data of fruit and vegetable intake were as follows:

## Time frame

- The study was relatively recent-defined as having been performed since 1980.


## Study sample

- The reference population was described.
- The sampling strategy was as close as possible to random sampling.
- The sample was representative of the reference population.
- The sample size was large (sample size calculation ideally included).
- As wide an age range as possible was included.
- The level of non-response was ideally documented.


## Study design

- Only population-based cross-sectional studies, baseline assessment of large cohort studies (sample representative of the general population) or large interventions (sample representative of the general population) were considered for inclusion.
- Case-control studies were excluded from the selection process.


## Validity of the methods

- The methods used to collect data were as free of bias as possible.
- Data were collected at the level of the individual.
- The statistical analysis of the data was appropriate.


## Type of dietary information

- Data on fruit and vegetable intake had to be available as grams per day and not as frequencies (e.g. $<1$ serving a day, $1-2$ servings a day, every day, etc.).


### 2.3 Search strategy for the identification of studies

Dietary intake data were identified using a comprehensive worldwide search which included computerized databases of published articles, library catalogues, hand-searching of bibliographies, an Internet search of possible sources of data, and extensive contact with experts in the field, national governments and nongovernmental organizations.

## Computerized Databases and library Search

The following computerized sources of information were included in the search process: Medline, CAB abstracts and Embase. MESH terms used to search in Medline and HealthStar included "fruit", "vegetables", "nutrition-surveys", "diet-surveys" and "food-habits" (each term included all subheadings). Similar search terms were used in the other databases but adapted to the specific database search facilities. The search was restricted to human studies published in all languages since 1980.

Articles were rejected on initial screen if the reviewer could determine from the title and abstract that the article did not provide estimates of fruit and vegetable intake of a population or did not report data from a representative population-based survey of dietary intake. When a title or abstract could not be rejected with certainty, the full text of the article was obtained for further evaluation. Citation lists in the articles retrieved were reviewed. A second reviewer performed random checks.

The following catalogues were searched for other publications and conference proceedings that could provide appropriate data: the University of London; the British Library; the Resource Centre of the Public Health Nutrition Unit at the London School of Hygiene and Tropical Medicine; Libraries at FAO, Rome and the Ministry of Agriculture, United Kingdom of Great Britain and Northern Ireland. Food and Fisheries were also consulted. Citation lists in the documents retrieved were reviewed.

## INTERNET SEARCHES

Internet searches (using the Google search engine-http://www. google.com) had two objectives: to locate original sources of food intake data available on the Internet, and to identify national and international organizations that could identify possible data sources including acade-
mic departments of nutrition or dietetics, food and nutrition agencies and ministries of health.

Messages requesting help in identifying data sources were also posted to four scientific mailing lists: NUTEPI@listserv.gmd.de (nutritional epidemiology); food-for-thought@jiscmail.ac.uk (nutrition); publichealth@jiscmail.ac.uk (public health); and epidemio-1@cc.umontreal.ca (epidemiology).

## Contacts with experts

Numerous direct contacts were made with nutrition officers from the World Health Organization (WHO) regional offices and experts for references to published or unpublished data sources or for the identification of appropriate contact persons. Experts were defined as contact authors for large population-based studies of dietary intake, or contact persons in governmental agencies or country-specific nutrition organizations (this included existing networks from WHO and the International Obesity Task Force, and other international nutritional networks).

### 2.4 Methods for obtaining estimates of national intake WHERE MORE THAN ONE DATA SOURCE EXISTS

The following hierarchy of data quality was used to select one source of data for a given country where more than one was available:

- national survey of individual dietary intake;
- large sample survey of good quality-its quality being assessed from how well the survey met the list of criteria for including sources of individual level data (section 2.2); and
- small sample survey of good quality-its quality being assessed as above.


### 2.5 Methods for obtaining estimates where no data SOURCE EXISTED <br> Data on mean intakes not available for some age or sex groups

Attempts were made to contact the original investigators to obtain data broken down by the required age categories. However, this was not always possible and so indirect estimates were made using the following approaches.

## Data not available for children

Few of the available dietary intake surveys had data for children aged $<18$ years. Extrapolations for the intakes of children were on the basis of two observations.

- Using data from the surveys collected for this project, it was estimated that boys and girls aged $5-14$ years and those aged 0-4 years
consume, respectively, about $20 \%$ and $45 \%$ less fruit and vegetables than adults aged 30-59 years.
- Published estimates on energy requirements for infants and children (WHO 1985, 1990) suggested that girls and boys aged 5-14 years require approximately $15 \%$ and $20 \%$ less dietary energy than adult women and men, respectively. The figures for girls and boys aged 0-4 years are about $40 \%$ and $50 \%$ less than adults of the same sex, respectively. These estimates, however, may vary among countries and they depend on the true energy expenditure of the children.

Assuming that fruit and vegetable consumption decreases proportionally to energy intake in children compared with adults, the two sources of information tend to agree. Thus, the following adjustment factors were used.

- Children 5-14 years: $20 \%$ lower fruit and vegetable intake than adults aged 30-59 years.
- Children $0-4$ years: $45 \%$ lower fruit and vegetable intake than adults aged 30-59 years.


## Data not available for the elderly

Many surveys were only for adults up to age 60-65 years. Once again, available survey data on fruit and vegetable intakes collected for this project and published estimates of energy requirements (WHO 1985, 1990) were used to derive an adjustment factor.

- Information on fruit and vegetable intakes from survey data indicate that men and women aged 70-79 years consume daily approximately the same amount of fruit and vegetables, on average, as their counterparts aged $30-59$ years, while individuals aged $\geq 80$ years consume approximately $10 \%$ less fruit and vegetables than middle-aged adults.
- Figures based on energy requirements suggest that men and women in older age groups require approximately $10-15 \%$ less energy than middle-aged adults do.
Based on these observations, the following assumptions were made.
- Individuals aged 70-79 years consume the same amount of fruit and vegetables as individuals in the closest age group (60-69 years).
- Individuals aged $\geq 80$ years consume $10 \%$ less fruit and vegetables than those aged $30-59$ years. However, when the resulting estimates were greater than the reported intakes of survey participants aged 70-79 years, a different approach was taken: it was assumed that individuals aged $\geq 80$ years had an intake of fruit and vegetables similar to the intake observed in the 70-79-year age group.

Cases where the age groups did not correspond to the CRA AGE GROUPS
In these cases, the results available for the most similar age categories (greatest overlap of ages) were applied, weighing for population sizes when necessary.

## Data on mean intakes available for only one sex or for men and WOMEN TAKEN TOGETHER

In the case of Mexico, adult data were available only for women. In the case of France, only the overall mean intakes by age group (men and women taken together) were accessible.

Using data available from the surveys gathered for this project, it was estimated that, on average, men consume approximately only $1 \%$ more fruit and vegetables than women. It was thus assumed that Mexican and French males consume similar amounts of fruit and vegetables as do their female counterparts.

## Data on standard deviations not available

In some cases, survey information did not include standard deviation estimates. The authors of the studies were contacted and, for some countries, the required estimates were provided. When this was not possible, the following assumptions were made.

- When standard deviations were missing for one or more age groups and for one or more countries within a subregion (usually for children or the elderly), data were pooled based on the information available (all countries with information for these age groups).
- When standard deviations for all age groups were missing for a country, the standard deviations of the country-within the same sub-region-displaying the most similar mean intakes and method of data collection were applied. For example, for the United Kingdom, the standard deviations from Germany were used. However, since data on sample size are required for the estimation of the pooled standard deviation (and its confidence interval), pooled estimates were based only on the information available from the surveys.


## Data on standard deviation and sample size not available for all COUNTRIES IN A SUBREGION

For two subregions (SEAR-D and EMR-B), the survey data available included only mean intakes. As a result, it was not possible to extrapolate standard deviations from other countries within the same subregion. Thus, the pooled standard deviations of the subregion displaying the closest subregional mean intakes were applied (EUR-C for SEAR-D and EUR-B for EMR-B).

Another approach for the extrapolation of missing standard deviations may have been to use the standard deviations of a subregion that
is close geographically, that has similar economic characteristics, and for which data were available. For SEAR-D, for example, a possible choice may have been to use the standard deviations of WPR-B. However, as the standard deviations obtained for WPR-B are smaller than those of EUR-C, we preferred to choose the largest standard deviations and thus keep those obtained for EUR-C. For EMR-B, there was no obvious choice among the subregions for which data were available and thus the standard deviations of EUR-B were kept.

## Data on mean intakes unavailable for a subregion

When survey data were unavailable for all countries within a subregion, it was originally planned to apply the results obtained for another subregion displaying the most similar fruit and vegetable availability (from FAO food balance sheets information) and demographic and health characteristics (using data from the World bealth report 2000 [WHO 2000], the World Bank classification of economies based on gross national product [World Bank 2000], and the CIA World Factbook [Directorate of Intelligence 2000]).

However, it became clear the dietary patterns and socioeconomic characteristics of the African country mortality groupings (AFR-D and AFRE), and of the EMR-D, SEAR-B and AMR-D groupings are too different from those found in the other subregions to allow for any correct extrapolation. We have thus used FAO food balance sheet data combined with survey information to obtain FAO-derived proxy mean intakes by age group and sex for the five subregions for which no individual survey data were identified (as described in the next section). We concluded that this approach was likely to provide more representative estimates of mean subregional fruit and vegetable intakes than extrapolations from other subregions, as the basis of the calculations was information collected directly from within each country within the subregion, that is, countryspecific FAO food balance sheet data.

## Obtaining FAO-derived proxy mean intakes

Country-specific data on availability of fruit (excluding wine-FAO code 2919) and vegetables (excluding potato-FAO code 2918) and population size estimates were downloaded from the FAOstat database on the FAO web site (FAO 1998a). Three-year averages (1996-1998) were calculated in order to reduce the effect of yearly variations. These data were then used to calculate subregional population-weighted average fruit and vegetable availability in grams/person per day. For seven relatively small countries, no estimates were available (Bahrain, Bhutan, Equatorial Guinea, Oman, Qatar, Samoa, Singapore). Estimates of calculated subregional food availability (1996-1998) are listed in Table 9.2.

As mentioned in section 2.2, food balance sheets provide information on the amounts of foods available for the consumers and are thus a reflection of both intake and waste at the retail, food service and house-

Table 9.2 Fruit and vegetable availability by subregion ${ }^{\text {a }}$

| Subregion with no available survey data | Fruit and vegetable availability (g/person per day) |
| :--- | :---: |
| AFR-D | 291 |
| AFR-E | 194 |
| AMR-D | 317 |
| EMR-D | 323 |
| SEAR-B | 205 |

a 1996-1998 averages.
Source: FAO (2001).
hold levels. Therefore, it has been reported that the balance sheets tend to overestimate intakes in developed market economies (Joffe and Robertson 2001). In developing countries, such as those included in the five subregions for which no survey data were obtained (World Bank 2000), it has been suggested that food balance sheets are likely to underestimate food availability as they do not take account of food grown for home consumption or wild food collected. However, few studies have tested this hypothesis. In Nepal and Pakistan, the average energy consumption from intake surveys was found to be about $10 \%$ higher than that from FAO food balance sheets (FAO 1998a, 1998b).

Using survey data obtained for this project, it was estimated that food balance sheet data tend to overestimate the national mean fruit and vegetable intake by $20 \%$ in India (FAO $=252 \mathrm{~g} /$ day vs survey $=209 \mathrm{~g} /$ day ), and underestimate intakes by $74 \%$ in Bangladesh (overall weighed mean $=+10 \%)$. However, the results for Bangladesh were based on very small availability estimates ( $\mathrm{FAO}=59 \mathrm{~g} /$ day vs survey $=226 \mathrm{~g} /$ day ) that appear peculiar. In view of these contradictory results and because of the lack of further available information, we did not apply any correction factor to the FAO estimates in our calculations of the FAO-derived proxy mean intakes for the five subregions with no available survey data.

As food balance sheets do not provide information on food intake by sex and age group, an attempt was then made to estimate how the total availability of fruit and vegetables in a subregion would be distributed among the different sex and age groups. To reach this objective, a twostep process was used.

Step 1: We calculated the proportion of total fruit and vegetable intake consumed by the different age-sex groups for each subregion with available survey data. As expected, the distributions of intakes were strongly influenced by the population structures of the subregions.

Step 2: For each subregion with no available data (AFR-D, AFR-E, AMR-D, EMR-D and SEAR-B), we then applied to the FAO availability data the calculated distributions of intakes (Step 1) of the sub-

## Table 9.3 Details of subregional extrapolation of age-sex intake distribution for subregions where no survey data were available

| Subregion with no available survey data | Distribution of intakes extrapolated from |
| :--- | :---: |
| AFR-D | EMR-B |
| AFR-E | EMR-B |
| AMR-D | EMR-B |
| EMR-D | EMR-B |
| SEAR-B | SEAR-D |

Table 9.4 Details of subregional extrapolation of standard deviations for subregions where no survey data were available

| Subregion using FAO-derived proxy intake estimates | Subregion from which standard <br> deviations were extrapolated |
| :--- | :---: |
| AFR-D | EMR-D |
| AFR-E | SEAR-D |
| AMR-D | EMR-B |
| EMR-D | EMR-B |
| SEAR-B | SEAR-D |

region displaying the most similar population structure (Table 9.3).
As a result, we obtained FAO-derived proxy mean intakes by age and sex.

## Obtaining standard deviation estimates for FAO-derived proxy mean intakes

In order to obtain estimates of standard deviations when FAO-derived proxy mean intakes were used, we took the following approach. We first compared proxy intakes with all other subregional mean intakes (stratified by age and sex). We then applied to AFR-D, AFR-E, AMR-D, EMRD and SEAR-D, the standard deviations of the subregion displaying the most similar subregional intakes and closest level of socioeconomic development (see Table 9.4).

### 2.6 Description of Studies, including METHODOLOGICAL QUALITIES

The 26 countries for which data were obtained are shown in Table 9.5, along with the proportion of the subregional population covered by these countries. This proportion is generally high or acceptable except for two subregions (EMR-B $=1.4 \%$; EUR-B $=3.8 \%$ ).

Details of the studies included in this project are described in Table 9.6. Twenty-two were from national surveys. For Argentina, a compila-

Table 9.5 $\begin{aligned} & \text { Proportion of subregional population for which survey data } \\ & \text { were obtained }\end{aligned}$

| Subregion | Countries for which survey data were available | \% of subregional population ${ }^{\text {a }}$ |
| :---: | :---: | :---: |
| AFR-D | - | - |
| AFR-E | - | - |
| AMR-A | United States | 87.5 |
| AMR-B | Argentina, Mexico | 32.0 |
| AMR-D | - | - |
| EMR-B | Kuwait | 1.4 |
| EMR-D | - | - |
| EUR-A | Belgium, Denmark, Finland, France, Germany, Ireland, Israel, Italy, Norway, United Kingdom | 71.3 |
| EUR-B | Bulgaria | 3.8 |
| EUR-C | Estonia, Kazakhstan, Latvia, Lithuania, Russian Federation | 69.2 |
| SEAR-B | - | - |
| SEAR-D | Bangladesh, India | 93.7 |
| WPR-A | Australia, Japan, Singapore | 97.6 |
| WPR-B | China | 84.0 |

- No data.
a Percentage of subregional population covered by countries for which survey data were obtained.
tion of small representative surveys was provided-these cover the majority of the country. All but two studies were from the 1990s. Most studies used information from one 24-hour dietary recall or food diary. Other methods of data collection included multiple 24-hour recalls, 7day weighed food records, food-frequency questionnaires and food history. The majority of the surveys attempted to provide nationally representative samples; most used multistage random sampling techniques. Sample sizes ranged from about 1000 people (Argentina) to over 22000 (Belgium).


### 2.7 Characteristics of excluded studies

Due to the paucity of available information on fruit and vegetable intake at the individual level, few studies were excluded. Reasons for exclusion included the following.

- Another source of data was used for the country (e.g. more representative sample of the population or better method of data collection).
- The amounts consumed in grams per day could not be derived from the survey.
Table 9.6 Details of the dietary intake studies used

| Subregion | Country | Contact/ reference | Name of survey (if any) | Sample | Dietary data collection method | Data collection period | Sample <br> size | Sex | Age range (years) | Limitation |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AMR-A | United States | J.S. Hampl, personal communication, 2001; Taylor et al. (2000) | USDA <br> Continuing Survey of Food Intakes of Individuals | Multistage stratified random sample | Two 24-hour recall-non-consecutive days | 1994-1996 | 15303 | M+F | $\geq 0$ | $\ldots$ |
| AMR-B | Argentina | M. Rio et al., personal communication, 2001 | Collection of several dietary surveys in Argentina | Random samples in greater Buenos Aires, Province of Buenos Aires, West Areas (Mendoza) | 7-day food record | 1999-2000 | 1068 | M+F | $\geq 0$ | Collection of several small surveys. Very small sample size in $>60$ years excluded ( $n=35$ ). In West Areas (Salta), recruitment through a nutrition programme |
|  | Mexico | Rivera Dommarco (2001); J. Rivera Dommarco et al., personal communication, 2001 | National <br> Nutrition Survey | Multistage stratified random sample |  | Oct 1998- <br> March 1999 | 2646 | F | 12-49 | No data on adult males. Restricted age range |

Table 9.6 Details of the dietary intake studies used (continued)

| Subregion | Country | Contact/ reference | Name of survey (if any) | Sample | Dietary data collection method | Data collection period | Sample size | Sex | Age range (years) | Limitation |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| EMR-B | Kuwait | Sawaya et al. (1999) | Kuwait Total Diet study | Multistage sample of Kuwati nationals | One 24-hour recall | 1997 | 6700 | M+F | $\geq 0$ | No standard deviations |
| EUR-A | Belgium | S. De Henauw, personal communication, 2001; De Henauw and De Backer (1999) | Belgian Interuniversity Research on Nutrition and Health | Random sample form voting lists in 42 out of 43 Belgian districts | One 24-hour recall | 1980-1984 | 22224 | M+F | 25-74 | Restricted age range |
|  | Denmark | Andersen et al. (1996); S. Fagt, personal communication, 2001 | Dietary Habits in Denmark | Random sample from Danish civil registration system | 7-day food record | Feb/March/ <br> April 1995 | 3098 | M+F | I-79 | Restricted age range |
|  | Finland | Findiet Study Group (1998) | Dietary survey of Finnish adults | Age-stratified random sample in 5 regions | One 24-hour recall | $\begin{aligned} & \text { Jan-April } \\ & \text { I } 997 \end{aligned}$ | 3153 | M+F | 25-74 | Restricted age range |
|  | France | Volatier (1999) | INCA: Enquête Individuelle et Nationale sur les <br> Consommations Alimentaires | Stratified national multistage random sample | 7-day food record | Aug 1998June 1999 | 3003 | M+F | $\geq 3$ | Means for males and females jointly. No standard deviations |

Table 9.6 Details of the dietary intake studies used (continued)

| Subregion | Country | Contact/ reference | Name of survey (if any) | Sample | Dietary data collection method | Data collection period | Sample size | Sex | Age range (years) | Limitation |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| EUR-B | Bulgaria | S. Petrova, personal communication, 2001; Petrova et al. (2000) | National Dietary and Nutritional Survey of the population of Bulgaria | Multistage random sample stratified by region/size of settlement | One 24-hour food record | March $1998$ | 2800 | M+F | $\geq 1$ | $\ldots$ |
| EUR-C | Estonia | J. Pomerleau, personal communication, 2001 | Baltic Nutrition Survey | Random sample from the National Population Register | One 24-hour recall | Summer of 1997 | 2108 | M+F | 18-65 | Restricted age range |
|  | Kazakhstan | T. Sharmanov, personal communication, 2001 | National Survey of the State of nutrition within the 1995 Kazakhstan Demographic and Health Survey | Multistage random sample | One 24-hour recall | 1996 | 3480 | M+F | 15-80 | No data on children |
|  | Latvia | J. Pomerleau, personal communication, 2001 | Baltic Nutrition Survey | Random sample from the National Population Register | One 24-hour recall | Summer of 1997 | 2308 | M+F | 18-65 | Restricted age range |


|  | Lithuania | J. Pomerleau, personal communication, 2001 | Baltic Nutrition Survey | Random sample from the National Population Register | One 24-hour recall | Summer of 1997 | 2153 | M+F | 18-65 | Restricted age range |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Russian Federation | B. Popkin, personal communication, 2001 | Russian <br> Longitudinal Monitoring Survey | Multistage random cluster sample | One 24-hour recall | $\begin{aligned} & \text { Oct 1998- } \\ & \text { Jan } 1999 \end{aligned}$ | 9593 | M+F | $\geq 0$ | $\ldots$ |
| SEAR-D | Bangladesh | Ahmad and Hassan (1982) | Nutrition Survey of Rural Bangladesh | Multistage random cluster sample | One 24-hour weighed record by trained dietary investigator | 1981-1982 | 4904 | M+F | $\mathrm{I}-\geq 70$ | No standard deviation. Age groups different from CRA groups |
|  | India | Government of India (1998) | National <br> Nutrition Monitoring Bureau surveys (1994) and District Nutrition Profiles (1995-1996) | Varied survey designs | One 24-hour recall | 1994-1996 | Compiled surveys of 18 states, 4 regions | M+F | $\geq 1$ | No standard deviation. Age groups different from CRA groups |
| WPR-A | Australia | K. Baghurst and <br> S. Record, personal communication, 2001 | National Dietary Survey in Australia | Multistage stratified random sample | One 24-hour recall | Feb 1995- <br> March <br> 1996 | 13858 | M+F | $\geq 2$ | $\ldots$ |

Table 9.6 Details of the dietary intake studies used (continued)

| Subregion | Country | Contact/ reference | Name of survey (if any) | Sample | Dietary data collection method | Data collection period | Sample size | Sex | Age range (years) | Limitation |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Japan | Y. Matsumura and N. Yoshiike, personal communication, 2001; Ministry of Health and Welfare (2000); S. Mizushima, personal communication, 2001 | National Nutrition Survey | Multistage random cluster sample | Semi-weighed I-day food record | Nov 1995 | 14240 | M+F | $\geq 1$ | $\ldots$ |
|  | Singapore | M. DeurenbergYap et al., personal communication, 2001 | National Nutrition Survey | Random sample | Food frequency questionnaire | 1998 | 2388 | M+F | 18-69 | Restricted age range |
| WPR-B | China | B. Popkin, personal communication, 2001 | China Health and Nutrition Survey | Multistage random cluster sample | 3 contiguous 24-hour recall | Fall of 1997 | 12194 | M+F | $\geq 0$ | $\ldots$ |

- The data were not representative of the population of the country.
- Estimates included potato.


### 2.8 Estimates by subregion, age and sex

Obtaining subregional estimates from dietary survey data
The following approach was used to obtain subregional estimates of fruit and vegetable intake using available data from individual dietary surveys.

Obtaining estimates for subregions where data are available for two or more countries

In order to obtain subregional means and standard deviations (and obtain $95 \%$ confidence intervals for these estimates) when data were available for two or more countries within a subregion, means and standard deviations were pooled. It is assumed that each subregion is a stratified sample with the strata being countries. ${ }^{2}$ Because of the lack of information on the shape of the distributions of intakes, it was also assumed that intakes follow a normal distribution (this assumption is discussed in more detail in section 2.8).

It is important to note that if there is substantial heterogeneity among countries in a subregion, these methods tend to underestimate the true standard error of the pooled mean and pooled standard deviation. In addition, pooling is based on only a few countries within a subregion. We have thus assumed that the pooled subregional mean intakes and standard deviations are representative of the true estimates, and that differences between the pooled estimates and estimates for which data are not available would tend to cancel each other out. Finally, using data for only a few countries may underestimate the true variation of intakes within a subregion. However, for most subregions with available data, a large proportion of the total subregional population was covered by the surveys (see Table 9.5).

## Obtaining estimates for subregions where data are available for only one country

In four subregions (AMR-A, EMR-B, EUR-B and WPR-B), only one source of intake data was available. For AMR-A and WPR-B, the surveys were conducted in the United States of America and China, respectively. As these countries represent $84-88 \%$ of the total subregional population (Table 9.4), it was assumed that intake data from these countries were representative of subregional intakes. For EMR-B and EUR-B, however, the surveys were conducted in countries that represented only a very small proportion of the total subregional population $(1.4 \%$ and $3.8 \%$ respectively). For this reason, a different approach based on pooling survey and FAO food balance sheet data was used.

EMR-B: First, FAO-derived proxy mean intakes for the subregion were calculated using the method described previously (when no survey data were available for a subregion). As it has been reported that FAO food balance sheet data tend to overestimate intakes in developed countries, a correction factor was applied to the subregional three-year average FAO fruit and vegetable supply. The chosen correction factor corresponded to the difference between FAO food balance sheet data and subregional intake data for AMR-A. The subregional fruit and vegetable supply in AMR-A is the closest to that observed in EMR-B. It was assumed that the age-sex subregional distribution of intakes was similar to that observed in Kuwait. Second, the FAO-derived proxy mean intakes were pooled with intake data from Kuwait to obtain mean intakes for EMR-B.

EUR-B: First, FAO-derived proxy mean intakes for the subregion was calculated using the methods described previously (when no survey data were available for a subregion). As for EMR-B, a correction factor was applied to the subregional three-year average FAO fruit and vegetable supply. The chosen correction factor corresponded to the difference between FAO food balance sheet data and pooled survey data for the subregion EUR-C (assumed to be the most similar in terms of intakes, demographic and health characteristics). It was assumed that the distribution of intakes among age-sex groups was similar to that observed in Bulgaria. Second, the FAO-derived proxy mean intakes were pooled with intake data from Bulgaria to obtain pooled mean intakes for EUR-B.

## Final estimates

Estimates of fruit and vegetable intakes stratified by subregion, age and sex are given in Tables 9.8 to 9.11 . Results are presented as means with $95 \%$ CI for the mean, and as standard deviations with $95 \%$ CI for the standard deviation.

Estimates of fruit and vegetable intakes were highest in EUR-A followed by WPR-A (Table 9.7). Surprisingly, reported intakes in AMR-A-the other highly economically developed subregion-are on average only $74-82 \%$ of those observed in EUR-A and WPR-A. The lowest intakes were found in AMR-B, EUR-C, SEAR-B, SEAR-D and AFR-E. As expected, intakes varied by age group, with children and the elderly generally having lower intakes than middle-aged adults. However, in a few subregions elderly individuals appeared to have higher intakes than younger adults do in our calculations. This was the case particularly for AFR-E, AFR-D and EMR-D, three groupings where FAO-derived proxy mean intakes were calculated using the distribution of total intakes from another grouping with available data. Because the true age-sex distribution of AFR-E, AFR-D and EMR-D is slightly different from that of the chosen proxy grouping, the calculation of mean intake within each
Table 9.7 Mean intake of fruit and vegetables ${ }^{\text {a }}(95 \% \mathrm{Cl})$ by subregion, sex and age

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | $\begin{gathered} 144 \\ (115-173) \end{gathered}$ | $\begin{gathered} 296 \\ (272-320) \end{gathered}$ | $\begin{gathered} 288 \\ (256-3 \mid 9) \end{gathered}$ | $\begin{gathered} 413 \\ (378-448) \end{gathered}$ | $\begin{gathered} 419 \\ (386-452) \end{gathered}$ | $\begin{gathered} 439 \\ (403-476) \end{gathered}$ | $\begin{gathered} 446 \\ (404-488) \end{gathered}$ | $\begin{gathered} 476 \\ (406-546) \end{gathered}$ |
|  | Female | $\begin{gathered} 140 \\ (113-167) \end{gathered}$ | $\begin{gathered} 279 \\ (255-304) \end{gathered}$ | $\begin{gathered} 302 \\ (275-328) \end{gathered}$ | $\begin{gathered} 345 \\ (308-38 \mathrm{I}) \end{gathered}$ | $\begin{gathered} 305 \\ (27 I-340) \end{gathered}$ | $\begin{gathered} 355 \\ (320-390) \end{gathered}$ | $\begin{gathered} 349 \\ (306-392) \end{gathered}$ | $\begin{gathered} 382 \\ (302-462) \end{gathered}$ |
| AFR-E | Male | $\begin{gathered} 94 \\ (82-105) \end{gathered}$ | $\begin{gathered} 193 \\ (181-205) \end{gathered}$ | $\begin{gathered} 192 \\ (178-206) \end{gathered}$ | $\begin{gathered} 278 \\ (266-290) \end{gathered}$ | $\begin{gathered} 294 \\ (279-309) \end{gathered}$ | $\begin{gathered} 325 \\ (309-34 I) \end{gathered}$ | $\begin{gathered} 333 \\ (306-361) \end{gathered}$ | $\begin{gathered} 380 \\ (316-443) \end{gathered}$ |
|  | Female | $\begin{gathered} 91 \\ (78-103) \end{gathered}$ | $\begin{gathered} 181 \\ (170-192) \end{gathered}$ | $\begin{gathered} 201 \\ (194-209) \end{gathered}$ | $\begin{gathered} 236 \\ (229-243) \end{gathered}$ | $\begin{gathered} 214 \\ (205-223) \end{gathered}$ | $\begin{gathered} 257 \\ (245-268) \end{gathered}$ | $\begin{gathered} 244 \\ (229-259) \end{gathered}$ | $\begin{gathered} 245 \\ (225-266) \end{gathered}$ |
| AMR-A | Male | $\begin{gathered} 278 \\ (265-291) \end{gathered}$ | $\begin{gathered} 247 \\ (235-259) \end{gathered}$ | $\begin{gathered} 257 \\ (240-274) \end{gathered}$ | $\begin{gathered} 305 \\ (288-321) \end{gathered}$ | $\begin{gathered} 338 \\ (321-354) \end{gathered}$ | $\begin{gathered} 369 \\ (349-390) \end{gathered}$ | $\begin{gathered} 387 \\ (36\|-4\| 3) \end{gathered}$ | $\begin{gathered} 364 \\ (323-404) \end{gathered}$ |
|  | Female | $\begin{gathered} 262 \\ (25 I-274) \end{gathered}$ | $\begin{gathered} 236 \\ (224-248) \end{gathered}$ | $\begin{gathered} 234 \\ (22 \mathrm{I}-248) \end{gathered}$ | $\begin{gathered} 261 \\ (248-274) \end{gathered}$ | $\begin{gathered} 307 \\ (292-32 \mathrm{I}) \end{gathered}$ | $\begin{gathered} 335 \\ (318-352) \end{gathered}$ | $\begin{gathered} 346 \\ (325-367) \end{gathered}$ | $\begin{gathered} 348 \\ (316-380) \end{gathered}$ |
| AMR-B | Male | $\begin{gathered} 72 \\ (42-103) \end{gathered}$ | $\begin{gathered} 147 \\ (104-189) \end{gathered}$ | $\begin{gathered} 148 \\ (\|24-17\|) \end{gathered}$ | $\begin{gathered} 168 \\ (143-194) \end{gathered}$ | $\begin{gathered} 208 \\ (148-268) \end{gathered}$ | $\begin{gathered} 220 \\ (160-280) \end{gathered}$ | $\begin{gathered} 230 \\ (17 \mid-290) \end{gathered}$ | $\begin{gathered} 180 \\ (120-239) \end{gathered}$ |
|  | Female | $\begin{gathered} 82 \\ (5\|-\| \| 2) \end{gathered}$ | $\begin{gathered} 134 \\ (78-191) \end{gathered}$ | $\begin{gathered} 167 \\ (153-182) \end{gathered}$ | $\begin{gathered} 218 \\ (1 \mid 1-324) \end{gathered}$ | $\begin{gathered} 204 \\ (153-255) \end{gathered}$ | $\begin{gathered} 220 \\ (\mid 68-27 I) \end{gathered}$ | $\begin{gathered} 235 \\ (183-286) \end{gathered}$ | $\begin{gathered} 230 \\ (178-28 \mid) \end{gathered}$ |
| AMR-D | Male | $\begin{gathered} 193 \\ (165-222) \end{gathered}$ | $\begin{gathered} 352 \\ (328-376) \end{gathered}$ | $\begin{gathered} 299 \\ (268-330) \end{gathered}$ | $\begin{gathered} 408 \\ (372-443) \end{gathered}$ | $\begin{gathered} 392 \\ (360-425) \end{gathered}$ | $\begin{gathered} 387 \\ (35 I-424) \end{gathered}$ | $\begin{gathered} 353 \\ (311-395) \end{gathered}$ | $\begin{gathered} 306 \\ (236-377) \end{gathered}$ |
|  | Female | $\begin{gathered} 192 \\ (165-220) \end{gathered}$ | $\begin{gathered} 339 \\ (315-363) \end{gathered}$ | $\begin{gathered} 316 \\ (289-342) \end{gathered}$ | $\begin{gathered} 332 \\ (295-368) \end{gathered}$ | $\begin{gathered} 287 \\ (253-321) \end{gathered}$ | $\begin{gathered} 328 \\ (293-363) \end{gathered}$ | $\begin{gathered} 287 \\ (244-330) \end{gathered}$ | $\begin{gathered} 24 \mid \\ (\|6\|-322) \end{gathered}$ |
| EMR-B | Male | $\begin{gathered} 218 \\ (189-247) \end{gathered}$ | $\begin{gathered} 335 \\ (3 \mid 1-359) \end{gathered}$ | $\begin{gathered} 296 \\ (265-327) \end{gathered}$ | $\begin{gathered} 368 \\ (333-404) \end{gathered}$ | $\begin{gathered} 374 \\ (341-407) \end{gathered}$ | $\begin{gathered} 392 \\ (355-428) \end{gathered}$ | $\begin{gathered} 350 \\ (308-392) \end{gathered}$ | $\begin{gathered} 334 \\ (264-404) \end{gathered}$ |
|  | Female | $\begin{gathered} 218 \\ (190-245) \end{gathered}$ | $\begin{gathered} 327 \\ (303-35 I) \end{gathered}$ | $\begin{gathered} 323 \\ (297-350) \end{gathered}$ | $\begin{gathered} 362 \\ (325-398) \end{gathered}$ | $\begin{gathered} 346 \\ (3 \mid I-380) \end{gathered}$ | $\begin{gathered} 392 \\ (357-427) \end{gathered}$ | $\begin{gathered} 336 \\ (293-378) \end{gathered}$ | $\begin{gathered} 319 \\ (238-399) \end{gathered}$ |
| EMR-D | Male | $\begin{gathered} 174 \\ (145-203) \end{gathered}$ | $\begin{gathered} 342 \\ (318-367) \end{gathered}$ | $\begin{gathered} 312 \\ (28 \mid-343) \end{gathered}$ | $\begin{gathered} 388 \\ (353-424) \end{gathered}$ | $\begin{gathered} 409 \\ (376-442) \end{gathered}$ | $\begin{gathered} 446 \\ (410-482) \end{gathered}$ | $\begin{gathered} 442 \\ (400-485) \end{gathered}$ | $\begin{gathered} 420 \\ (350-490) \end{gathered}$ |
|  | Female | $\begin{gathered} 174 \\ (147-201) \end{gathered}$ | $\begin{gathered} 333 \\ (308-357) \end{gathered}$ | $\begin{gathered} 348 \\ (322-375) \end{gathered}$ | $\begin{gathered} 352 \\ (316-389) \end{gathered}$ | $\begin{gathered} 319 \\ (284-353) \end{gathered}$ | $\begin{gathered} 385 \\ (350-420) \end{gathered}$ | $\begin{gathered} 372 \\ (329-415) \end{gathered}$ | $\begin{gathered} 409 \\ (329-489) \end{gathered}$ |

Table 9.7 Mean intake of fruit and vegetables ${ }^{\text {a }}$ ( $95 \% \mathrm{CI}$ ) by subregion, sex and age (continued)

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EUR-A | Male | $\begin{gathered} 232 \\ (204-260) \end{gathered}$ | $\begin{gathered} 299 \\ (274-324) \end{gathered}$ | $\begin{gathered} 423 \\ (401-445) \end{gathered}$ | $\begin{gathered} 450 \\ (433-468) \end{gathered}$ | $\begin{gathered} 488 \\ (467-508) \end{gathered}$ | $\begin{gathered} 511 \\ (487-535) \end{gathered}$ | $\begin{gathered} 515 \\ (473-556) \end{gathered}$ | $\begin{gathered} 469 \\ (407-530) \end{gathered}$ |
|  | Female | $\begin{gathered} 233 \\ (2 \mid I-255) \end{gathered}$ | $\begin{gathered} 299 \\ (279-318) \end{gathered}$ | $\begin{gathered} 423 \\ (406-439) \end{gathered}$ | $\begin{gathered} 448 \\ (435-461) \end{gathered}$ | $\begin{gathered} 483 \\ (469-497) \end{gathered}$ | $\begin{gathered} 488 \\ (467-509) \end{gathered}$ | $\begin{gathered} 479 \\ (45 \mathrm{I}-507) \end{gathered}$ | $\begin{gathered} 446 \\ (4 \mid I-48 ।) \end{gathered}$ |
| EUR-B | Male | $\begin{gathered} 263 \\ (234-292) \end{gathered}$ | $\begin{gathered} 374 \\ (349-398) \end{gathered}$ | $\begin{gathered} 396 \\ (365-427) \end{gathered}$ | $\begin{gathered} 352 \\ (317-388) \end{gathered}$ | $\begin{gathered} 396 \\ (363-428) \end{gathered}$ | $\begin{gathered} 366 \\ (330-403) \end{gathered}$ | $\begin{gathered} 358 \\ (316-400) \end{gathered}$ | $\begin{gathered} 300 \\ (230-370) \end{gathered}$ |
|  | Female | $\begin{gathered} 238 \\ (211-265) \end{gathered}$ | $\begin{gathered} 372 \\ (348-396) \end{gathered}$ | $\begin{gathered} 344 \\ (317-370) \end{gathered}$ | $\begin{gathered} 333 \\ (296-369) \end{gathered}$ | $\begin{gathered} 383 \\ (348-4 I 7) \end{gathered}$ | $\begin{gathered} 352 \\ (3 \mid 7-387) \end{gathered}$ | $\begin{gathered} 358 \\ (3\|5-40\|) \end{gathered}$ | $\begin{gathered} 303 \\ (223-383) \end{gathered}$ |
| EUR-C | Male | $\begin{gathered} 134 \\ (122-\mid 46) \end{gathered}$ | $\begin{gathered} 198 \\ (185-210) \end{gathered}$ | $\begin{gathered} 233 \\ (218-247) \end{gathered}$ | $\begin{gathered} 237 \\ (225-249) \end{gathered}$ | $\begin{gathered} 246 \\ (23 I-26 I) \end{gathered}$ | $\begin{gathered} 254 \\ (237-270) \end{gathered}$ | $\begin{gathered} 233 \\ (206-260) \end{gathered}$ | $\begin{gathered} 233 \\ (169-297) \end{gathered}$ |
|  | Female | $\begin{gathered} 133 \\ (\|2\|-\mid 46) \end{gathered}$ | $\begin{gathered} 182 \\ (17 I-193) \end{gathered}$ | $\begin{gathered} 196 \\ (188-204) \end{gathered}$ | $\begin{gathered} 187 \\ (180-194) \end{gathered}$ | $\begin{gathered} 202 \\ (193-2 \mid I) \end{gathered}$ | $\begin{gathered} 200 \\ (\|89-2\| 1) \end{gathered}$ | $\begin{gathered} 209 \\ (194-224) \end{gathered}$ | $\begin{gathered} 190 \\ (170-2 \mid I) \end{gathered}$ |
| SEAR-B | Male | $\begin{gathered} 108 \\ (96-120) \end{gathered}$ | $\begin{gathered} 198 \\ (185-210) \end{gathered}$ | $\begin{gathered} 245 \\ (231-259) \end{gathered}$ | $\begin{gathered} 243 \\ (232-255) \end{gathered}$ | $\begin{gathered} 258 \\ (243-273) \end{gathered}$ | $\begin{gathered} 248 \\ (23 \mathrm{I}-264) \end{gathered}$ | $\begin{gathered} 244 \\ (217-272) \end{gathered}$ | $\begin{gathered} 225 \\ (16 \mid-288) \end{gathered}$ |
|  | Female | $\begin{gathered} 107 \\ (94-120) \end{gathered}$ | $\begin{gathered} 183 \\ (172-195) \end{gathered}$ | $\begin{gathered} 201 \\ (194-209) \end{gathered}$ | $\begin{gathered} 195 \\ (188-202) \end{gathered}$ | $\begin{gathered} 202 \\ (193-2 \mid I) \end{gathered}$ | $\begin{gathered} 201 \\ (190-212) \end{gathered}$ | $\begin{gathered} 201 \\ (187-216) \end{gathered}$ | $\begin{gathered} 173 \\ (153-194) \end{gathered}$ |
| SEAR-D | Male | $\begin{gathered} 94 \\ (82-106) \end{gathered}$ | $\begin{gathered} 177 \\ (165-190) \end{gathered}$ | $\begin{gathered} 258 \\ (244-272) \end{gathered}$ | $\begin{gathered} 262 \\ (250-274) \end{gathered}$ | $\begin{gathered} 262 \\ (247-277) \end{gathered}$ | $\begin{gathered} 259 \\ (243-275) \end{gathered}$ | $\begin{gathered} 259 \\ (232-286) \end{gathered}$ | $\begin{gathered} 234 \\ (170-298) \end{gathered}$ |
|  | Female | $\begin{gathered} 95 \\ (82-108) \end{gathered}$ | $\begin{gathered} 170 \\ (159-182) \end{gathered}$ | $\begin{gathered} 224 \\ (217-232) \end{gathered}$ | $\begin{gathered} 229 \\ (222-236) \end{gathered}$ | $\begin{gathered} 227 \\ (218-236) \end{gathered}$ | $\begin{gathered} 229 \\ (218-240) \end{gathered}$ | $\begin{gathered} 228 \\ (2 \mid 3-243) \end{gathered}$ | $\begin{gathered} 205 \\ (185-226) \end{gathered}$ |
| WPR-A | Male | $\begin{gathered} 264 \\ (253-275) \end{gathered}$ | $\begin{gathered} 345 \\ (333-356) \end{gathered}$ | $\begin{gathered} 366 \\ (355-378) \end{gathered}$ | $\begin{gathered} 376 \\ (367-386) \end{gathered}$ | $\begin{gathered} 450 \\ (439-462) \end{gathered}$ | $\begin{gathered} 491 \\ (474-509) \end{gathered}$ | $\begin{gathered} 446 \\ (428-463) \end{gathered}$ | $\begin{gathered} 415 \\ (398-433) \end{gathered}$ |
|  | Female | $\begin{gathered} 232 \\ (222-242) \end{gathered}$ | $\begin{gathered} 342 \\ (332-35 I) \end{gathered}$ | $\begin{gathered} 352 \\ (342-362) \end{gathered}$ | $\begin{gathered} 383 \\ (374-392) \end{gathered}$ | $\begin{gathered} 486 \\ (475-497) \end{gathered}$ | $\begin{gathered} 485 \\ (469-501) \end{gathered}$ | $\begin{gathered} 440 \\ (424-456) \end{gathered}$ | $\begin{gathered} 386 \\ (370-402) \end{gathered}$ |
| WPR-B | Male | $\begin{gathered} 204 \\ (\|87-22\|) \end{gathered}$ | $\begin{gathered} 274 \\ (266-282) \end{gathered}$ | $\begin{gathered} 344 \\ (336-352) \end{gathered}$ | $\begin{gathered} 346 \\ (338-354) \end{gathered}$ | $\begin{gathered} 360 \\ (350-370) \end{gathered}$ | $\begin{gathered} 335 \\ (320-350) \end{gathered}$ | $\begin{gathered} 304 \\ (285-323) \end{gathered}$ | $\begin{gathered} 258 \\ (22 I-294) \end{gathered}$ |
|  | Female | $\begin{gathered} 190 \\ (170-209) \end{gathered}$ | $\begin{gathered} 270 \\ (26 \mathrm{I}-279) \end{gathered}$ | $\begin{gathered} 317 \\ (308-325) \end{gathered}$ | $\begin{gathered} 334 \\ (326-34 I) \end{gathered}$ | $\begin{gathered} 345 \\ (336-355) \end{gathered}$ | $\begin{gathered} 304 \\ (292-317) \end{gathered}$ | $\begin{gathered} 273 \\ (257-288) \end{gathered}$ | $\begin{gathered} 250 \\ (221-278) \end{gathered}$ |

[^36]age-sex strata and the balance among strata was affected. This is observed particularly in the smaller age strata (particularly the elderly in these subregions). As a result, the FAO-proxy mean intakes are less reliable in population strata with relatively smaller sample size and must be interpreted with caution.

Pooled standard deviation estimates were available from only seven subregions (AMR-A, AMR-B, EUR-A, EUR-B, EUR-C, WPR-A and WPR-B). For the other subregions, we applied data from the subregion displaying the most similar intakes by age and sex, and when appropriate, method of data collection. These extrapolations need to be taken with caution, as the standard deviations of one subregion may not represent well the standard deviations of another subregion despite similarity in overall mean intakes. The results shown in Table 9.8 indicate that standard deviations varied considerably by subregion, sex and age group, with an overall median standard deviation of $223 \mathrm{~g} / \mathrm{day}$. Estimates tended to be lower in women than in men on average (but with variations by age group), and they were generally lower in young children. In some subregions, standard deviations were also slightly smaller in the elderly.

It is assumed that the reported fruit and vegetable intakes were normally distributed, due to the general lack of available information on the skewness of the distributions (except for the United States). However, this is unlikely to be true as dietary intakes are typically skewed towards higher values (Willett 1998b). Assuming a normal distribution creates the problem that some individuals will be recorded as having negative consumption when estimating impact fractions. As this is impossible, the normal distribution is truncated at zero, with all those falling below this value allocated a value of zero. The results of a sensitivity analysis, described below, based on data from the United States suggest that the approach used is likely to be conservative.

## Sensitivity analysis: skewed distributions and calculation of the attributable fraction

Data from the United States were used to evaluate the possible effects of skewness in the distribution of fruit and vegetable intake on the calculation of the attributable fraction for AMR-A. The data indicated a positive skewness ranging from 1.5 to 3 . To approximate this type of skewed distribution, the Weibull probability distribution function (PDF) was utilized by varying the shape and scale parameters (decreasing the shape parameter of a Weibull increases positive skewness away from a normal distribution).

Figure 9.1 illustrates a normal distribution (dashed line) with a fruit and vegetable mean intake of $300 \mathrm{~g} /$ day and standard deviation of $300 \mathrm{~g} /$ day. A significant part of the population with a normal distribution is truncated at zero consumption (approximately $10 \%$ of the population in this example). The skewed distribution (solid line) is the
Table 9.8 Standard deviations of fruit and vegetables ${ }^{\text {a }}(95 \% \mathrm{Cl})$ by subregion, sex and age

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | $\begin{gathered} 175.0 \\ (156.7-198.2) \end{gathered}$ | $\begin{gathered} 244.8 \\ (228.7-263.3) \end{gathered}$ | $\begin{gathered} 293.1 \\ (272.6-316.9) \end{gathered}$ | $\begin{gathered} 225.0 \\ (202.4-253.3) \end{gathered}$ | $\begin{gathered} 220.7 \\ (199.6-246.8) \end{gathered}$ | $\begin{gathered} 213.4 \\ (190.4-242.8) \end{gathered}$ | $\begin{gathered} 235.1 \\ (208.6-269.3) \end{gathered}$ | $\begin{gathered} 214.6 \\ (174.1-279.9) \end{gathered}$ |
|  | Female | $\begin{gathered} 163.3 \\ (146.0-185.3) \end{gathered}$ | $\begin{gathered} 240.8 \\ (224.8-259.3) \end{gathered}$ | $\begin{gathered} 247.6 \\ (230.2-267.9) \end{gathered}$ | $\begin{gathered} 224.5 \\ (201.2-253.9) \end{gathered}$ | $\begin{gathered} 237.4 \\ (215.3-264.6) \end{gathered}$ | $\begin{gathered} 210.5 \\ (188.4-238.5) \end{gathered}$ | $\begin{gathered} 251.5 \\ (224.4-286.1) \end{gathered}$ | $\begin{gathered} 239.0 \\ (192.8-314.6) \end{gathered}$ |
| AFR-E | Male | $\begin{gathered} 96.2 \\ (88.6-105.3) \end{gathered}$ | $\begin{gathered} 178.6 \\ (170.3-187.8) \end{gathered}$ | $\begin{gathered} 254.9 \\ (247.5-262.7) \end{gathered}$ | $\begin{gathered} 220.7 \\ (214.7-227.1) \end{gathered}$ | $\begin{gathered} 231.5 \\ (224.3-239.3) \end{gathered}$ | $\begin{gathered} 192.6 \\ (183.5-202.6) \end{gathered}$ | $\begin{gathered} 176.3 \\ (159.9-196.4) \end{gathered}$ | $\begin{gathered} 165.8 \\ (130.0-228.9) \end{gathered}$ |
|  | Female | $\begin{gathered} 105.5 \\ (97.3-115.3) \end{gathered}$ | $\begin{gathered} 155.9 \\ (148.4-164.2) \end{gathered}$ | $\begin{gathered} 163.4 \\ (206.3-222.2) \end{gathered}$ | $\begin{gathered} 157.6 \\ (189.6-202.3) \end{gathered}$ | $\begin{gathered} 171.6 \\ (166.9-176.5) \end{gathered}$ | $\begin{gathered} 168.2 \\ (161.2-175.1) \end{gathered}$ | $\begin{gathered} 153.5 \\ (144.2-164.1) \end{gathered}$ | $\begin{gathered} 115.4 \\ (102.5-132.0) \end{gathered}$ |
| AMR-A | Male | $\begin{gathered} 239.0 \\ (230.2-248.4) \end{gathered}$ | $\begin{gathered} 221.3 \\ (213.0-230.3) \end{gathered}$ | $\begin{gathered} 297.0 \\ (285.3-309.8) \end{gathered}$ | $\begin{gathered} 299.3 \\ (288.0-3 \mid 1.5) \end{gathered}$ | $\begin{gathered} 297.8 \\ (286.5-310.0) \end{gathered}$ | $\begin{gathered} 295.8 \\ (282.0-310.9) \end{gathered}$ | $\begin{gathered} 295.8 \\ (278.5-3 \mid 5.5) \end{gathered}$ | $\begin{gathered} 318.7 \\ (292.5-350.1) \end{gathered}$ |
|  | Female | $\begin{gathered} 222.5 \\ (2\|4.4-23\| .2) \end{gathered}$ | $\begin{gathered} 209.1 \\ (201.1-2 \mid 7.8) \end{gathered}$ | $\begin{gathered} 230.4 \\ (221.2-240.5) \end{gathered}$ | $\begin{gathered} 236.3 \\ (227.3-246.0) \end{gathered}$ | $\begin{gathered} 262.4 \\ (252.4-273.2) \end{gathered}$ | $\begin{gathered} 243.3 \\ (231.6-256.2) \end{gathered}$ | $\begin{gathered} 222.8 \\ (209.1-238.4) \end{gathered}$ | $\begin{gathered} 243.4 \\ (222.8-268.1) \end{gathered}$ |
| AMR-B | Male | $\begin{gathered} 153.3 \\ (134.4-178.4) \end{gathered}$ | $\begin{gathered} 294.3 \\ (229.8-409.4) \end{gathered}$ | $\begin{gathered} 470.1 \\ (438.2-507.0) \end{gathered}$ | $\begin{gathered} 260.0 \\ (210.9-339.2) \end{gathered}$ | $\begin{gathered} 390.3 \\ (312.9-518.9) \end{gathered}$ | $\begin{gathered} 390.3 \\ (3 \mid 2.9-518.9) \end{gathered}$ | $\begin{gathered} 390.3 \\ (3\|2.9-5\| 8.9) \end{gathered}$ | $\begin{gathered} 390.3 \\ (312.9-518.9) \end{gathered}$ |
|  | Female | $\begin{gathered} 160.2 \\ (141.2-185.2) \end{gathered}$ | $\begin{gathered} 341.8 \\ (250.2-539.1) \end{gathered}$ | $\begin{gathered} 293.6 \\ (272.8-317.9) \end{gathered}$ | $\begin{gathered} 718.2 \\ (515.0-1185.6) \end{gathered}$ | $\begin{gathered} 260.5 \\ (190.7-410.8) \end{gathered}$ | $\begin{gathered} 260.5 \\ (190.7-410.8) \end{gathered}$ | $\begin{gathered} 260.5 \\ (190.7-410.8) \end{gathered}$ | $\begin{gathered} 260.5 \\ (190.7-410.8) \end{gathered}$ |
| AMR-D | Male | $\begin{gathered} 175.0 \\ (156.7-198.2) \end{gathered}$ | $\begin{gathered} 244.8 \\ (228.7-263.3) \end{gathered}$ | $\begin{gathered} 293.1 \\ (272.6-316.9) \end{gathered}$ | $\begin{gathered} 225.0 \\ (202.4-253.3) \end{gathered}$ | $\begin{gathered} 220.7 \\ (199.6-246.8) \end{gathered}$ | $\begin{gathered} 213.4 \\ (190.4-242.8) \end{gathered}$ | $\begin{gathered} 235.1 \\ (208.6-269.3) \end{gathered}$ | $\begin{gathered} 214.6 \\ (174.1-279.9) \end{gathered}$ |
|  | Female | $\begin{gathered} 163.3 \\ (146.0-185.3) \end{gathered}$ | $\begin{gathered} 240.8 \\ (224.8-259.3) \end{gathered}$ | $\begin{gathered} 247.6 \\ (230.2-267.9) \end{gathered}$ | $\begin{gathered} 224.5 \\ (201.2-253.9) \end{gathered}$ | $\begin{gathered} 237.4 \\ (215.3-264.6) \end{gathered}$ | $\begin{gathered} 210.5 \\ (188.4-238.5) \end{gathered}$ | $\begin{gathered} 251.5 \\ (224.4-286.1) \end{gathered}$ | $\begin{gathered} 239.0 \\ (192.8-3 \mid 4.6) \end{gathered}$ |
| EMR-B | Male | $\begin{gathered} 175.0 \\ (156.7-198.2) \end{gathered}$ | $\begin{gathered} 244.8 \\ (228.7-263.3) \end{gathered}$ | $\begin{gathered} 293.1 \\ (272.6-316.9) \end{gathered}$ | $\begin{gathered} 225.0 \\ (202.4-253.3) \end{gathered}$ | $\begin{gathered} 220.7 \\ (199.6-246.8) \end{gathered}$ | $\begin{gathered} 213.4 \\ (190.4-242.8) \end{gathered}$ | $\begin{gathered} 235.1 \\ (208.6-269.3) \end{gathered}$ | $\begin{gathered} 214.6 \\ (174.1-279.9) \end{gathered}$ |
|  | Female | $\begin{gathered} 163.3 \\ (146.0-185.3) \end{gathered}$ | $\begin{gathered} 240.8 \\ (224.8-259.3) \end{gathered}$ | $\begin{gathered} 247.6 \\ (230.2-267.9) \end{gathered}$ | $\begin{gathered} 224.5 \\ (201.2-253.9) \end{gathered}$ | $\begin{gathered} 237.4 \\ (215.3-264.6) \end{gathered}$ | $\begin{gathered} 210.5 \\ (188.4-238.5) \end{gathered}$ | $\begin{gathered} 251.5 \\ (224.4-286.1) \end{gathered}$ | $\begin{gathered} 239.0 \\ (192.8-3 \mid 4.6) \end{gathered}$ |
| EMR-D | Male | $\begin{gathered} 175.0 \\ (156.7-198.2) \end{gathered}$ | $\begin{gathered} 244.8 \\ (228.7-263.3) \end{gathered}$ | $\begin{gathered} 293.1 \\ (272.6-316.9) \end{gathered}$ | $\begin{gathered} 225.0 \\ (202.4-253.3) \end{gathered}$ | $\begin{gathered} 220.7 \\ (199.6-246.8) \end{gathered}$ | $\begin{gathered} 213.4 \\ (190.4-242.8) \end{gathered}$ | $\begin{gathered} 235.1 \\ (208.6-269.3) \end{gathered}$ | $\begin{gathered} 214.6 \\ (174.1-279.9) \end{gathered}$ |
|  | Female | $\begin{gathered} 163.3 \\ (146.0-185.3) \end{gathered}$ | $\begin{gathered} 240.8 \\ (224.8-259.3) \end{gathered}$ | $\begin{gathered} 247.6 \\ (230.2-267.9) \end{gathered}$ | $\begin{gathered} 224.5 \\ (201.2-253.9) \end{gathered}$ | $\begin{gathered} 237.4 \\ (215.3-264.6) \end{gathered}$ | $\begin{gathered} 210.5 \\ (188.4-238.5) \end{gathered}$ | $\begin{gathered} 251.5 \\ (224.4-286.1) \end{gathered}$ | $\begin{gathered} 239.0 \\ (192.8-314.6) \end{gathered}$ |


| EUR-A | Male | $\begin{gathered} 347.9 \\ (333.9-363.1) \end{gathered}$ | $\begin{gathered} 284.7 \\ (273.9-296.4) \end{gathered}$ | $\begin{gathered} 350.3 \\ (341.5-359.7) \end{gathered}$ | $\begin{gathered} 312.2 \\ (306.3-318.3) \end{gathered}$ | $\begin{gathered} 345.4 \\ (338.6-352.5) \end{gathered}$ | $\begin{gathered} 283.7 \\ (275.9-292.0) \end{gathered}$ | $\begin{gathered} 344.6 \\ (332.0-358.2) \end{gathered}$ | $\begin{gathered} 289.7 \\ (269.1-3 \mid 4.2) \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Female | $\begin{gathered} 280.4 \\ (269.2-292.6) \end{gathered}$ | $\begin{gathered} 226.2 \\ (217.5-235.8) \end{gathered}$ | $\begin{gathered} 290.2 \\ (283.2-297.4) \end{gathered}$ | $\begin{gathered} 254.2 \\ (252.0-258.9) \end{gathered}$ | $\begin{gathered} 262.4 \\ (257.2-267.7) \end{gathered}$ | $\begin{gathered} 265.8 \\ (258.3-273.8) \end{gathered}$ | $\begin{gathered} 317.1 \\ (305.1-330.1) \end{gathered}$ | $\begin{gathered} 266.6 \\ (250.1-285.8) \end{gathered}$ |
| EUR-B | Male | $\begin{gathered} 175.0 \\ (156.7-198.2) \\ 163.3 \\ (146.0-185.3) \end{gathered}$ | $\begin{gathered} 244.8 \\ (228.7-263.3) \end{gathered}$ | $\begin{gathered} 293.1 \\ (272.6-316.9) \end{gathered}$ | $\begin{gathered} 225.0 \\ (202.4-253.3) \end{gathered}$ | $\begin{gathered} 220.7 \\ (199.6-246.8) \end{gathered}$ | $\begin{gathered} 213.4 \\ (190.4-242.8) \end{gathered}$ | $\begin{gathered} 235.1 \\ (208.6-269.3) \end{gathered}$ | $\begin{gathered} 214.6 \\ (174.1-279.9) \end{gathered}$ |
|  | Female |  | $\begin{gathered} 240.8 \\ (224.8-259.3) \end{gathered}$ | $\begin{gathered} 247.6 \\ (230.2-267.9) \end{gathered}$ | $\begin{gathered} 224.5 \\ (201.2-253.9) \end{gathered}$ | $\begin{gathered} 237.4 \\ (215.3-264.6) \end{gathered}$ | $\begin{gathered} 210.5 \\ (188.4-238.5) \end{gathered}$ | $\begin{gathered} 251.5 \\ (224.4-286.1) \end{gathered}$ | $\begin{gathered} 239.0 \\ (192.8-314.6) \end{gathered}$ |
| EUR-C | Male | $\begin{gathered} 96.2 \\ (88.6-105.3) \end{gathered}$ | $\begin{gathered} 178.6 \\ (170.3-187.8) \end{gathered}$ | $\begin{gathered} 254.9 \\ (247.5-262.7) \end{gathered}$ | $\begin{gathered} 220.7 \\ (214.7-227.1) \end{gathered}$ | $\begin{gathered} 231.5 \\ (224.3-239.3) \end{gathered}$ | $\begin{gathered} 192.6 \\ (183.5-202.6) \end{gathered}$ | $\begin{gathered} 176.3 \\ (159.9-196.4) \end{gathered}$ | $\begin{gathered} 165.8 \\ (130.0-228.9) \end{gathered}$ |
|  | Female | $\begin{gathered} 105.5 \\ (97.3-115.3) \end{gathered}$ | $\begin{gathered} 155.9 \\ (148.4-164.2) \end{gathered}$ | $\begin{gathered} 163.4 \\ (206.3-222.2) \end{gathered}$ | $\begin{gathered} 157.6 \\ (189.6-202.3) \end{gathered}$ | $\begin{gathered} 171.6 \\ (166.9-176.5) \end{gathered}$ | $\begin{gathered} 168.2 \\ (161.2-175.1) \end{gathered}$ | $\begin{gathered} 153.5 \\ (144.2-164.1) \end{gathered}$ | $\begin{gathered} 115.4 \\ (102.5-132.0) \end{gathered}$ |
| SEAR-B | Male | $\begin{gathered} 96.2 \\ (88.6-105.3) \end{gathered}$ | $\begin{gathered} 178.6 \\ (170.3-187.8) \end{gathered}$ | $\begin{gathered} 254.9 \\ (247.5-262.7) \end{gathered}$ | $\begin{gathered} 220.7 \\ (214.7-227.1) \end{gathered}$ | $\begin{gathered} 231.5 \\ (224.3-239.3) \end{gathered}$ | $\begin{gathered} 192.6 \\ (183.5-202.6) \end{gathered}$ | $\begin{gathered} 176.3 \\ (159.9-196.4) \end{gathered}$ | $\begin{gathered} 165.8 \\ (130.0-228.9) \end{gathered}$ |
|  | Female | $\begin{gathered} 105.5 \\ (97.3-\mid 15.3) \end{gathered}$ | $\begin{gathered} 155.9 \\ (148.4-164.2) \end{gathered}$ | $\begin{gathered} 163.4 \\ (206.3-222.2) \end{gathered}$ | $\begin{gathered} 157.6 \\ (189.6-202.3) \end{gathered}$ | $\begin{gathered} 171.6 \\ (166.9-176.5) \end{gathered}$ | $\begin{gathered} 168.2 \\ (161.2-175.1) \end{gathered}$ | $\begin{gathered} 153.5 \\ (144.2-164.1) \end{gathered}$ | $\begin{gathered} 115.4 \\ (102.5-132.0) \end{gathered}$ |
| SEAR-D | Male | $\begin{gathered} 96.2 \\ (88.6-105.3) \end{gathered}$ | $\begin{gathered} 178.6 \\ (170.3-187.8) \end{gathered}$ | $\begin{gathered} 254.9 \\ (247.5-262.7) \end{gathered}$ | $\begin{gathered} 220.7 \\ (214.7-227.1) \end{gathered}$ | $\begin{gathered} 231.5 \\ (224.3-239.3) \end{gathered}$ | $\begin{gathered} 192.6 \\ (183.5-202.6) \end{gathered}$ | $\begin{gathered} 176.3 \\ (159.9-196.4) \end{gathered}$ | $\begin{gathered} 165.8 \\ (130.0-228.9) \end{gathered}$ |
|  | Female | $\begin{gathered} 105.5 \\ (97.3-1 \mid 5.3) \end{gathered}$ | $\begin{gathered} 155.9 \\ (148.4-164.2) \end{gathered}$ | $\begin{gathered} 163.4 \\ (206.3-222.2) \end{gathered}$ | $\begin{gathered} 157.6 \\ (189.6-202.3) \end{gathered}$ | $\begin{gathered} 171.6 \\ (166.9-176.5) \end{gathered}$ | $\begin{gathered} 168.2 \\ (161.2-175.1) \end{gathered}$ | $\begin{gathered} 153.5 \\ (144.2-164.1) \end{gathered}$ | $\begin{gathered} 115.4 \\ (102.5-132.0) \end{gathered}$ |
| WPR-A | Male | $\begin{gathered} 201.4 \\ (190.9-213.2) \end{gathered}$ | $\begin{gathered} 204.9 \\ (198.7-206.7) \end{gathered}$ | $\begin{gathered} 255.5 \\ (249.4-261.9) \end{gathered}$ | $\begin{gathered} 239.6 \\ (234.2-245.3) \end{gathered}$ | $\begin{gathered} 268.0 \\ (261.4-275.1) \end{gathered}$ | $\begin{gathered} 278.1 \\ (268.1-288.8) \end{gathered}$ | $\begin{gathered} 249.8 \\ (237.8-263.1) \end{gathered}$ | $\begin{gathered} 238.7 \\ (220.0-260.9) \end{gathered}$ |
|  | Female | $\begin{gathered} 158.1 \\ (149.8-167.4) \end{gathered}$ | $\begin{gathered} 190.4 \\ (184.6-196.7) \end{gathered}$ | $\begin{gathered} 234.2 \\ (228.7-240.1) \end{gathered}$ | $\begin{gathered} 229.8 \\ (224.8-235.1) \end{gathered}$ | $\begin{gathered} 260.0 \\ (253.7-266.6) \end{gathered}$ | $\begin{gathered} 262.0 \\ (252.9-271.7) \end{gathered}$ | $\begin{gathered} 241.4 \\ (231.6-252.1) \end{gathered}$ | $\begin{gathered} 217.3 \\ (203.7-232.8) \end{gathered}$ |
| WPR-B | Male | $\begin{gathered} 110.1 \\ (99.2-123.7) \end{gathered}$ | $\begin{gathered} 136.1 \\ (130.7-142.0) \end{gathered}$ | $\begin{gathered} 161.5 \\ (155.9-167.5) \end{gathered}$ | $\begin{gathered} 157.3 \\ (151.8-163.2) \end{gathered}$ | $\begin{gathered} 167.7 \\ (161.2-174.7) \end{gathered}$ | $\begin{gathered} 167.1 \\ (157.1-178.4) \end{gathered}$ | $\begin{gathered} 141.3 \\ (129.1-\mid 56.0) \end{gathered}$ | $\begin{gathered} 147.1 \\ (125.0-178.8) \end{gathered}$ |
|  | Female | $\begin{gathered} 107.5 \\ (95.4-123.1) \end{gathered}$ | $\begin{gathered} 146.0 \\ (139.9-152.7) \end{gathered}$ | $\begin{gathered} 150.2 \\ (144.6-156.3) \end{gathered}$ | $\begin{gathered} 153.2 \\ (148.1-158.7) \end{gathered}$ | $\begin{gathered} 161.9 \\ (155.6-168.8) \end{gathered}$ | $\begin{gathered} 148.4 \\ (140.0-157.9) \end{gathered}$ | $\begin{gathered} 130.9 \\ (120.6-143.2) \end{gathered}$ | $\begin{gathered} 136.2 \\ (118.6-159.9) \end{gathered}$ |

Figure 9.I Illustration of skewed and normal distribution based on data from the United States

approximation of what the actual intake data resemble (skewness is 2 in this illustration). Note that all data in the skewed distribution are nonzero (even though it appears that there are zero values).

The attributable fraction was then calculated, for the two different distributions, for IHD. The result for the truncated normal distribution with probability mass at zero was $35 \%$. Incorporating a skewness value of 2 resulted in an attributable fraction of $38 \%$, thus suggesting that our general approach is conservative.

### 2.9 Quantitative and qualitative sources of uncertainty in measuring dietary intake

One major source of uncertainty is that the collective term "fruit and vegetables" comprises a very heterogeneous group of foods in different countries or cultures. In a Western-type diet alone, fruit and vegetables include roots, leaves, stems, fruit and seeds from more than 40 botanical families (Lampe 1999). They can be consumed fresh or cooked in many different ways that will influence their biochemical content. Bio-
chemical composition also varies among different types of the same fruit. For example, the vitamin C content of different types of apple varies tenfold. Composition is also subject to differences in growing conditions, such as soil composition, and storage conditions, a factor of increasing importance as commodities are transported globally to ensure yearround supply in developed countries. It was decided to keep fruit and vegetables as a single entity for two main reasons. First, there remains uncertainty as to which components of fruit and vegetables would confer a protective effect. Even if the relevant effective constituents had been correctly identified, the nature of their relationship to disease risk would still need to be correctly specified. Second, obtaining intake data for specific foods (for this project) would have been even more difficult than for fruit and vegetables taken together.

Seasons also influence the amounts and variety of fruit and vegetables consumed. Indeed, evidence is emerging to link seasonality of consumption of fresh fruit and vegetables to the pattern of cardiovascular disease mortality in some countries (Powles et al. 1996). It is possible that the consequences for disease of an annual cycle of seasonal excesses and out-of-season shortages (as in the less economically developed countries of the former Soviet Union) may be quite different to the effects of consuming a similar annual level where counter-seasonal supplies ensure that there is no period of very low consumption (as in the affluent countries of north-west Europe). However, in the absence of information on national variations in fruit and vegetable intake and epidemiological evidence, it was assumed that it is the long-term annualized average of fruit and vegetable intake that best predicts disease risk. The need for caution is illustrated by the case of alcohol, where risk of cardiovascular disease appears to be more sensitive to the pattern of alcohol consumption over time as well as the total amount consumed (Kauhanen et al. 1999; Rehm et al. 2001). It is also assumed that the estimates obtained for this project represent annualized mean intakes, although some surveys collected data during only one or two seasons.

The choice of data sources may also have influenced the final estimates. It was decided that dietary surveys of representative population samples would be used as the primary source of information for this project. However, the quality and validity of individual level data depend on the ability (and willingness) of each individual to provide accurate information on his/her dietary intake (Johansson et al. 2001; Nelson and Bingham 1997). If the aim is to assess current diet, the procedure involved in measuring dietary intake may lead to changes in behaviour. If the aim is to measure past diet, then the reliability of the information provided will depend on memory and on the conceptual abilities of the respondents. The high intakes observed in AMR-A and WPR-A suggest that reported consumption could have been inflated by conscious (social desirability bias) or unconscious over-reporting of fruit and vegetable intake by the survey respondents (Hebert et al. 1995). The reported
intakes in some countries within these subregions are greater than expected. This is particularly the case for the United Kingdom and Germany where the estimated mean national fruit and vegetable consumption was higher than in Mediterranean countries such as Italy and Israel. It is possible that recent public health campaigns, such as those that took place in Finland (Puska 2000), coupled with changes in the retail trade, and thus in marketing and distribution of fruit and vegetables, have improved the dietary habits and increased the fruit and vegetable intake of these populations. This would be consistent with the striking improvements in cardiovascular mortality that they have experienced. Conversely, it is possible that the inclusion of fruit juices in the estimates of fruit and vegetable intakes made the estimates appear larger than expected. Other difficulties include the conversion of food frequencies into mean intakes in surveys that used food-frequency questionnaires, and the limitations and completeness of the various food analysis software programs used in different countries. Finally, it is possible that the survey respondents were not entirely representative of the reference populations, even though most data were from national surveys of dietary intakes.

In dietary surveys, variation as measured by standard deviations is influenced by the method used to collect data (see section 2.2). Most of the surveys used in this study were based on only one day (sometimes two days) of information. It is thus expected that standard deviations were overestimated. However, as described earlier, the method used to pool data from two or more surveys tends to underestimate the level of uncertainty surrounding the pooled standard deviation for the subregion based on a subsample of countries if there is substantial between-country variation.

Although we aimed to obtain dietary survey data for each country, these do not exist and thus food availability statistics were used for subregions where no or few data were available. The validity of food balance sheet statistics depends on the availability and validity of the basic national data on which they are based, including statistics of population, production, stock, import and export. These are known to vary among countries, and from one year to another, both in terms of coverage and accuracy (Kelly et al. 1991). The net availability of vegetables is affected by factors such as non-commercial production and uncertain losses to animal feed, spoilage and waste. However, the FAO performs external consistency checking using supplementary information such as household survey results as well as the application of relevant technical, nutritional and economic expertise in an attempt to eliminate these potential deficiencies. In this study we have used at least three years of FAO data in order to try to reduce the effect of potential yearly variations in coverage and accuracy. However, the current lack of information on adjustment factors to apply to FAO food balance sheet data in developing countries is a source of uncertainty. Finally, extrapolation among coun-
tries would be an important source of uncertainty, especially in the presence of inter-country heterogeneity.

## 3. Estimating risk factor-disease RELATIONSHIPS

### 3.1 OUtcomes to be assessed, evidence of causality, EXCLUDED OUTCOMES AND REASONS FOR EXCLUSION

## Selected outcomes

The outcomes to be assessed are those for which the best evidence of a risk factor-disease relationship is available for fruit and vegetable intake. These include:

- IHD;
- cerebrovascular disease;
- lung cancer;
- stomach cancer;
- colon and rectum cancers; and
- oesophageal cancer.


## EVIDENCE OF CAUSALITY AND EXCLUDED OUTCOMES

A general discussion of the evidence of causality for the association of fruit and vegetable intake with health outcomes follows. Specific details of evidence for each outcome are included in section 3.7.

The choice of outcomes was guided mainly by previous reviews of the literature, including those of Ness and Powles (1997) and Law and Morris (1998) that suggested a protective effect of fruit and vegetables for IHD and stroke. The review from the World Cancer Research Fund (WCRF) and the American Institute for Cancer Research (AICR) for a wide range of cancers (1997) concluded that the evidence for fruit and vegetables decreasing cancer risk was convincing for lung and digestive tract cancers. In the present study, cancers of the lung, oesophagus, stomach, colon and rectum were examined, leaving cancers of the mouth and pharynx for future work. Cancers for which the WCRF/AICR review reported only a probable association (larynx, pancreas and bladder cancers) or limited evidence of an association (cancers which may have a hormonal etiology including ovary, endometrium, thyroid and prostate) were not included in the CRA project at this stage.

Although there is also limited evidence for other health outcomes such as type II diabetes, chronic obstructive pulmonary disease and cataract (Sargeant et al. 2001; Smit et al. 1999; Tavani et al. 1996b), it was decided not to include these in the 2000 revision of the GBD
project because the number of published studies is currently too limited to draw conclusions on the size of the effect for these health problems. The evidence for these will be reconsidered in the future. Although other cardiovascular disease outcomes such as peripheral vascular disease share most risk factors with IHD and occlusive stroke, these outcomes were also excluded from the present report due to the current lack of information on a possible relationship with fruit and vegetable intake.

GENERAL DISCUSSION OF CAUSALITY FOR THE SELECTED EXPOSURE VARIABLE
Bradford Hill's criteria of causality, comprising biological plausibility, temporality, strength, consistency, dose-response, and experimental evidence, were considered in order to determine the likelihood of causality for the association of fruit and vegetable intake with the six selected health outcomes (Hill 1965).

## Biological plausibility

Evidence of causality for the relationship between fruit and vegetable consumption and health can be obtained from the identification of possible biological mechanisms. A number of mechanisms have been proposed (Lampe 1999). They generally involve specific nutrient and non-nutrient constituents of fruit and vegetables, including antioxidants and various other micronutrients in fruit and vegetables. Nutrients are substances in food that the body can use to obtain energy or synthesize tissues (Insel et al. 2003). Non-nutrients are substances, which may not have a known nutrient function but are still important for biological mechanisms such as regulatory functions. Few attempts have been made to simultaneously explore a combination of potential mechanisms.

The issue of biological plausibility is thus extremely complex and evidence remains fragmentary. There are several reasons for this. First, it is very difficult, in conventional epidemiological studies, to specify precisely exposure to different types of foods or food components over a prolonged period of time. As mentioned, obtaining valid information on individual dietary intake is very difficult. Even where this is possible, it is likely that the true content of the reported foods will have varied according to variations within particular types of food (such as different brands of oranges), differences in methods of food preparation, and seasonal variation in food composition. Second, while the growth in understanding of molecular mechanisms of disease is identifying many new factors, in particular non-nutrient components of food which may be important in preventing specific diseases, in many cases their existence, let alone their possible importance, was not known at the time when cohort studies now reporting results were established (e.g. glucosinolates in brassicas [van Poppel et al. 1999] and isoflavones in soya [Messina 1999], both of which appear to reduce the incidence of some types of cancer). The incompleteness of current food composition tables is a
major limitation to the assessment of the possible effect of varying intakes of these substances.

Consequently, much of the available research for the assessment of biological mechanisms is based on studies in animals and often involves the administration of pharmacological, rather than physiological, doses of various substances. This raises important questions about the applicability of such studies to humans. These reservations must therefore be borne in mind when interpreting the evidence discussed below.

## Cancer

The immediate cause of cancer is damage to DNA at some stage during the cell cycle (Weinberg 1996). At the risk of over-simplification, this can arise from one of three broad mechanisms: genetic (e.g. certain childhood cancers), cancers linked with endogenous hormonal patterns such as breast cancer and the action of exogenous carcinogens. This last group includes compounds produced from tobacco and a wide range of other chemical agents, such as asbestos or benzene, ionizing radiation, and, as is being increasingly recognized, many infectious agents (e.g. Helicobacter pylori as the leading cause of stomach cancer).

From this brief review it is apparent that fruit or vegetable consumption might only be expected to influence directly the risk of certain cancers and not others, and even where they do have a role, this is likely to be modulated by a wide variety of other factors, the importance of which will vary in different populations.

The substances present in fruit and vegetables that might have an impact on cancer incidence can be divided into agents that block the action of carcinogens (Table 9.9), agents that suppress carcinogenesis (Table 9.10) and antioxidants, which can prevent oxidative DNA damage. Some of these agents have both complementary and overlapping mechanisms of action.

Antioxidants include certain vitamins, such as vitamins C (Gershoff 1993) and E (Brigelius-Flohe and Traber 1999), carotenoids (Haegele et al. 2000) (including beta-carotene and other compounds such as

## Table 9.9 Selected carcinogen-blocking agents present in fruit and vegetables

| Component | Fruit/vegetable |
| :--- | :--- |
| Terpenes (Schut et al. 1997) | Citrus fruit |
| Organosulfides (Singh et al. 1996) | Allium vegetables: onion, leek, garlic, scallion, chives |
| Indoles (Oganesian et al. I999) | Cruciferous vegetables |
| Flavonoids (Malaveille et al. 1996) | Onion, apple, citrus fruit, tea, coffee, cola, alcoholic <br> beverages <br> Yellow and orange fruit and vegetables, green leafy <br> vegetables |
| Carotenoids (Sengupta and Das 1999) | (Sand |

Table 9.10 Selected carcinogenesis-suppressing agents present in fruit and vegetables

| Component | Fruit/vegetable |
| :--- | :--- |
| Protease inhibitors (Kennedy 1998) | Legumes, potato, spinach, broccoli, cucumber |
| Terpenes | Citrus fruit |
| Isothiocyanates | Cruciferous vegetables |
| Plant sterols (Awad et al. 2000) | Vegetables, beans, seed |
| Carotenoids | Yellow and orange fruit and vegetables, green <br> leafy vegetables |

flavonoids) and selenium. These act by scavenging free radicals that would otherwise damage DNA. In doing so, they would reduce the impact of certain exogenous carcinogens.

In general, any protective effect that fruit and vegetables might exert is more likely to be apparent with cancers where exogenous carcinogens play a major part. Examples include lung, stomach and colorectal cancer. Evidence from observational studies seems to support this. The WCRF/AICR review of the literature (World Cancer Research Fund and American Institute for Cancer Research 1997) concluded that the evidence for fruit and vegetables decreasing cancer risk was convincing for oral-pharyngeal, lung and digestive tract cancers; that there was a probable association for larynx, pancreas and bladder cancers; and that the evidence was limited for cancers which may have a hormonal etiology including ovary, endometrium, thyroid and prostate.

Most human research investigating cancer risk has examined either the effect of a specific compound or of overall fruit and vegetable consumption. Research into the former has yielded mixed results. While many studies have shown an association between high beta-carotene intake and reduced risk of cancer, especially lung and stomach cancer (van Poppel and Goldbohn 1995), a highly publicized study among smokers receiving vitamin supplements, including beta-carotene, found that supplementation was associated with an increased rate of lung cancer. Similarly, while a recent meta-analysis found a small reduction in the risk of breast cancer with high levels of fruit and vegetable consumption (Gandini et al. 2000), some large studies looking at vitamin supplements have found no effect (Watkins et al. 2000; Wu et al. 2000). However, the effect of dietary composition on cancers with hormonal etiologies is likely to be greatest in the pre-pubertal years, as diet during this phase of growth may influence breast cancer risk substantially via its effect on body size, age of menarche, etc.

Given the many substances potentially involved in protecting against cancer, and the diverse mechanisms involved, these mixed results have highlighted the need to look at non-nutrient components of fruit and
vegetables. More research is now investigating, among others, the potential impact of other food components such as bioactive compounds (allium compounds, dithiolthiones, isothiocyanates, terpenoids, isoflavones, protease inhibitors, phytic acid, polyphenols, glucosinolates and indoles, flavonoids, plant sterols, saponins and coumarins) (Verhagen et al. 1995).

In summary, there are many possible mechanisms by which fruit and vegetable consumption might reduce the risk of cancer. However, our knowledge is handicapped by the uncertainty with regard to the many pathways involved in carcinogenesis and the relative quantitative importance of each of the mechanisms that, on current knowledge, could plausibly be involved. However, it appears that the impact of diet is likely to be greatest for cancers caused by specific external carcinogens, such as gastrointestinal and lung cancer, and less important for cancers where endocrine factors play a greater role, such as breast and prostate cancer. Furthermore, the overall importance of diet in a given population will clearly reflect the prevalent pattern of exposure to specific carcinogens as well as differences in genetic susceptibility. Thus, an agent that acts to protect against the effects of a particular carcinogen will have less of an effect in a population where exposure to that carcinogen is rare than where it is common.

## Cardiovascular disease

As with cancer, the multiple mechanisms by which fruit and vegetable consumption might act on the risk of cardiovascular disease are difficult to disentangle because of the inadequate understanding of the determinants of disease. Most research has concentrated on atherosclerosis. Other potential mechanisms by which fruit and vegetables could impact indirectly on cardiovascular risk include a link with blood pressure modulation, through the high potassium content of some fruit and vegetables (Ascherio et al. 1998; Lampe 1999), or with chronic respiratory disease (associated with fruit and vegetables-and their constituents) and vascular disease risk. For the sake of simplicity, this short review will focus on IHD, although some issues are also relevant to cerebrovascular disease.

Atheroma is thought to arise as a result of monocytes adhering to endothelial cells and migrating into the arterial intima where they become macrophages, taking up low density lipoprotein, subsequently becoming foam cells. The role of fruit and vegetable constituents in monocyte adhesion is increasingly well understood. Methionine, derived from dietary protein, is converted within cells to homocysteine. Deficiency of folic acid, vitamins $B_{12}$ and $B_{6}$ will give rise to an elevated level of homocysteine (Chambers et al. 2001). High levels of homocysteine contribute to the generation of free radicals, and thus oxidative damage, in enthothelial cells, leading to the aggregation of monocytes and platelets, as well as vasoconstriction (Kojda and Harrison 1999). These,
in turn, promote atherogenesis. There is now compelling epidemiological evidence to link homocysteine and vascular disease. A recent metaanalysis showed that the risk of cardiovascular disease increased with plasma homocysteine, with odds ratios of 1.6 (1.4-2.3) and 1.8 (1.4-2.3) per $5 \mathrm{~mol} / \mathrm{l}$ increment in plasma homocysteine in men and women, respectively. The relationship was similar for cerebrovascular disease (Boushey et al. 1995).

It has been suggested that other components of fruit and vegetables, in particular antioxidants, act at other stages in the process of atherogenesis. Such compounds could act by reducing the oxidation of lowdensity lipoprotein, thus reducing the formation of fatty streaks and plaques. These antioxidant compounds include vitamin C, which is involved in free radical scavenging, haemostasis and in the stabilization of lipid membranes (Witztum 1994), and beta-carotene, which neutralizes singlet oxygen molecules and prevents chain formation, so reducing oxidative processes that are important in atherogenesis (Gaziano and Hennekens 1993). Flavonoids also inhibit the oxidation of low density lipoprotein and reduce thrombotic tendencies.

Observational studies support a strong inverse association between plasma levels of vitamins C and E and cardiovascular mortality (Daviglus et al. 1997; Gale et al. 1995; Gey 1995; Manson et al. 1992; Yokoyama 2000). However, only a few studies have reported a significant inverse relationship between vitamin C specifically and cardiovascular risk (Khaw et al. 2001). Several reviews have found no significant benefit from vitamin C after controlling for other antioxidant intake or multivitamin use (Price and Fowkes 1997; Rexrode and Manson 1996). Even those that reported a benefit from vitamin C differ with regard to the point at which an effect appears and the potential magnitude of the relationship. Some studies indicated an increased risk of cardiovascular disease only at very low levels of plasma vitamin C, with no effect within the range seen in most populations. Other studies have reported a significantly reduced risk only in persons with the highest levels or with supplemental intake. However, a recent study showed a significant decrease in cardiovascular and IHD risk throughout the normal plasma range (Khaw et al. 2001). Similarly, although results from many observational studies suggest that higher serum levels of beta-carotene reduce the risk of cardiovascular disease, systematic reviews have concluded that evidence for a protective effect is inconsistent (Jha et al. 1995; Rexrode and Manson 1996).

As with studies of cancer, there is no clear evidence from intervention trials that antioxidant supplements reduce the risk of cardiovascular disease (Collins et al. 2002; Gaziano 1996; Greenberg and Sporn 1996; Hennekens et al. 1996; Lonn and Yusuf 1999). In the case of betacarotene there is even some evidence of harm. Several trials were not, however, designed specifically to assess cardiovascular disease risk and did not provide data on non-fatal cardiovascular end-points. Some
authors have suggested that the apparent protective association found in observational studies would be due to residual confounding by differences in socioeconomic status, health behaviour and dietary intake (Ness 2001).

Many of the same challenges that arise with studies of the etiology of cancer also apply to IHD. First, if fruit and vegetables do affect atherogenesis, then their effect will be modulated by other important factors that are involved in atherogenesis and are also influenced by diet. This is particularly the case for high density and low density lipoproteins. These are determined primarily by the amount and nature of fat in the diet, but are also influenced by alcohol consumption, with the precise effect determined by the pattern of drinking. In addition, it is important to remember that atherogenesis is only one factor involved in myocardial infarction. Another is thrombosis, which may also be influenced by certain dietary factors (Chen et al. 2000).

Second, and less well recognized, cardiovascular disease embraces a wide variety of different processes. In particular, it is becoming clear that some myocardial infarctions in young people have a different etiology. Even within the more conventional understanding of IHD, there are clearly differences between those whose atheroma predominantly takes the form of plaques that are lipid-rich, and thus likely to lead to plaque rupture and so to acute infarction, and those that are predominantly fibrous, with smooth muscle proliferation, which are more likely to cause progressive angina (Kharbanda and Vallance 2001).

## EXPERIMENTAL EVIDENCE

## Trials of dietary changes

There is little experimental evidence for the health effects of increasing fruit and vegetables in the diet. Obviously foods that are part of a usual diet are not easily amenable to traditional trials in the general population. Although no trial examined just giving advice to eat more fruit and vegetables on the disease outcomes considered in this report, a few trials of dietary advice in secondary prevention of IHD have included advice to eat more fruit and vegetables (Burr et al. 1989; de Lorgeril et al. 1994, 1999; Rinzler 1968; Singh et al. 1992).

The Diet and Reinfarction Trial (DART) was set up to examine the effect of diet on the secondary prevention of myocardial infarction. Participants were randomized to receive advice or no advice on each of three dietary factors: a reduction of fat intake and an increase in the ratio of polyunsaturated to saturated fat; an increase in fatty fish intake; or an increase in cereal fibre intake. Although the fat advice arm of the trial was associated with an increased fruit and vegetable consumption of about $50 \mathrm{~g} /$ day, no effect on total mortality at two years was observed ( $\mathrm{RR}=1.00,95 \% \mathrm{CI} 0.77-1.30$ ) and there was no effect on IHD events (IHD deaths plus non-fatal myocardial infarction: $\mathrm{RR}=0.91,95 \% \mathrm{CI}$
0.71-1.16) (Burr et al. 1989). It is possible, however, that the follow-up period was too short to allow for an effect to be detected or that the increase in fruit and vegetable consumption was too small to produce a detectable effect.

The Lyon Diet Heart Study investigated whether a Mediterraneantype diet compared with a prudent Western-type diet could reduce the rate of recurrence after a first myocardial infarction. Intermediate analysis showed a marked protective effect after 27 months of follow-up ( $73 \%$ reduction in rate of recurrence and death from cardiovascular causes; $R R=0.27,95 \%$ CI $0.11-0.65)$, which was maintained for four years after infarction ( $\mathrm{RR}=0.28,95 \% \mathrm{CI} 0.15-0.53$ ) (de Lorgeril et al. 1999). The increase in fruit and vegetable consumption in the intervention group was thought to be an important factor in risk reduction. However, as in the DART study, diet changed in a number of ways during the trial and it is thus impossible to estimate the specific influence of increased fruit and vegetable intake in either trial.

Evidence that increasing fruit and vegetable intake alone may be important as a dietary intervention in reducing cardiovascular disease risk comes from the Indian Experiment of Infarct Survival (IEIS) (Singh et al. 1992). This randomized controlled trial showed that the consumption of a low-fat diet enriched with fruit and vegetables compared with a standard low-fat diet was associated with about $40 \%$ ( $R \mathrm{R}=0.60$, $95 \%$ CI $0.31-0.75)$ reduction in cardiac events and $45 \%(R R=0.55$, $95 \%$ CI $0.34-0.75$ ) reduction in mortality in 406 men with acute myocardial infarction after one year. These findings suggest a very rapid effect of dietary change on incidence and mortality from IHD that would appear to be difficult to explain.

Some recent trials also assessed the impact of increased fruit and vegetable intake on blood pressure. In the Dietary Approaches to Stop Hypertension (DASH) trial, hypertensive participants were fed a control diet for three weeks, and then were randomized to receive for eight weeks either the control diet, a diet rich in fruit and vegetables, or a combination diet rich in fruit and vegetables, and reduced in saturated fat, fat and cholesterol (Conlin et al. 2000; Obarzanek et al. 2001). Both the combination diet and the fruit-and-vegetables diet significantly reduced systolic and diastolic blood pressure. After eight weeks, $70 \%$ of the participants on the combination diet had a normal blood pressure, $45 \%$ of those on the fruit and vegetable diet, and $23 \%$ of those on the control diet. Unsurprisingly the fruit and vegetable diet produced few changes in blood lipids and had less effect on blood pressure reduction than the combination diet. Both diets showed that they could potentially help reduce IHD risk. However, studies with a longer follow-up would be required to assess the true long-term effect of such changes.

Another trial assessed the specific effect of increased guava intake in hypertensive individuals (Singh et al. 1993b). After four weeks, the diet rich in guava ( $0.5-1 \mathrm{~kg} /$ day) was associated with $7.5 / 8.5 \mathrm{mmHg}$ net
decrease in mean systolic and diastolic pressures compared with the control group, a significant decrease in serum total cholesterol ( $7.9 \%$ ), triglycerides $(7.0 \%)$, and an insignificant increase in high density lipoprotein (HDL) cholesterol ( $4.6 \%$ ) with a mild increase in the ratio of total to HDL cholesterol. The authors suggested that an increased consumption of guava fruit could cause a substantial reduction in blood pressure and blood lipids without a decrease in HDL cholesterol. These changes were attributed to its high potassium and soluble fibre content, respectively. Further research is needed to confirm this hypothesis with more easily applicable dietary changes.

## Nutrient supplement trials

Due to the lack of trials of increased fruit and vegetable intake on health outcomes, most data from intervention studies relate to studies of nutrient supplements. Unfortunately, these trials have generally been of small sample size and relatively short duration (Lampe 1999).

In contrast to the results of observational studies, findings from trials of antioxidant and vitamin supplementation have mostly shown no effect on mortality, cardiovascular events or cancer (Blot et al. 1995; Correa et al. 2000; Egger et al. 1998; Hooper et al. 2001; Ness et al. 1999a). There has even been some concern following two trials that showed an increased risk of lung cancer mortality with beta-carotene and vitamin A supplements in the Alpha-Tocopherol, Beta-Carotene and Cancer Prevention (ATBC) study (Anonymous 1994) and the Beta-Carotene and Retinol Efficacy Trial (CARET) (Omenn et al. 1996a, 1996b). However, the recent Heart Protection study (HPS) showed neither benefit nor harm of supplementation with antioxidant vitamins after 5.5 years follow-up. This double-blind randomized trial with a $2 \times 2$ factorial design investigated, in over 20500 persons, the use of simvastatin and antioxidant vitamins (vitamin E, vitamin C and beta-carotene) (Collins et al. 2002).

One exception is the Linxian trial in China (Blot et al. 1995). This trial showed reduced total mortality in the group supplemented with beta-carotene, alpha-tocopherol and selenium compared with the placebo group after six years. However, it is not possible to identify which antioxidant contributed most to the trend towards lower mortality. Some trials have also suggested that vitamin E supplementation may prevent ischaemic stroke in high-risk hypertensive patients (Leppala et al. 2000).

These generally null findings, while initially surprising, are not entirely unexpected given the large number of potentially active compounds in food and the scope for interactions, both with other exogenous substances and genetic status. Given that there are very few randomizedcontrolled trials that investigated the association of fruit and vegetable consumption with disease outcomes, current evidence of causality is mainly based on observational studies.

## Strength of association

The review of the evidence from cohort and case-control studies for this project generally supports the hypothesis that a high dietary intake of fruit and vegetables is protective for cardiovascular disease and the selected cancer (see results of the systematic review in sections 3.6 and 3.7). We consider this to be a relatively strong association for the outcomes presented in this chapter, especially taking into account the potential dilution inherent in dietary exposure measurement.

## Consistency

The review of the literature performed for this project shows that most studies of fruit and vegetable consumption demonstrated a generally consistent inverse relationship with the six disease outcomes in different populations (see section 3.7). There were virtually no studies of whole foods (thus excluding nutritional supplements) that showed harmful associations, and many of the studies that reported a null association reported in fact insignificant inverse trends. The major caveat to this statement is that there have been few studies in populations from developing countries.

## Temporality

It is virtually axiomatic that fruit and vegetable consumption will precede disease outcomes. The many cohort studies reviewed here that have long follow-up periods provide more convincing evidence for temporality, as they are less likely to have been affected by information bias, a major source of bias in case-control studies (e.g. recall bias).

## Dose response

Evidence from the literature shows that in general, people in the highest categories of fruit and vegetable consumption have lower risk of cardiovascular disease and cancers compared with those in the lowest consumption categories. Many of the studies also reported a significant trend between the quartiles, quintiles or tertiles of consumption and disease risk, and a few studies have reported significant effects with fruit and vegetable treated as a continuous variable (see sections 3.6 and 3.7).

## Summary

There are still many uncertainties with regard to the mechanisms that lead to common diseases, to the roles that fruit and vegetables could play in these mechanisms, and to the specific substances in fruit and vegetables that are particularly important. Different studies have suggested that flavonoids, carotenoids, vitamin C, folic acid, and fibre could play a protective role. However, it must be kept in mind that studies based on single food constituents may underestimate the effects of exposures such as foods, which are chemically complex. Also, single constituents can be a
marker for other active constituents (as the conflicting results between observational studies and trials have suggested for beta-carotene) (Egger et al. 1998), or even for a combination of constituents that are responsible for the protective effect. Until these mechanisms are better understood, it will not be possible to determine, with any certainty, what precise role specific components of fruit and vegetables might play. What can be said with some confidence is that there are a wide variety of substances within fruit and vegetables that appear to play a role in the etiology of cardiovascular disease and some cancers.

### 3.2 OvERVIEW OF METHODS FOR ESTIMATING RISK FACTOR-DISEASE RELATIONSHIPS

The associations reported in this study were based on a systematic review of the literature. This provided evidence for the direction and size of the relationship between fruit and vegetable consumption and the selected disease outcomes. This was complemented with meta-analyses for four disease outcomes as described in section 3.7.

### 3.3 Criteria for identifying relevant studies

All studies that satisfied the following criteria were included in the systematic review of the literature:

- studies that measured dietary intake of fruit and/or vegetables;
- studies of vegetarians that measured food intake; and
- a special focus was placed on studies that explored associations of fruit and vegetable intake with diseases. However, for completeness we have also included studies that used as their exposure variable proxy measures of intake derived from the measurement of intermediate variables (such as dietary fibre) or biological markers (such as carotenoids, folate, flavonoids, vitamins A and C not due to supplements) where there was a high correlation with the specific food type.


### 3.4 Search strategy and inclusion criteria for the LITERATURE REVIEW

Studies were identified through a systematic review of the literature. The following databases were searched: Medline, Embase, Cochrane On-line, CABHealth and CABAbstracts using the keywords "fruit" or "vegetables" and "coronary heart disease", "cerebrovascular disorder", "lung", "colorectal", "stomach" and "esophageal", "neoplasms" and "cancer". All search terms were linked to MESH headings and exploded. Searches were limited to human studies in English from 1980 onwards (for the CABabstracts, however, the search was from 1987 to 2000).

The outcomes included in this systematic review were the following.

- Cardiovascular diseases: symptomatic heart disease, cerebrovascular disease and total circulatory disease. Studies of peripheral vascular
disease, all-cause mortality and cardiovascular risk factors were excluded.
- Cancer: all histological types of the site-specific cancers were included but not reviewed separately.

Relevant studies for inclusion were determined through the review of titles and abstracts. If there was any doubt regarding study relevance, the full text of the study was retrieved.

This search was complemented with a hand search of citations from books, reviews and citations of references already located. Authors who had published key studies and reviews in the field were also approached to help us identify any other studies, published or unpublished.

### 3.5 Characteristics of excluded studies

Studies were excluded if any one of the following criteria was satisfied.

- The measurement of risk was based solely on blood biochemical markers with no measure or estimate of dietary or nutrient intake.
- The study focus was on investigating the effect of non-dietary supplements.
- The outcome measure was prognosis, pre-cancerous lesions or predisease markers rather than incident cases or mortality.
- The statistical analyses of the study were not adjusted for major confounding factors such as age, sex and smoking.


### 3.6 Description of studies

## Characteristics of included studies

A short summary of the number of studies included in the systematic review is given in Table 9.11. All studies included in the review of the literature are described in section 3.7, where the assessment of causality for each outcome is discussed. Details of the studies included for each specific outcome are provided in Tables 9.21 to 9.28.

## Limitations of the studies included in the review of the literature

The studies included in the review of the literature differed in many ways, including:

- the type of study design;
- the sex, age range and ethnicity of the study population;
- the method and validity of measurement of the dietary exposure;
- the method of reporting the dietary exposure (qualitative vs quantitative);

Table 9.II Summary of the studies included in the review of the literature

| Outcome | Case-control studies |  | Cohort studies |  |
| :---: | :---: | :---: | :---: | :---: |
|  |  |  | Number | Countries/areas covered |
|  | Number | Countries/areas covered |  |  |
| Ischaemic heart disease | Not assessed | $\ldots$ | 27 | Japan, USA and Europe (north) |
| Ischaemic stroke | Not assessed | $\ldots$ | 21 | Japan, USA and Europe (north) |
| Lung cancer | 32 | Canada, China, Brazil, India, Japan, USA and Europe (east, north and south) | 21 | Japan, USA and Europe (north and south) |
| Colorectal cancer | 33 | Argentina, Australia, Canada, China, Japan, Russian Federation, Singapore, Uruguay, USA and Europe (north and south) | 15 | Japan, USA and Europe (north) |
| Gastric cancer | 32 | Canada, China, Japan, Mexico, Republic of Korea, Turkey, USA, Venezuela and Europe (north and south) | 14 | Japan, USA and Europe (north) |
| Oesophageal cancer | 28 | China, India, Japan, USA, Europe (north and south) and South America | 4 | China, Japan and Europe (north) |

- the period of follow up;
- the outcome measured;
- the range of intake of fruit and vegetables of the study population;
- the underlying disease risk of the population (i.e. high vs low); and
- the potential confounders that were adjusted for.

These differences often made results incomparable among studies. The fact that the majority of studies came from Japan, the United States and Europe is another limitation as it restricts the generalizability of the data for use in the CRA project.

## Steps to assess and reduce random error

Most measures of association presented in this report include confidence intervals as a measure of uncertainty. The number of outcome events on which the measures of association are based is also given in the tables.

## Methods used to obtain the estimates of relative risks

To date, there have been few reported meta-analyses of the association of fruit and vegetable intake with disease. In 1998, Law and Morris
reported the results of a meta-analysis of published cohort studies of the relationship of different markers of fruit and vegetable consumption, including dietary intake of fruit and vegetables, on the risk of IHD, adjusted for other factors (Law and Morris 1998). However, the results of this study were criticized by some researchers who suggested that potential residual confounding and heterogeneity among studies could have influenced the results (Ness et al. 1999b).

More recently, a group of researchers reported the results of metaanalysis of previously published case-control and cohort studies. This estimated the association of total fruit or total vegetable consumption with oesophageal, gastric and colorectal cancer (Norat et al. 2001). The methodology used had limitations: studies were included if there was information necessary for the statistical analysis, but there was no assessment of study quality or potential confounding. The studies had measured fruit and vegetable intake in a range of ways, both quantitative and qualitative. If intake data were only available as a qualitative amount (i.e. a subjective categorization into high vs low consumption), the amount of fruit and vegetables consumed in grams was estimated from average consumption in other studies or data sources, including FAO food balance sheets. The methodology used in that meta-analysis highlights the difficulties in obtaining an accurate summary measure of association for studies of fruit and vegetable intake.

## Selection of studies

Considering the large variations among studies with regard to study design, study quality and measurement of both exposure and outcome, it would be methodologically inappropriate, and potentially misleading, to pool results statistically from all the separate studies identified in the systematic review to obtain summary measures of association. This view has also been taken by other researchers who believe, given the quality and heterogeneity of the evidence for fruit and vegetable consumption and the substantial potential error in the measurement of diet, that metaanalyses are not appropriate for pooling observational studies and will only serve to attenuate the error without exploring the heterogeneity which may be important in diet-disease relationships (Ness and Powles 1997, 1999; Wiseman 2002).

In this study, it was decided first to apply strict criteria to select only the best quality and most representative studies for the association of fruit and vegetable intake with each disease outcome. Only the studies meeting these criteria were then eligible for inclusion in a meta-analysis.

The following selection criteria were applied to the studies identified in the systematic review.

- Results from cohort studies were considered as more reliable evidence of association than results from case-control studies. So case-control studies were excluded from the analysis.
- The sample size of the study was large and ideally representative of the population.
- The study population ideally included a broad age range.
- The methodology for data collection and analysis was robust and clearly documented.
- The study collected data on total fruit and vegetable consumption and not just by selected groups of fruit or vegetables (e.g. citrus fruit, green leafy vegetables, raw and cooked vegetables).
- Dietary measurement had been validated and was detailed enough to quantify fruit and vegetable consumption accurately (e.g. a food frequency questionnaire having more than 40 items of fruit and vegetables is likely to be better than one that includes only four items).
- Dietary assessment performed using one 24 -hour recall or food record/diary was excluded.
- The statistical analyses were adjusted for important potential confounders.
- The information was available to enable the estimation of relative risk and confidence intervals with intake treated as a continuous variable for meta-analysis.

The number of studies that met the selection criteria for each outcome is given in Table 9.12.

## Data preparation

The final relative risk estimates are expressed as the unit of change in relative risk associated with an 80 g increase in fruit and vegetable con-sumption-this amount representing a recognized standard serving size (World Cancer Research Fund and American Institute for Cancer

Table 9.I2 Number of cohort studies meeting the selection criteria for inclusion in a meta-analysis

| Outcome | Number of cohort studies reviewed | Number of studies meeting <br> selection criteria |
| :--- | :---: | :---: |
| Ischaemic heart disease | 27 | 4 |
| Ischaemic stroke | 21 | 2 |
| Lung cancer | 21 | 4 |
| Colorectal cancer | 15 | 3 |
| Gastric cancer | 14 | 1 |
| Oesophageal cancer | 4 | 0 |

Research 1997). When data from the selected studies were not presented in this format, the following methods were used.

- Where food consumption was expressed in frequencies (e.g. number of servings per day), these were multiplied by 80 g to give daily intake in grams per day.
- Where the relative risk estimates were reported for various increments in intake (e.g. for 100 g or 1 g increase in intake), the relative risk estimates were first transformed onto a log scale and then divided by the comparison difference to give the log relative risk/gram per day; these were then multiplied by 80 to give final estimates expressed as per 80 g increase.
- Where an overall relative risk was not reported for consumption over the entire population range, two methods were used to obtain the relative risk estimates. In method 1 , we estimated the additional gram per day for which the relative risks given applied (details are given later in the text for each selected study). In method 2, the method of Greenland and Longnecker (1992), implemented in Stata 7, was used to estimate the weighted regression slope over the published relative risks, allowing for correlations due to a common reference category. This method uses all the published relative risks, and should coincide approximately with method 1 if the log relative risks are linear on consumption. When there is non-linearity, the two methods will differ, with the second giving the best "average" slope over the whole consumption range, while the first gives a better estimate of slope over a smaller consumption range. Where there is a tendency for risk reduction to be less marked at higher consumption levels, method 2 will give a more conservative estimate of the relative risk per consumption increase.
- Standard errors were calculated on the log scale by taking the upper (log) confidence limit minus the (log) estimated relative risk and dividing this by 1.96 ; standard errors were also scaled in the same way as estimates to apply to an $80 \mathrm{~g} /$ day comparison difference.


## Meta-analysis

Where more than one study was identified using our strict selection criteria, a meta-analysis was conducted to combine estimates and obtain a summary estimate of the relationship between fruit and vegetable intake and the selected outcome (DerSimonian and Laird 1986). Meta-analysis was performed using study log relative risks and the corresponding standard errors and implemented in Stata 7 ("meta" macro). Heterogeneity between studies was tested using the chi-squared statistic. The randomeffects result was pre-specified conditional on evidence of heterogeneity. When only two studies were available, fixed-effect meta-analysis was
used. Forest plots, showing the results for individual studies, were prepared.

## Extrapolations of the relative risk estimates

Subregion. For each outcome, the same relative risk estimates were applied to all subregions, assuming no interaction between the level of intake and subregion on the associations. However, it is not possible to verify whether this assumption is true, as the study populations covered by the literature reviews were from limited geographical areas, which did not allow subregional comparisons.

Sex. Another issue is the fact that several studies pooled data for men and women. Due to the limitations of the evidence for men and women separately, it was decided to apply the same relative risk estimates to both sexes for each outcome.

Age group. Many of the studies covered only limited age ranges, with most being from middle-aged or elderly populations. None of the studies included in the review were of children aged $<16$ years. Results from the Boyd Orr cohort suggest that childhood fruit consumption may have a long-term protective effect on cancer risk in adults (Maynard et al. 2003). This study measured intake of fruit and vegetables, energy, vitamins C and E, carotene, and retinol during a study of family diet and health in 16 rural and urban areas of England and Scotland from 1937 to 1939 , and followed participants for up to 60 years. However, apart from this cohort, there are currently very few studies which have looked at the impact of childhood intake. A previous review for the New Zealand burden of disease and injury study proposed that the agespecific relative risks for fruit-and-vegetable disease associations describe an inverted u-curve, which assumes that the relative risk is one at the extremes of age ( $<25$ years and $>75$ years) (Tobias 2001). The authors of the review argued that they do not expect individuals aged $<25$ years to be at risk given that the outcomes are chronic diseases, that such outcomes are rare in children, and that children have probably had insufficient duration of exposure. They also applied reduced relative risks to older age groups (i.e. applying a relative risk of 1 to everyone aged $>75$ years), as there are high competing mortality risks at these ages. In this project, it was decided to apply the same relative risks to all age groups between the ages of 15 and 69 years. It is reasonable to postulate that there may be age attenuation in the relative risks at both extremes of age. However, due to the current lack of information on how this would influence the relative risks at varying intakes of fruit and vegetables, approximate age attenuations were applied as follows: at older age groups, the excess risks were reduced by a quarter for ages $70-79$ years, and by half for the age group of $\geq 80$ years. For those
aged $<15$ years, a relative risk of 1 was applied as in the New Zealand study.

Steps to assess and reduce bias and to assess causality
There are a number of generic methodological issues that could lead to the introduction of bias in nutritional epidemiological studies. The following paragraphs briefly describe confounding and the major sources of measurement, recall and selection biases common to these studies, with a particular focus on issues that are specific to studying fruit and vegetables as a risk factor.

## Confounding

Well-known potential confounders of the association of fruit and vegetable intake with cardiovascular disease and cancer include, among others, sex, age and smoking. It is possible that a high intake of fruit and vegetables may be associated with other healthy behaviour, for example lower consumption of saturated fat or non-smoking that also has a protective effect on the selected outcomes.

High intakes of fruit and vegetables may also displace other foods from the diet, causing reduced intake of potentially harmful substances such as saturated fat and salt. Results from the DASH trial suggested that changes in dietary fats do not necessarily accompany an increase in fruit and vegetable intake. In this trial, hypertensive participants were fed a control diet for three weeks and then randomized to receive for eight weeks either the control diet, a diet rich in fruit and vegetables, or a combination diet rich in fruit and vegetables and reduced in saturated fat, fat and cholesterol (Conlin et al. 2000; Obarzanek et al. 2001). Both the combination diet and the fruit-and-vegetables diet significantly reduced systolic and diastolic blood pressure. After eight weeks, 70\% of the participants on the combination diet had a normal blood pressure, $45 \%$ of those on the fruit-and-vegetables diet did and $23 \%$ of those on the control diet did as well. The fruit- and-vegetables diet produced few changes in blood lipids but was still likely to reduce IHD risk independently. Sodium/salt is perhaps an underacknowledged potential confounder for IHD and stroke. Persons who consume more salads may consume less salt. The lack of evidence of confounding by salt mainly relates to the difficulty of measuring sodium exposure in individualsnot to its intrinsic importance.

In order to account for the potential effect of confounding on the relative risk estimates, all studies that were identified in the review of the literature must have performed some statistical adjustment for potential confounders. Most studies adjusted for the basic confounding factors, age and sex. The majority of recent studies also statistically controlled for a range of other variables including smoking, alcohol consumption, total energy intake, other foods and food constituents (including saturated fat intake for heart disease), body mass index (BMI) and vitamin
supplementation. Some studies also adjusted for socioeconomic status, educational level, ethnicity, occupation and place of residence. The potential confounders that are specific to each of the selected disease outcomes are discussed in more detail in section 3.7. It is important to note that statistical adjustment for potential confounding implies accepting that the instruments used to measure these potential confounders did this reasonably well. This may not be the case for all potential confounders (e.g. energy intake or physical activity level). In addition, even where there is a high degree of statistical control for potential confounding, the possibility remains that part of the association estimated is due to uncontrolled (residual) confounding (Ness et al. 1999b).

## Selection bias

The issues related to selection bias in the studies reviewed for this project are similar to those of studies investigating other risk factor-disease relationships. For example, it is generally accepted that the selection of controls in case-control studies is likely to influence the study results. Study participation is usually high for cases but lower for controls; those who participate are more likely to be more health conscious, and thus perhaps consume more fruit and vegetables (Michels et al. 2000).

There were a variety of approaches used to reduce selection bias in the studies reviewed. For example, many studies tried to match controls to their cases as closely as possible in terms of age and sex. The generalizability of the results is also influenced by the source of controls, with population-based controls being better than hospital-based controls.

## Information bias

Exposure. In the studies included in the review of the literature, data on dietary intake were collected using some form of diet history, food frequency questionnaire, 24-hour dietary recall(s) or diary/food record. Measurement of exposure will thus be influenced by the different limitations, sources of error and bias affecting each method. Some of these were described in sections 2.2 and 2.9.

Measurement error is an issue in all studies of dietary exposure (Smith-Warner et al. 1997). In general, this imprecision leads to a substantial attenuation of diet-disease associations, and an underestimation of potential thresholds in these associations (Marshall and Chen 1999). This is the case, for example when a single 24 -hour recall is used as the method of dietary assessment (as in a few studies included in our review of the literature). A single 24 -hour recall has a high degree of intra- to inter-individual variability and cannot accurately represent an individual's usual intake (Bingham et al. 1988). This may lead to important misclassification (likely to be non-differential) error.

Food frequency questionnaires are more commonly used in nutritional epidemiology. However, most food frequency questionnaires can be
criticized because of their limitations in collecting detailed accurate information on the intake of fruit and vegetables (Bingham et al. 1994; Michels et al. 2000). Little is known of the measurement error structure for reported fruit and vegetable intake in food frequency questionnaires. Many early estimates based on comparisons with different questionnaires or diet records had problems, underestimating both the degree of error and the correlation between the sources of errors (Day and Ferrari 2002; Thompson et al. 1997). Various problems are apparent. The level of measurement error is large compared with the true variation of intake in many study populations; there may be systematic bias in reporting at the individual level; and lack of independence of the measurement errors between the food frequency questionnaire and reference instrument (Day et al. 2001; Willett 2001). This may lead to considerably greater relative risk attenuation than has been previously realized, making modest decreases in relative risks difficult to detect (Kipnis et al. 1999). Null results of the diet-disease relationship in studies may thus be misleading, and controlling for a number of correlated dietary variables when exploring the diet-disease association of a specific dietary item can lead to uninterpretable or unpredictable results.

Other factors that can bias diet-disease relationships include the frequent assumption of unchanged dietary intake over long follow-up periods, particularly affecting long-term cohort studies, and recall bias, which is a major problem in case-control studies (Willett 1998b). In some cases, surrogate interviewees (spouses or immediate family members) are asked to provide information for the cases and controls in case-control studies; but this might also lead to misclassification bias. Another important issue to consider is the validity and reliability of extrapolating results from studies based on micronutrient intakes or status to the effect of intakes of fruit and vegetables. In this review, we have thus tried to select only studies with data based on food consumption rather than biomarkers of intakes. Where no appropriate foodexposure data existed, it was decided to include proxy nutrients (such as vitamin C from diet) when this was validated by consumption studies. This approach was necessary, as there were a large number of studies that framed their hypotheses in terms of nutrients, and reported only associations with nutrients. For example, studies such as the EPICNorfolk study (Bingham et al. 1997) have tried to reduce the subjective nature of dietary assessment by using biological indicators such as plasma ascorbic acid, which they correlated with food intake. The use of plasma ascorbic acid measurement is thought to represent dietary intake in the preceding few weeks and may overcome some of the issues involved with dietary assessment (Bates et al. 1991). However, biomarkers are also prone to measurement error that could also explain the lack of consistency in studies in which such biomarkers have been used. In addition, these proxy measures of intake are not ideal, as it is clear that any beneficial effect of fruit and vegetables involves many nutrient
and non-nutrient factors. They would tend to underestimate the impact of a mixed intake of fruit and vegetables.

Finally, there is substantial variability among studies in the categorization of exposure groups, not only in terms of what constitutes the "fruit and vegetables" measured, but also the actual levels of intake within these groups. Exactly what level of intake represents a high or a low intake will vary significantly among populations and will be influenced by the method of data collection. This was discussed in section 2.9. This literature review does not comment on the association of disease risk with specific fruit and vegetables, although several studies have attempted to do this in their analyses.

Outcome. The best studies reviewed for this project were those that utilized more than one method to identify cases to avoid any losses to follow up in the final analysis. The methods used included death certificates, hospital records, living relatives, self-report and cancer registries.

### 3.7 Structured assessment of causality for each OUTCOME CHOSEN

## ISCHAEMIC HEART DISEASE

A detailed description of the studies included in the review of the literature follows.

## Previous reviews of the literature

Three recent reviews of the association between fruit and vegetable consumption and IHD were identified (Klerk et al. 1998; Law and Morris 1998; Ness and Powles 1997). The review by Klerk et al. (1998) concluded that high vs low consumption of fruit and vegetables (increasing from 250 to $400 \mathrm{~g} /$ day) is likely to reduce the risk of IHD by $20-40 \%$ in men and women; however, the methods used to derive the final estimates are unclear.

Law and Morris (1998) performed a meta-analysis of cohort studies of the relationship between IHD and markers of fruit and vegetable consumption, namely dietary intake of fruit, vegetables, carotenoids, vitamin C, fruit fibre and vegetable fibre, and serum concentration of carotenoids and vitamin C, adjusted for other factors. They estimated that the risk of IHD is about $15 \%$ lower at the 90 th than at the 10th percentile of fruit and vegetable consumption.

The review by Ness and Powles (1997) identified 10 ecological, three case-control and 16 cohort studies investigating IHD. Of these, nine ecological studies, two case-control studies, and six cohort studies reported a statistically significant negative relationship between IHD and the consumption of fruit and vegetables or proxy nutrients. Ness and Powles did not attempt to arrive at a summary statistic for the association as the measures of exposure and disease varied considerably between
studies. They concluded that the results are consistent with a protective effect of fruit and vegetables for IHD.

## Current review of the literature

The literature review identified 27 references of prospective studies that investigated the association of IHD risk with fruit and vegetable consumption. Details of the study characteristics are given in Table 9.13. In summary, the study populations were all from Japan, the United States and Europe (Finland, the Netherlands, Norway, Sweden and the United Kingdom). Five of the studies each gave rise to more than one report; 14 studies included men and women, nine studied men only and four studied women only. The follow-up time varied between four and 25 years. The method used to measure fruit and vegetable intake differed considerably among studies, ranging from one 24 -hour dietary recall and a seven-day prospective weighed food diary to a variety of food-frequency questionnaires.

Sixteen of the 27 studies reported a statistically significant inverse association between intake of fruit and vegetables and IHD. Thirteen of these showed an association with food intake, while the other three showed an association with a proxy diet measure that was correlated with fruit and vegetable intake. Nine further studies also reported an inverse association; seven of these were not statistically significant, two others not reporting confidence intervals or measures of statistical significance. The actual relative risks for the studies included in the metaanalysis are given in Table 9.23. Relative risks for the studies in the systematic review are not given here due to the wide variation in how fruit and vegetables were reported and analysed.

## Confounding

Most of the studies reviewed adjusted for some potential confounding factors shown to be associated with the risk of cardiovascular disease. All adjusted for age, and most studies adjusted for sex and smoking. Very few of the older studies had adequately addressed the issue of confounding, and this cannot be discounted as an explanation for an observed association in some studies. However, most recent studies have dealt with a comprehensive range of confounding factors, including the majority of the following: smoking, alcohol, total energy intake, saturated fat intake, cholesterol, BMI, hypertension, type II diabetes, physical activity, hormone replacement therapy, educational status or social class and nutritional supplement use. Measurement of some of these candidate confounders will potentially have substantial error (e.g. energy intake).

## Summary

In summary, the review of the literature suggests that there is a strong inverse relationship between fruit and vegetable intake and
Table 9.13 Summary of cohort studies reporting association between intake of fruit and vegetables and ischaemic heart disease

| Country | Study population (reference) | Sex | Age range (years) | Study <br> size | No. of cases | Follow-up (years) | Exposure measure | Association with fruit or vegetables | Association with diet proxy |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Finland | Finnish Mobile Clinic Health Examination survey cohort (Knekt et al. I994, I996) | Males and females | 30-69 | 5133 | Mortality (244) | 14 | Repeated diet history | Inverse, statistically significant | $\ldots$ |
| Finland | Smokers in ATBC study (Pietinen et al. 1996) | Males | 50-69 | 21930 | Incidence (818), mortality (58I) | 6.1 | Diet history | Inverse, statistically significant | $\ldots$ |
| Japan | Japanese general population survey (1965 census cohort) (Hirayama 1990) | Males and females | $\geq 40$ | 265118 | Mortality (NA) | 16 | Crude-not clear | Inverse, statistically significant (unadjusted) | $\ldots$ |
| Netherlands | Rotterdam study (Klipstein-Grobusch et al. 1999) | Males and females | 55-95 | 4802 | MI (124) | 4 | Food frequency questionnaire | $\ldots$ | Inverse, statistically significant (beta-carotene) |
| Netherlands | Zutphen Elderly study (Hertog et al. 1993) | Males | 65-84 | 805 | MI (38), mortality (43) | 5 | Cross-check of dietary history | Inverse, not statistically significant | Inverse (flavonoids) |
| Sweden | Gothenburg women (Lapidus et al. 1986) | Females | 38-60 | 1462 | MI (23), mortality (75) | 12 | 24-hour recall | $\ldots$ | No association (vitamin C) |
| United Kingdom | Elderly cohort, <br> Department of Health and Social Security survey (Gale et al. 1995) | Males and females | $\geq 65$ | 730 | Mortality (182) | 20 | 7-day weighed food record | $\ldots$ | Inverse, not statistically significant (vitamin C) |

Table 9.13 Summary of cohort studies reporting association between intake of fruit and vegetables and ischaemic heart disease (continued)

| Country | Study population (reference) | Sex | Age range (years) | Study size | No. of cases | Follow-up (years) | Exposure measure | Association with fruit or vegetables | Association with diet proxy |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| United Kingdom | Vegetarians and health conscious individuals study II (Mann et al. 1997) | Males and females | 16-79 | 10802 | Mortality (64) | 13.3 | Food frequency questionnaire | Inverse, statistically significant | $\ldots$ |
| United Kingdom | Vegetarians and health conscious people (Key et al. 1996) | Males and females | $\geq 16$ | 10771 | Mortality (350) | 16.8 | Crude food frequency questionnaire | Inverse, statistically significant | $\ldots$ |
| United Kingdom | EPIC-Norfolk (Khaw et al. 2001) | Males and females | 45-79 | 19496 | Mortality (123) | 4 | Food frequency questionnaire and plasma ascorbic acid analysis | $\ldots$ | Inverse, statistically significant (vitamin C) |
| United Kingdom | Bus and bank workers, London and south England (Morris et al. 1977) | Males | 30-67 | 337 | Mortality (26), incidence (45) | 10-20 | 7-day weighed diary | $\ldots$ | Inverse, not statistically significant (fibre) |
| United Kingdom Scotland | Scottish Heart Health study (Todd et al. 1999) | Males and females | 40-59 | 11629 | Incidence (296) | 6-9 | 60-item food frequency questionnaire | $\ldots$ | Inverse, statistically significant (vitamin C, beta-carotene in men, fibre) |
| United Kingdom Wales | Caerphilly Ischaemic Heart Disease study (Elwood et al. 1996; Fehily et al. 1993) | Males | 45-59 | 2423 | Incidence (148), mortality (132) | 5 (10) | 7-day weighed diet intake | $\cdots$ | Inverse, no Cl (vitamin C, magnesium) |


| USA | Rancho Bernardo cohort, California (Khaw and Barrett-Connor 1987a) | Males and females | 50-79 | 859 | MI and mortality | 12 | 24-hour recall | ... | Positive, not statistically significant (dietary potassium) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| USA | The Adventist Health study, California (Fraser et al. 1992) | Males and females | $\geq 25$ | 26473 | MI (I34), mortality (260) | 6 | 65-item food frequency questionnaire | Positive, not statistically significant | ... |
| USA | Western Electric Company study, Chicago (Pandey et al. 1995) | Males | 40-55 | 1556 | Mortality (231) | 24 | $2 x$ cross-check of diet history and food frequency questionnaire (to participant and homemaker) |  | Inverse, not statistically significant (vitamin C and beta-carotene) |
| USA | Massachusetts Health Care Panel study (Gaziano et al. 1995) | Males and females | $\geq 66$ | 1299 | Mortality (48) | 4.75 | 43-item food frequency questionnaire | Inverse, statistically significant | $\ldots$ |
| USA | lowa Women's Health study (Kushi et al. 1996; Yochum et al. 1999) | Females | 55-69 | 34486 | Mortality (242) | 7 | 127-item food frequency questionnaire | Inverse, statistically significant for some vegetables | Inverse, not statistically significant (vitamin C) |
| USA | NHANES I epidemiologic follow-up study (Bazzano et al. 2000) | Males and females | 24-74 | 11924 | Mortality (793) | 19 | 24-hour recall and food frequency questionnaire | Inverse, statistically significant | ... |
| USA | NHS (Liu et al. 2000) | Females | 34-59 | 39876 | MI (126) | 5 | Food frequency questionnaire | Inverse, not statistically significant | $\cdots$ |
| USA | HPFS (Liu et al. 2001) | Males | 40-75 | 15220 | MI (387) | 12 | Food frequency questionnaire repeated twice yearly | Inverse, statistically significant | $\ldots$ |

Table 9.13 Summary of cohort studies reporting association between intake of fruit and vegetables and ischaemic heart disease

| Country | Study population (reference) | Sex | Age range (years) | Study <br> size | No. of cases | Follow-up (years) | Exposure measure | Association with fruit or vegetables | Association with diet proxy |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| USA | HPFS (Rimm et al. 1996) | Males | 40-75 | 43757 | MI: nonfatal (5II), fatal (229) | 6 | Repeated food frequency questionnaire | Inverse, statistically significant | $\ldots$ |
| USA | NHS/HPFS (Joshipura et al. 200I) | Males Females | $\begin{aligned} & 40-75 \\ & 34-59 \end{aligned}$ | $\begin{aligned} & 42148 \\ & 84251 \end{aligned}$ | Incidence (I I27) Incidence (1063) | 8-14 | \|3|-item food frequency questionnaire at intervals | Inverse, statistically significant | $\ldots$ |
| Multiple countries | Seven Countries study, Japan, USA and Europe (Menotti et al. 1999) | Males | 40-59 | 12763 | Mortality (I555) | 25 | Weighed diet intake (I-7 day records) | Inverse, statistically significant (vegetables) | $\ldots$ |
| NA Not applicable. |  |  |  |  |  |  |  |  |  |
| MI Myocardial infarction. |  |  |  |  |  |  |  |  |  |

cardiovascular disease in over half the prospective observational studies in many different populations. The relationship remains after some adjustment for confounding. This result needs to be interpreted in the light of findings with regard to the more favoured diet hypotheses. In a review of the evidence for the classic diet-heart hypothesis, Willett found a positive association with saturated fat intake in only two of the 12 cohort studies reviewed, and a positive association with cholesterol intake in two (Willett 1998b).

## IsCHAEMIC STROKE

This analysis is confined to ischaemic stroke. Most of the studies in the systematic review had only considered outcomes of ischaemic stroke. Although a few studies had analysed ischaemic and haemorrhagic stroke separately, there was insufficient evidence to draw any conclusions on the association of differential outcomes with fruit and vegetable intake. As it is more biologically plausible that the relationship of fruit and vegetable protection is with ischaemic stroke, we decided to limit our analyses to this outcome. For the attributable burden calculations, stroke was divided into ischaemic and haemorrhagic using the method described in chapter 7.

## Previous reviews of the literature

Three recent reviews that previously studied the association between fruit and vegetable consumption and stroke were identified (Klerk et al. 1998; Ness and Powles 1997, 1999). The review by Klerk et al. (1998) concluded that the risk of stroke is reduced by $0-25 \%$ with higher intakes of fruit and vegetables. The reviews by Ness and Powles (1997, 1999) identified five ecological, one case-control and eight cohort studies reporting measures of association between the intake of fruit and vegetables and stroke. Of these, three ecological studies and six cohort studies reported a statistically significant negative association with the consumption of fruit and vegetables or proxy nutrients. The authors concluded that the results were consistent with a strong protective effect of fruit and vegetables for stroke, but they did not calculate a summary statistic for the association as the measures of exposure and outcome varied considerably among studies.

## Current review of the literature

As with IHD, evidence from case-control and ecological studies was not reviewed because the number of cohort studies identified was sufficiently high (cohort studies represent a stronger study design), and because the general pattern of findings observed was consistent among studies. Overall, the evidence suggests a strong protective effect of fruit and vegetable consumption on ischaemic stroke risk.

Twenty-one references of prospective studies of the association between stroke and the consumption of fruit and vegetables were iden-
tified. Details of the study characteristics are described in Table 9.14. In summary, the study populations were all from China, Japan, the United States or Europe (Finland, the Netherlands, Norway, Sweden and the United Kingdom). Three of the studies gave rise to more than one report. Ten studies had populations of men and women, eight studied men only and four studied women only. The follow-up period varied between five and 28 years. The method used to measure dietary intake of fruit and vegetables also varied considerably, including postal diet survey, 24-hour dietary recall, seven-day prospective weighed diet record, and various food-frequency questionnaires of differing length and quality.

Thirteen studies showed a statistically significant inverse association between the intake of fruit and vegetables and stroke. Six studies showed an association with food, while the other eight showed an association with a nutrient considered to be a proxy for fruit and vegetable intake.

## Confounding

As discussed for IHD, the observed protective association for fruit and vegetables and stroke could, in theory, be explained by confounding. All studies included in the literature review adjusted for age and sex, and most recent studies reviewed also dealt with a comprehensive range of major measured confounders.

## Summary

There is a strong inverse relationship between the level of fruit and vegetable consumption and stroke risk in prospective observational studies in many populations. The relationship persists after adjustment for major confounders.

## LUNG CANCER

A detailed description of the studies included in the review of the literature follows.

## Previous reviews of the literature

Five recent comprehensive reviews of the association of fruit and vegetable intake with lung cancer risk were identified. Three concluded that there was convincing evidence that a diet rich in fruit and vegetables decreases the risk of lung cancer.

One of these three reviews was that of the WCRF/AICR (1997), which reviewed seven cohort and 17 case-control studies. Of the seven cohort studies, after adjustment for smoking, all showed a protective association for some fruit or vegetable. Most of the relative risks (23 of 31) they presented were for a protective association, although not all were statistically significant. No studies showed a statistically significant increase in risk for any type of fruit or vegetable. Sixteen of the case-control studies reported statistically significant inverse associations for one or more vegetable or fruit categories. The evidence was most abundant for
Table 9.14 Summary of cohort studies reporting measures of association between intake of fruit and vegetables and stroke

| Country | Study population (reference) | Sex | Age range (years) | Study size | No. of cases | Follow-up (years) | Exposure measure | Association with fruit or vegetables | Association with diet proxy |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| China | Men in Shanghai (Ross et al. 1997) | Males | 45-64 | 18244 | CVA mortality (245) | 5-8 | Food frequency questionnaire | No association | $\ldots$ |
| Finland | Smokers in ATBC study (Hirvonen et al. 2000) | Males | 50-69 | 26593 | Cerebral infarction (736), haemorrhagic stroke (178) | Diet <br> 6.1 | questionnaire | Inverse, statistically | significant (beta-carotene) |
| Finland | Finnish Mobile Clinic Health Examination survey cohort (Knekt et al. 2000) | Males and females | 30-69 | 9208 | Mortality (244) | 28 | Repeated diet history | $\ldots$ | Inverse, statistically significant <br> (subgroups only) |
| Japan | Japanese general population survey (1965 census cohort) (Hirayama 1990) | Males and females | $\geq 40$ | 265118 | Mortality (NA) | 16 | Crude (not specified) | Positive, not statistically significant | $\ldots$ |
| Japan | Shibata study, rural Japan (Yokoyama 2000) | Males and females | $\geq 40$ | 2\|21 | CVA (109) | 20 | Food frequency questionnaire and serum vitamin C | $\ldots$ | Inverse, statistically significant (vitamin C) infarction and haemorrhagic stroke |
| Netherlands | Zutphen study (Keli et al. 1996) | Males | 50-69 | 552 | CVA (42) | 15 | Repeated crosscheck of dietary history | Inverse, not statistically significant | Inverse statistically significant (flavonoids) |
| Norway | Norwegian dietary postal survey (Vollset and Bjelke 1983) | Males and females | 45-74 | 16713 | Mortality (438) | 11.5 | Postal diet survey | . | Inverse, statistically significant (vitamin C) |

Table 9.14 Summary of cohort studies reporting measures of association between intake of fruit and vegetables and stroke (continued)

| Country | Study population (reference) | Sex | Age range <br> (years) | $\begin{aligned} & \text { Study } \\ & \text { size } \end{aligned}$ | No. of cases | Follow-up (years) | Exposure measure | Association with fruit or vegetables | Association with diet proxy |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sweden | Gothenburg women (Lapidus et al. 1986) | Females | 38-60 | 1462 | CVA (13) | 12 | 24-hour recall | $\ldots$ | No significant correlation (vitamin C) |
| United Kingdom | Vegetarians and health conscious people (Key et al. 1996) | Males and females | $\geq 16$ | 10771 | Mortality (147) | 17 | Crude food frequency questionnaire | Inverse, statistically significant | $\ldots$ |
| United Kingdom | Elderly cohort, <br> Department of Health and Social Security survey <br> (Gale et al. 1995) | Males and females | $\geq 65$ | 730 | Mortality (124) | 20 | 7-day weighed food record | $\ldots$ | Inverse, statistically significant (vitamin C) |
| USA | Rancho Bernardo cohort, <br> California (Khaw and <br> Barrett-Connor 1987b) | Males and females | 50-79 | 859 | Mortality (24) | 12 | 24-hour recall | $\ldots$ | Inverse, statistically significant (dietary potassium) |
| USA | Hawaiian men of Japanese descent (Lee et al. 1988) | Males | 45-68 | 7591 | CVA (408) | 16 | 24-hour recall | $\ldots$ | Inverse (dietary potassium) |
| USA | Framingham study (Gillman et al. 1995) | Males | 45-65 | 832 | Incident stroke (97) | 20 | 24-hour recall | Inverse, statistically significant | $\ldots$ |


| USA | Western Electric Company study, Chicago (Daviglus et al. 1997) | Males | 40-55 | 1556 | CVA (222) | 24 | $2 x$ cross-check diet history and food frequency questionnaire (to participant and homemaker) | $\ldots$ | Inverse, not statistically significant (vitamin C and betacarotene) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| USA | Iowa Women's Health study (Yochum et al. 2000) | Females | 55-69 | 34492 | Mortality (131) | 10 | 127-item food frequency questionnaire | $\ldots$ | No association (flavonoids) |
| USA | NHANES I epidemiologic follow-up study (Bazzano et al. 2000, 2001) | Males and females | 24-74 | 9805 | Stroke events (927) | 19 | 24-hour recall and food frequency questionnaire | Inverse, statistically significant | Inverse, statistically significant (dietary potassium) |
| USA | NHS (Liu et al. 2000) | Females | 34-59 | 39876 | CVA incidence (I60) | 5 | Food frequency questionnaire | Inverse, not statistically significant (all CVD) | $\ldots$ |
| USA | HPFS (Ascherio et al. 1998, 1999) | Males | 40-75 | 43738 | CVA (328): <br> Ischaemic (210), haemorrhagic (70) | 8 | Repeated food frequency questionnaire | $\ldots$ | Inverse, statistically significant (dietary potassium); positive not significant (vitamin C) |
| USA | NHS/HPFS (Joshipura et al. 1999) | Males <br> Females | $\begin{aligned} & 40-75 \\ & 34-59 \end{aligned}$ | $\begin{aligned} & 38683 \\ & 75596 \end{aligned}$ | Ischaemic stroke incidence: F (366), M (204) | 8-14 | Repeated food frequency questionnaire at intervals | Inverse, statistically significant | ... |
| NA | licable. |  |  |  |  |  |  |  |  |
| CVA | vascular accident. |  |  |  |  |  |  |  |  |

green vegetables and carrots. Results of an analysis of dose-response relationship between vegetable intake and risk of lung cancer estimated that the relative risk decreases by about $50 \%$ as intake increases from $150 \mathrm{~g} /$ day to $400 \mathrm{~g} /$ day. An intake of $>400 \mathrm{~g} /$ day is always associated with a lower risk than $100 \mathrm{~g} /$ day or less.

In their review, Ziegler et al. (1996b) asserted that the results of observational studies of diet and lung cancer suggest strongly that an increased fruit and vegetable intake is associated with a reduced risk in men and women; in various countries; in smokers, ex-smokers, and neversmokers; and for all types of lung cancer.

The review by Klerk et al. (1998) concluded that high vs low consumption of fruit and vegetables (an average difference of $150 \mathrm{~g} /$ day ) is likely to reduce the risk of lung cancer by $35-55 \%$ in men and women.

Koo (1997), in contrast, concluded that epidemiological studies performed over the last 20 years do not provide overwhelming evidence of an inverse association between fruit and vegetable consumption and lung cancer risk. Koo proposed the imperfect control of smoking-associated dietary correlates and "lifestyle" differences as the major problems with the perceived associations between diet and lung cancer.

## Current review of the literature

The literature review identified 21 cohort and 32 case-control studies that examined the association of fruit and vegetable intake with the risk of lung cancer incidence and mortality. Details of the studies are provided in Tables 9.15 and 9.16 . Overall, the evidence appears to support an inverse relationship between fruit and vegetable consumption and lung cancer risk (both incidence and mortality).

The cohort studies identified were conducted in a range of countries including Finland, Japan, the Netherlands, the United States and other southern and northern European countries included in the Seven Countries study. The study populations were not necessarily nationally representative as some studies were limited to religious groups or those with particular lifestyle characteristics (Fraser et al. 1991; Key et al. 1996), specific occupational groups (Feskanich et al. 2000; Michaud et al. 2000) or very narrow age groups such as the elderly (Shibata et al. 1992). Three studies from the United States analysed national survey data and are therefore perhaps more nationally representative. Very few studies were of young people.

It should be noted that, of the 21 cohort studies, five of the study populations were each used in more than one study. The studies differed in their analysis by reporting different risk factors (i.e. carotenoids vs fruit and vegetables) or using different outcome measures (mortality vs incidence).

Twelve studied incidence of lung cancer as the outcome measure, while eight studied lung cancer mortality. Follow-up periods varied between
Table 9.15 Summary of cohort studies reporting association between intake of fruit and vegetables and lung cancer

| Country | Study population (reference) | Sex | Age range (years) | Study size | No. of cases | Follow-up (years) | Exposure measure | Association with fruit or vegetable intake | Association with diet proxy |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Finland | Finnish Mobile Clinic Health Examination survey cohort (Knekt et al. 1991) | Males | 20-69 | 4538 | Incidence (117) | 20 | 100-item food frequency questionnaire | Inverse, statistically significant | $\ldots$ |
| Finland | Finnish Mobile Clinic Health Examination survey cohort (Knekt et al. 1997) | Males and females | 15-99 | 9959 | Incidence (15\|) | 24 | 100-item food frequency questionnaire | Inverse, statistically significant | Inverse, statistically significant (flavonoids) |
| Finland | Smokers in ATBC study (Hirvonen et al. 200I) | Males | 50-69 | 27110 | Incidence (791) | 6.1 | Diet history | Inverse, statistically significant | $\ldots$ |
| Japan | Japanese general population survey (1965 census cohort) (Hirayama 1986, 1990) | Males and females | $\geq 40$ | 265118 | Mortality (1917) | 16 | Crude-not clear | Inverse, statistically significant (men) | $\ldots$ |
| Netherlands | Zutphen study (Kromhout 1987; Ocke et al. 1997) | Males | 40-59 | 561 | Mortality (54) | 25 | Repeated crosscheck of dietary history | Inverse, statistically significant | $\ldots$ |
| Netherlands | The Netherlands Cohort study (Voorrips et al. 2000b) | Males and females | 55-69 | 120852 | Incidence $\text { ( } 1074 \text { ) }$ | 6.3 | 150-item food frequency questionnaire | Inverse, statistically significant | $\cdots$ |
| Norway | 3 cohorts (Kvale et al. 1983) | Males and females | NA | 16713 | Incidence (168) | 11.5 | Food frequency questionnaire | Inverse, not statistically significant | $\ldots$ |

Table 9.I5 Summary of cohort studies reporting association between intake of fruit and vegetables and lung cancer (continued)

| Country | Study population (reference) | Sex | Age range (years) | Study <br> size | No. of cases | Follow-up (years) | Exposure measure | Association with fruit or vegetable intake | Association with diet proxy |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| United Kingdom | Vegetarians and health conscious people (Key et al. 1996) | Males and females | $\geq 16$ | 10771 | Mortality (59) | 16.8 | Crude food frequency questionnaire | Inverse, not statistically significant | $\ldots$ |
| USA | National Health Interview survey (Breslow et al. 2000) | Males and females | Nationally representative | 20004 | Mortality (158) | 8.5 | 59-item food frequency questionnaire | Inverse, not statistically significant | $\ldots$ |
| USA | Volunteers from 25 states (American Cancer Society cohort) (Wang and Hammond 1985) | Males and females |  | 1000000 | NA | 11 | Not clear | Inverse, not statistically significant | $\ldots$ |
| USA | lowa Women's Health study (Steinmetz et al. 1993) | Females | 55-69 | 41387 | Incidence (179) | 4 | 127-item food frequency questionnaire | Inverse, statistically significant | $\ldots$ |
| USA | Lutheran Brotherhood Insurance cohort (Chow et al. 1992) | Males |  | 17818 | Mortality (219) | 20 | Diet questionnaire | Inverse, not statistically significant | $\ldots$ |
| USA | Leisure World cohort, California (Shibata et al. 1992) | Males and females | 50-79 | 11580 | Incidence (164) | 8 | 24-hour recall | Inverse, not statistically significant (women) | $\cdots$ |


| USA | The Adventist Health study, California (Fraser et al. 1991) | Males and females | $\geq 25$ | 34198 | Incidence (61) | 6 | 65-item food frequency questionnaire | Inverse, not statistically significant |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| USA | NHANES I <br> Epidemiologic follow-up study (Yong et al. 1997) | Males and females | 25-74 | 10068 | Incidence (248) | 19 | 24-hour recall and food frequency questionnaire | Inverse, statistically significant | $\ldots$ |
| USA | NHS (Speizer et al. 1999) | Females | 34-59 | 121700 | Incidence (593) | 16 | Repeated food frequency questionnaire | Inverse, statistically significant | $\ldots$ |
| USA | NHS/HPFS (Feskanich et al. 2000; Michaud et al. 2000) | Males Females | $\begin{aligned} & 40-75 \\ & 34-59 \end{aligned}$ | $\begin{aligned} & 47778 \\ & 77283 \end{aligned}$ | Incidence (793) | 10-12 | \|3|-item food frequency questionnaire at intervals | Inverse, not statistically significant | Inverse, statistically significant (carotenoids) |
| Multiple countries | Seven Countries study, Japan, USA and Europe (Mulder et al. 2000) | Males | 40-59 | 12763 | Mortality (424) | 25 | Various methods in each country | Inverse, not statistically significant | $\ldots$ |
| Multiple countries | Men in Finland, Italy and Netherlands, part of Seven Countries study (Jansen et al. 2001) | Males | 40-59 | 3108 | Mortality (187) | 25 | Cross-check of diet history | Inverse, not statistically significant | $\ldots$ |
| NA Not a | cable |  |  |  |  |  |  |  |  |

Table 9.16 Summary of case-control studies reporting an association between fruit and vegetable intake and lung cancer

| Country | Study population/area (reference) | Sex | Association with fruit or vegetables | Association with specific risk factors only |
| :---: | :---: | :---: | :---: | :---: |
| Brazil | Rio de Janeiro (Suzuki et al. 1994) | Males and females | Null | - |
| Canada | Toronto (Jain et al. 1990) | Males and females | Inverse (vegetables) | - |
| China | (Hu et al. 1997) | Males and females | Null | - |
| China | Yunnan miners (Forman et al. 1992) | Males | Null | - |
| China | Yunnan miners (Swanson et al. 1992) | Males | Inverse (vegetables) | - |
|  |  |  | Null (fruit) |  |
| China | North-east China, women (Wu-Williams et al. 1990) | Females | Null | - |
| China |  |  |  |  |
| Hong Kong SAR | Never-smokers (Koo 1988) | Females | Null | - |
| Greece | Never-smokers (Kalandidi et al. 1990) | Females | Inverse (fruit) | - |
| India | Kerala (Sankaranarayanan et al. 1994) | Males and females | Null | Inverse (pumpkin, onion) |
| Italy | Lombardy (Pisani et al. 1986) | Males and females | - | Inverse (carrot) |
| Japan | Nagoya (Takezaki et al. 2001) | Males and females | Inverse | - |
| Japan | Tokai (Gao et al. 1993) | Males and females | Inverse | - |
| Poland | (Pawlega et al. 1997) | Males | Inverse | Null (green vegetables) |
| Poland | (Rachtan and Sokolowski 1997) | Females | Inverse-no smoking adjustment | - |
| Spain | Barcelona (Agudo et al. 1997) | Females | Inverse (vegetables), null (fruit) | - |
| Sweden | Stockholm, never-smokers (Nyberg et al. 1998) | Males and females | Null | - |


| Sweden | West Sweden (Axelsson et al. 1996) | Males | Inverse (vegetables) | - |
| :---: | :---: | :---: | :---: | :---: |
| United Kingdom | Oxford (Harris et al. 1991) | Males | Null | - |
| United Kingdom | South-west England (Darby et al. 2001) | Males and females | Null | Inverse (carrot, tomato) |
| USA | (Pillow et al. 1997) | Males and females | Null (vegetables), inverse (fruit) | - |
| USA | Hawaii (Goodman et al. 1992) | Males and females | Null | - |
| USA | Hawaii (Le Marchand et al. 1989) | Males and females | Inverse | - |
| USA | Florida, female never-smokers (Candelora et al. 1992) | Females | Inverse (vegetables) | - |
|  |  |  | Null (fruit) |  |
| USA | Louisiana (Fontham et al. 1988) | Males and females | Inverse | - |
| USA | New Jersey (Dorgan et al. 1993) | Males and females | Inverse (vegetables) | - |
| USA | New Jersey (Ziegler et al. 1986) | Males | Inverse | - |
| USA | New Jersey (Ziegler et al. 1996a) | Males and females | Inverse | - |
| USA | New York (Mayne et al. 1994) | Males and females | Inverse | - |
| USA | New York (Byers et al. 1987) | Males and females | Null (carotene) | Inverse, males only |
| USA | New York, Buffalo (Mettlin 1989) | Males and females | Inverse | - |
| USA | Texas (Bond et al. 1987) | Males | Null | - |
| Multiple countries | Europe, never-smokers (Brennan et al. 2000) | Males and females | Null | Inverse (tomato) |
| Inverse Statistically significant protective association of high vs low fruit/vegetable consumption. |  |  |  |  |
| Null Non-significant. |  |  |  |  |
| - No data. |  |  |  |  |

four and 25 years, with 16 studies having follow-up periods longer than 10 years. Twelve of the studies investigated populations of men and women, seven studies had male-only cohorts and two studies were entirely female. Most studies have pooled the results for men and women, the explanation being that the number of lung cancer cases in women is too small to justify a meaningful separate analysis. Only a few studies have analysed men and women separately, usually where the entire study cohort was either men or women. More detailed characteristics of all studies are given in Table 9.15.

Eleven studies showed a statistically significant inverse association between a diet high in fruit or vegetables (one of these was significant for dietary carotenoids) and lung cancer. The remainder of the studies showed an inverse association that was not statistically significant. In some studies with non-significant results for total fruit and vegetables, subcohort analyses were reported as having significant associations.

Of particular interest are studies of lung cancer incidence and mortality that considered non-smokers, former smokers and smokers separately in the analyses. It appears that the benefit conferred through a high intake of fruit and vegetables more often reaches statistical significance in current smokers than in non-smokers (however, confidence intervals are often large and overlap). In summary, eight of the cohort studies reviewed for this project stratified the analyses by smoking status. Three of these showed a significant relationship between fruit and vegetable consumption and lung cancer incidence in smokers but not in nonsmokers (but for one of these, the relative risks of both groups were similar) (Fraser et al. 1991; Knekt et al. 1991; Yong et al. 1997), while only one study showed an inverse relationship in non-smokers only (Knekt et al. 1997). The other studies showed non-significant results (Feskanich et al. 2000; Mulder et al. 2000; Steinmetz et al. 1993; Voorrips et al. 2000b) for both smokers and non-smokers. These results would tend to agree with the hypothesized biological mechanisms for the benefits of fruit and vegetables in lung cancer through late stage modification of carcinogenesis following an initial carcinogen exposure. Further research is needed to consider the current inconsistencies in findings, including exposure among non-smokers to environmental tobacco smoke and the range of other lung cancer risk factors.

The literature review identified 32 case-control studies that satisfied the inclusion criteria. These are summarized in Table 9.16. As the case-control studies were not used in the meta-analyses, fewer study details are provided here for conciseness. These studies were conducted across a range of populations including Brazil (1), Canada (13), China (4), Greece (1), India (1), Italy (1), Japan (2), Poland (2), Spain (1), Sweden (2), the United Kingdom (2) and the United States. Of the studies, 19 collected data from men and women, seven from men only and six from women only.

Eighteen found an inverse relationship between a high intake of fruit or vegetables and lung cancer. When separate associations with fruit and vegetables were analysed, there seemed to be a greater number of studies finding significant associations with vegetables as a group compared to fruit as a group. Four of the studies found significant associations only with specific types of vegetables, such as pumpkin and onion (Sankaranarayanan et al. 1994), or tomato (Brennan et al. 2000), or carrot (Pisani et al. 1986) or in specific age groups such as men aged 60-79 years (Byers et al. 1987). The results of these subgroup analyses should be treated with caution.

Of the four case-control studies that specifically collected and analysed data on non-smokers, two found no association, while two found a significant inverse association between fruit or vegetable consumption and lung cancer. But due to the inherent limitations of case-control studies, the results do not necessarily imply that dietary modification after quitting smoking is effective in reducing risk. Exsmokers with elevated vegetable and fruit consumption could have been high consumers while they were smokers as well. However, the results of case-control studies seem consistent with the results of the cohort studies.

Although outside the criteria of this review, it is important to mention some of the evidence from experimental studies that may appear to contradict the protective relationship of fruit and vegetables on lung cancer. The CARET study (Anonymous 1994) and the ATBC study (Omenn et al. 1996b) were randomized control trials designed to investigate the effect of beta-carotene high-dose supplementation on lung cancer. The results of these trials suggested a harmful effect (increase in incidence and mortality) of beta-carotene supplementation in current smokers. This suggests that raised beta-carotene may be a marker associated with other protective factors found in foods (Lonn and Yusuf 1999), and that the status of the antioxidant hypothesis might need to be re-evaluated critically (Ness et al. 1999a).

## Confounding and interactions

All studies of lung cancer adjusted for age, sex and smoking in their analyses, except for one case-control study from Poland (Rachtan and Sokolowski 1997), which did not adjust for smoking. Other potential confounders dealt with statistically included: environmental tobacco smoke; previous lung disease; occupational exposure (arsenic, asbestos, chloromethyl ethers and nickel); radon; air pollution; total energy intake; intake of other macronutrients; BMI; physical activity; and socioeconomic status (using educational level or occupation used as a proxy). The actual adjustment varied among studies.

Smoking is the most important potential confounder to consider given the strength of the association between smoking and lung cancer. There is also the possibility of smoking being an effect modifier but published
results are contradictory. The studies reviewed here have dealt with smoking in a number of ways. Some studies have not only adjusted for current smoking status, but have also considered the intensity of current smoking behaviour (as the number of cigarettes smoked per day), age of starting smoking and duration of smoking. For current non-smokers, several studies have also adjusted for time since quitting (including intensity) in ex-smokers. A couple of studies also adjusted for time since quitting for subjects who quit during follow-up.

An important issue, highlighted by Ziegler et al. (1996b), is that several studies have shown that consumption of fruit, vegetables and carotenoids is higher in non-smokers compared with current smokers, and that consumption is inversely related to smoking intensity in smokers. Thus, those studies that only considered smoking status without smoking intensity might have generated inflated estimates of the protective effect of fruit and vegetables.

The interpretation of these studies is difficult given the strong misclassification bias and the strong association between lung cancer and tobacco use, making it difficult to ensure that all the confounding effects from smoking have been removed (Boshuizen et al. 2002).

## Summary

There is some evidence of an inverse relationship in many populations between lung cancer risk and fruit and vegetable consumption. There is currently not enough evidence to justify stratifying the results by smoking status.

## GASTRIC CANCER

A detailed description of the studies included in the review of the literature follows.

## Previous reviews of the literature

Four recent reviews of the literature concluded that epidemiological evidence shows a consistent protective effect of fruit and vegetable intake on stomach cancer risk (Klerk et al. 1998; Kono and Hirohata 1996; Steinmetz and Potter 1996; World Cancer Research Fund and American Institute for Cancer Research 1997).
The report from the WCRF/AICR (1997) reviewed six cohort and 32 case-control studies. Three of the six cohort studies and 27 of the 32 case-control studies reported a statistically significant protective association for one or more vegetable or fruit categories. The evidence for raw vegetables, allium vegetables and citrus fruit in particular is consistent with a protective effect. Any contradictory evidence related entirely to salted and pickled vegetables. Analyses of dose-response relationships suggested that the risk of stomach cancer decreases by about $50 \%$ as fruit and vegetable intake increases from $50 \mathrm{~g} /$ day to $300 \mathrm{~g} / \mathrm{day}$. An intake of $>150 \mathrm{~g} /$ day is always associated with a lower risk than $100 \mathrm{~g} /$ day or
less. In comparison, the review by Klerk et al. (1998) concluded that high vs low consumption of fruit and vegetables (an average difference of $150 \mathrm{~g} /$ day ) is likely to reduce the risk of stomach cancer by $40-55 \%$ in men and women.

A recent meta-analysis by Norat et al. (2001) of previously published case-control and cohort studies examined the association of total fruit or total vegetable consumption with gastric, colorectal and oesophageal cancer. It included all studies published in English from 1973-2000 and referenced in Medline that provided data on total fruit or vegetable intake. There was no assessment of study quality, adjustment for confounders was not assessed and studies were included as long as they could provide the information necessary for the statistical analysis. For gastric cancer, 32 studies were included that analysed total fruit intake and 22 studies were included for total vegetable intake. The pooled relative risks associated with an increase of consumption of $100 \mathrm{~g} /$ day were 0.75 ( $95 \%$ CI $0.67-0.83$ ) for fruit and $0.80(0.74-0.86)$ for vegetables.

## Current review of the literature

The current literature review identified 14 cohort and 32 case-control studies that investigated the association between gastric cancer risk (incidence and mortality) and the consumption of fruit and/or vegetables. Overall, the evidence seems to support an inverse association between fruit and vegetable consumption and gastric cancer risk.

The 14 cohort studies were conducted in populations from both developed and developing countries including one from China, two from Japan, two from mainland USA, three of Japanese descendants in Hawaii and six from Europe. The Asian populations had a relatively high risk of stomach cancer. The results of the studies are summarized in Table 9.17.

Of these studies, nine investigated the effect of fruit and vegetable intake on gastric cancer incidence, while the remaining five studied mortality. Two cohorts were investigated in more than one study (the Netherlands Cohort study, and Japanese descendants in Hawaii in the Honolulu Heart Program). Seven cohorts consisted of both men and women, six studies investigated men only and one study focused exclusively on post-menopausal women. Follow-up of cohorts ranged from five to 25 years.

A statistically significant inverse association between gastric cancer and fruit and vegetable intake was reported in four studies. This inverse relationship was found in populations of both European and Japanese origin. The 10 other studies showed inverse relationships, but they were not statistically significant for fruit and vegetable, or nutrient intake.

The 32 case-control studies, summarized in Table 9.18, represented a wider range of populations and geographical areas than the cohort studies. Studies were conducted in Canada and the United States (5), China (3), Japan (4), Mexico (1), Poland (1), the Republic of Korea (1),
Table 9.17 Summary of cohort studies reporting a measure of association between intake of fruit and vegetables and gastric

| Country | Study population (reference) | Sex | Age range (years) | Study size | No. of cases | Follow-up (years) | Exposure measure | Association with fruit or vegetables | Association with diet proxy |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| China | Linxian Nutrition Intervention Trial cohort (Guo et al. 1994) | Males and females | NA | 29584 | Incidence (539) | 5 | Dietary interview | Inverse, not statistically significant (fruit) | - |
| Finland | Smokers in ATBC study (Hirvonen et al. 2001) | Males | 50-69 | 27110 | Incidence (111) | 6.1 | Food frequency questionnaire | - | Inverse, not statistically significant (flavonoids) |
| Japan | Japanese general population survey (1965 census cohort) (Hirayama 1986, 1990) | Males and females | $\geq 40$ | 265118 | Mortality (5 247) | 17 | Crude-not clear | Inverse, statistically significant trend | - |
| Japan | Rural cohort (Kato et al. 1992) | Males and females | NA | 9753 | Mortality (57) | 6 | Food frequency questionnaire | Inverse, not statistically significant | - |
| Netherlands | The Netherlands Cohort study (Botterweck et al. 1998, 2000) | Males and females | 55-69 | 120852 | Incidence (282) | 6.3 | 150-item food frequency questionnaire | Inverse, statistically significant | Non-significant trends: inverse (vitamin C), positive (beta-carotene retinol) |
| Sweden | Cohort of Swedish twins (Terry et al. 1998) | Males and females | NA | 11546 | Incidence (116) | 21 | Crude food frequency questionnaire | Inverse, statistically significant | - |


| United Kingdom Wales | The Caerphilly study (Hertog et al. 1996) | Males | 45-69 | 2112 | Mortality (45) | 13.8 | Food frequency questionnaire | Inverse, not statistically significant | - |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| USA | Iowa Women's Health study (Zheng et al. I995) | Females | 55-69 | 34691 | Incidence (26) | 7 | 127-item food frequency questionnaire | Inverse, not statistically significant | - |
| USA | Men of German and Scandinavian origin (Kneller et al. 1991) | Males | NA | 17633 | Mortality | 20 | Food frequency questionnaire | Inverse, not statistically significant | - |
| USA | Japanese residents, random survey of Hawaiian households (Galanis et al. 1998) | Males and females | $\geq 18$ | 11907 | Incidence (108) | 14.8 | Food frequency questionnaire | Inverse, statistically significant | - |
| USA | Cohort of Hawaiian men of Japanese Ancestry, Honolulu Heart Program (case cohort study) (Chyou et al. 1990) | Males | 49-68 | 8006 | Incidence (111) | 18 | Food frequency questionnaire and 24 -hour recall | Inverse, not statistically significant | - |
| USA | Cohort of Hawaiian men of Japanese ancestry, Honolulu Heart Program (Nomura et al. 1990) | Males | 49-68 | 7990 | Incidence (150) | 19 | Food frequency questionnaire | Inverse, not statistically significant | - |
| Multiple countries | Seven Countries study, Japan, USA and Europe (Ocke et al. 1998) | Males | 40-59 | 12763 | Mortality (NA) | 25 | Various methods (population intake) | Inverse, not statistically significant | - |
| NA Not applicable. |  |  |  |  |  |  |  |  |  |
| No data. |  |  |  |  |  |  |  |  |  |

Table 9.18 Summary of case-control studies reporting a measure of association between intake of fruit and vegetables and

| Country | Study population/area (reference) | Sex | Association with fruit and vegetables | Association with specific risk factors only |
| :---: | :---: | :---: | :---: | :---: |
| Belgium | (Tuyns et al. 1992) | Males and females | Inverse | - |
| Canada | (Risch et al. 1985) | Males and females | Null | Inverse (citrus fruit) |
| China | (Hu et al. 1988) | Males and females | Null | Inverse (spinach) |
| China | North-east China (You et al. 1988) | Males and females | Inverse | - |
| China | Shanghai (Ji et al. 1998) | Males and females | Inverse | Null (vegetables) in females |
| France | (Cornee et al. 1995) | Males and females | Null | - |
| Germany | (Boeing et al. 1991a) | Males and females | Inverse (fruit) | - |
| Greece | (Trichopoulos et al. 1985) | Males and females | Inverse (vegetables), null (fruit) | - |
| Italy | (Buiatti et al. 1989) | Males and females | Inverse | Null (cooked vegetables) |
| Italy | (Munoz et al. 1997) | Males and females | Null | Inverse (citrus fruit) |
| Italy | Milan (La Vecchia et al. 1987) | Males and females | Inverse | - |
| Japan | (Hoshiyama and Sasaba 1992) | Males and females | Null | Inverse (raw vegetables) |
| Japan | (Kato et al. 1990) | Males and females | Null | Inverse (raw vegetables) in males |
| Japan | (Kono et al. 1988) | Males and females | Null | Inverse (mandarins) |
| Japan | (Tajima and Tominaga 1985) | Males and females | Null | Inverse (spinach, onion) |
| Mexico | Mexico City (Ward and Lopez-Carrillo 1999) | Males and females | Inverse | - |
| Poland | (Boeing et al. 1991b) | Males and females | Inverse | - |

Cracow (Jedrychowski et al. 1986)
Inverse
Inverse
Inverse
Inverse (vegetables), null (fruit)
Inverse
Inverse
ull
Inverse
Inverse Inverse

Null
Inverse (only for fruit)
Positive for fruit
Null (vegetables), inverse (fruit) Inverse (vegetables) Males and females Males and females Males and females Males and females Males and females
 Males and females Males and females Males and females Males and females Males and females Males and females Males

Males and females Males and females

[^37](Munoz et al. 2001)

Turkey (2), Venezuela (1) and northern and southern Europe (14). All but two of these studies were carried out on both men and women. Not all the populations were at high risk of stomach cancer.

Of the 32 studies, 20 reported a significant inverse association between total fruit or vegetable consumption and gastric cancer risk. A further nine studies found an inverse association only with specific food types or within a subcohort. Particular attention was placed on the consumption of allium vegetables (onion, leek, garlic and chives). One study (from Venezuela) found that the risk of gastric cancer incidence was inversely related (protective effect) with vegetable intake but directly related (harmful effect) with fruit intake. This is the only study reporting a significant positive (harmful) relationship (Munoz et al. 2001). Despite the inherent limitations of information bias in case-control studies, the evidence is strong and consistent for a protective effect from a diet high in fruit and vegetables, and supports the findings of the cohort studies.

## Confounding

Most studies appear to have adjusted adequately for potential confounding factors. All adjusted for age and sex, and most studies adjusted for smoking and alcohol consumption. Other factors, considered particularly in prospective studies, were previous history of stomach illness, family history of stomach cancer, other dietary components and socioeconomic status. Protective associations remained even after adjustment for other dietary factors such as salty foods or starchy foods in some studies (Coggon et al. 1989; Hansson et al. 1993; La Vecchia et al. 1987; Lee et al. 1995; Ramon et al. 1993; Risch et al. 1985).
H. pylori infection is a major risk factor for gastric cancer, yet few studies were able to collect this information given the retrospective nature of many studies and the recent ability of testing to confirm infection.

## Summary

There is an inverse relationship between fruit and vegetable intake and gastric cancer risk in both cohort and case-control studies in different populations worldwide. The relationship remains after adjustment for confounding. Some of the variations in the findings from separate studies could be due to the between-country differences in the varieties of fruit and vegetables consumed, the methods of consumption (raw or cooked), the number of specific fruit or vegetable items included in the questionnaires used or the validity of the dietary assessment methods.

Although H. pylori infection is an established risk factor, its relationship with fruit and vegetable consumption remains inadequately understood. A multistage model of gastric carcinogenesis is now accepted, according to which different dietary and non-dietary factors, involving genetic susceptibility, are involved at different stages in the cancer
process. A protective effect of supplementation of vitamin $C$ and betacarotene in the progression of pre-malignant gastric lesions was found in Latin America (Correa et al. 2000), but alpha-tocopherol and betacarotene supplement trials in Finland showed no effect. In spite of this, results from the observational studies suggest a protective effect of diets rich in fruit and vegetables.

## Colorectal cancer

A detailed description of the studies included in the review of the literature follows.

## Previous reviews of the literature

Four recent comprehensive reviews of the literature investigating fruit and vegetable consumption and risk of colorectal cancer were found. They all concluded that the evidence is consistent in supporting a decreased risk of colorectal cancer with higher consumption of vegetables, and that data for an association with fruit consumption are inconsistent (Klerk et al. 1998; Norat et al. 2001; Potter 1996; World Cancer Research Fund and American Institute for Cancer Research 1997).

The review by Klerk et al. (1998) estimated that high vs low consumption of fruit and vegetables (an average difference of $150 \mathrm{~g} /$ day) is likely to reduce the risk of colorectal cancer by $20-45 \%$ in men and women.

The meta-analysis by Norat et al. (2001) (see description above in the section on gastric cancer) included 13 studies assessing the effect of total fruit intake and 28 studies assessing the effect of total vegetable intake. The pooled relative risks associated with an increased intake of $100 \mathrm{~g} /$ day were 0.94 ( $95 \%$ CI $0.90-0.98$ ) for fruit and 0.90 ( $95 \%$ CI $0.84-0.96$ ) for vegetables (subanalyses found similar relative risks for men and women, and for European and USA populations).

## Current review of the literature

The systematic review identified 15 cohort studies and 34 case-control studies that examined the association between colorectal cancer risk (both incidence and mortality) and the consumption of fruit and vegetables. Details of these studies are given in Tables 9.19 and 9.20. Overall, the evidence supports an inverse relationship between fruit and vegetable intake and colorectal cancer risk (incidence and mortality), although it is not as strong as that for gastric cancer.

The studies investigating colorectal cancer also included a number of studies that looked separately at colon and rectal cancers. For the purposes of this review, we have combined the results of all these outcomes.

The cohort studies were conducted on a limited range of populations with most studies being from the United States (10), the others being
Table 9.19 Summary of cohort studies reporting a measure of association between intake of fruit and vegetables and colorectal

| Country | Study population (reference) | Sex | Age range (years) | Study <br> size | No. of cases | Follow-up (years) | Exposure measure | Association with fruit or vegetable | Association with diet proxy intake |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Finland | Male smokers in ATBC study (Pietinen et al. 1999) | Males | 50-69 | 27111 | Incidence (185) | 8 | Food frequency questionnaire | Positive, not statistically significant | $\ldots$ |
| Japan | Japanese general population survey (1965 census cohort) (Hirayama 1986, 1990) | Males and females | $\geq 40$ | 265118 | Mortality: colon (552), rectum (563) | 17 | Crude-not clear | Inverse (colon), positive (rectum), not statistically significant | $\ldots$ |
| Netherlands | The Netherlands Cohort study (Voorrips et al. 2000a) | Males and females | 55-69 | 120852 | Incidence (910) | 6.3 | 150-item food frequency questionnaire | Inverse, not statistically significant | Non-significant trends: inverse (vitamin C), positive (beta-carotene, retinol) |
| Sweden | Swedish Mammography study (Terry et al. 2001) | Females | NA | 61463 | Incidence (460) | 9.6 | Food frequency questionnaire | Inverse, statistically significant | $\ldots$ |
| USA | NHS (Fuchs et al. 1999) | Females | 34-59 | 88757 | Incidence (787) | 16 | Repeated food frequency questionnaire |  | Inverse, not statistically significant (fibre) |
| USA | HPFS (Giovannucci et al. 1994) | Males | 40-75 | 47949 | Incidence (205) | 6 | Repeated food frequency questionnaire | Inverse (fruit), positive (vegetables), not statistically significant | $\ldots$ |


| USA | NHS/HPFS (Michels et al. 2000, 2002) | Males Females | $\begin{aligned} & 40-75 \\ & 30-55 \end{aligned}$ | $\begin{aligned} & 88764 \\ & 47325 \end{aligned}$ | Incidence: colon (937), rectum (244) | 10-16 | Repeated food frequency questionnaire | Positive, not statistically significant | $\ldots$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| USA | The Adventist Health study (Phillips and Snowdon 1985) | Males and females | $\geq 16$ | 25493 | Mortality (182) | 21 | Food frequency questionnaire | Inverse (rectum), positive (colon), not statistically significant | $\ldots$ |
| USA | Leisure World cohort, California (Paganini-Hill et al. 1987) | Males and females | Elderly | 10473 | Incidence (110) | 6 | Food frequency questionnaire |  | No association (beta-carotene) |
| USA | Leisure World cohort, California (Shibata et al. 1992) | Males and females | Elderly | 11580 | Incidence (202) | 9 | Food frequency questionnaire | Inverse, not statistically significant (females) | $\ldots$ |
| USA | Cohort of Hawaiian men of Japanese ancestry, Honolulu Heart Program (Heilbrun et al. 1989) | Males | 49-68 | 8006 | Incidence: colon (102), rectum (60) | 18 | Food frequency questionnaire |  | Inverse, statistically significant (vitamin C) |
| USA | Iowa Women's Health study (Steinmetz et al. 1994) | Females | 55-69 | 34691 | Incidence (212) | 5 | 127-item food frequency questionnaire | Inverse, not statistically significant | $\ldots$ |
| USA | New York University women's health study (Kato et al. 1997) | Females | 34-65 | 14727 | Incidence (100) | 7.1 | Food frequency questionnaire | Positive, not statistically significant | $\ldots$ |
| USA | American Cancer Society Prevention study, USA and Puerto Rico (Thun et al. 1992) | Males and females | $\geq 30$ | 764343 | Mortality-colon (1 150) | 6 | 32-item food frequency questionnaire | Inverse, statistically significant | $\cdots$ |
| NA | able. |  |  |  |  |  |  |  |  |

Summary of case-control studies reporting a measure of association between intake of fruit and vegetables and colorectal cancer

| Country | Study population (reference) | Sex | Association with fruit or vegetables | Association with specific risk factors |
| :---: | :---: | :---: | :---: | :---: |
| Argentina | (Iscovich et al. 1992) | Males and females | Inverse (vegetables) | $\ldots$ |
| Australia | (Kune et al. 1987, 1992) | Males and females | Inverse (vegetables) |  |
| Australia | (Steinmetz and Potter 1993) | Males and females | Null |  |
| Belgium | (Tuyns et al. 1988) | Males and females | Inverse (fruit) |  |
| Canada | French Canadians (Ghadirian et al. 1997) | Males and females | Inverse (vegetables) | $\ldots$ |
| China | (Hu et al. 1991) | Males and females | Inverse (vegetables) | $\ldots$ |
| France | (Faivre et al. 1997) | Males and females | Null | $\ldots$ |
| France | (Macquart-Moulin et al. 1986) | Males and females | Inverse | $\ldots$ |
| Italy | (Bidoli et al. 1992) | Males and females | Null | Inverse (spinach) |
| Italy | (Franceschi et al. 1994) | Males and females | Inverse | Tomato |
| Italy | (Francecshi et al. 1997) | Males and females | Inverse | $\ldots$ |
| Italy | (Franceschi et al. 1998) | Males and females | Inverse |  |
| Italy | (La Vecchia et al. 1987) | Males and females | Inverse (vegetables) | $\ldots$ |
| Italy | (Negri et al. 1998) | Males and females | Inverse | $\ldots$ |
| Japan | (Kotake et al. 1995) | Males and females | Null | $\ldots$ |
| Japan | (Tajima and Tominaga 1985) | Males and females | Null | Inverse (spinach, onion, pumpkin) |
| Netherlands | (Kampman et al. 1995) | Males and females | Inverse (vegetables) | ... |
| Russian Federation | (Zaridze et al. 1993) | Males and females | Null | Positive (dried fruit) |


from Europe (3) and Japan (1). It should be noted that three of the study cohorts were common to more than one study.

The majority of studies measured cancer incidence as the outcome, with mortality being used as the study outcome in only four studies. Eight studies investigated both men and women, while three investigated men only and four looked at women only.

Although a number of the studies reported a protective effect of fruit and vegetable consumption on colorectal cancer risk, only three of the 15 showed a statistically significant inverse trend. This is summarized in Table 9.19. A further eight studies did show some inverse association between vegetable consumption and colorectal cancer risk for certain subgroup analyses of fruit, vegetables or colon or rectal cancer separately; three studies showed a statistically insignificant positive association between fruit and vegetable intake and colorectal cancer.

The case-control study populations came from a wider range of countries than those of the cohort studies. They included studies from Argentina, Australia, Belgium, Canada, China, France, Italy, Japan, the Russian Federation, Singapore, Spain, Switzerland, the United States and Uruguay. All studies collected data on both men and women. It should be noted that two of the population cohorts (Australia and New York) were each common to two studies.

Results from 21 out of the 33 populations showed an inverse association with fruit or vegetables, while three found an inverse association only with specific food types (tomato; or spinach, onion and pumpkin; or tomato, pepper and celery). This is summarized in Table 9.20.

It should be noted that although the majority of colorectal cancers are adenocarcinomas, some studies found differences depending on tumour location (proximal vs distal, colon vs rectum). This suggests that pooling of results from different anatomical sites and cell types in many of the cohort and case-control studies may have obscured a true relationship for subgroups.

## Confounding and interaction

All studies adjusted for the effect of sex and age. Other confounders taken into account in the statistical analyses included total energy, meat, fat and protein intakes, BMI, smoking, alcohol consumption, physical activity, family history, dietary supplements, education and area of residence. By analogy with the evidence of lung and stomach cancer, it is conceivable that an effect might only be anticipated in those with particular genetic states (e.g. fast acetylators) that also consume significant quantities of meat. Many studies have excluded cases with strong family history of colorectal cancer or with conditions such as familial polyposis coli, and inflammatory bowel diseases such as ulcerative colitis or Crohn disease from the final analyses.

## Summary

On their own, the prospective epidemiological studies have yet to produce conclusive evidence to support an association of colorectal cancer with fruit and vegetable intake. However, the hypothesis is still worthy of further consideration, as there appears to be an inverse relationship in many cohort studies and in two thirds of the case-control studies in a wide range of different populations. The relationships remained after adjustment for confounding.

The uncertainty in the studies of colorectal cancer is relatively unsurprising in view of the complex biological mechanisms involved in its etiology. The incidence can be expected to reflect a complex combination of genetic factors (including factors that affect the colonic mucosa and those that determine whether certain toxins are excreted in bile or urine), diet (both carcinogenic and protective factors) and hormonal status.

## OESOPHAGEAL CANCER

A detailed description of the studies included in the review of the literature follows.

## Previous reviews of the literature

Three recent reviews of the literature (Cheng and Day 1996; Klerk et al. 1998; World Cancer Research Fund and American Institute for Cancer Research 1997) concluded that there is convincing evidence that diets high in fruit and vegetables decrease the risk of oesophageal cancer.

WCRF/AICR reviewed 22 case-control studies (1997). Of these, 18 showed a statistically significant protective association with at least one category of fruit or vegetables. The protective association in the studies remained after controlling for smoking and alcohol consumption. The Dutch review (Klerk et al. 1998) concluded that high vs low consumption of fruit and vegetables (an average difference of $150 \mathrm{~g} /$ day ) is likely to reduce the risk of oesophageal cancer by $40-55 \%$ in men and women.

Finally, the meta-analysis by Norat et al. (2001) pooled the results of 10 studies assessing the effect of fruit intake and 11 studies assessing the effect of vegetable intake on oesophageal cancer (see more details of methods described in the section on stomach cancer). This estimated that an increase in food intake of $100 \mathrm{~g} /$ day is associated with a relative risk for oesophageal cancer of 0.79 ( $95 \%$ CI $0.65-0.95$ ) for fruit; and 0.92 ( $95 \%$ CI $0.85-1.01$ ) for vegetables.

## Current review of the literature

This systematic review identified four cohort and 28 case-control studies investigating the association of fruit and vegetable intake with oesophageal cancer risk (incidence and mortality). Most studies looked
at oesophageal cancer mortality or at the effect of fruit and vegetable consumption on incidence and mortality considered jointly. In this review, we have also considered these outcomes together, assuming that, since oesophageal cancer survival rates are very poor, incidence is closely correlated with mortality.

The four cohort studies considered are from China (2), Japan and Norway. Details of the studies are given in Table 9.21. Both Chinese studies were based on the same population in the Linxian province, which has one of the highest incidence rates of oesophageal cancer in the world. The study population is part of a Nutrition Intervention Trial.

Three of the studies showed a statistically non-significant inverse trend with fruit and vegetable intake. One Chinese study found an inverse relationship with fresh vegetable consumption (Yu et al. 1993), while the other found an inverse relationship between the intake of both fruit and vegetables and oesophageal cancer incidence (Guo et al. 1994). The most recent study of Norwegian men also found a non-significant inverse relationship between fruit and vegetable intake and oesophageal cancer incidence, but this was based on only 22 cancer events.

Case-control studies provide the majority of evidence for the association of fruit and vegetable consumption and oesophageal cancer risk. The 28 case-control studies identified are summarized in Table 9.22.

The studies came from a wide range of populations worldwide including China (4), Hong Kong Special Administrative Region of China (Hong Kong SAR) (2), India (2), Japan (2) the United States (4), Europe (11) and South America (3). Most of these studies investigated populations of men and women, but five investigated only men and two studied only women. Not all of the study populations were from populations at a high risk of oesophageal cancer.

Of the 28 studies, 20 reported significant inverse associations between oesophageal cancer and fruit and/or vegetable consumption. Interestingly, more of these inverse associations were with fruit rather than vegetables.

Although many studies did not specifically look at the histological type of oesophageal cancer, the majority of the results presented here are likely to be limited to squamous cell cancer. There was one study, in women from the United Kingdom (Cheng et al. 2000), which focused on oesophageal adenocarcinoma. This study found an inverse association of cancer risk with a higher consumption of fruit but not vegetables.

## Confounding

Most studies adjusted for age and sex. Other potential confounders considered in several studies, included smoking, total energy intake, pickled vegetable intake, BMI, marital status, occupational status, educational or socioeconomic status and ethnicity. Supplement intakes, such as vitamins, and intake of hot teas, were also considered in a few studies.
Table 9.2 I Summary of cohort studies reporting a measure of association between intake of fruit and vegetables and

| Country | Study population (reference) | Sex | Age range (years) | Study size | No. of cases | Follow-up (years) | Exposure measure | Association with fruit or vegetable intake |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| China | Linxian Nutrition Intervention Trial cohort (Guo et al. 1994) | Males and females | 40-69 | 29584 | Incidence (640) | 5.25 | Structured questionnaire interview | Inverse, not statistically significant |
| China | Linxian Nutrition Intervention Trial cohort (Yu et al. 1993) | Males and females | 40-69 | 12693 | Incidence and mortality (I 162) | 15 | Structured questionnaire interview | Inverse, not statistically significant |
| Japan | Japanese general population survey (1965 census cohort) (Hirayama 1986, 1990) | Males and females | $\geq 40$ | 265118 | Mortality (585) | 17 | Crude-not clear | Positive, not statistically significant |
| Norway | Cohort of Norwegian men (Kjaerheim et al. 1998) | Males | 35-74 | 10960 | Incidence (22) | 25 | 32-item food frequency questionnaire | Inverse, not statistically significant |

Table 9.22 Summary of case-control studies reporting a measure of association between intake of fruit and vegetables and

| Country | Study population/area (reference) | Sex | Association with fruit and vegetables group (or proxy) | Association with specific risk factors only |
| :---: | :---: | :---: | :---: | :---: |
| Brazil | (Victora et al. 1987) | Males and females | Inverse (fruit) | - |
| China | Heilongjiang (Hu et al. 1994) | Males and females | Null | - |
| China | Linxian (Li et al. 1989) | Males and females | Null | - |
| China | Shanghai (Gao et al. 1994) | Males and females | Inverse (fruit), null (vegetables) | - |
| China | Shanxi (Wang et al. 1992) | Males and females | Null | Inverse (cabbage) |
| China, |  |  |  |  |
| Hong Kong SAR | (Cheng et al. 1992) | Males and females | Inverse | - |
| China, |  |  |  |  |
| Hong Kong SAR | Never-smokers/drinkers (Cheng et al. 1995) | Males and females | Inverse | - |
| France | (Tuyns et al. 1987) | Males and females | Inverse | - |
| France | Multicentre (Launoy et al. 1998) | Males | Inverse | - |
| Greece | Athens (Garidou et al. 1996) | Males and females | Null | Inverse (vegetables) in adenocarcinoma type |
| India | (Nayar et al. 2000) | Males and females | Null | - |
| India | (Notani and Jayant 1987) | Males and females | Inverse (vegetables), null (fruit) | - |
| Italy | (Decarli et al. 1987) | Males and females | Inverse (fruit), null (vegetables) | - |


| Italy | (Tavani et al. 1996a) | Males and females | Null | Inverse (fruit) in smokers |
| :---: | :---: | :---: | :---: | :---: |
| Italy | Northern Italy (Bosetti et al. 2000) | Males and females | Inverse | Null (cooked vegetables) |
| Italy | Milan (Negri et al. 1991) | Males and females | Inverse | - |
| Italy | Milan (Tavani et al. 1993) | Females | Inverse (fresh fruit), null (green vegetables) | - |
| Italy | Non-smokers (Tavani et al. 1994) | Males and females | Inverse | - |
| Japan | (Hanaoka et al. 1994) | Males | Null | - |
| Japan | (Takezaki et al. 2000) | Males | Inverse | - |
| Switzerland | (Levi et al. 2000) | Males and females | Inverse | - |
| United Kingdom | England; Scotland (Cheng et al. 2000) | Females | Null (vegetables), inverse (fruit) | Adenocarcinoma only |
| Uruguay | (DeStefani et al. 1990) | Males and females | Null | - |
| USA | (Brown et al. 1995) | Males and females | Inverse | - |
| USA | (Brown et al. 1998) | Males | Inverse | Squamous cell carcinoma only |
| USA | California (Yu et al. 1988) | Males and females | Inverse | - |
| USA | South Carolina (Brown et al. 1988) | Males | Inverse (fruit), null (vegetables) | - |
| Multiple countries | South America (5 countries) (Castellsague et al. 2000) | Males and females | Inverse | - |
| Null Association not statistically significant. |  |  |  |  |
| Inverse Statistically significant protective association. |  |  |  |  |
| - No data. |  |  |  |  |

## Summary

The epidemiology of oesophageal cancer shows wide geographical variation in incidence and mortality rates. Striking differences have been reported not only between regions of the world and countries, but also within smaller geographical areas (Muir and McKinney 1992; Walker et al. 2002). Oesophageal cancer risk also varies with race, ethnicity and sex. Changes in incidence patterns observed in migrant studies appear to indicate that environmental factors, among which diet may be important, play an important role in the etiology of oesophageal cancer.

The main risk factors for oesophageal cancer in Europe were previously thought to be tobacco smoking and alcohol consumption. However, time trends of oesophageal mortality from 1950-1985 in 17 European countries showed that oesophageal cancer had either decreased or increased only slightly (Cheng et al. 1992). This trend differed from that of lung cancer, cirrhosis and alcohol consumption, which had increased substantially during the same period. The results suggest that other populationwide changes in protective risk factors, such as improvements in diet after World War II, had mitigated against the effect of tobacco and alcohol and resulted in a reduction of oesophageal cancer risk. In China, which has a very high incidence of oesophageal cancer, an ecological study of 65 counties showed that oesophageal cancer mortality was significantly associated with fruit consumption, but no correlation was observed with tobacco smoking or alcohol consumption (Guo et al. 1990). There is strong evidence from ecological studies that dietary factors, especially fruit and vegetable consumption, may affect rates of oesophageal cancer.

This review of the literature showed that there are few prospective observational studies of the effect of fruit and vegetable consumption on oesophageal cancer, and that most of these show non-significant inverse associations. Evidence from supplement intervention trials is also inconclusive. Two randomized supplement intervention trials in Linxian, China tested the effect of nutrient/vitamin supplementation in a population with very high rates of oesophageal cancer (Blot et al. 1995). Modest protective effects were seen for mortality in the supplemented group in both trials, but none of the results was statistically significant for oesophageal or gastric cancer. In spite of this, evidence from case-control studies appears to support an inverse relationship between fruit and vegetable consumption and oesophageal cancer risk (incidence and mortality), although this is not as strong as that for stomach cancer.

### 3.8 Estimates of RISK Factor-disease relationships by SUbregion, SEX AND AGE

## Estimates of relative risks

This section describes, for each selected outcome, the studies that were chosen based on the strict selection criteria for meta-analysis outlined in section 3.6, and the final estimates of relative risks derived for CRA.

## Ischaemic heart disease

Details of the cohort studies that most closely met our selection criteria are given in Table 9.23. The EPIC-Norfolk study (Khaw et al. 2001) was included in spite of the fact that it presents results in relation to plasma vitamin C, because plasma vitamin C measurements (available for the whole cohort) were relatively well correlated with fruit and vegetable intake (available from a 7-day food record analysed for a subset of the cohort) and because of the high quality of the methods used to collect and analyse data.

Fruit and vegetable intake in the Nurses' Health and Health Professionals' follow-up studies (NHS/HPFS), EPIC-Norfolk and the Finnish Mobile Clinic studies were treated as continuous variables. The NHS/HPFS studies gave the relative risk in terms of one additional serving per day, which we converted as $80 \mathrm{~g} /$ day (assumed to be one standard serving); the EPIC-Norfolk and Finnish Mobile Clinic studies reported relative risk estimates for 50 g and 1 g increase in intake respectively. Thus, the relative risk estimates were transformed to give final estimates expressed as per $80 \mathrm{~g} /$ day increase.

For the Massachusetts Health Care Panel study (data analysed as quartiles of intakes), the two methods described in section 3.6 were used to estimate the relative risks. With method 1, we estimated the additional $\mathrm{g} / \mathrm{day}$ for which the relative risks given applied. The study gave exposure quartiles with midpoints at $32,88,138$ and $190 \mathrm{~g} /$ day (the last estimated as the lower limit, 164, plus half the previous category interval), and reported relative risks for the two latter categories compared with the first quartile. A difference of 55 g between adjacent quartiles was

Table 9.23 Relative risk estimates for the association between ischaemic heart disease and fruit and vegetable consumption

| Country | Study population (reference) | Sex and age range (years) | Outcome | $R R$ (95\% CI) per $80 \mathrm{~g} /$ day increase in fruit and vegetable intake |
| :---: | :---: | :---: | :---: | :---: |
| Finland | Finnish Mobile Clinic Health Examination study (Knekt et al. 1994) | Males and females (30-69) | Mortality | $\begin{gathered} 0.964 \\ (0.930-0.999) \end{gathered}$ |
| United Kingdom | EPIC-Norfolk (Khaw et al. 2001) | Males and females (45-79) | Mortality | $\begin{gathered} 0.54 \\ (0.40-0.74) \end{gathered}$ |
| USA | NHS/HPFS (Joshipura et al. 200I) | Males and females (34-75) | Myocardial infarction incidence | $\begin{gathered} 0.96 \\ (0.94-0.99) \end{gathered}$ |
| USA | Massachusetts Health Care Panel study (Gaziano et al. 1995) | Males and females (>66) | Mortality | $\begin{gathered} 0.54 \\ (0.35-0.84) \end{gathered}$ |

assumed, which gives exposures of approximately 30, 85, 140 and $195 \mathrm{~g} /$ day. This approximation yielded comparable relative risks per 55 g difference whichever of the two reported relative risks was converted, and the relative risk reported for the 4th vs 1 st quartile was used, which was expressed as a difference in exposure of $3 \times 55=165 \mathrm{~g} / \mathrm{day}$. The method of Greenland and Longnecker (method 2) was then used to estimate the weighted regression slope over the published relative risks, allowing for correlations due to common reference category. Because of its advantages, method 2 was chosen to obtain the final estimates for inclusion in the meta-analysis.

Meta-analysis was used to pool relative risk estimates using the method described in section 3.6. The test of heterogeneity gave a chisquared value of $19.044(d f=3 ; P<0.001)$. The resulting variation between studies suggests that it is inappropriate to pool estimates according to the fixed effects method. Using random effects meta-analysis, the pooled relative risk estimate was 0.903 ( $95 \%$ CI $0.824-0.989$ ) for an $80 \mathrm{~g} /$ day increase in fruit and vegetable consumption. The random effects results are shown in Figure 9.2.

Figure 9.2 Random effects meta-analysis of the association of fruit and vegetable intake with ischaemic heart disease


[^38]The figure shows that there is marked heterogeneity in the best available evidence on the effects of fruit and vegetable consumption on the risk of IHD. The sources of this heterogeneity are currently not understood scientifically, and there is, therefore, no fully satisfactory means for arriving at a summary effect estimate. The random effects model used here provides a pragmatic interim solution to summarizing this evidence, pending better scientific understanding of the underlying relationships. The derived effect size seems plausible in the light of the consistency of the study findings set out in Table 9.13, but it remains subject to substantial uncertainty-only some of that derives from the statistical uncertainty associated with the four studies included in Figure 9.2.

## Ischaemic stroke

Only two cohort studies met the selection criteria for meta-analysis. These are summarized in Table 9.24.

The NHS/HPFS gave the relative risk in terms of one additional serving per day, which we converted as $80 \mathrm{~g} / \mathrm{day}$. As the Zutphen study did not give a single estimate of linear (log) relative risk per consumption, the relative risk estimates were derived using the two methods described earlier for IHD. Since vegetable and fruit consumption were separately analysed, the relative risk for vegetable consumption was used: it seems to us likely that vegetable and fruit consumption would confound each other, so that what is reported as a purely "vegetable" effect includes the effect of fruit eaten by vegetable consumers. Since a diet of only fruit is rare, the vegetable effect reported was assumed to be similar to that which would have been reported for a combined fruit and vegetable diet (although this may underestimate the effect). Method 2, of Greenland and Longnecker (1992), gave an almost identical estimate to Method 1, which used a consumption difference of the 4th vs 1st quartile ( $249.9-128.1=121 \mathrm{~g} / \mathrm{day}$ ) for the 0.82 relative risk (vegetables) as the basis for the estimate. Results based on method 2 are therefore reported here.

Table 9.24 Relative risk estimates for the association between stroke and fruit and vegetable consumption

|  | Study population <br> (reference) | Sex and <br> age range <br> (years) | Outcome | RR (95\% CI) per <br> $80 \mathrm{~g} /$ day increase <br> in fruit and <br> vegetable intake |
| :--- | :--- | :--- | :--- | :---: |
| Country | Zutphen study (Keli <br> et al. 1996) | Males <br> $(50-69)$ | Incidence <br> cerebrovascular <br> accident | 0.87 <br> $(0.49-1.53)$ |
| Netherlands | NHS/HPFS (Joshipura <br> et al. 200I) | Males and <br> females <br> $(34-75)$ | Incidence <br> ischaemic <br> stroke | 0.96 |

Figure 9.3 Fixed effects meta-analysis of the association of fruit and vegetable intake with ischaemic stroke


Key: I, Zupthen study; 2, NHS/HPFS.

Relative risk estimates were combined using fixed-effect metaanalysis, as the random effects could not be estimated. The pooled relative risk estimate was 0.939 ( $95 \%$ CI $0.892-0.989$ ) for an $80 \mathrm{~g} /$ day increase in fruit and vegetable consumption. The results are shown in Figure 9.3.

## Lung cancer

Four studies met our selection criteria for meta-analysis. These are shown in Table 9.25.

Relative risks in the NHS/HPFS and the Finnish Mobile Clinic Health Examination study were available with fruit and vegetable intake treated as a continuous variable. The NHS/HPFS gave the relative risk in terms of one additional serving per day, which we converted as $80 \mathrm{~g} / \mathrm{day}$. The authors of the Finnish Mobile Clinic study provided relative risks per $1 \mathrm{~g} /$ day which were transformed on a log scale to give estimates expressed as an $80 \mathrm{~g} /$ day increase (as discussed earlier in this section).

For the Netherlands Cohort study and the National Health Interview study, data had been analysed as quintiles and quartiles of intake respectively. The method of Greenland and Longnecker (1992) was used to

Table 9.25 Relative risk estimates for the association between lung cancer and fruit and vegetable consumption

|  |  |  | RR (95\% CI) per <br> 80 g/day increase <br> in fruit and |
| :--- | :--- | :--- | :--- | :--- |
| Country | Study population (reference) |  |  |$\quad$| Sex and |
| :--- |
| age range (years) |$\quad$ Outcome | vegetable intake |
| :--- | :--- |

estimate the weighted regression slope over the published relative risks. This gave estimates of relative risks as a continuous variable.

As there was no evidence of heterogeneity (chi-squared value of 3.553, $d f=3, P=0.314$ ), the results from fixed effect meta-analysis were used (estimates were very similar to those obtained with the random effects methods). The pooled relative risk estimate was 0.926 ( $0.886-0.968$ ) for an $80 \mathrm{~g} /$ day increase in fruit and vegetable consumption. The forest plot of the results is shown in Figure 9.4.

These results are consistent with those found in the only other pooled analysis for an association between lung cancer and fruit and vegetable consumption (S.A. Smith-Warner et al. personal communication, 2002). In this study, they found strong, inverse associations in the age-adjusted analyses, and analyses adjusted for education, BMI, alcohol intake and energy intake for total fruit, total vegetable, and total fruit and vegetable intakes. Associations were similar among never, past, and current smokers.

Smith-Warner et al. estimated that the pooled results for total fruit and vegetable intake gave a slightly lower reduction in lung cancer risk than the meta-analysis results for this project. Due to the potential for residual confounding, in our analysis we have decided to use the more conservative estimate of pooled relative risk which is equivalent to 0.96 (0.93-0.99) per $80 \mathrm{~g} /$ day increase in fruit and vegetable intake (S.A. Smith-Warner et al. personal communication, 2002). However, we should note that because of the limitations in our ability to accurately specify fruit and vegetable exposure, we might be underestimating any effect.

## Gastric cancer

Only one study, the Netherlands Cohort study on diet and cancer, met our selection criteria for meta-analysis. This is shown in Table 9.26.

Figure 9.4 Fixed effects meta-analysis of the association of fruit and vegetable intake with lung cancer


Key: I, The Netherlands Cohort study; 2, National Health Interview study; 3, NHS/HPFS; 4, Finnish Mobile Clinic Health Examination study.

Table 9.26 Relative risk estimate for the association between gastric cancer and fruit and vegetable consumption

|  | Sex and <br> age range (years) | Outcome | RR (95\% CI) per <br> $80 \mathrm{~g} /$ day increase in fruit <br> and vegetable intake |
| :--- | :--- | :--- | :---: |
| Study population (reference) | Males and females <br> $(55-69)$ | Incidence | 0.94 (0.86-I.03) |
| The Netherlands Cohort <br> study (Botterweck et al. 1998) |  |  |  |

The relative risks obtained from the Netherlands Cohort study are consistent with the effect estimates obtained from most case-control studies reviewed, but they are more conservative than the estimates of the meta-analysis from Norat et al. (2001).

Data from the Netherlands Cohort study had been analysed as quintiles of intake of fruit and vegetables. The method of Greenland and Longnecker (1992) was used to estimate the weighted regression slope over the published relative risks to estimate relative risks as a continuous variable for an $80 \mathrm{~g} /$ day increase in total fruit and vegetable consumption. This gave a final relative risk estimate for gastric cancer
incidence and mortality for the CRA project of 0.94 (0.86-1.03) for an $80 \mathrm{~g} /$ day increase in fruit and vegetable consumption.

## Colorectal cancer

Three studies met our selection criteria for meta-analysis. These are shown in Table 9.27.

Relative risk in the NHS/HPFS was available with fruit and vegetable intake treated as a continuous variable. The NHS/HPFS gave the relative risk in terms of one additional serving per day, which was converted as $80 \mathrm{~g} / \mathrm{day}$.

For the Netherlands Cohort study and the Swedish Mammography study, data had been analysed as quintiles and quartiles of intake respectively. The method of Greenland and Longnecker (1992) was used to estimate the weighted regression slope over the published relative risks to give relative risks as a continuous variable.

The NHS/HPFS and the Netherlands Cohort study had presented results separately for colon and rectal cancers. The Netherlands Cohort study had also analysed data for men and women separately. These subgroup analyses were combined in the meta-analysis as if they were separate studies. The test for heterogeneity gave a chi-squared value of 8.849 $(d f=6 ; P=0.182)$. As there was no evidence of heterogeneity, the fixed effects results were used (these were very similar to the results obtained with the random effects method). The pooled relative risk estimate was 0.997 (0.973-1.021) for an $80 \mathrm{~g} /$ day increase in fruit and vegetable consumption. The results are shown in Figure 9.5.

The Swedish Mammography study showed that the inverse association was stronger, and the dose-response more evident, among individuals who consumed the lowest amounts of fruit and vegetables. In a subgroup analysis of the population in the lowest quartile of fruit and vegetable intake, the relative risk was $0.77(0.67-0.97)$ for an increase

Table 9.27 Relative risk estimates for the association between colorectal cancer and fruit and vegetable consumption

|  | Study population <br> (reference) | Sex and <br> age range <br> (years) |  | $R R(95 \%$ Cl) per $80 \mathrm{~g} / \mathrm{day}$ <br> increase in fruit and vegetable <br> intake |
| :--- | :--- | :--- | :--- | :--- |
| Country |  |  |  |  |

Figure 9.5 Fixed effects meta-analysis of the association of fruit and vegetable intake with colorectal cancer


Key: I, The Netherlands Cohort study (males, colon cancer); 2, The Netherlands Cohort study (women, colon cancer); 3, The Netherlands Cohort study (males, rectal cancer); 4, The Netherlands Cohort study (females, rectal cancer); 5, Swedish Mammography study (colorectal cancer); 6, NHS/HPFS (males and females, colon cancer); 7, NHS/HPFS (males and females, rectal cancer).
of one serving per day (Terry et al. 2001). It may be that individuals who consume very low amounts of fruit and vegetables have the greatest risk of colorectal cancer. This is consistent with the presence of a plateau effect and highlights the need to study populations with a wide range of dietary exposures. This finding was not replicated in the NHS/HPFS (Michels et al. 2000), although it should be noted that this study had a high reference exposure category and might not have been able to examine very low intake of fruit and vegetables as in the Swedish study.

## Oesophageal cancer

None of the cohort studies of oesophageal cancer in the systematic review met our selection criteria. Studies were either of high-risk populations, had a small number of events, had a relatively short follow-up time or did not cover total fruit and vegetable intake. Very few of the case-control studies had calculated risk for total fruit and vegetable consumption, and none had estimated relative risks for quantified levels of fruit and vegetable consumption that would allow estimation of a continuous variable. Thus, it was decided to use the results of the meta-
analysis from Norat et al. (2001) to derive the final RR estimates. This meta-analysis calculated risk estimates for fruit and vegetables separately per $100 \mathrm{~g} /$ day intake. The pooled relative risks associated with an increase of consumption of $80 \mathrm{~g} /$ day were calculated. These were 0.828 (0.708-0.960) for fruit, and $0.935(0.878-1.008)$ for vegetables. The results suggested risk reductions of between $17.2 \%$ and $6.5 \%$ per $80 \mathrm{~g} /$ day increase in fruit intake and vegetable intake, respectively. For this study, the most conservative relative risks were used, that is, those for vegetables only, 0.935 ( $95 \%$ CI $0.878-1.008$ ).

## Summary of the estimates of relative risks

The relative risk estimates and $95 \%$ confidence intervals are summarized in Table 9.28. Estimates are expressed as the change in relative risk associated with an $80-\mathrm{g}$ increase in daily fruit and vegetable intake.

As discussed earlier, the relative risks are applied to all subregions and to both men and women. Assuming age attenuation at the extremes of age, these relative risks apply for individuals aged 15-69 years. For older adults, the relative risks were reduced by a quarter for ages 70-79 years, and by half for the age group $\geq 80$ years. A relative risk of 1 was applied for those aged $<15$ years.

## Estimates of risk reversibility

There is very little published evidence of risk reversibility for an increased fruit and vegetable intake, mainly due to the difficulty of conducting dietary intervention trials. Hence, the period of the latency between increased fruit and vegetable intake and risk reduction are difficult to predict.

## Cardiovascular diseases

There is some limited evidence from randomized-controlled trials, which suggests that the effect of fruit and vegetables is not immediate for IHD. As was described in section 3.1, the DART study and the Lyon Diet Heart study examined the effect of dietary changes on the secondary prevention of myocardial infarction. The DART study showed no effect on mortality after two years, in the fat advice arm of the trial (advice to reduce fat intake and increase polyunsaturated to saturated fat ratio vs no advice) which was associated with an increased fruit and vegetable intake of about $50 \mathrm{~g} /$ day. However, the Lyon Diet Heart study showed that a Mediterranean-style diet had a marked protective effect on the rate of recurrence of myocardial infarction after 27 months, with an effect still maintained after four years of follow-up. The main disadvantage of these trials was that the interventions encompassed several dietary changes. As a result, it is not possible to estimate the specific influence of increasing fruit and vegetables intake.

Evidence for the specific benefits of an increase in fruit and vegetable intake on cardiovascular health comes from the IEIS and the DASH trial.
Table 9.28 Summary of relative risks ( $95 \% \mathrm{Cl}$ ), for selected diseases with increased fruit and vegetable consumption by

| Outcome | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| Ischaemic heart disease | 1.00 | 1.00 | $\begin{gathered} 0.90 \\ (0.82-0.99) \end{gathered}$ | $\begin{gathered} 0.90 \\ (0.82-0.99) \end{gathered}$ | $\begin{gathered} 0.90 \\ (0.82-0.99) \end{gathered}$ | $\begin{gathered} 0.90 \\ (0.82-0.99) \end{gathered}$ | $\begin{gathered} 0.93 \\ (0.85-1.01) \end{gathered}$ | $\begin{gathered} 0.95 \\ (0.87-1.03) \end{gathered}$ |
| Ischaemic stroke | 1.00 | 1.00 | $\begin{gathered} 0.94 \\ (0.89-0.99) \end{gathered}$ | $\begin{gathered} 0.94 \\ (0.89-0.99) \end{gathered}$ | $\begin{gathered} 0.94 \\ (0.89-0.99) \end{gathered}$ | $\begin{gathered} 0.94 \\ (0.89-0.99) \end{gathered}$ | $\begin{gathered} 0.95 \\ (0.91-1.00) \end{gathered}$ | $\begin{gathered} 0.97 \\ (0.92-1.02) \end{gathered}$ |
| Lung cancer | 1.00 | 1.00 | $\begin{gathered} 0.96 \\ (0.93-0.99) \end{gathered}$ | $\begin{gathered} 0.96 \\ (0.93-0.99) \end{gathered}$ | $\begin{gathered} 0.96 \\ (0.93-0.99) \end{gathered}$ | $\begin{gathered} 0.96 \\ (0.93-0.99) \end{gathered}$ | $\begin{gathered} 0.97 \\ (0.91-1.02) \end{gathered}$ | $\begin{gathered} 0.98 \\ (0.92-1.03) \end{gathered}$ |
| Gastric cancer | 1.00 | 1.00 | $\begin{gathered} 0.94 \\ (0.86-1.03) \end{gathered}$ | $\begin{gathered} 0.94 \\ (0.86-1.03) \end{gathered}$ | $\begin{gathered} 0.94 \\ (0.86-1.03) \end{gathered}$ | $\begin{gathered} 0.94 \\ (0.86-1.03) \end{gathered}$ | $\begin{gathered} 0.95 \\ (0.87-1.04) \end{gathered}$ | $\begin{gathered} 0.97 \\ (0.89-1.06) \end{gathered}$ |
| Colorectal cancer | 1.00 | 1.00 | $\begin{gathered} 0.99 \\ (0.97-1.02) \end{gathered}$ | $\begin{gathered} 0.99 \\ (0.97-1.02) \end{gathered}$ | $\begin{gathered} 0.99 \\ (0.97-1.02) \end{gathered}$ | $\begin{gathered} 0.99 \\ (0.97-1.02) \end{gathered}$ | $\begin{gathered} 0.99 \\ (0.97-1.02) \end{gathered}$ | $\begin{gathered} 1.00 \\ (0.97-1.02) \end{gathered}$ |
| Oesophageal cancer | 1.00 | 1.00 | $\begin{gathered} 0.94 \\ (0.88-1.01) \end{gathered}$ | $\begin{gathered} 0.94 \\ (0.88-1.01) \end{gathered}$ | $\begin{gathered} 0.94 \\ (0.88-1.01) \end{gathered}$ | $\begin{gathered} 0.94 \\ (0.88-1.01) \end{gathered}$ | $\begin{gathered} 0.95 \\ (0.89-1.02) \end{gathered}$ | $\begin{gathered} 0.97 \\ (0.91-1.04) \end{gathered}$ |
| Note: Unit of change in risk is change per $80 \mathrm{~g} /$ day increase in fruit and vegetable intake. |  |  |  |  |  |  |  |  |

As was described in section 3.1, the IEIS showed that after only one year, cardiac events and mortality were reduced by $40 \%$ and $45 \%$, respectively, in a group of men who had previously suffered an acute myocardial infarction and who followed a low-fat fruit and vegetable-rich diet compared with men on a standard low-fat diet (Singh et al. 1993a). The DASH trial showed that hypertensive individuals achieved a reduction in their blood pressure after eight weeks of dietary changes (Conlin et al. 2000; Obarzanek et al. 2001). Both a diet rich in fruit and vegetables and a combination diet rich in fruit and vegetables and low in saturated fat, fat and cholesterol reduced systolic and diastolic blood pressure. After eight weeks, $70 \%$ of the participants on a combination diet rich in fruit and vegetables and low in saturated fat, fat and cholesterol, had a normal blood pressure, with $45 \%$ of those on a diet rich in fruit and vegetables, and $23 \%$ of those on a control diet. The true long-term effect of such dietary changes needs to be assessed using studies of longer duration.

Further evidence for risk reversibility of cardiovascular disease comes from time series analyses in countries of central and eastern Europe. One natural experiment in terms of rapid dietary change occurred with the reunification of Germany in 1989. After reunification, fruit and vegetable availability increased markedly, doubling between 1989 and 1992 (J.W. Powles, personal communication, 2001). But although the previous decline in vascular disease mortality accelerated, the change of trend was not noticeably greater than that in Poland and the Czech and Slovak republics where the increase in fruit and vegetable consumption was almost certainly less marked. However, the results of an ecological study performed in Poland suggested that the observed increase in supply of fresh fruit and vegetables and changes in the type of dietary fat between 1986-1990 and 1994 may have had a short-term positive effect on mortality risk (Zatonski et al. 1998). This study investigated the reasons for the decline in deaths attributed to IHD in Poland since 1991 after two decades of rising rates. For most of the potentially explanatory variables studied (i.e. alcohol consumption, cigarette smoking, socioeconomic indices and medical services) there were no corresponding changes in trend. The authors concluded that dietary changes could provide the best explanation to the reduced IHD mortality rates, with an effect noticeable within four years. While these findings cannot be dismissed, due to the limitations of ecological studies, these exploratory analyses would need to be confirmed with stronger study designs.

## Cancer

The evidence for cancer outcomes is more limited and comes from one major source. The NHS/HPFS studies have both collected dietary intake data repeatedly during a long follow-up period. This enables the examination of the temporal aspects of diet-cancer associations, although this has not yet been published for all the cancer outcomes covered in this
report. For lung cancer risk, the findings support a relatively long latency period for carotenoids (used as a proxy for fruit and vegetable intake), consistent with the known natural history of the disease, and suggest that the critical period for exposure to carotenoids to maximize protection may be between four to eight years before diagnosis (Michaud et al. 2000). This may explain why some observational studies and short supplement intervention trials may not have detected a significant association.

## Proposed risk reversibility for the CRA project

Assuming that the future population intake of fruit and vegetables is increased to the proposed counterfactual, we propose the following estimates of risk reduction:

- ten per cent reduction in relative risk in four years for IHD and stroke; and
- ten per cent reduction in relative risk in eight years for cancer outcomes.

Due to the current limitations of the evidence, it is assumed that risk reversibility following an increase in fruit and vegetable consumption is age independent and that it does not vary among subregions. As this may not be the case, the estimates should be seen as first approximations. Due to the high level of uncertainty, it is proposed that a $\pm 10 \%$ range of error is included in the calculations of avoidable burden for the CRA project.

### 3.9 Quantitative and qualitative sources of uncertainty in MEASURING THE DIET-DISEASE RELATIONSHIP

There are several sources of uncertainty in risk assessment. Some of these will be described in this section with regard to the study of the relationship between fruit and vegetable intake and health outcomes. It is likely that these underestimate the true level of uncertainty that affected the estimates presented in this report.

Little is known of the relative importance of early and late dietary exposure of fruit and vegetables for the risk of the chronic diseases examined in this report. Results from the Boyd Orr cohort suggest that fruit and vegetable intake in childhood could have a protective effect on the risk of cancer in adulthood (Maynard et al. 2003). However, we do not know whether an increased intake of fruit and vegetables is more important in young or middle age for greatest risk reversibility. Hence, it is not possible to comment on how the time trends of risk reversibility may vary among age groups.

In this project, it was assumed that there was no difference in relative risk across populations. However, it is not possible to verify whether this assumption is true as the study populations covered by the literature review were from limited geographical areas and ethnic groups. The
majority of cohort studies were from China, Japan, the United States and Europe, although there was at least one case-control study from the following countries: Australia, Brazil, China (Hong Kong SAR), India, Mexico, the Republic of Korea, Singapore, Turkey, Uruguay and Venezuela.

Relative risks in the developed world may not provide good estimates of the effect in the less developed world. For example, the impact of increasing fruit and vegetable intake may be stronger in undernourished, micronutrient-poor (e.g. vitamin A deficiency) populations in developing countries. However, this has to be considered in the wider context of nutritional requirements (energy and protein may be more important for survival in some populations) and relative causes of deaths (noncommunicable diseases may not be as important killers as in developed countries).

One possibility refers to the ascertainment of exposure. There remains considerable uncertainty as to which constituents, or combination of constituents, of fruit and vegetables would confer a protective effect. Several components have been suggested, but it is unlikely that any single compound will capture the benefits, which seem to arise from a varied diet rich in fruit and vegetables. The sources of uncertainty described for exposure measurement in section 2.9 also need to be taken into account here. Errors in measurement of dietary intake mean that the strength of true associations, if they exist, is likely to be reduced. In addition, the definition of the exposure as "fruit and vegetables" introduces a substantial amount of non-specificity with the resultant potential attenuation of underlying causal relations. That is, the currently unknown active constituents contained within the category "fruit and vegetables" are likely to have a varying quantitative relationship to the less specific category of fruit and vegetables. This is especially important across food cultures with substantial differences in the make-up of foods that constitute the fruit and vegetable group, but also applies even within cultures because of differences in food composition. At a minimum, this suggests that findings of protective associations between fruit and vegetable consumption and a specific disease within one food culture may not generalize well to another culture where the fruit and vegetables in the diet are constituted by different foods-in which the causally active constituents may be present to substantially differing amounts.

Another major source of uncertainty refers to the fact that confounding is a real possibility in observational studies of the relationship between fruit and vegetable intake and health outcomes. Although the studies considered in this project have attempted to adjust for confounding, measurement error in the assessment of potential confounders and the possibility of residual confounding need to be considered.

## 4. Discussion of estimates of attributable burden

The results of the analysis showed that a lack of dietary fruit and vegetables contributes an important share of the worldwide disease burden. It was estimated that increasing individual fruit and vegetable consumption up to the theoretical-minimum-risk distribution could reduce the worldwide burden of disease for IHD and ischaemic stroke by about $31 \% ~(30 \%$ for men and $31 \%$ for women) and $19 \% ~(18 \%$ for men and $19 \%$ for women), respectively. For cancers of the stomach and oesophagus, the potential reduction in disease attributable to an increase in fruit and vegetable intake was $19 \%$ and $20 \%$, respectively. Attributable risk fractions were lower for lung and colorectal cancers ( $12 \%$ and $2 \%$ ).

There were relatively few sex and age differences in the attributable fractions of disease attributable to low fruit and vegetable consumption, mainly because relative risk estimates were assumed to be constant in all adult men and women. However, larger variations were observed in the estimates of attributable DALYs due to low fruit and vegetable intake, as this is a function of both the attributable fractions and the amount of burden of disease accounted for by each health outcome studied in the various age-sex groups, itself affected by other risk factors. The number of DALYs was generally higher in men (except for stroke). As expected for chronic diseases, it tended to be lower in young adults.

There were some subregional variations in the attributable risk fractions and number of DALYs attributable to low fruit and vegetable intake. These tended to reflect differences in exposure levels and disease patterns among subregions.

The total worldwide mortality attributable to inadequate fruit and vegetable consumption was estimated to be 2.726 million deaths or 26.662 million DALYs in 2000. These results need to be interpreted in the light of the current limitations of the methods used, as described earlier in the chapter (e.g. residual confounding, misclassification of exposure). However, they highlight that increasing fruit and vegetable consumption could play an important role in improving public health worldwide.

## 5. Methods for projection of FUTURE EXPOSURE

An aggregate approach based on subregional trends in fruit and vegetable availability was used to project future exposure levels. The methods used are described below.

### 5.1 Trends in fruit and vegetable availability over time

The projection of future fruit and vegetable consumption for the years 2000 to 2030 was based primarily on the observation of subregional
trends in fruit and vegetable availability. The FAO food balance sheet data (FAO 2001) show that there has been an overall worldwide increase in the availability of fruit and vegetables during the past decades. However, large variations exist, among subregions and countries, in the observed trends and amounts available. These variations are described below.

FAO food balance sheet statistics (1961-1999) for each country were used to calculate three-year average subregional fruit and vegetable availability. These calculations show that there has been a relatively steady increasing trend in availability in AMR-A, AMR-B, EMR-D, EUR-A, SEAR-B, SEAR-D and WPR-B between 1961 and 1999. In contrast, downward trends have been observed since the late 1970s or early 1980s in WPR-A and AFR-E. However, as the trend has been slowing down in WPR-A during the last decade, we also assumed that the downward trend would stop. Based on the observed trends in AFR-D, we also assumed that the downward trend would stop in AFR-E. Large fluctuations were observed in AFR-D between 1961 and 1981, followed by a relatively steady increase in fruit and vegetable availability. In AMR-D, the trend also was downwards between the late 1960s until the early 1990s, but this has been followed by a small recovery. In EMR-B, a steep increase was seen between the early 1970s until the mid-1980s. This was followed by a decrease in fruit and vegetable availability between about 1985 and 1990. Thereafter, the trend has been upward again. Finally, in EUR-B and EUR-C, the general increase in fruit and vegetable availability observed during the 1960 s, 1970 s and early 1980 s was followed by a 10 -year decrease in availability. This coincided with the significant political and economic changes that occurred in the subregion before and after the collapse of the former Soviet Union. Around 1993/1994, the trend then changed and a recovery was observed. The current upward trend is relatively gradual in EUR-C, but relatively steep in EUR-B.

### 5.2 Predicting subregional fruit and vegetable intake то 2030

In order to predict subregional fruit and vegetable intakes by age and sex between 2000 and 2030, a two-step process was used. First, the predicted subregional availability figures were obtained using linear regression analyses with fruit and vegetable availability and time (years) as the dependent and independent variables, respectively. For subregions where a relatively steadily increasing trend was observed (AMR-A, AMR-B, EMR-D, EUR-A, SEAR-B, SEAR-D and WPR-B), FAO data from 1961 onwards were used. For WPR-A and AFR-E, it was conservatively assumed that a plateau was reached during the prediction period. For the other subregions, a case-by-case approach was used based on the trends observed. For AFR-D and AMR-D, 1978 and 1992 were used as respective cut-off points for the regression analyses, as these years rep-
resented a point in time where a steep downward trend was stopped. Both EUR-B and EUR-C have experienced large fluctuations recently, and it is thus very difficult to predict future trends. For EUR-C, 1993 was selected as a cut-off since the downward trend slowed down from that point in time. The conservative increasing trend from 1993 onwards was thus used to predict future fruit and vegetable availability in that subregion. In EUR-B, a steep upward trend was observed from 1993. However, it is assumed that this trend will slow down and thus the average of the trend from 1993 onwards and the trend during the last two decades (1980+) was arbitrarily chosen. A similar approach was used for EMR-B where it is assumed that the observed trend since 1988 is too steep to reflect future projections.

The second step calculated the proportional changes in fruit and vegetable availability between 1997 (arbitrarily selected as representing the time when intake data were obtained) and 2000, 2005, 2010, 2015, 2020, 2025 and 2030. These were then applied to the intake data presented in section 2.8 , assuming the same proportional increase in intakes in both genders and all age groups within each subregion.

The results obtained for AMR-D, EMR-B, EUR-B and EUR-C, which are based on smaller numbers of data points, should be treated with particular caution.

### 5.3 Limitations of the methods

The results obtained for the projected fruit and vegetable intakes need to be treated with caution as the approach used has several limitations. One major limitation is related to the lack of time trend data on fruit and vegetable intake around the world, and the resulting need to base the calculations on food balance sheet data. Availability statistics can be a useful tool to observe trends in food availability in the past. However, predicting availability in the future (outside the range of available data) using linear regression analysis and making the assumption that trends will not change, may lead to false conclusions as current trends may change in the next 30 years for many reasons. For example, had we undertaken this analysis for EUR-C 10 years previously, the results would have been completely different. The values obtained for EMR-B, for example, appear to be too high compared with those obtained for the other subregions. In addition, as food balance sheets reflect food availability and not food intake, it is not possible to know whether the relative changes in availability are explained by variations in intakes or by other changes such as an increase in waste at the household level.

Predicting fruit and vegetable consumption is especially difficult because of the range of factors influencing dietary intakes, including food availability, economic situation, seasons, cultural preferences, fashions and knowledge and potential interactions between these factors. To base future estimates of fruit and vegetable intake on the predicted levels of any one of these factors (e.g. predicted gross national product [GNP] of
a country in 20 years) is not justified, in part due to the lack of information on future trends in these factors themselves. Furthermore, assumptions of such relationships would overlook the multiple determinants of diet. For example, we have seen that intakes vary considerably among countries with similar GNP (e.g. Ireland, Italy and the United Kingdom). Thus, the predictions presented in this chapter simply assume a continuation of complex patterns of social, economic and agricultural development of these countries in recent years.

## Acknowledgements

We would like to thank Dan Altmann, Majid Ezzati and Steve Vander Hoorn for their statistical advice. The Stata program used to implement the method of Greenland and Longnecker was written and tested by Ian White.

We also wish to thank Aileen Robertson for her contributions, and to acknowledge the help of many people worldwide who provided us with data on fruit and vegetable intake, those who helped us identify potential sources of intake data and research evidence, and all those who offered advice on methods (see list below).

We would like to thank M.E. Rio, L. Cappelen, M.C. Perez Somigliana, N.M. de Parada, N. Piazza, B. Houssay, H. Lareyna, S. Closa, K. Baghurst, S. Record, S. De Henauw, S. Petrova, S. Fagt, M. Lahti Koski, G.B.M. Mensink, S. Friel, D. Nitzan Kaluski, R. Goldberg, A. Turrini, Y. Matsumura, N. Yoshiike, T. Sharmanov, J. Rivera Dommarco, L. Johansson, M. Deurenberg-Yap, T-Bee Yian, C. Suok-Kai, J.S. Hampl and B.M. Popkin for providing us with data stratified according to the GBD age-sex categories.

Further acknowledgements go to: M.A.G. Bendech, M.K. Gabr, C. Fatoumata, F. Staubli Asobayire, P. Bovet, D. Labadarios, N. Stein, N.A. Ouedraogo, A. Pacin, L. Cappelen, M.C. Perez Somigliana, N. Marcilla, N. Piazza, H. Lareina, S. Closa, M.E. Gomez del Rio, D.R. Petteta, J. Montero, M. Esther, A. Achutti, R. Costa Ribeiro, M. Pastor, C.A. Monteiro, F. Begin, E. Boy, S. Alary, B. Choi, C. Koeune, P. Fisher, M. Houde Nadeau, S. Kirkland, C. Leinweber, L. Marrett, J. MacDonald, T. Wolever, O.F. Herran Falla, A. Rodriguez-Ojea, J. Gonzalez, F. Henry, J. Rivera Dommarco, C.P. Sanchez-Castillo, B. Popkin, T. Forrester, W. Freire, J. Potter, C. Garza, N. Scrimshaw, E. Murray, L.B. Dixon, J. Hampl, J. Hart, J. Kause, J. McClelland, S. McPherson, Lee Soo-Kyung, P. Terry, J.R. Hebert, P. Kaufmann, C. Junshi, W. Zhao, C. Chumming, J. Chen, J. Woo Wong, S. Sung, R. Gupta, H. Soori, Y. Matsumura, S. Mizushima, K. Ogawa, N.M. Al Hamad, N. Hwalla Baba, S. Jabbour, S. Reijo, J.A. Al Lawati, L.T. Cavalli-Sforza, M.O.O. Alcantara, C. Suok-Kai, M. Deurenberg-Yap, T. Bee-Yian, K. Cho-il Kim, W.H. Pan, S. Durongkadaj, P. Rawdaree, P. Chompook, L. Georgieva, S. Petrova, A. Kaic-Rak, Z. Brázdová, A.M. Beck, L. Ovesen, S. Fagt, J.

Haraldsdóttir, J.M. Lauritsen, A. Kasmel, S. Vaask, M. Lahti-Koski, R. Prattala, P. Knekt, M. Chauliac, J.L. Volatier, C. Cohidon, S. Hercberg, P. Wild, T. Norat, E. Riboli, K. Nemsadze, L. Bregvadze, G. Mensink, A. Potz, W. Sichert-Hellert, M. Thamm, G. Biro, P. Józan, J. Fodor, M. Szabo, T. Varga, G. Zajkás, J.M. Kearney, S. Friel, D. Kaluski, R. Goldsmith, F. Branca, A. Turrini, A. Ferro-Luzzi, A. Zucchi, P. Merrikin, B. Sarbayev, T.S. Sharmanov, G. Abuova, I. Pudule, M. Ellul, K .Hulshof, C.A.J. van Mierlo, H. Boshuizen, B. Blaker, L. Johansson, L. Szponar, P. Graça, P. Moreira, A.D. Deev, J. Maucec Zakotnik, B. Tivadar, A. Agudo, J. De Irala Estevez, O. Moreiras, I. López Azpiazu, K. Manderbacka, S. Hess, F.S. Asobayire, P. Bovet, E. Brunner, N. Day, A. Draper, S. Green, L. Hayes, P. James, K.T. Khaw, T. Lobstein, A. Ness, J. Powles, I. Sharp, H. Teomainen, M. Thorogood, P. Warner, T. Vos, K. Baghurst, S. Record, M. Coad, M. Anthea Magarey, E. Rutishauser Engel, I. Coles-Rutishauser, M. Watson, D. Woodward, D. Schulz, L. Andrew, K. Sladden and M. Tobias.

This work was supported by a grant from the English Department of Health. However, the English Department of Health cannot accept responsibility for any information provided or views expressed.

## Notes

1 See preface for an explanation of this term.
2 Estimating the subregional mean

- Estimation of the subregional (pooled) mean:
$\hat{\mu}=\frac{\sum N_{i} \times \bar{x}_{i}}{\sum N_{i}}$ where $i=1, \ldots, k$ sampled countries, $N_{\mathrm{i}}$ is the
population of the $i$ th country and $\bar{x}_{i}$ is the mean of the $i$ th country.
- $95 \%$ confidence interval (CI) for this estimator:

The variance of this estimator can be derived using:

$$
\operatorname{Var}(\hat{\mu})=\left(\frac{1}{\sum N_{i}}\right)^{2} \operatorname{Var}\left(\sum N_{i} \times \bar{x}_{i}\right)
$$

and assuming the means are independent

$$
=\left(\frac{1}{\sum N_{i}}\right)^{2} \sum \operatorname{Var}\left(N_{i} \times \bar{x}_{i}\right)=\left(\frac{1}{\sum N_{i}}\right)^{2} \sum N_{i}^{2} \operatorname{Var}\left(\bar{x}_{i}\right)
$$

Where now:
$\operatorname{Var}\left(\bar{x}_{i}\right)=\frac{s_{i}^{2}}{n_{i}}\left(1-\frac{n_{i}}{N_{i}}\right)$, with $s_{i}$ the standard deviation for the $i$ th country, sample size $n_{i}$.
The term $\left(1-\frac{n_{i}}{N_{i}}\right)$ is the finite population correction.
The standard error of the estimator is the square root of the variance.
The $95 \%$ CI for the subregional (pooled) mean is calculated as:
$=$ Subregional mean $\pm$ (1.96 x standard error of this estimator $)$
Estimating the subregional standard deviation

- Estimation of the subregional (pooled) variance and standard deviation:

$$
\hat{\sigma}^{2}=\frac{\sum N_{i} \times s_{i}^{2}}{\sum N_{i}}=\hat{\sigma}^{2}=\frac{\sum N_{i} \times s_{i}^{2}}{N}
$$

where $\sigma^{2}=$ pooled variance for the subregion, $I=1, \ldots, k$ sampled countries, $N_{\mathrm{i}}$ is the population of the $i$ th country and $s_{i}^{2}$ is the variance (square of SD ) of the $i$ th country; N is the sum of the $\mathrm{N}_{\mathrm{i}}$, in other words the sum of the populations of the sampled countries. This is unbiased because the expected value of a sample country's variance is the subregional variance, that is, $\mathrm{E}\left(s_{i}^{2}\right)=\sigma^{2}$,
then $\mathrm{E}\left(\hat{\sigma}^{2}\right)=\frac{1}{N} \sum N_{i} E\left(s_{i}^{2}\right)=\frac{1}{N} \sigma^{2} \sum N_{i}=\sigma^{2}$

- $95 \%$ CI for this estimator:

The $95 \%$ CI for the subregional (pooled) variance $\left(\hat{\sigma}^{2}\right)$ is approximated using:
Lower CI $=$ Subregional variance $\mathrm{x}(n-1) / \chi^{2}(n-1,0.025)$
Upper CI $=$ Subregional variance $\mathrm{x}(n-1) / \chi^{2}(n-1,0.975)$
where $n=\left(\sum n_{i}\right)$ and $n_{i}$ is the sample size for the $i$ th country: $n$ is thus the total size of the sample taken from the subregion.

## References

Agudo A, Esteve MG, Pallares C et al. (1997) Vegetable and fruit intake and the risk of lung cancer in women in Barcelona, Spain. European Journal on Cancer, 33:1256-1261.
Ahmad K, Hassan N (1982) Nutrition survey of rural Bangladesh. Institute of Nutrition and Food Science, Dhaka University, Dhaka.
Andersen NL, Fagt S, Groth MV et al. (1996) Danskernes kostvaner 1995 [Danish dietary patterns 1995]. National Food Agency, Copenhagen. (In Danish; English summary).
Anonymous (1994) The effect of vitamin E and beta-carotene on the incidence of lung cancer and other cancers in male smokers. The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study group. New England Journal of Medicine, 330:1029-1035.

Ascherio A, Rimm EB, Hernan MA et al. (1998) Intake of potassium, magnesium, calcium and fiber and risk of stroke among US men. Circulation, 98:1198-1204.
Ascherio A, Rimm EB, Hernan MA et al. (1999) Relation of consumption of vitamin E, vitamin C and caretonoids to risk for stroke among men in the United States. Annals of Internal Medicine, 130:963-970.

Awad AB, Downie A, Fink CS, Kim U (2000) Dietary phytosterol inhibits the growth and metastasis of MDA-MB-231 human breast cancer cells grown in SCID mice. Anticancer Research, 20:821-824.

Axelsson G, Liljeqvist T, Andersson L, Bergman B, Rylander R (1996) Dietary factors and lung cancer among men in west Sweden. International Journal of Epidemiology, 25:32-39.
Bates C, Thurnham S, Bingham S, Margetts B, Nelson M (1991) Biochemical markers of nutrient intake. In: Design concepts in nutritional epidemiology. Margetts B, Nelson M, eds. Oxford Medical Publications, Oxford.
Bazzano LA, He J, Ogden LG et al. (2001) Dietary potassium intake and risk of stroke in US men and women: National Health and Nutrition Examination Survey I epidemiologic follow-up study. Stroke, 32:1473-1480.

Bazzano L, Ogden LG, Vupputur IS, Loria C, Meyers L, Whelton PK (2000) Fruit and vegetable intake reduces cardiovascular disease mortality: results from the NHANES I epidemiologic follow-up study (NHEFS). American Heart Association, 40th Annual Conference on Cardiovascular Disease Epidemiology and Prevention. San Diego, CA.
Benito E, Obrador A, Stiggelbout A et al. (1990) A population-based case-control study of colorectal cancer in Majorca. I. Dietary factors. International Journal of Cancer, 45:69-76.

Bidoli E, Franceschi S, Talamini R, Barra S, La Vecchia C (1992) Food consumption and cancer of the colon and rectum in north-eastern Italy. International Journal of Cancer, 50:223-229.

Bingham SA, Gill, C, Welch, A et al. (1994) Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, foodfrequency questionnaires and estimated-diet records. British Journal of Nutrition, 72:619-643.
Bingham SA, Gill C, Welch A et al. (1997) Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. International Journal of Epidemiology, 26:S137-151.
Bingham SA, Nelson M, Paul AA, Haraldsdottir J, Løken EB, van Staveren WA (1988) Methods of data collection at an individual level. In: Manual on methodology for food consumption studies. Cameron ME, van Staveren WA, eds. Oxford University Press, Oxford.
Blot WJ, Li JY, Taylor PR, Guo W, Dawsey SM, Li B. (1995) The Linxian trials: mortality rates by vitamin-mineral intervention group. American Journal of Clinical Nutrition, 62:S1424-1426.
Boeing H, Frentzel-Beyme R, Berger M et al. (1991a) Case-control study on diet and stomach cancer in Germany. International Journal of Cancer, 47: 858-864.

Boeing H, Jedrychowski W, Wahrendorf J, Popiela T, Tobiasz-Adamczyk B, Kulig A (1991b) Dietary risk factors in intestinal and diffuse types of stomach cancer: a multicenter case-control study in Poland. Cancer Causes and Control, 2:227-233.

Bond GG, Thompson FE, Cook RR (1987) Dietary vitamin A and lung cancer: results of a case-control study among chemical workers. Nutrition and Cancer, 9:109-121.

Bosetti C, La Vecchia C, Talamini R et al. (2000) Food groups and risk of squamous cell esophageal cancer in northern Italy. International Journal of Cancer, 87:289-294.
Boshuizen HC, Bueno-de-Mesquita HB, Altenburg HP et al. (2002) Adjustment for smoking in lung cancer analyses in the EPIC cohort. IARC Scientific Publications, 156:59-61.
Botterweck AA, van den Brandt PA, Goldbohm RA (1998) A prospective cohort study on vegetable and fruit consumption and stomach cancer in the Netherlands. American Journal of Epidemiology, 148:842-853.

Botterweck AA, van den Brandt PA, Goldbohm RA (2000) Vitamins, carotenoids, dietary fiber, and the risk of gastric carcinoma. Cancer, 88:737-748.

Boushey CJ, Beresford SA, Omenn GS, Motulsky AG (1995) A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. Journal of the American Medical Association, 274:1049-1057.
Brennan P, Fortes C, Butler J et al. (2000) A multicenter case-control study of diet and lung cancer among non-smokers. Cancer Causes and Control, 11:49-58.

Breslow RA, Graubard BI, Sinha R, Subar AF (2000) Diet and lung cancer mortality: a 1987 National Health Interview Survey cohort study. Cancer Causes and Control, 11:419-431.

Brigelius-Flohe R, Traber MG (1999) Vitamin E: function and metabolism. FASEB Journal, 13:1145-1155.

Brown LM, Blot WJ, Schuman SH et al. (1988) Environmental factors and high risk of esophageal cancer among men in coastal South Carolina. Journal of the National Cancer Institute, 80:1620-1625.
Brown LM, Swanson CA, Gridley C et al. (1995) Adenocarcinoma of the esophagus: role of obesity and diet. Journal of the National Cancer Institute, 87:104-109.

Brown LM, Swanson CA, Gridley et al. (1998) Dietary factors and the risk of squamous cell esophageal cancer among black and white men in the United States. Cancer Causes and Control, 9:467-474.

Buiatti E, Palli D, Decarli A et al. (1989) A case-control study of gastric cancer and diet in Italy. International Journal of Cancer, 44:61-616.

Burr ML, Fehily AM, Gilbert JF et al. (1989) Effects of changes in fat, fish and fibre intakes on death and myocardial infarction: diet and reinfarction trial (DART). The Lancet, 2:757-761.

Byers TE, Graham S, Haughey BP, Marshall JR, Swanson MK (1987) Diet and lung cancer risk: findings from the Western New York Diet Study. American Journal of Epidemiology, 125:351-363.

Cameron M, van Staveren WA (1988) Manual on methodology for food consumption studies. Oxford University Press, Oxford.

Candelora EC, Stockwell HG, Armstrong AW, Pickham PA (1992) Dietary intake and risk of lung cancer in women who never smoked. Nutrition and Cancer, 17:263-270.

Castellsague X, Munoz N, De Stefani E, Victoria CG, Castelletto R, Rolon PA (2000) Influence of mate drinking, hot beverages and diet on esophageal cancer risk in South America. International Journal of Cancer, 88:658-664.
Chambers JC, Seddon MD, Shah S, Kooner JS (2001) Homocysteine-a novel risk factor for vascular disease. Journal of Royal Society of Medicine, 94:10-13.

Chen LY, Jokela R, Li DY et al. (2000) Effect of stable fish oil on arterial thrombogenesis, platelet aggregation, and superoxide dismutase activity. Journal of Cardiovascular Pharmacology, 35:502-505.
Cheng KK, Day NE (1996) Nutrition and esophageal cancer. Cancer Causes and Control, 7:33-40.

Cheng KK, Day NE, Davies TW (1992) Oesophageal cancer mortality in Europe: paradoxical time trend in relation to smoking and drinking. British Journal of Cancer, 65:613-617.
Cheng KK, Day NE, Duffy SW, Lam TH, Fok M, Wong J (1992) Pickled vegetables in the aetiology of oesophageal cancer in Hong Kong Chinese. The Lancet, 339:1314-1318.
Cheng KK, Duffy SW, Day NE, Lam TH (1995) Oesophageal cancer in never-smokers and never-drinkers. International Journal of Cancer, 60: 820-822.

Cheng KK, Sharp L, McKinney PA et al. (2000) A case-control study of oesophageal adenocarcinoma in women: a preventable disease. British Journal of Cancer, 83:127-132.

Chow WH, Schuman LM, McLaughlin JK et al. (1992) A cohort study of tobacco use, diet, occupation, and lung cancer mortality. Cancer Causes and Control, 3:247-254.
Chyou PH, Nomura AM, Hankin JH, Stemmermann GN (1990) A case-cohort study of diet and stomach cancer. Cancer Research, 50:7501-7504.
Coggon D, Barker DJ, Cole RB, Nelson M (1989) Stomach cancer and food storage. Journal of the National Cancer Institute, 81:1178-1182.
Collins R, Peto R, Armitage J (2002) The MRC/BHF Heart Protection study: preliminary results. International Journal of Clinical Practice, 56:53-56.

Conlin PR, Chow D, Miller ER 3rd et al. (2000) The effect of dietary patterns on blood pressure control in hypertensive patients: results from the Dietary Approaches to Stop Hypertension (DASH) trial. American Journal of Hypertension, 13:949-955.

Cornee J, Pobel D, Riboli E, Guyader M, Hemon B (1995) A case-control study of gastric cancer and nutritional factors in Marseille, France. European Journal of Epidemiology, 11:55-65.

Correa P, Fontham ET, Bravo JC et al. (2000) Chemoprevention of gastric dysplasia: randomized trial of antioxidant supplements and anti-helicobacter pylori therapy. Journal of the National Cancer Institute, 92:1881-1888.

Correa P, Fontham E, Pickle LW, Chen V, Lin YP, Haenszel W (1985) Dietary determinants of gastric cancer in south Louisiana inhabitants. Journal of the National Cancer Institute, 75:645-654.

Darby S, Whitley E, Doll R, Key TJA, Silcocks P (2001) Diet, smoking and lung cancer: a case control study of 1000 cases and 1500 controls in south-west England. British Journal of Cancer, 84:728-735.

Daviglus ML, Orencia AJ, Dyer AR et al. (1997) Dietary vitamin C, betacarotene and 30-year risk of stroke: results from the Western Electric study. Neuroepidemiology, 16:69-77.

Day NE, Ferrari P (2002) . Some methodological issues in nutritional epidemiology. IARC Scientific Publications, 156:5-10.

Day N, McKeown N, Wong M, Welch A, Bingham S (2001) Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. International Journal of Epidemiology, 30:309-317.
De Henauw S, De Backer G (1999) Nutrient and food intakes in selected subgroups of Belgian adults. British Journal of Nutrition, 81:S37-42.
de Lorgeril M, Renaud S, Mamelle N et al. (1994) Mediterranean alpha-linolenic acid-rich diet in secondary prevention of coronary heart disease. The Lancet, 343:1454-1459.
de Lorgeril M, Salen P, Martin JL, Monjaud I, Delaye J, Mamelle N (1999) Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction: final report of the Lyon Diet Heart study. Circulation, 99:779-785.

Decarli A, Liati P, Negri E, Franceschi S, La Vecchia C (1987) Vitamin A and other dietary factors in the etiology of esophageal cancer. Nutrition and Cancer, 10:29-37.
Demirer T, Icli F, Uzunalimoglu O, Kucuk O (1990) Diet and stomach cancer incidence. A case control study in Turkey. Cancer, 65:2344-2348.
Deneo-Pellegrini PH, De Stephani E, Ronco A (1996) Vegetables, fruits, and risk of colorectal cancer: a case-control study from Uruguay. Nutrition and Cancer, 25:297-304.

DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Controlled Clinical Trials, 7:177-188.

Directorate of Intelligence (2000) World fact book 2000. Central Intelligence Agency, Washington, DC.

Dorgan JF, Ziegler RG, Schoenberg JB et al. (1993) Race and sex differences in associations of vegetables, fruits and carotenoids with lung cancer risk in New Jersey (United States). Cancer Causes and Control, 4:273-281.
Egger M, Schneider M, Davey Smith G (1998) Spurious precision? Metaanalysis of observational studies. British Medical Journal, 316:140-144.

Ekstrom AM, Serafini M, Nyren O, Hansson LE, Ye W, Wolk A (2000) Dietary antioxidant intake and the risk of cardia cancer and noncardia cancer of the interstitial and diffuse types: a population-based case-control study in Sweden. International Journal of Cancer, 87:133-140.
Elwood PC, Fehily AM, Ising H, Poor DJ, Pickering J, Kamel F (1996) Dietary magnesium does not predict ischaemic heart disease in the Caerphilly cohort. European Journal of Clinical Nutrition, 50:694-697.
Faivre J, Boutron MC, Senesse P, Couillault C, Belighiti C, Meny B (1997) Environmental and familial risk factors in relation to the colorectal-carcinoma sequence: results of a case-control study in Burgundy (France). European Journal of Cancer Prevention, 6:127-131.

Fehily AM, Yarnell JW, Sweetnam PM, Elwood PC (1993) Diet and incident ischaemic heart disease: the Caerphilly study. British Journal of Nutrition, 69:303-314.

Feskanich D, Ziegler R, Michaud DS et al. (2000) Prospective study of fruit and vegetable consumption and risk of lung cancer among men and women. Journal of the National Cancer Institute, 92:1812-1823.
Finch S, Doyle, W, Lowe C, Bates C, Prentice A, Smithers G, Clarke PC (1998) National diet and nutrition survey-people aged 65 years and over. HMSO, London.

Findiet Study Group (1998) Dietary survey of Finnish adults 1997. National Public Health Institute, Helsinki.

Fontham ET, Pickle LW, Haenszel W, Correa P, Lin YP, Falk RT (1988) Dietary vitamin A and C and lung cancer risk in Louisiana. Cancer, 62:2267-2273.

FAO (1983) A comparative study of food consumption data from food balance sheets and household surveys. Food and Agriculture Organization of the United Nations, Rome.
FAO (1998a) Nutrition country profile: Nepal. Food and Agriculture Organization of the United Nations, Rome.
FAO (1998b) Nutrition country profile: Pakistan. Food and Agriculture Organization of the United Nations, Food and Nutrition Division, Rome.
FAO (2001) FAOSTAT. Available at http://apps.fao.org/. Food and Agriculture Organization of the United Nations, Rome (checked February 28, 2003).
Forman MR, Yao SX, Graubard BI et al. (1992) The effect of dietary intake of fruits and vegetables on the odds ratio of lung cancer among Yunnan tin miners. International Journal of Epidemiology, 21:437-441.
Franceschi S, Bidoli E, La Vecchia C, Talamini R, D’Avanzo B, Negri E (1994) Tomatoes and risk of digestive-tract cancers. International Journal of Cancer, 59:181-184.

Francecshi S, Favero A, La Vecchia C et al. (1997) Food groups and risk of colorectal cancer in Italy. International Journal of Cancer, 72:56-61.
Franceschi S, Parpinel M, La Vecchia C, Favero A, Talamini R, Negri E (1998) Role of different types of vegetables and fruit in the prevention of cancer of the colon, rectum and breast. Epidemiology, 9:338-341.

Fraser GE, Beeson WL, Phillips RL (1991) Diet and lung cancer in California Seventh-day Adventists. American Journal of Epidemiology, 133:683-693.

Fraser GE, Sabate J, Beeson WL, Strahan TM (1992) A possible protective effect of nut consumption on risk of coronary heart disease. The Adventist health study. Archives of Internal Medicine, 152:1416-1424.
Freudenheim JL, Graham S, Horvath PJ, Marshall JR, Haughey BP, Wilkinson G (1990) Risk associated with source of fiber and fiber components in cancer of the colon and rectum. Cancer Research, 50:3295-3300.
Friel S, NicGabhann S, Kelleher C (1999) The National Health and Lifestyle Survey (SLÁN)—Survey of lifestyles, attitudes and nutrition. Health Promotion Unit/NUI Galway, Dublin.

Fuchs CS, Giovannucci, EL, Colditz GA et al. (1999) Dietary fiber and the risk of colorectal cancer and adenoma in women. New England Journal of Medicine, 340:169-176.

Galanis DJ, Kolonel LN, Lee J, Nomura A (1998) Intakes of selected foods and beverages and the incidence of gastric cancer among the Japanese residents of Hawaii: a prospective study. International Journal of Epidemiology, 27:173-180.
Gale CR, Martyn CN, Winter PD, Cooper C (1995) Vitamin C and risk of death from stroke and coronary heart disease in cohort of elderly people. British Medical Journal, 310:1563-1566.
Gandini S, Merzenich H, Robertson C, Boyle P (2000) Meta-analysis of studies on breast cancer risk and diet: the role of fruit and vegetable consumption and the intake of associated micronutrients. European Journal of Cancer, 36:636-646.

Gao C, Tajima K, Kuroishi T, Hirose K, Inoue M (1993) Protective effects of raw vegetables and fruit against lung cancer among smokers and ex-smokers: a case-control study in the Tokai area of Japan. Japanese Journal of Cancer Research, 84:594-600.
Gao YT, McLaughlin JK, Gridley G et al. (1994) Risk factors for esophageal cancer in Shanghai, China. II. Role of diet and nutrients. International Journal of Cancer, 58:197-202.
Garcia-Closas R, Gonzales CA, Agudo A, Riboli E (1999) Intake of specific carotenoids and flavonoids and the risk of gastric cancer in Spain. Cancer Causes and Control, 10:71-75.

Garidou A, Tzonou A, Lipworth L, Signorello LB, Kalapothaki V, Trichopoulos D (1996) Life-style factors and medical conditions in relation to esophageal cancer by histologic type in a low-risk population. International Journal of Cancer, 68:295-299.
Gaziano JM (1996) Antioxidants in cardiovascular disease: randomized trials. Nutrition, 12:583-588.
Gaziano JM, Hennekens CH (1993) The role of beta-carotene in the prevention of cardiovascular disease. Annals of the New York Academy of Science, 691:148-155.

Gaziano JM, Manson JE, Branch LG, Colditz GA, Willett WC, Buring JE (1995) A prospective study of consumption of carotenoids in fruits and vegetables and decreased cardiovascular mortality in the elderly. Annals of Epidemiology, 5:255-260.
Gershoff SN (1993) Vitamin C (ascorbic acid): new roles, new requirements? Nutrition Reviews, 51:313-326.
Gey K (1995) Cardiovascular disease and vitamins. Concurrent correction of 'suboptimal' plasma antioxidant levels may, as important part of 'optimal' nutrition, help to prevent early stages of cardiovascular disease and cancer, respectively. Bibliotheca Nutritio et Dieta, 52:75-91.

Ghadirian P, Lacroix A, Maisonneuve P et al. (1997) Nutritional factors and colon carcinoma: a case-control study involving French Canadians in Montreal, Quebec, Canada. Cancer, 80:858-864.

Gillman MW, Cupples LA, Gagnon D et al. (1995) Protective effect of fruits and vegetables on development of stroke in men. Journal of the American Medical Association, 273:1113-1117.
Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Willett WC (1994) Intake of fat, meat, and fiber in relation to risk of colon cancer in men. Cancer Research, 54:2390-2397.
Gonzales CA, Sanz JM, Marcos G et al. (1991) Dietary factors and stomach cancer in Spain: a multi-centre case-control study. International Journal of Cancer, 49:513-519.

Goodman MT, Hankin JH, Wilens LR, Kolonel LN (1992) High-fat foods and the risk of lung cancer. Epidemiology, 3:288-299.

Government of India (1998) India nutrition profile. Ministry of Human Resource Development, Delhi.
Graham S, Haughey B, Marshall J et al. (1990) Diet in the epidemiology of gastric cancer. Nutrition and Cancer, 65:2344-2348.
Graham S, Marshall J, Haughey B et al. (1988) Dietary epidemiology of cancer of the colon in western New York. American Journal of Epidemiology, 128:490-503.

Greenberg E, Sporn MB (1996) Antioxidant vitamins, cancer and cardiovascular disease. New England Journal of Medicine, 334:1189-1190.
Greenland S, Longnecker MP (1992) Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. American Journal of Epidemiology, 135:1301-1309.
Gregory J (1990) The dietary and nutritional survey of British adults a survey of the dietary behaviour, nutritional status and blood pressure of adults aged 16 to 64 living in Great Britain. HMSO, London.

Gregory J (1995) National diet and nutrition survey children aged 1 1/2 to 4 1/2 years. HMSO, London.

Gregory J (2000) National diet and nutrition survey: young people aged 4 to 18 years. Report of the diet and nutrition survey. HMSO, London.

Guo W, Blot WJ, Li JY et al. (1994) A nested case-control study of oesophageal and stomach cancers in the Linxian nutrition intervention trial. International Journal of Epidemiology, 23:444-450.
Guo WD, Li JY, Blot WJ, Hsing AW, Chen JS, Fraumeni JF Jr (1990) Correlations of dietary intake and blood nutrient levels with esophageal cancer mortality in China. Nutrition and Cancer, 13:121-127.
Haegele AD, Gillette C, O’Neill C et al. (2000) Plasma xanthophyll carotenoids correlate inversely with indices of oxidative DNA damage and lipid peroxidation. Cancer Epidemiology, Biomarkers and Prevention, 9:421-425.
Hanaoka T, Tsugane S, Ando N et al. (1994) Alcohol consumption and risk of esophageal cancer in Japan: A case-control study in seven hospitals. Japanese Journal of Clinical Oncology, 24:241-246.
Hansson LE, Nyren O, Bergstrom R et al. (1993) Diet and risk of gastric cancer. A population-based case-control study in Sweden. International Journal of Cancer, 55:181-189.

Harris RW, Key TJ, Silcocks PB, Bull D, Wald NJ (1991) A case-control study of dietary carotene in men with lung cancer and in men with other epithelial cancers. Nutrition and Cancer, 15:63-68.
Harrison LE, Zhang ZF, Karpeh MS, Sun M, Kurtz RC (1997) The role of dietary factors in the intestinal and diffuse histological subtypes of gastric adenocarcinoma: a case-control study in the US. Cancer, 80:1021-1028.
Hebert JR, Clemow L, Pbert L, Ockene IS, Ockene JK (1995) Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. International Journal of Epidemiology, 24:389-398.

Heilbrun LK, Nomura A, Hankin JH, Stemmermann GN (1989) Diet and colorectal cancer with special reference to fiber intake. International Journal of Cancer, 44:1-6.

Hennekens CH, Buring JE, Manson JE et al. (1996) Lack of effect of long-term supplementation with beta-carotene on the incidence of malignant neoplasms and cardiovascular disease. New England Journal of Medicine, 334: 1145-1149.
Hertog M, Bueno-de-Mesquita HB, Fehili AM et al. (1996) Fruit and vegetable consumption and cancer mortality in the Caerphilly study. Cancer Epidemiology, Biomarkers and Prevention, 5:673-677.

Hertog MG, Feskens EJ, Holliman PC, Katan MB, Kromhout D (1993) Dietary antioxidant flavonoids and risk of coronary heart disease: the Zutphen Elderly study. The Lancet, 342:1007-1011.

Hill AB (1965). The environment and disease: Association or causation? Proceedings of the Royal Society of Medicine, 58:295-300.
Hirayama T (1986) Nutrition and cancer-a large scale cohort study. Progress in Clinical and Biological Research, 206:299-311.
Hirayama T (1990) Diet and mortality. In: Lifestyle and mortality. Hirayama T, ed. Karger, Basel.

Hirvonen T, Virtamo J, Korhonen P, Albanes D, Pietinen P (2000) Intake of flavonoids, carotenoids, vitamin C and E , and risk of stroke in male smokers. Stroke, 31:2301-2306.

Hirvonen T, Virtamo J, Korhonen P, Albanes D, Pietinen P (2001) Flavonol and flavone intake and the risk of cancer in male smokers (Finland). Cancer Causes and Control, 12:789-796.
Hooper L, Ness AR, Smith GD (2001) Antioxidant strategy for cardiovascular diseases. The Lancet, 357:1705-1706.
Hoshiyama Y, Sasaba T (1992) A case-control study of single and multiple stomach cancers in Saitama Prefecture, Japan. Japanese Journal of Cancer Research, 83:937-943.
Hu J, Johnson KC, Mao Y et al. (1997) A case-control study of diet and lung cancer in northeast China. International Journal of Cancer, 71:924-931.
Hu JF, Liu YY, Yu YK, Zhao TZ, Lui SD, Wang QQ (1991) Diet and cancer of the colon and rectum: a case-control study in China. International Journal of Epidemiology, 20:362-367.
Hu J, Nyren O, Wolk A et al. (1994) Risk factors for oesophageal cancer in northeast China. International Journal of Cancer, 57:38-46.
Hu JF, Zhang SF, Jia EM et al. (1988) Diet and cancer of the stomach: a casecontrol study in China. International Journal of Cancer, 41:331-335.
Insel P, Turner RE, Ross D (2003) Discovering nutrition-interactive glossary. Jones and Bartlett Publishers, Boston, MA. http://nutrition.jbpub.com/discovering/.
Iscovich JM, L'Abbe KA, Castelleto R et al. (1992) Colon cancer in Argentina. I. Risk from intake of dietary items. International Journal of Cancer, 51:851-857.

Jain M, Burch JD, Howe GR, Risch HA, Miller AB (1990) Dietary factors and risk of lung cancer: results from a case-control study, Toronto, 1981-1985. International Journal of Cancer, 45:287-293.
Jansen MC, Bueno-de-Mesquita HB, Rasanen L et al. (2001) Cohort analysis of fruit and vegetable consumption and lung cancer mortality in European men. International Journal of Cancer, 92:913-918.
Jedrychowski W, Wahrendorf J, Popiela T, Rachtan J (1986) A case-control study of dietary factors and stomach cancer risk in Poland. International Journal of Cancer, 37:837-842.
Jha P, Flather M, Lonn E, Farkouh M, Yusuf S (1995) The antioxidant vitamins and cardiovascular disease. A critical review of epidemiologic and trial data. Annals of Internal Medicine, 123:860-872.
Ji BT, Chow WH, Yang G et al. (1998) Dietary habits and stomach cancer in Shanghai, China. International Journal of Cancer, 76:659-664.

Joffe M, Robertson A (2001) The potential contribution of increased vegetable and fruit consumption to health gain in the European Union. Public Health Nutrition, 4:893-901.

Johansson G, Wikman A, Ahren AM, Hallmans G, Johansson I (2001) Underreporting of energy intake in repeated 24 -hour recalls related to gender, age, weight status, day of interview, educational level, reported food intake, smoking habits and area of living. Public Health Nutrition, 4:919-927.
Joshipura KJ, Ascherio A, Manson JE et al. (1999) Fruit and vegetable intake in relation to risk of ischemic stroke. Journal of the American Medical Association, 282:1233-1239.
Joshipura KJ, Hu FB, Manson JE et al. (2001) The effect of fruit and vegetable intake on risk for coronary heart disease. Annals of Internal Medicine, 134:1106-1114.

Kalandidi A, Katsouyanni K, Voropoulou N, Bastas G, Saracci R, Trichopoulos D (1990) Passive smoking and diet in the etiology of lung cancer and among nonsmokers. Cancer Causes and Control, 1:15-21.

Kampman E, Verhoeven D, Sloots L, van't Veer P (1995) Vegetable and animal products as determinants of colon cancer in Dutch men and women. Cancer Causes and Control, 6:225-234.
Kantor LS, Kipton K, Manchester A, Oliveria V (1997) Estimating and addressing America's food losses. Food Review U.S. Department of Agriculture, Economic Research Service, Washington, DC.
Kato I, Akhmedkhanov A, Koenig K, Toniolo PG, Shore RE, Riboli E (1997) Prospective study of diet and female colorectal cancer: the New York University Women's Health study. Nutrition and Cancer, 28:276-281.
Kato I, Dnistrian, A, Schwartz M et al. (1999) Serum folate, homocysteine and colorectal cancer risk in women: a nested case-control study. British Journal of Cancer, 79:1917-1922.

Kato I, Tominaga S, Ito Y et al. (1990) A comparative case-control analysis of stomach cancer and atrophic gastritis. Cancer Research, 50:6559-6564.

Kato I, Tominaga S, Matsumoto K (1992) A prospective study of stomach cancer among a rural Japanese population: a 6 -year survey. Japanese Journal of Cancer Research, 83:568-575.
Kauhanen J, Kaplan GA, Goldberg DE, Salonen R, Salonen JT (1999) Pattern of alcohol drinking and progression of atherosclerosis. Arteriosclerosis, Thrombosis and Vascular Biology, 19:3001-3006.
Keli SO, Hertog MG, Feskens EJ, Kromhout D (1996) Dietary flavonoids, antioxidant vitamins, and incidence of stroke: the Zutphen study. Archives of Internal Medicine, 156:637-642.

Kelly A, Becker W, Helsing E (1991) Food balance sheets. In: Food and health data. Their use in nutrition policy-making. Becker W, Helsing E, eds. (European Series No. 34.) WHO Regional Publications, World Health Organization, Copenhagen.
Kennedy AR (1998) Chemopreventive agents: protease inhibitors. Pharmacology and Therapeutics, 78:167-209.
Key TJ, Thorogood M, Appleby PN, Burr ML (1996) Dietary habits and mortality in 11000 vegetarians and health conscious people: results of a 17-year follow-up. British Medical Journal, 313:775-779.

Kharbanda R, Vallance P (2001) Coronary artery disease-from bench to bedside. Journal of Royal Society of Medicine, 94:61-64.

Khaw KT, Barrett-Connor E (1987a) Dietary fiber and reduced ischemic heart disease mortality rates in men and women: a 12-year prospective study. American Journal of Epidemiology, 126:1093-1102.
Khaw KT, Barrett-Connor E (1987b) Dietary potassium and stroke-associated mortality. A 12-year prospective population study. New England Journal of Medicine, 316:235-240.
Khaw KT, Bingham S, Welch A et al. (2001) Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: a prospective population study. European Prospective Investigation into Cancer and Nutrition. The Lancet, 357:657-663.

Kipnis V, Carroll RJ, Freedman LS, Li L (1999) Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies. American Journal of Epidemiology, 150:642-651.
Kjaerheim K, Gaard M, Andersen A (1998) The role of alcohol, tobacco, and dietary factors in upper aerodigestive tract cancers: a prospective study of 10900 Norwegian men. Cancer Causes and Control, 9:99-108.
Klerk M, Jansen MCJF, van't Veer P, Kok FJ (1998) Fruits and vegetables in chronic disease prevention. Wageningen Agricultural University, Wageningen.
Klipstein-Grobusch K, Geleijnse JM, den Breeijen JH et al. (1999) Dietary antioxidants and risk of myocardial infarction in the elderly: the Rotterdam study. American Journal of Clinical Nutrition, 69:261-266.
Knekt P, Isotupa S, Rissanen H et al. (2000) Quercetin intake and the incidence of cerebrovascular disease. European Journal of Clinical Nutrition, 54: 415-417.

Knekt P, Jarvinen R, Reunanen A, Maatela J (1996) Flavonoid intake and coronary mortality in Finland: a cohort study. British Medical Journal, 312: 478-481.

Knekt P, Jarvinen R, Seppanen R et al. (1991) Dietary antioxidants and the risk of lung cancer. American Journal of Epidemiology, 134:471-479.
Knekt P, Jarvinen R, Seppanen R et al. (1997) Dietary flavonoids and the risk of lung cancer and other malignant neoplasms. American Journal of Epidemiology, 146:223-230.
Knekt P, Reunanen A, Jarvinen R, Seppanen R, Heliovaara M, Aromaa A (1994) Antioxidant vitamin intake and coronary mortality in a longitudinal population study. American Journal of Epidemiology, 139:1180-1189.

Kneller RW, McLaughlin JK, Bjelke E et al. (1991) A cohort study of stomach cancer in a high risk American population. Cancer, 68:672-678.

Kojda G, Harrison D (1999) Interactions between NO and reactive oxygen species: pathophysiological importance in atherosclerosis, hypertension, diabetes and heart failure. Cardiovascular Research, 43:562-571.

Kono S, Hirohata T (1996) Nutrition and stomach cancer. Cancer Causes and Control, 7:41-55.

Kono S, Ikeda M, Tokudome S, Kuratsune M (1988) A case-control study of gastric cancer and diet in northern Kyushu, Japan. Japanese Journal of Cancer Research, 79:1067-1074.

Koo L (1988) Dietary habits and lung cancer risk among Chinese females in Hong Kong who never smoked. Nutrition and Cancer: An International Journal, 11:155-172.
Koo L (1997) Diet and lung cancer 20+ years later: more questions than answers? International Journal of Cancer, Suppl. 10:S22-29.
Kotake K, Koyama Y, Nasu J, Fukutomi T, Yamaguchi N (1995) Relation of family history of cancer and environmental factors to the risk of colorectal cancer: a case-control study. Japanese Journal of Clinical Oncology, 25: 195-202.

Kristenson M, Zieden B, Kucinskiene Z et al. (1997) Antioxidant state and mortality from coronary heart disease in Lithuanian and Swedish men: concomitant cross sectional study of men aged 50. British Medical Journal, 314: 629-633.

Kromhout D (1987) Essential micronutrients in relation to carcinogenesis. American Journal of Clinical Nutrition, 45:S1361-1367.
Kune GA, Bannerman S, Watson LF (1992) Attributable risk to diet, alcohol and family history in the Melbourne Colorectal Cancer study. Nutrition and Cancer, 18:231-235.

Kune S, Kune GA, Watson LF (1987) Case-control study of dietary etiological factors: the Melbourne Colorectal Cancer study. Nutrition and Cancer, 9:21-42.

Kushi LH, Folsom AR, Prinaes RJ, Mink PJ, Wu Y, Bostick RM (1996) Dietary antioxidant vitamins and deaths from coronary heart disease in postmenopausal women. New England Journal of Medicine, 334:1156-1162.

Kvale G, Bjelke E, Gart JJ (1983) Dietary habits and lung cancer risk. International Journal of Cancer, 31:397-405.
La Vecchia C, Negri E, Decarli A, D'Avanzo B, Franceschi S (1987) A casecontrol study of diet and gastric cancer in northern Italy. International Journal of Cancer, 40:484-489.
Lampe J (1999) Health effects of vegetables and fruits: assessing mechanisms of action in human experimental studies. American Journal of Clinical Nutrition, 70:S475-490.

Lapidus L, Andersson H, Bengtsson C, Bosaeus I (1986) Dietary habits in relation to incidence of cardiovascular disease and death in women: a 12 -year follow-up of the participants in the population study of women in Gothenberg, Sweden. American Journal of Clinical Nutrition, 44:444-448.

Launoy G, Milan C, Day NE, Pienkowski MP, Gignoux M, Faivre J (1998) Diet and squamous-cell cancer of the oesophagus: a French multicentre casecontrol study. International Journal of Cancer, 76:7-12.
Law M, Morris J (1998) By how much does fruit and vegetable consumption reduce the risk of ischaemic heart disease? European Journal of Clinical Nutrition, 52:549-556.

Le Marchand L, Hankin JH, Wilkens LR, Kolonel LN, Englyst HN, Lyu LC (1997) Dietary fibre and colorectal cancer risk. Epidemiology, 8:658-665.

Le Marchand L, Yoshizawa CN, Kolonel LN, Hankin JH, Goodman MT (1989) Vegetable consumption and lung cancer risk: a population-based case-control study in Hawaii. Journal of the National Cancer Institute, 81:1158-1164.

Lee CN, Reed DM, MacLean CJ, Yano K, Chiu D (1988) Dietary potassium and stroke. New England Journal of Medicine, 318:995-996.
Lee HP, Gourley L, Duffy SW, Esteve J, Lee J, Day NE (1989) Colorectal cancer and diet in an Asian population: a case-control study among Singapore Chinese. International Journal of Cancer, 43:1007-1016.
Lee JK, Park BJ, Yoo KY, Ahn YO (1995) Dietary factors and stomach cancer: case-control study in Korea. International Journal of Epidemiology, 24: 33-41.

Leppala JM, Virtamo J, Fogelholm R, Albanes D, Taylor PR, Heinonen OP (2000) Vitamin E and beta-carotene supplementation in high risk for stroke: a subgroup analysis of the Alpha-Tocopherol, Beta-Carotene Cancer Prevention study. Archives of Neurology, 57:1503-1509.
Levi F, Pasche C, La Vecchia C, Lucchini F, Franceschi S (1999) Food groups and colorectal cancer risk. British Journal of Cancer, 79:1283-1287.
Levi F, Pasche C, Lucchini F et al. (2000) Food groups and oesophageal cancer risk in Vaud, Switzerland. European Journal of Cancer Prevention, 9: 257-263.
Li JY, Ershow AG, Chen ZJ et al. (1989) A case-control study of cancer of the esophagus and gastric cardia in Linxian. International Journal of Cancer, 43:755-761.

Liu S, Lee IM, Ajani U, Cole SR, Buring JE, Manson JE (2001) Intake of vegetables rich in carotenoids and risk of coronary heart disease in men: The Physicians' Health study. International Journal of Epidemiology, 30:130-135.

Liu S, Manson JE, Lee IM et al. (2000) Fruit and vegetable intake and risk of cardiovascular disease: the Women's Health study. American Journal of Clinical Nutrition, 72:922-928.
Lonn EM, Yusuf S (1999) Evidence-based cardiology: emerging approaches in preventing cardiovascular disease. British Medical Journal, 318:1337-1341.
Macquart-Moulin G, Riboli E, Cornee J, Charnay B, Berthezene P, Day N (1986) Case-control study on colorectal cancer and diet in Marseilles. International Journal of Cancer, 38:183-191.

Malaveille C, Hautefeuille A, Pignatelli B, Talaska G, Vineis P, Bartsch H (1996) Dietary phenolics as anti-mutagens and inhibitors of tobacco-related DNA adduction in the urothelium of smokers. Carcinogenesis, 17:2193-200.

Mann JI, Appleby PN, Key TJ, Thorogood M (1997) Dietary determinants of ischemic heart disease in health conscious individuals. Heart, 78:450-455.
Marshall JR, Chen Z (1999) Diet and health risk: risk patterns and diseasespecific associations. American Journal of Clinical Nutrition, 69:S1351-1356.

Mathers C, Vos T, Stevenson C (1999) The burden of disease and injury in Australia. Australian Institute of Health and Welfare, Canberra.

Maynard M, Gunnell D, Emmett P, Frankel S, Davey Smith G (2003) Fruit, vegetables, and antioxidants in childhood and risk of adult cancer: the Boyd Orr cohort. Journal of Epidemiology and Community Health, 57:218-225.
Mayne ST, Janerich DT, Greenwald P et al. (1994) Dietary beta-carotene and lung cancer risk in U.S. nonsmokers. Journal of the National Cancer Institute, 86:33-38.
Memik F, Nak SG, Gulten M, Ozturk M (1992) Gastric carcinoma in northwestern Turkey: epidemiologic characteristics. Journal of Environmental Pathology, Toxicology and Oncology, 11:335-338.

Menotti A, Kromhout D, Blackburn H, Fidanza F, Buzina R, Nissinen A (1999) Food intake patterns and 25 -year mortality from coronary heart disease: cross cultural correlations from the Seven Countries study. The Seven Countries study research group. European Journal of Epidemiology, 15:507-515.
Mensink G, Hermann-Kunz E, Thamm M (1998) Der Ernährungssurvey. Das Gesundheitswesen, 60:S83-86.
Messina MJ (1999) Legumes and soybeans: overview of their nutritional profiles and health effects. American Journal of Clinical Nutrition, 70:S439-450.
Mettlin C (1989) Milk drinking, other beverage habits, and lung cancer risk. International Journal of Cancer, 43:608-612.
Meyer F, White E (1993) Alcohol and nutrients in relation to colon cancer in middle-aged adults. American Journal of Epidemiology, 138:225-236.
Michaud DS, Feskanich D, Rimm ED et al. (2000) Intake of specific carotenoids and risk of lung cancer in 2 prospective US cohorts. American Journal of Clinical Nutrition, 72:990-997.

Michels KB, Giovannucci E, Joshipura KJ et al. (2000) Prospective study of fruit and vegetable consumption and incidence of colon and rectal cancers. Journal of the National Cancer Institute, 92:1740-1752.

Michels KB, Giovannucci E, Joshipura KJ et al. (2002). Fruit and vegetable consumption and colorectal cancer incidence. IARC Scientific Publications, 156:139-140.

Ministry of Health and Welfare (2000) Annual report of the national nutrition survey in 1998. Daiichi Publishing Co, Tokyo.
Morris JN, Marr JW, Clayton DG (1977) Diet and heart: a postscript. British Medical Journal, 2:1307-1314.
Muir C, McKinney PA (1992) Cancer of the oesophagus: a global overview. European Journal of Cancer Prevention, 1:259-264.
Mulder I, Jansen MC, Smit HA et al. (2000) Role of smoking and diet in the cross-cultural variation in lung cancer mortality: the Seven Countries study. Seven Countries study research group. International Journal of Cancer, 88: 665-671.

Munoz N, Plummer M, Vivas J et al. (2001) A case-control study of gastric cancer in Venezuela. International Journal of Cancer, 93:417-423.

Munoz SE, Ferraroni M, La Vecchia C, Decarli A (1997) Gastric cancer risk factors in subjects with family history. Cancer Epidemiology, Biomarkers and Prevention, 6:137-140.

National Institute of Public Health (1997) Determinants of the burden of disease in the EU. National Institute of Public Health, Stockholm.
Nayar D, Kapil U, Joshi YK et al. (2000) Nutritional risk factors in esophageal cancer. Journal of the Association of Physicians in India, 48:781-787.
Negri E, Franceschi S, Parpinel M, La Vecchia C (1998) Fiber intake and risk of colorectal cancer. Cancer Epidemiology, Biomarkers and Prevention, 7: 667-671.
Negri E, La Vecchia C, Franceschi S, D’Avanzo B, Parazzini F (1991) Vegetable and fruit consumption and cancer risk. International Journal of Cancer, 48:350-354.
Nelson M, Bingham S (1997) Assessment of food consumption and nutrient intake. In: Design concepts in nutritional epidemiology. Margetts BM, Nelson M, eds. Oxford University Press, Oxford.

Ness A (2001) Commentary: beyond beta-carotene-antioxidants and cardiovascular disease. International Journal of Epidemiology, 30:143-144.
Ness A (2002) Is olive oil a key ingredient in the Mediterranean recipe for health? International Journal of Epidemiology. 31:481-482.
Ness A, Egger M, Davey Smith G (1999a) Role of antioxidant vitamins in prevention of cardiovascular diseases. Meta-analysis seems to exclude benefit of vitamin C supplementation. British Medical Journal, 319:577.
Ness A, Egger M, Powles J (1999b) Fruit and vegetables and ischaemic heart disease: systematic review or misleading meta-analysis? European Journal of Clinical Nutrition, 53:900-902.

Ness AR, Powles JW (1997) Fruit and vegetables, and cardiovascular disease: a review. International Journal of Epidemiology, 26:1-13.

Ness AR, Powles JW (1999) The role of diet, fruit and vegetables and antioxidants in the aetiology of stroke. Journal of Cardiovascular Risk, 6:229-234.

Nomura AM, Grove JS, Stemmermann GN, Severson RK (1990) A prospective cohort study on stomach cancer and its relation to diet, cigarettes and alcohol consumption. Cancer Research, 50:627-631.
Norat T, Van Bakel M, Riboli E (2001) Fruit and vegetable consumption and risk of cancer of the digestive tract: meta-analysis of published case control and cohort studies. European Conference on Diet and Cancer. Lyon.
Notani PN, Jayant K (1987) Role of diet in upper aerodigestive tract cancers. Nutrition and Cancer, 10:103-113.
Nyberg F, Agrenius V, Svartengren K, Svensson C, Pershagen G (1998) Dietary factors and risk of lung cancer in never-smokers. International Journal of Cancer, 78:430-436.

Obarzanek E, Sacks FM, Vollmer VM et al. (2001) Effects on blood lipids of a blood pressure-lowering diet: the Dietary Approaches to Stop Hypertension (DASH) trial. American Journal of Clinical Nutrition, 74:80-89.

Ocke M, Bueno-de-Mesquita HB, Feskens EJ, Kromhout D, Menotti A, Blackburn H (1998) Adherence to the European Code Against Cancer in relation to long-term cancer mortality: intercohort comparisons from the Seven Countries study. Nutrition and Cancer, 30:14-20.
Ocke M, Bueno-de-Mesquita HB, Feskens EJ, van Staveren WA, Kromhout D (1997) Repeated measurements of vegetables, fruits, beta-carotene, and vitamins C and E in relation to lung cancer. The Zutphen study. American Journal of Epidemiology, 145:358-365.
Oganesian A, Hendricks JD, Pereira CB, Orner GA, Bailey GS, Williams DE (1999) Potency of dietary indol-3-carbinol as a promoter of aflatoxin B1initiated hepatocarcinogenesis: results from a 9000 animal tumor study. Carcinogenesis, 20:453-458.

Omenn G, Goodman GE, Thornquist MD et al. (1996a) Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Retinol Efficacy Trial. Journal of the National Cancer Institute, 88:1550-1559.
Omenn GS, Goodman G, Thornquist MD et al. (1996b) Effects of a combination of beta-carotene and vitamin A on lung and cardiovascular disease. New England Journal of Medicine, 334:1150-1155.
Paganini-Hill GM, Ross RK, Gray GE, Henderson BE (1987) Vitamin A, betacarotene and risk of cancer: a prospective study. Journal of the National Cancer Institute, 79:443-448.

Pandey DK, Shekelle R, Selwyn BJ, Tangney C, Stamler J (1995) Dietary vitamin C and beta-carotene and risk of death in middle-aged men. The Western Electric study. American Journal of Epidemiology, 142:1269-1278.
Pawlega J, Rachtan J, Dyba T (1997) Evaluation of certain risk factors for lung cancer in Cracow (Poland)-a case-control study. Acta Oncologica, 36: 471-476.
Peters RK, Pike MC, Garabrant D, Mack TM (1992) Diet and colon cancer in Los Angeles County, California. Cancer Causes and Control, 3:457-473.
Petrova S, Angelova K, Ivanova L et al. (2000) National dietary and nutritional status survey of the population in Bulgaria, 1998. Food consumption. Hygiene and Public Health, XLIII:55-62.

Phillips RL, Snowdon DN (1985) Dietary relationships with fatal colorectal cancer among Seventh-Day Adventists. Journal of the National Cancer Institute, 74:307-317.

Pietinen P, Malila N, Virtene M et al. (1999) Diet and risk of colorectal cancer in a cohort of Finnish men. Cancer Causes and Control, 10:387-396.

Pietinen P, Rimm EB, Korhonen P et al. (1996) Intake of dietary fiber and risk of coronary heart disease in a cohort of Finnish men. The Alpha-Tocopherol, Beta-Carotene Cancer Prevention study. Circulation, 94:2720-2727.
Pillow PC, Hursting SD, Duphorne CM et al. (1997) Case-control assessment of diet and lung cancer risk in African Americans and Mexican Americans. Nutrition and Cancer, 29:169-173.

Pisani P, Berrino F, Macaluso M, Pastorino U, Crosignani P, Baldasseroni A (1986) Carrots, green vegetables and lung cancer: a case-control study. International Journal of Epidemiology, 15:463-468.

Potter JD (1996) Nutrition and colorectal cancer. Cancer Causes and Control, 7:127-146.

Powles JW, Day NE, Sanz MA, Bingham SA (1996) Protective foods in winter and spring: a key to lower vascular mortality? The Lancet, 348:898-899.
Price JE, Fowkes FG (1997) Antioxidant vitamins in the prevention of cardiovascular disease. The epidemiological evidence. European Heart Journal, 18:719-727.
Puska P (2000) Nutrition and mortality: the Finnish experience. Acta Cardiologica, 55:213-220.
Rachtan J, Sokolowski A (1997) Risk factors for lung cancer among women in Poland. Lung Cancer, 18:137-145.

Ramon JM, Serra L, Cerdo C, Oromi J (1993) Dietary factors and gastric cancer risk. A case control study in Spain. Cancer, 71:1731-1735.

Rehm J, Monteiro M, Room R et al. (2001) Steps towards constructing a global comparative risk analysis for alcohol consumption: determining indicators and empirical weights for patterns of drinking, deciding about theoretical minimum, and dealing with different consequences. European Addiction Research, 7:138-147.
Rexrode KM, Manson JE (1996) Antioxidants and coronary heart disease: observational studies. Journal of Cardiovascular Risk, 3:363-367.
Rimm EB, Ascherio A, Giovannucci E, Speigelman D, Stampfer MJ, Willett WC (1996) Vegetable, fruit, and cereal fiber intake and risk of coronary heart disease among men. Journal of the American Medical Association, 275: 447-451.

Rinzler SH (1968) Primary prevention of coronary heart disease by diet. Bulletin of the New York Academy of Medicine, 44:936-949.
Risch HA, Jain M, Choi NW et al. (1985) Dietary factors and the incidence of cancer of the stomach. American Journal of Epidemiology, 122:947-959.
Rivera Dommarco J, Shamah Levy T, Villalpando Hernándes S, González de Cosío T, Hernández Prado B, Sepúlveda J, eds. (2001b) Encuesta nacional de nutrición 1999. Estado nutricio de niños y mujeres en México. [National nutrition survey 1999. Nutritional status of women and children in Mexico.] Instituto Nacional de Salud Pública National Institute of Public Health, Cuernavaca, Mexico.

Ross RK, Yuan JM, Henderson BE, Park J, Gao YT, Yu MC (1997) Prospective evaluation of dietary and other predictors of fatal stroke in Shanghai, China. Circulation, 96:50-55.

Sankaranarayanan R, Varghese C, Duffy SW, Padmakumary G, Day NE, Nair MK (1994) A case-control study of diet and lung cancer in Kerala, south India. International Journal of Cancer, 58:644-649.

Sargeant LA, Khaw KT, Bingham S et al. (2001) Fruit and vegetable intake and population glycosylated haemoglobin levels: the EPIC-Norfolk study. European Journal of Clinical Nutrition, 55:342-348.

Sawaya WN, al-Awadi FA, Saeed T et al. (1999) Kuwait's total diet study, dietary intake of organochlorine, carbamate, benzimidazole and pheylurea pesticide residues. Journal of AOAC International, 82:1458-1465.
Schut HA, Cummings DA, Smale MH, Josyula S, Friesen MD (1997) DNA adducts of heterocyclic amines: formation removal and inhibition by dietary components. Mutation Research, 376:185-194.
Sekula W, Becker W, Trichopoulou A, Zajkas G (1991) Comparison of dietary data from different sources: some examples. In: Food and health data. Their use in nutrition policy-making. Becker W, Helsing E, eds. (European Series No. 34.) WHO Regional Publications, World Health Organization, Copenhagen.
Sengupta A, Das S (1999) The anti-carcinogenic role of lycopene, abundantly present in tomatoes. European Journal of Cancer Prevention, 8:325-330.
Shannon J, White E, Shattuck AL, Potter JD (1996) Relationship of food groups and water intake to colon cancer risk. Cancer Epidemiology, Biomarkers and Prevention, 5:495-502.
Shibata A, Paganini-Hill A, Ross RK, Henderson BE (1992) Intake of vegetables, fruits, beta-carotene, vitamin C and vitamin supplements and cancer incidence among the elderly: a prospective study. British Journal of Cancer, 66:673-680.

Singh RB, Niaz MA, Gosh S, Singh R, Rastogi SS (1993a) Effect on mortality and reinfaction of adding fruits and vegetables to a prudent diet in the Indian experiment of infarct survival (IEIS). Journal of the American College of Nutrition, 12:255-261.
Singh RB, Rastogi SS, Singh NK, Ghosh S, Gupta S, Niaz MA (1993b) Can guava fruit decrease blood pressure and blood lipids? Journal of Human Hypertension, 7:33-38.
Singh RR, Rastogi SS, Verma R et al. (1992) Randomised controlled trial of cardioprotective diet in patients with recent acute myocardial infarction: results of one year follow up. British Medical Journal, 304:1015-1019.

Singh SV, Mohan RR, Agarwal R et al. (1996) Novel anti-carcinogenic activity of an organosulfide from garlic: inhibition of H-RAS oncogene transformed tumor growth in vivo by diallyl disulfide is associated with inhibition of p21H-ras processing. Biochemical Biophysical Research Communications, 225:660-665.
Slattery ML, Potter JD, Coates A et al. (1997) Plant foods and colon cancer: an assessment of specific foods and their related nutrients (United States). Cancer Causes and Control, 8:575-590.

Slattery ML, Schumacher MC, Smith KR, West DW, Abd-Elghany N (1988) Physical activity, diet, and risk of colon cancer in Utah. American Journal of Epidemiology, 128:989-999.

Smit HA, Grievink L, Tabak C (1999) Dietary influences on chronic obstructive lung disease and asthma: a review of the epidemiological evidence. Proceedings of the Nutrition Society, 58:309-319.
Smith-Warner SA, Elmer PJ, Fosdick L, Tharp TM, Randall B (1997) Reliability and comparability of three dietary assessment methods for estimating fruit and vegetable intakes. Epidemiology, 8:196-201.
Speizer FE, Colditz GA, Hunter DJ, Rosner B, Hennekens C (1999) Prospective study of smoking, antioxidant intake and lung cancer in middle-aged women. Cancer Causes and Control, 10:475-482.

Steinmetz KA, Kushi LH, Bostick RM, Folsom AR, Potter JD (1994) Vegetables, fruits, and colon cancer in the Iowa women's health study. American Journal of Epidemiology, 139:1-15.

Steinmetz KA, Potter JD (1993) Food-group consumption and colon cancer in the Adelaide Case-Control study. I. Vegetables and fruit. International Journal of Cancer, 53:711-719.

Steinmetz KA, Potter JD (1996) Vegetables, fruit, and cancer prevention: a review. Journal of the American Dietetic Association, 96:1027-1039.
Steinmetz KA, Potter JD, Folsom AR (1993) Vegetables, fruit, and lung cancer in the Iowa Women's Health study. Cancer Research, 53:536-543.
Suzuki I, Hamada GS, Zamboni MM, de Biasi Cordeiro P, Watanabe S, Tsugane S (1994) Risk factors for lung cancer in Rio de Janeiro, Brazil: a case-control study. Lung Cancer, 11:179-190.
Swanson CA, Mao BL, Li JY et al. (1992) Dietary determinants of lung cancer risk: results from a case-control study in Yunnan Province, China. International Journal of Cancer, 50:876-880.

Tajima K, Tominaga S (1985) Dietary habits and gastro-intestinal cancers: a comparative case-control study of stomach and large intestinal cancers in Nagoya, Japan. Japanese Journal of Cancer Research, 76:705-716.

Takezaki T, Hirose K, Inoue M et al. (2001) Dietary factors and lung cancer risk in Japanese: with special reference to fish consumption and adenocarcinomas. British Journal of Cancer, 84:1199-1206.
Takezaki T, Shinoda M, Hatooka et al. (2000) Subsite-specific risk factors for hypopharyngeal and esophageal cancer (Japan). Cancer Causes and Control, 11:597-608.

Tavani A, Negri E, Francesch, La Vecchia C (1993) Risk factors for esophageal cancer in women in northern Italy. Cancer, 72:2531-2536.

Tavani A, Negri E, Franceschi S, La Vecchia C (1994) Risk factors for esophageal cancer in lifelong nonsmokers. Cancer Epidemiology, Biomarkers and Prevention, 3:387-392.

Tavani A, Negri E, Franceschi S, La Vecchia C (1996a) Tobacco and other risk factors for oesophageal cancer in alcohol non-drinkers. European Journal of Cancer Prevention, 5:313-318.

Tavani A, Negri E, La Vecchia C (1996b) Food and nutrient intake and risk of cataract. Annals of Epidemiology, 6:41-46.

Taylor CA, Hampl JS, Johnston CS (2000) Low intakes of vegetables and fruits, especially citrus fruits, lead to inadequate vitamin C intakes among adults. European Journal of Clinical Nutrition, 54:573-578.

Terry P, Giovanucci E, Michels KB et al. (2001) Fruit, vegetables, dietary fiber, and risk of colorectal cancer. Journal of the National Cancer Institute, 93:525-533.

Terry P, Nyren, O, Yuen J (1998) Protective effect of fruit and vegetables on stomach cancer in a cohort of Swedish twins. International Journal of Cancer, 76:35-37.

Thompson FE, Moler JE, Freedman LS, Cliffors CK, Stables GJ, Willet WC (1997) Register of dietary assessment calibration-validation studies: a status report. American Journal of Clinical Nutrition, 65:S1142-1147.

Thun MJ, Calle EE, Namboodiri MM et al. (1992) Risk factors for fatal colon cancer in a large prospective study. Journal of the National Cancer Institute, 84:1491-1500.

Tobias M (2001) The burden of disease and injury in New Zealand. New Zealand Ministry of Health, Wellington.
Todd S, Woodward M, Tunstall-Pedoe H, Bolton-Smith C (1999) Dietary antioxidant vitamins and fiber in the etiology of cardiovascular disease and allcauses mortality: results of the Scottish Heart Health study. American Journal of Epidemiology, 150:1073-1080.
Trichopoulos D, Ouranos G, Day N et al. (1985) Diet and cancer of the stomach: a case-control study in Greece. International Journal of Cancer, 36:291-297.

Turrini A, Saba A, Perrone D, Cialfa E, D’Amicis A (2001) Food consumption patterns in Italy: the INN-CA study 1994-1996. European Journal of Clinical Nutrition, 55:571-588.

Tuyns AJ, Kaaks R, Haelterman M (1988) Colorectal cancer and the consumption of foods: a case-control study in Belgium. Nutrition and Cancer, 11:189-204.

Tuyns AJ, Kaaks R, Haelterman M, Riboli E (1992) Diet and gastric cancer. A case-control study in Belgium. International Journal of Cancer, 51:1-6.
Tuyns AJ, Riboli E, Doornbos G, Pequignot G (1987) Diet and esophageal cancer in Calvados (France). Nutrition and Cancer, 9:81-92.
van Poppel G, Goldbohm RA (1995) Epidemiologic evidence for beta-carotene and cancer prevention. American Journal of Clinical Nutrition, 62: S1393-1402.
van Poppel G, Verhoeven DT, Verhagen H, Goldbohm RA (1999) Brassica vegetables and cancer prevention. Epidemiology and mechanisms. Advances in Experimental Medicine and Biology, 472:159-168.

Verhagen H, Poulson HE, Loft S, van Poppel G, Willems MI, van Bladeren PJ (1995) Reduction of oxidative DNA damage in humans by brussel sprouts. Carcinogenesis, 16:969-970.
Victora CG, Munoz N, Day NE, Barcelos LB, Peccin DA, Braga NM (1987) Hot beverages and oesophageal cancer in southern Brazil: a case-control study. International Journal of Cancer, 39:710-716.

Volatier J (1999) Enquête individuelle et nationale sur les consummations alimentaires. Editions TEC et DOC, Paris.
Vollset SE, Bjelke E (1983) Does consumption of fruit and vegetables protect against stroke? The Lancet, 2:742.

Voorrips LE, Goldbohm RA, van Poppel G, Sturmans F, Hermus RJ, van den Brandt PA (2000a) Vegetable and fruit consumption and risks of colon and rectal cancer in a prospective cohort study: the Netherlands Cohort Study on diet and cancer. American Journal of Epidemiology, 152:1081-1092.
Voorrips LE, Goldbohm RA, Verhoeven DT et al. (2000b) Vegetable and fruit consumption and lung cancer risk in the Netherlands Cohort study on diet and cancer. Cancer Causes and Control, 11:101-115.
Walker A, Adam F, Walker J, Walker BF (2002) Cancer of the oesophagus in Africans in sub-Saharan Africa: any hopes for its control? European Journal of Cancer Prevention, 11:413-418.
Wang LD, Hammond EC (1985) Lung cancer, fruit, green salad and vitamin pills. Chinese Medical Journal, 98:206-210.
Wang YP, Han XY, Su W et al. (1992) Esophageal cancer in Shanxi Province, People's Republic of China: a case-control study in high and moderate risk areas. Cancer Causes and Control, 3:107-113.
Ward MH, Lopez-Carrillo L (1999) Dietary factors and the risk of gastric cancer in Mexico. American Journal of Epidemiology, 149:925-932.
Watkins ML, Erickson JD, Thun MJ, Mulinare J, Heath CW Jr (2000) Multivitamin use and mortality in a large prospective study. American Journal of Epidemiology, 152:149-162.

Weinberg RA (1996) How cancer arises. Scientific American, 275:62-70.
West DW, Slattery, ML, Robinson, LM et al. (1989) Dietary intake and colon cancer: sex- and anatomic site-specific associations. American Journal of Epidemiology, 130:883-894.
Willett W (1998a) Nature of variation in diet. In: Nutritional epidemiology. Willett W, ed. Oxford University Press, Oxford.

Willett W, ed. (1998b) Nutritional epidemiology. Oxford University Press, Oxford.

Willett W (2001) Commentary: dietary diaries versus food frequency question-naires-a case of undigestible data. International Journal of Epidemiology, 30:317-319.
Wiseman M (2002) The World Cancer Research Fund Expert Report: the next steps. IARC Scientific Publications, 156:523-524.
Witztum JL (1994) The oxidation hypothesis of atherosclerosis. The Lancet, 344:793-795.
World Bank (2000) Classification of economies. World Bank, Washington DC. Available at http://www.worldbank.org/data/databytopic/class.htm\#Low_ income.

World Cancer Research Fund, American Institute for Cancer Research (1997) Food, nutrition and the prevention of cancer: a global perspective. American Institute for Cancer Research, Washington.
WHO (1985) Energy and protein requirements. (Technical Report Series No. 724). World Health Organization, Geneva.

WHO (1990) Diet, nutrition and the prevention of chronic diseases. (Technical Report Series No. 797). World Health Organization, Geneva.
WHO (2000) World health report 2000. Health Systems: Improving performance. World Health Organization, Geneva.
Wu K, Helzlsouer KJ, Alberg AJ, Comstock GW, Norkus EP, Hoffman SC (2000) A prospective study of plasma ascorbic acid concentrations and breast cancer (United States). Cancer Causes and Control, 11:279-283.
Wu-Williams AH, Dai XD, Blot W et al. (1990) Lung cancer among women in north-east China. British Journal of Cancer, 62:982-987.
Yochum LA, Folsom AR, Kushi LH (2000) Intake of antioxidant vitamins and risk of death from stroke in postmenopausal women. American Journal of Clinical Nutrition, 72:476-483.

Yochum L, Kushi, LH, Meyer K, Folsom AR (1999) Dietary flavonoid intake and risk of cardiovascular disease in postmenopausal women. American Journal of Epidemiology, 149:943-949.
Yokoyama T, Date C, Kokubo Y, Yoshiike N, Matsumura Y, Tanaka H (2000) Serum vitamin C concentration was inversely associated with subsequent 20year incidence of stroke in a Japanese rural community. The Shibata study. Stroke, 31:2287-2294.
Yong LC, Brown CC, Schatzkin A et al. (1997) Intake of vitamins E, C and A and risk of lung cancer. The NHANES I epidemiologic follow-up study. First National Health and Nutrition Examination Survey. American Journal of Epidemiology, 146:231-243.
You WC, Blot WJ, Chang YS et al. (1988) Diet and high risk of stomach cancer in Shandong, China. Cancer Research, 48:3518-3523.
Young TB, Wolf DA (1988) Case-control study of proximal and distal colon cancer and diet in Wisconsin. International Journal of Cancer, 42:167-175.
Yu MC, Garabrant DH, Peters JM, Mack TM (1988) Tobacco, alcohol, diet, occupation, and carcinoma of the esophagus. Cancer Research, 48: 3843-3848.

Yu Y, Taylor PR, Li JY et al. (1993) Retrospective cohort study of risk-factors for esophageal cancer in Linxian, People's Republic of China. Cancer Causes and Control, 4:195-202.
Zaridze D, Filipchenko V, Kustov V, Serdyuk V, Duffy S (1992) Diet and colorectal cancer: results of two case-control studies in Russia. European Journal of Cancer, 29A:112-115.
Zatonski WA, McMichael AJ, Powles JW (1998) Ecological study of reasons for sharp decline in mortality from ischaemic heart disease in Poland since 1991. British Medical Journal, 316:1047-1051.

Zheng W, Sellers TA, Doyle TJ, Kushi LH, Potter JD, Folsom AR (1995) Retinol, antioxidant vitamins, and cancers of the upper digestive tract in a prospective cohort study of postmenopausal women. American Journal of Epidemiology, 142:955-960.
Ziegler RG, Colavito EA, Hartge P et al. (1996a) Importance of alpha-carotene, beta-carotene, and other phytochemicals in the etiology of lung cancer. Journal of the National Cancer Institute, 88:612-615.
Ziegler RG, Mason TJ, Stemhagen, A et al. (1986) Carotenoid intake, vegetables, and the risk of lung cancer among white men in New Jersey. American Journal of Epidemiology, 123:1080-1093.

Ziegler RG, Mayne ST, Swanson CA (1996b) Nutrition and lung cancer. Cancer Causes and Control, 7:157-177.

# Chapter io 

# Physical inactivity 

Fiona C. Bull, Timothy P. Armstrong, Tracy Dixon, Sandra Ham, Andrea Neiman and Michael Pratt

## Summary

Physical inactivity is recognized as an important risk factor for multiple causes of death and chronic morbidity and disability.

Physical activity was chosen rather than physical fitness as the measure of exposure because it is through increases in the behaviour (physical activity) that health benefits accrue and improvements in cardiorespiratory fitness can be achieved. Moreover, there were insufficient data available worldwide to consider fitness as the exposure. Exposure was assessed as a trichotomous variable to avoid limiting the assessment of total burden to only that associated with the highest risk, namely the most inactive (a dichotomous approach). However due to a lack of data on physical inactivity, use of a more detailed (continuous) exposure variable was not possible nor was the use of a fourth category of "high activity". Therefore, our estimates of burden are likely to underestimate the total attributable burden to inactivity because of limitations with measures of exposure. Level 1 exposure (inactive) was defined as "doing no or very little physical activity at work, at home, for transport or in discretionary time". Level 2 exposure (insufficiently active) was defined as "doing some physical activity but less than 150 minutes of moderateintensity physical activity or 60 minutes of vigorous-intensity physical activity a week accumulated across work, home, transport or discretionary domains".

We found a wide range of survey instruments and methodologies have been used for collecting, analysing and reporting data on physical activity. Most data were available for discretionary-time activity, some data were found on occupational activity and little and no national data were available for transport- and domestic-related activity, respectively. A comprehensive literature search and contact with key agencies and known researchers uncovered over 50 data sets on physical inactivity in adult populations covering 43 countries across 13 subregions. ${ }^{1}$ However,
only 21 data sets covering 32 countries met our inclusion criteria. Hierarchical modelling techniques were used to predict discretionary-time activity using age, sex, geographic region and a measure of tertiary education. Linear regression was used to predict occupational activity and transport-related activity using two World Bank indicators (\% employed in agriculture and car ownership, respectively). We used these estimates to compute the level of total inactivity for 145 countries and aggregated these data to create estimates for 14 subregions.

The final global estimate for total inactivity (level 1 exposure) was $17.1 \%$ and this ranged from $10.3 \%$ in AFR-D to $24.8 \%$ in EUR-C. Across most but not all subregions females were slightly more inactive than males and younger adults were less inactive than older adults (range $9.6-46.8 \%$ across the 12 age-sex categories). The final global estimate for insufficient activity (level 2 exposure) was $40.6 \%$ and this ranged from $31.7 \%$ in AMR-D to $51.5 \%$ in WPR-A.

The independent causal relationship between physical inactivity and ischaemic heart disease, ischaemic stroke, type II diabetes, colon cancer and breast cancer is well established; we provided new estimates of the magnitude of risk associated with inactivity. A comprehensive search of literature from 1980 onwards identified well over 100 studies assessing the relationship between physical inactivity and the set of health outcomes that met our criteria. Also, several quantitative and qualitative reviews of the association between physical inactivity and ischaemic heart disease and stoke were found but there were no quantitative meta-analyses for breast cancer, colon cancer and type II diabetes. Most of the epidemiological studies meeting our inclusion criteria measured discretionary-time activity, some studies assessed occupational activity but only a few studies incorporated transport-related activity. No study included domestic-related physical activity. Given these data and differences between previous work and our definition of exposure, we completed a series of new meta-analyses for each health outcome. To address concerns regarding measurement error associated with physical activity, an adjustment factor was incorporated into the meta-analyses. All risk estimates were attenuated for ages 70 and over. There is emerging consensus on the protective effects of activity in regards to preventing falls, osteoarthritis and osteoporosis and impaired mental health but these disease end-points did not meet our inclusion criteria.

Globally physical inactivity accounted for $21.5 \%$ of ischaemic heart disease, $11 \%$ of ischaemic stroke, $14 \%$ of diabetes, $16 \%$ of colon cancer and $10 \%$ of breast cancer. The results show small differences between males and females, due in part to differences in level of exposure and to different distribution of events between men and women. In summary, physically inactive lifestyles accounted for $3.3 \%$ of deaths and 19 million disability-adjusted life years (DALYs) worldwide. There were small, nonsignificant differences in the attributable fractions across subregions. Due
to our conservative methods and a number of important limitations, our global estimates are likely to be an underestimate of the true burden attributable to inactive lifestyles.

## 1. Introduction

Physical inactivity is associated with many of the leading causes of death, chronic morbidity and disability. The apparent protective effect of being more active, and consequently less inactive, was identified first through studies of occupational activity over 50 years ago. Subsequent research has investigated different types, duration, frequency and intensity of activity in association with various cardiovascular, musculoskeletal and mental health outcomes. Today, there is a significant amount of literature quantifying and qualifying the role of physical inactivity as a risk factor and worldwide interest and efforts to increase levels of participation.

### 1.1 Choice of the exposure variable - physical activity or PHYSICAL FITNESS

There is substantial epidemiological evidence for the protective effects of both a physically active lifestyle and of various levels of physical fitness. Yet, to date, it is still not possible to determine whether one of these exposure variables is more important than the other (Blair et al. 2001). Thus, in theory, either variable could have been selected as the measure of exposure for this project. The final selection of physical activity over physical fitness was based on several reasons. Firstly, physical fitness is primarily determined by patterns of physical activity, particularly activity undertaken in recent weeks or months (Blair et al. 2001). Secondly, there is a genetic contribution to physical fitness and while of some importance, genetic makeup is likely to account for less variation than the lifestyle behaviour (Bouchard 2001). Thirdly, assessment of physical fitness in large samples of adult populations is rare across the majority of countries and infrequent in those countries in which it has been undertaken. Moreover, studies of fitness often exclude adults with certain chronic conditions. Fourthly, most national and international recommendations specify public health targets in terms of reaching thresholds of physical activity not levels of physical fitness.

It is acknowledged that one advantage of choosing physical fitness as the exposure variable would be the opportunity to use an objective, physical measure such as maximum oxygen uptake. This would be desirable for several reasons, including correspondence with many epidemiological studies assessing relative risk. Disadvantages to this approach, however, stem from the lack of nationally representative populationbased estimates, which outweigh the benefits. The difficulties associated with measuring behavioural risk factors are well known and the specific
challenges pertaining to measuring physical inactivity are discussed in detail in the following sections.

Therefore, considering the scientific support, the availability of data, and consistency with public health initiatives, physical activity was selected in preference to physical fitness as the measure of exposure. Given that protective benefits come from undertaking physical activity, from here on, the risk factor is specified as physical inactivity. It is, however, impossible to limit the discussion of both conceptual issues and the reporting of data to solely the "absence" of physical activity. Therefore the reader is forewarned that both physical activity and physical inactivity are discussed. In places the term "physical (in)activity" is used to refer to either activity or inactivity.

### 1.2 CONCEPTUAL FRAMEWORK AND DEFINITION OF PHYSICAL ACTIVITY

The definition of the exposure of physical inactivity was determined after consideration of a number of factors. These issues included: what types of activity across what domains would be included/excluded; what cutpoints between inactive and active would provide the greatest availability of data and opportunity for comparability across data sets and countries; and what cut-points would be consistent with established and emerging scientific evidence. Currently these issues are under considerable debate within the scientific community and there are no definitive answers. Thus, it is in advance of any consensus that a framework and workable definition of physical inactivity as an exposure is presented.

Traditionally, physical activity research has been interested in exercise defined as "planned, structured and repetitive bodily movement done to improve or maintain one or more components of physical fitness"(Caspersen and Stephens 1994). This interest focused attention towards certain types of exercise, mostly vigorous-intensity activities that were undertaken usually outside of work in recreational or discretionary time. In general terms it was acknowledged that some occupational or work-related activities could reach the threshold of vigorous-intensity and thus could potentially qualify as exercise and be beneficial to health. However, because technology and industrial development were causing a rapid decline in these occupations, particularly in developed countries, considerably less interest was invested in this direction. Moreover the opportunity to intervene was more promising outside the workplace. In essence, for those interested, particularly in developed countries, the focus was on exercise and the domain of "leisure-time". Many of the measurement instruments used today reflect this perspective.

A shift away from solely focusing on exercise started after the results from several large prospective cohort studies were published in the late 1980s and early 1990s. These studies were significant because they identified the protective effects of less intense physical activity (Blair and Jackson 2001; Blair et al. 1989, 1992, 2001). By 1996 these findings
were endorsed by several leading institutions and scientific organizations (Pate et al. 1995) and formed the primary focus of the U.S. Surgeon General's Report on Physical Activity and Health (U.S. Department of Health and Human Services 1996). Combined, the evidence and the widespread endorsement shifted the paradigm towards the benefits of moderate-intensity activity such that efforts more recently have been directed towards increasing the amount of moderate-intensity activity undertaken by all adults and children. Such activities include brisk walking ( $>3 \mathrm{mph}$ ), recreational cycling and swimming.

Some have viewed this shift as premature and raised concerns about the detrimental impact on the promotion of the known additional benefits that come from enhanced physical fitness. Furthermore, the exact nature and possible differences between the dose-response relationship between physical fitness and physical activity with various disease outcomes is still under intense debate (Bouchard 2001; Kesaniemi et al. 2001; Macera and Powell 2001). Meanwhile, the shift in focus has raised questions on the definition of "moderate-intensity", on what activities constituent moderate-intensity and on how to measure the prevalence of this broader range of activities within populations. Critically, it opened up the debate on whether activities undertaken in domains of life other than leisure-time can or should be considered beneficial and thus "counted".

The answers to these important questions are often populationspecific because different types of physical activities are undertaken between and within communities that differ across social, economic, geographical and religious aspects of life. Moreover, the importance of, or necessity for, physical activity in different domains of life differs between cultures. It is these differences and the complexity of measuring physical activity that have hindered the development of any international consensus on definitions and common instruments for assessing physical inactivity. Growing concern over these issues has led to several recent initiatives and considerable progress has been made toward developing a common measurement instrument for use in the future. The International Physical Activity Questionnaire (IPAQ) is a new instrument assessing total physical activity and sedentary behaviour (time spent sitting) developed by a group of international experts and tested in collaboration with researchers from 12 countries across six continents. IPAQ has demonstrated good-very good reliability and moderate criterion validity (Craig et al. 2003). Further details are available at http://www.ipaq.ki.se.

For the purpose of this project, it was acknowledged that physical (in)activity can occur across four domains: work, domestic, transport and discretionary time (see Figure 10.1). Each domain represents a sphere of daily life that is common to most populations regardless of culture or economic development and within each domain it is possible to be more or less active. The opportunity for, and the level of, physical (in)activity in each domain for any national population is dependent on

Figure 10.I Generic framework for four domains of physical (in)activity

| Work domain | Transport <br> domain | Domestic domain | Discretionary- <br> time domain |
| :---: | :---: | :---: | :---: |

Figure I 0.2 Relative importance of domains of physical (in)activity in two hypothetical countries


In country A, the work, domestic and transport domains contribute very little opportunity to be active. For example, this country may have well-established and complete coverage of water and electricity supply to homes and dwellings thus reducing the demands to be physically active in the domestic domain. Extensive car ownership with poor public transportation systems may have reduced the opportunity and need for activity in the transport domain. Work may have become more sedentary due to technological innovation. Thus the discretionary (leisure) domain represents the predominate domain where the greatest opportunity to be active exists. This may be typical of a developed country.


In country B, the work, domestic and transport domains contribute the most opportunity or most necessity to be active. For example, this may represent a country that has a high proportion of the adult population employed in agriculture or heavy industry both of which may require large amounts of vigorous physical exertion. Domestic activities may still involve carrying water and/or other food preparation techniques that require moderate- or vigorous-intensity physical activity. Transportation may be dominated by walking or cycling with minimal public transport alternatives and low car ownership. Relative to the previous domains, the discretionary domain may be much less important in terms of providing opportunity or need to be active for the population of this country. This may be typical of a developing country.
economic, technological, social, cultural and religious factors interacting at the individual, community and national level. Moreover, the relative importance of each domain, in terms of either representing an opportunity or necessity to be (in)active, will vary between and within countries over time. A generic model of this framework is shown in Figure 10.1 and two hypothetical countries ( a and b ) are shown in Figure 10.2, each with a brief description.

Using four domains as a framework for the definition of exposure is in contrast to many previous approaches to defining and measuring physical (in)activity, which made the task of searching for and interpreting global data on physical (in)activity more complex. National data on physical activity are not readily available in many parts of the world. There is even less data available across the four domains of work, domestic, transport and discretionary-time activity. Despite this limitation, the framework was selected because it presents the exposure variable in a way that has relevance to all countries around the globe. The framework
can be applied both now and in the foreseeable future, when many countries and large numbers of people are expected to experience major transitions in economic, social and health terms (Chockalingam 2000; Reddy and Yusuf 1998). The underpinning assumptions of our framework include: that physical activity can take place in different domains; that some of this activity is of sufficient intensity and duration to provide protective effects from certain diseases; and that the pattern of physical activity in each domain is very likely to vary across countries. The measurement issues related to each domain are discussed below.

### 1.3 Measurement issues

## INSTRUMENTS

There is currently no universal or even commonly used measure or instrument for physical (in)activity. This poses a serious limitation on efforts to compare levels of exposure across populations, a problem frequently cited in the literature and by others attempting international comparisons (Caspersen and Stephens 1994; IARC 2002).

Physical (in)activity is typically assessed using a series of questions as part of a self-administered or interviewer-administered questionnaire. The wording of the questions, the examples used (if any) and the response format can and do vary. A recent compendium of instruments illustrates the diversity of approaches, many of which are more suitable to the research environment rather than public health surveillance systems (Kriska and Caspersen 1997).

We undertook a review of instruments currently used to gather data at a national or large-scale population level and found that, in general, there are five different formats for questions assessing discretionary-time and work-related activities. These range from items asking for a yes/no response regarding participation in listed activities (format 1) to detailed responses about frequency, duration and type of activity performed (format 5). The formats for questions assessing physical (in)activity in these two domains are outlined in Table 10.1, and the limitations are briefly mentioned below. Few instruments were found that assessed transport-related and domestic activity.

## MEASURES OF DISCRETIONARY-TIME ACTIVITY

Instruments vary in both their intention and ability to capture details on different aspects of discretionary-time physical activities, namely: the specific type of activity (e.g. swimming, tennis, gardening, cycling); frequency (how many times in a specified time frame); duration (usually in minutes); and intensity (e.g. light, moderate, vigorous). Moreover, they use a variety of referent time frames (e.g. last week, last month, usual week, past year).

Format 1 and format 2 (Table 10.1) are similar in assessing frequency and duration of specific types of activity (either prompted or

Table 10.I Summary of five formats of questions assessing physical (in)activity

| Format | Question/response format | Example |
| :--- | :--- | :---: |
| Discretionary (leisure) time |  |  |
| I | Question presents a list of sports/activities. Response | Europe (Institute of |
|  | options are usually Yes/No. If yes, then frequency and | European Food |
| duration (minutes) are usually assessed. Reference time | Studies 200I) |  |
| frame vary from one week to 12 months | Brazil (Datafohla I997) |  |
|  |  | New Zealand (Hillary |
|  |  | Commission I998) |
|  |  | Canada (C. Craig, |
|  |  | unpublished data, |
|  |  | 200I) |
|  |  | USA (CDC I998) | undertaken but no lists are shown/provided. Selected examples may or may not have been provided. If yes, type, frequency and duration (minutes) are usually assessed. Intensity of activity may be recorded also

3

4 All activities over a specific time frame are recorded (e.g. last 24 hours or last 7 days). Responses include duration (minutes) and can be recorded by intensity and/or by domain

5 A series of questions assess participation in categories of activities usually defined by intensity (e.g. moderate-intensity or vigorous intensity). Examples usually provided. Response is usually Yes/No and if yes, frequency and duration (minutes) recorded. Walking as a specific activity can be asked either as a separate question specifying which types of walking to include (all walking, not at work, for transport) or walking is given as an example within the category of moderate-intensity activity

## Work-related

I Question asks respondents to indicate which description best describes their work "mostly" or "usually".
Response categories are often: I) primarily sitting/standing; 2) a lot of walking; 3) hard physical (sweat) labour

Argentina (Instituto
Gallup De La
Argentina 2001)
Chile (Jadue et al. 1999)

Egypt (Herman et al. 1995)

Estonia, Latvia, Lithuania (Pomerleau et al. 2000)

Ethiopia (Alemu and Lindtjorn 1995)

Australia (Armstrong et al. 2000)

South Africa (Steyn et al. I99I)

Table 10.I Summary of five formats of questions assessing physical (in)activity (continued)

| Format | Question/response format | Example |
| :---: | :---: | :---: |
| 2 | Question asks respondents to indicate how much time spent doing specific categories of activity (e.g. sitting/standing, walking, hard physical labour). Response scale may vary and can include either hours per day or proportion of time at work | Canada (C. Craig, unpublished data, 2001) <br> Europe (Institute of European Food Studies 2001) |
| 3 | Singular or multiple questions assess participation in categories of work activities defined by intensity (e.g. light or very light, moderate, heavy or very heavy). Examples of tasks or occupations may be provided for each category. Hours or portions of time may or may not be provided | China (North Carolina Population Center 200 Ib) <br> Estonia, Latvia, Lithuania (Pomerleau et al. 2000) |
| 4 | Questions assess frequency and sometimes duration of specific tasks undertaken within a reference time period (e.g. 2-12 months) are recorded. Respondents are prompted with a list of activities | Japan (lwai et al. 2000) |
| 5 | Occupational activity is assessed by classification of type of employment (e.g. brick layer, nurse) or occupation (e.g. professional, blue collar) | Not used by any study included in this chapter |

unprompted). Format 5 assesses categories of activities defined by intensity; the number and type of examples provided can vary. Additional criteria on minimum duration may also be specified. The diary or day-by-day recall methods are grouped together as format 4. However, format 3 is the most distinct set of instruments focusing on the level of activity, rather than its type or purpose. As a group they represent instruments that present categories or descriptions and use various response options, such as: pick the best description, or a Likert scale assessing frequency or duration, or scales such as 1-3 times/week, 4-6 times/week, 1-2 times/month.

Within each of the five formats, questions can be structured to assess level of activity in the discretionary-time domain only or they can address multiple domains. The latter, of course, makes it difficult to compare results across instruments. In the former, respondents are usually instructed to limit the activities they consider. Exactly how each instrument partially or totally includes/excludes different activities can however vary, particularly in regard to the following activities: walking; gardening; domestic/yard tasks. In addition, the majority of instruments focus on assessing participation in endurance/aerobic activities, probably due to the extensive literature supporting the links to health. However, activities that build muscular strength and increase flexibility are also important and warrant further attention.

Currently there is insufficient evidence to classify any of the instruments as right or wrong, but they are clearly different and there is good evidence that different instruments will produce different estimates of particular behaviours (Pratt et al. 1999). Much research has been undertaken testing different measures of physical activity to assess their reliability and validity (Jacobs et al. 1993). The more frequently used instruments (such as the Minnesota Activity Questionnaire, Seven Day Activity Recall) have been tested for reliability and validity in more studies across diverse populations (Kriska and Caspersen 1997). However, the questions used to gather population health data are often not established instruments and often lack formal testing.

## Measures of work-RELATED ACTIVITY

In recent times work-related physical activity has received less attention than discretionary-time activity. Therefore, there are sparse data on patterns of activity at work and much speculation on the accuracy of recall of work activities, especially potential discrepancies between recall and actual intensity and duration of activity (Jacobs et al. 1993; Leenders et al. 2001). In particular, there is concern about the extent of overreporting because individuals may over-estimate intensity and/or duration of work-based activities.

The most frequently used approach to assessment in this domain is to ask about three types of activities chosen because they are common in many occupations (formats 1, 2 and 3 in Table 10.1). More specific assessment of work activity can be obtained using a detailed recall (diary) or prompted recall with a list of activities (format 4). Historically, job occupation has been used as a proxy measure for work-related physical (in)activity (format 5) but as technology and work practice differs between countries and over time this is deemed to be the least favoured approach. To date, there is little evidence on the reliability and validity of most of these instruments or approaches.

## MEASURES OF TRANSPORT-RELATED AND DOMESTIC ACTIVITY

Measurement of transport-related and domestic physical activities is the least well-developed area. There are few specific questions or instruments available and some activities in these domains are captured as part of questions aimed more at the discretionary or work-related domains. This confusion makes quantifying the prevalence of physical (in)activity in the transport and domestic domains very difficult.

Most of the available data on transport-related activity are from sources within the discipline of transportation. Usually these data are limited to trip origin and destination, distance, duration and mode (e.g. cycling, walking). How well they capture walking and cycling as part of multi-modal travel can vary. Also transportation surveys rarely capture the perceived intensity of the activity although a computed measure could be obtained if distance and duration were both assessed. Data on
patterns of transport are most often presented as modal split with total number of trips as the denominator; other metrics include miles of travel. Data in these formats are not readily integrated or compared with data from population health surveys on physical activity. Moreover, it is usually only in research projects that data on transport-related activities and physical activities in other domains are collected on the same individuals.

Assessment of domestic activities as a distinct domain is uncommon. If these activities are considered at all, they are usually included as examples within the structure of other questions. For example, washing floors or vacuuming can be given as examples of moderately-intense activities. However, within the field of diet and nutrition some researchers very carefully assess energy intake and energy expenditure. But for these purposes they often assess total energy expenditure using doubly-labelled water $\left(\mathrm{H}_{2} \mathrm{O}_{2}\right)$, direct observation or very detailed diary methods. There is currently no suitable, valid and reliable instrument for use in assessing activity in the domestic domain. It is, therefore, evident that further instrument development is needed to advance the assessment of physical (in)activity in both transport and domestic domains.

## Further issues related to measurement methodology

In addition to the use of a variety of instruments, the method of data collection can vary. Self-report instruments continue to be the most widely used method for assessing physical activity in adults (Sallis and Saelens 2000), particularly in developed countries. In these countries widespread coverage of telecommunications has led to increased use of telephone-based surveillance systems (e.g. the Behavioral Risk Factor Surveillance System in the United States of America). In contrast, household surveys remain more common in other regions (e.g. Central and South America, Asia, Africa). There is no scientific evidence to support one methodology over the other in terms of accuracy of data on physical activity. But it is possible that each method may obtain different estimates of the behaviour. If the magnitude and direction of this difference were known for different populations it would be possible to include some statistical adjustment. However, in the absence of this information the inclusion of data collected using different interview methods will be a limitation to our estimates.

Patterns of physical activity are known to vary across different climatic seasons (Pratt et al. 1999). Ideally national estimates would be based on data collected across all seasons (a 12 -month period) or data collection would be limited to a specified season. Cross-country and global comparisons could then be conducted either using only data collected in comparable seasons or only data collected over all seasons. Needless to say, with the overall dearth of data on physical (in)activity it was not possible to impose seasonality criteria within this project and this feature of the data will be another limitation to our final estimates.

### 1.4 Quantifying levels of exposure

Once data on physical (in)activity have been collected using any one of the approaches described above, the data can be treated as either a continuous or categorical variable. There are examples of both approaches with national data from around the world. Both approaches were considered for this project but due to the availability and comparability of data we chose to treat physical inactivity as a trichotomous categorical variable. A brief description of both approaches is provided below.

## Treating physical (in)activity as a continuous variable

Minutes of activity can be presented as a continuous variable either as mean minutes per specified time frame (e.g. per week or per day) or it can be used to derive an estimate of energy expenditure expressed as metabolic cost (metabolic equivalent, MET) or kilocalories/kilojoules ( $\mathrm{kcal} / \mathrm{kJ})$. One MET represents the metabolic rate of an individual at rest and is set as $3.5 \mathrm{ml} \mathrm{O} 2 / \mathrm{kg}$ per minute or approximately $1 \mathrm{kcal} / \mathrm{kg}$ per hour (Kriska and Caspersen 1997). With this approach, analysts use available lists of the energy requirements of specific activities (Ainsworth et al. 2000) or standard MET values for categories of activity. For example, the following MET values are frequently, but not consistently, applied to the following categories: vigorous-intensity activities $=>6$ METs; moderate-intensity activities $=3-6 \mathrm{METs}$ ( $=4.5$ ); walking $=3$ METs (Ainsworth et al. 2000). Energy requirements of a large number of tasks across all four domains are available (Ainsworth et al. 2000).

The continuous measures were not selected as the basis for the definition of the exposure, nor were data in these formats available for use in any other way for two reasons. Firstly, these estimates were derived from different instruments, and each analysis used different criteria to compute their estimates. Thus, large variability prohibited comparability despite the seemingly similar continuous measures of risk. Secondly, data obtained as mean minutes, mean kcals or mean MET minutes of activity could not be used to compute the distribution of exposure because the relationship between these values and the pattern of inactivity in a population is not known. In contrast, the relationship between mean body mass index (BMI) and the prevalence of obesity in populations has been explored.

## Treating physical (in)activity as a categorical variable

A current trend in analysing population data on physical (in)activity is to collapse the continuous data on minutes of activity into a categorical variable and report the prevalence estimates of each category. This is undertaken for each specific type of activity (e.g. walking, sports, lifting loads, climbing stairs) or more usually, the total amount of time spent doing physical activities assigned to different categories (e.g. total minutes of moderate-intensity or vigorous-intensity activity or heavy
physical labour). The total of all activity can be calculated by summing all minutes of each activity or each category. However, close inspection reveals that these general approaches can include or exclude some specific physical activities or some domains depending on the investigators' purview. For example, gardening activities may be asked about but not included in calculations of activity (Armstrong et al. 2000) or workrelated activity may be assessed but excluded or else reported separately from discretionary-time activity (Pomerleau et al. 2000). These specific analytic details necessitate very careful attention before comparisons between data sets can be attempted.

It has also become quite common among those dealing with national or regional surveillance data to present categorical data and to select the categories that correspond with various national public health guidelines or goals. A recent development has been the reporting of data as a measure of "recommended" or "sufficient" levels of activity in accordance with the U.S. Surgeon Generals' Report on Physical Activity and Health (U.S. Department of Health and Human Services 1996). This recommendation states that " 30 minutes of moderate-intensity activity on most, if not all days of the week" is recommended for all adults. It is often, but not always interpreted and analysed as 150 minutes of moderate-intensity activity over at least five sessions or day/week. But there is still sufficient variation between the ways in which this recommendation is interpreted to make seemingly simple comparisons somewhat difficult.

### 1.5 Correspondence between measures of exposure used in population surveys and epidemiological research

The majority of epidemiological studies have explored the relationship between activity and disease outcomes by dividing level of activity into two or more groups. This is true whether the exposure was physical activity or physical fitness. The nature of the dose-response relationship has been explored using dichotomous categories (Morris et al. 1980), tertiles (Bijnen et al. 1998), quartiles (Folsom et al. 1985) and quintiles (Singh et al. 1998). Although continuous data are usually preferred because this maximizes the opportunity to detect associations, use of categorical data is warranted when the association may not be linear. Indeed, the nature of the curvilinear or linear relationship between physical activity and disease outcomes is a major focus of current research efforts (Blair et al. 2001; Kesaniemi et al. 2001; Williams 2001). Three possible relationships are illustrated in Figure 10.3. One contemporary view supports the presence of a threshold as shown by curve B in Figure 10.3, particularly for cardiovascular end-points. Curve B indicates that the greatest reduction in risk comes from increases at the lower levels of activity. Experts involved in a recent evidence-based symposium concluded that there is "an inverse and generally linear relationship for the rates of all-cause mortality, total cardiovascular disease, and

Figure 10.3 A schema of three possible relationships between physical inactivity and disease end-points


Increasing level of physical activity
coronary heart disease incidence and mortality and for the incidence of type 2 diabetes mellitus", a position reflected in curve C (Kesaniemi et al. 2001).

While the debate on the exact shape of the relationship is likely to continue, the use of categorical data may be justified given the known difficulties (i.e. error) associated with both the measurement instruments used to assess physical (in)activity and respondent recall of the behaviour. Under these circumstances continuous data may present an artificial level of accuracy and specificity.

It is worth noting one additional difficulty that obscures the level of correspondence between measures and thus the accurate application of relative risk results from epidemiological studies to prevalence estimates in order to calculate population attributable risk and the burden of disease. Namely, the lack of detailed descriptions of the absolute values of activity represented by the investigators' chosen categories. For instance, it is difficult to compare the level of activity in the lowest tertile of activity from study A with study B unless the activity level is defined and reported. Likewise it is almost impossible to make accurate interpretations of tertiles and quartiles without appropriate descriptive statistics for each group. This omission necessitates qualitative judgements when comparing or pooling results across studies and may obscure the strength and nature of the disease-risk factor association.

Notwithstanding, these are important limitations; currently the greatest concordance between measures of physical activity used in epidemiological studies and population surveillance is found at the lower and
upper ends of the physical activity continuum. Application of the relative risk results for the least active group (be it tertile, quartile or quintile) to prevalence estimates for lowest level of activity in whole populations is possible. There is, however, less certainty surrounding the agreement between measures of low-mid levels and mid-upper levels of exposure.

Measurement error associated with assessing exposure to physical inactivity is a major limitation to this project. We addressed this problem in several ways. Our treatment of the exposure as a trichotomous variable avoids presenting a misleading level of accuracy. The definition of the categories corresponds with our greatest level of confidence in the accuracy of our data, particularly at the lower end. Our definitions and estimates of exposure within each domain accommodate the known or suspected direction of measurement error. Finally, as we describe in our review of epidemiology, our pooled meta-analyses included an adjustment for measurement error based on available evidence of reliability of the instruments used in each study.

### 1.6 Definition of exposure

In balancing the need for a conceptual framework with global relevance, concordance with the epidemiological evidence, current public health recommendations, as well as the limited availability of national data, the following definitions of a three level exposure of physical inactivity were developed:

## Exposed

Level 1 Inactive:
defined as doing no or very little physical activity at work, at home, for transport or during discretionary time.

Level 2 Insufficiently active:
defined as doing some physical activity but less than 150 minutes of moderate-intensity physical activity or 60 minutes of vigorous-intensity physical activity a week accumulated across work, home, transport or discretionary domains.

## Unexposed

Level 3 Sufficiently active:
defined as at least 150 minutes of moderate-intensity physical activity or 60 minutes of vigorous-intensity physical activity a week accumulated across work, home, transport or discretionary domains, which approximately corresponds to current recommendations in many countries.

In the work domain, inactive was defined as not doing any "hard" or "vigorous" physical activities at work. Where needed, we classified those that reported "mostly walking" or "mostly standing" or any equivalent category, as inactive. The effect of our definition is to potentially inflate
the proportion classified as inactive at work because some adults may perform these tasks during their work in combinations of duration and intensity sufficient to gain some health benefits. However this is perhaps counterbalanced by at least some likely misclassification of adults as active due to "heavy" activity. In the absence of data with sufficient detail and no prior evidence on which to base our classification, our definition reflects the limitations of measurement in this domain.

Inactive in the transport domain was defined as no walking or cycling trips. Therefore, a single walk or cycle trip was sufficient for an adult to be classified as active in this domain. With few data on trips per person and no data on trip intensity or duration we accepted all walking and cycling trips as activity that could provide some health benefits. In contrast to the work domain this definition is lenient and the likely effect is to overestimate transport-related activity. However, unlike walking at work, which may be performed in short bouts and undertaken intermittently, transport-related activity may be of longer duration and undertaken at moderate-intensity. Inactive in the domestic domain was defined as no moderate- or vigorous-intensity domestic-related activity. However, with no national data worldwide we were unable to include this domain of activity in our assessment.

Insufficient activity was assessed by distinguishing between those adults who met the threshold of at least 150 minutes of moderateintensity physical activity accumulated from one or more domains from those adults who did not. Given the lack of national data using the same or equivalent definition, we used the limited available data to develop an algorithm to estimate the prevalence of this level of exposure in each country. Full details of our methods for assessing inactive (level 1) and insufficiently active (level 2) are described in section 2.

### 1.7 Age groups included-Adults but not children

Very few data were available on levels of (in)activity in children and adolescents. Moreover, the literature on children also includes a diversity of instruments and measurement methodologies (Sallis and Saelens 2000). Therefore, due to the magnitude of the task and the project timelines, the scope of this project was limited to quantifying exposure in the adult population only. Furthermore, there is limited epidemiological evidence demonstrating the level of current risk associated with inactivity and the selected health outcomes in populations aged $<18$ years.

### 1.8 Theoretical-minimum-Risk exposure distribution

For physical inactivity the theoretical-minimum-risk exposure distribution could be set as equal to the estimated proportion of the total population that would be physically unable to meet the basic requirement of "at least some activity" in at least one domain. Data from Australia $(2 \%)$ and the United States (4-6\%) indicate that total disability (as defined by each country) is well below $10 \%$ of the population but it is
unknown as to what proportion of this sub-population would be unable to do at least some physical activity.

Conceptually the theoretical-minimum-risk exposure is that level of activity which could theoretically occur if we could remove all individual, social and environmental causes of inactivity. Under such a scenario it is possible to consider a minimum that reflects all but the congenital causes of inactivity. As these are few, if any, and are likely to affect less that $1 \%$ of the whole population, we have chosen to set a theoretical minima as zero-a value that is comparable to other risk factors.

## 2. Estimating the prevalence of exposure

There were large gaps in the availability of nationally representative data on physical inactivity within some subregions and no data for several subregions. This was true across all four domains of activity included in the definition of exposure and both levels of exposure. To address this problem multivariate and linear regression analyses were conducted to create predicted values for missing data.

### 2.1 Search strategy

An extensive search was undertaken to identify studies reporting prevalence of inactivity for male and female adults (aged $\geq 15$ years) worldwide. A particular effort was made to identify data in developing regions to avoid the need to derive estimates for these countries solely on data from developed countries.

Data were sought on physical inactivity across four domains, namely, discretionary time, work-related, transport-related and domestic. The initial search identified several studies in which population-based estimates of physical (in)activity were reported as a risk factor for specific diseases. Based upon these findings, a second search was conducted to include as keywords those diseases where physical inactivity has been shown to have a relationship, in both prevention and as a risk factor. For example, diabetes, cardiovascular disease, obesity and cancer were keywords used in the second search. Papers reporting data on physical inactivity in one or more domains were accepted for consideration. However, our search most frequently identified estimates for discretionary-time activity and rarely data on occupational, transportation or domestic (household) physical activity. Therefore separate searches for occupational and/or work, transportation and domestic domains were performed. For all searches, country names were included when global, international and/or world keywords were not used.

Publications reporting data collected between 1996 and 2000 were sought because the primary goal was to estimate prevalence of inactivity in the year 2000 .

## ELECTRONIC LITERATURE SEARCH

Medline, HealthStar and Chronic Disease Prevention searches were systematically conducted for studies published between 1996 and 2001 and limited to English (both United Kingdom and American English spellings) and Spanish languages. Additionally, multiple queries for each keyword were used. For example diabetes AND physical activity AND country X.

The following keywords were used:

- exercise, physical fitness, physical exercise, physical (in)activity, sedentary, energy expenditure;
- chronic disease prevention, diabetes, cancer, cardiovascular disease, hypertension, obesity;
- international, global, world, developing countries, country names;
- monograph, statistics, survey(s), prevalence;
- transportation (and physical activity/energy expenditure);
- occupation (and physical activity/energy expenditure and activity); and
- domestic (and physical activity/energy expenditure and activity).


## Electronic Listserve

A request for assistance in locating "gray" or fugitive literature (e.g. government and non-government agency reports) as well as unpublished data was posted on a cardiovascular listserve (procor-dialogue@ healthnet.org).

## Government and non-government agencies

World Health Organization (WHO) Headquarters, WHO Regional Offices, WHO collaborating centres and partner organizations worldwide were contacted to assist in identifying authors, researchers, organizations and institutions in obtaining national data on prevalence of physical activity, either at the country, regional or municipal level.

Authors of relevant papers, ReLEVANT ACADEMIC INSTITUTIONS AND EXPERTS IN THE FIELD
Direct contacts were made with national governments, private and public institutions and individual authors to enquire whether unpublished or published data were available in English or other languages.

### 2.2 CRITERIA FOR CONSIDERING SOURCES AND STUDY INCLUSION

 All identified studies were reviewed and included if they fulfilled the following criteria.Time frame-data were obtained between 1996 and 2001. However, if there were no other data available for the country/region, studies reporting data collected after 1990 were considered.

Sample-studies reporting data from large, randomly selected, nationally representative samples with a wide age range were preferred. However, given this was not available in the majority of countries, studies with smaller samples, regional representative samples or narrow age ranges were considered.

Measure of exposure-the proportion of the population exposed (inactive) in one or more domains (discretionary time, occupational, domestic, transport-related) was required. In cases where data were presented as mean minutes, METs, total energy expenditure (kcals) or in any other way, the authors were contacted and asked to provide the data expressed as a proportion of the sample population. Where possible authors were asked to use our definition of exposure (overall inactivity) and/or the definition of inactivity relevant to each domain. If these data were obtained, they were considered for inclusion.

Measurement instrument—studies reported (or later provided) the instrument used to assess physical activity. As no standard measure of physical activity exists each instrument or set of questions was reviewed in this work. Those studies using previously published instruments assessing physical activity in one or more domain were accepted. Studies using questions unknown to the authors were assessed for face validity. A study was included only if it was deemed that the question(s) would provide reasonably comparable estimates.

Data collection-studies were included if data were collected by telephone survey, face-to face interview or during a physical examination.

### 2.3 Methods for selecting estimates where more than one DATA SOURCE EXISTED

The following criteria were used to select data sources in cases where more than one source was available for a country or group of countries:

- the most recent data source-in some cases this meant the data were not yet in the public domain and specific analyses were undertaken for this project;
- nationally representative data or the study that allowed the best approximation of a national sample considering the studies representativeness and coverage of the desirable age range;
- a single study with common methods and representative population data from many countries was preferred over equal quality studies from individual countries; and
- the study providing the most comparable measure of physical inactivity as defined in this study.


### 2.4 Description of studies, including METHODOLOGICAL QUALITIES

Overall our search identified over 50 studies containing measures of physical (in)activity in one or more domain. These data covered 43 countries and 13 subregions. However, not all studies met the inclusion criteria and for some studies we were unable to obtain data in the desired format within the study time frame. Thus, our final analyses were conducted using 17 data sets for discretionary-time activity, two data sets for work-related activity and five data sets for transport-related physical activity representing 34 countries across 10 subregions. Figure 10.4 illustrates data coverage by country showing those countries that had data that met our inclusion criteria. Table 10.2 shows the proportion of the adult population for which we had data for each subregion (calculated by 12 age and sex categories).

Figure 10.4 Exposure data coverage by country ${ }^{\text {a }}$


Table I0.2 Proportion of the adult population for which there were data on physical inactivity, by subregion

| Subregion | Population in subregion (000s) | \% population within <br> subregion covered with data |
| :---: | :---: | :---: |
| AFR D | 164917 | 0 |
| E | 192766 | 15 |
| AMR A | 255419 | 96 |
| B | 297674 | 48 |
| D | 44658 | 34 |
| EMR | 86854 | 7 |
|  | D | 204038 |
| EUR | 342220 | 21 |
|  | B | 161213 |

Table 10.3 provides an overview of the studies reviewed with details on the data source, sample characteristics and the domains of physical activity assessed by country and subregion. Overall, over $90 \%$ of the data sources comprised nationally representative samples and all data were collected using a random sampling methodology. Sample sizes ranged from 226 in Ethiopia (Alemu and Lindtjorn 1995) to 16000 in China (Du et al. 2002) with the majority of studies including between 2000 and 4000 subjects. Most data were collected as part of either national surveys or ongoing monitoring systems or as part of a research project. There is one example of data from a commercial marketing company (Instituto Gallup De La Argentina 2001) and one example of a large multi-country study (Martinez-Gonzalez et al. 2000). The foci of research studies were most often cardiovascular disease, hypertension or diabetes. All data were collected after 1990.

The majority of data were collected using a questionnaire assessing multiple lifestyle risk factors although several studies collected data on physical activity only (e.g. Australia). Only a few studies assessed occupational activity as a separate domain and typically these questions assessed prevalence of three types of occupational activity, for example "mostly walking", "mostly standing" or "mostly heavy labour". Very few data were found on transport-related activity. Specific items on trans-port-related activity were included in only one of the health surveys (China) (North Carolina Population Center 2001b) although several other countries included walking or cycling for transport within the context of their physical activity questions. No studies were found reporting national data on physical activity in the domestic domain
Table I0.3 Summary of sources of exposure data, by subregion

Eastern Mediterranean Region

Table 10.3 Summary of sources of exposure data, by subregion (continued)

| Subregion | Country | \% ${ }^{\text {a }}$ | Author | Year of data | Sample size | Sampling frame | Response rate | Data obtained by domain ${ }^{\text {b }}$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  | L | 0 | M |
| European Region |  |  |  |  |  |  |  |  |  |  |
| EUR-C | Estonia | 1 | Pomerleau et al. (2000) | Summer 1997 | 2018 | Simple random stratified sample from register | 67\% | $\checkmark$ | $\times$ |  |
|  | Latvia | 1 | Pomerleau et al. (2000) | Summer 1997 | 2303 | two stage sampling | 78\% | $\checkmark$ | $\times$ |  |
|  | Lithuania | 2 | Pomerleau et al. (2000) | Summer 1997 | 2140 | Simple random sample from register | 73\% | $\checkmark$ | $\times$ |  |
|  | Russian Federation | 63 | Cockerham and Sneed (200I) | 1995 | 8402 | Random national sample | Not available | $\checkmark$ |  |  |
| South-East Asia Region |  |  |  |  |  |  |  |  |  |  |
| SEAR-D | Bangladesh | 10 | Hypertension Study Group (2001) | Dec 1999- <br> Feb 2000 | 723 | Random multistage cluster sampling within selected site | Not reported |  |  | $\times$ |
|  | India | 82 | Hypertension Study Group (2001) | Dec 1999- <br> Feb 2000 | 723 | Random multistage cluster sampling within selected site | Not reported |  |  | $\times$ |
|  | India | 82 | Singh et al. (1998) | 1997 | $\begin{aligned} & \text { I } 769 \text { rural } \\ & \text { I } 806 \text { urban } \end{aligned}$ | Random multistage cluster sampling within selected sites | 81\% rural; 91\% urban |  |  | $\times$ |
|  | India | 82 | Gupta et al. (1994) | 1993 | 1150 | Random sampling | Not reported |  |  | $\times$ |

Western Pacific Region

| WPR-A | Australia | 12 | Armstrong et al. (2000) | 1999-2000 | 3841 | Random sample of household from white pages | Not reported | $\checkmark$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Australia | 12 | Transport-Travel Demand Management (2000) | 1997-1998 | $\ldots$ | Randomly selected households | Not reported |  | $\checkmark$ |
|  | Australia | 12 | Bull et al. (2000) | 1999-2000 | 3178 | Stratified random sampling of household from white pages | 46\% |  | $\checkmark$ |
|  | Japan | 84 | Iwai et al. (2000) | 1992-1994 | 1893 | Stratified random sample from registry of 9 rural and 4 urban selected communities | 60\% | $\checkmark$ |  |
|  | New <br> Zealand | 2 | Hillary Commission (1998) | 1997 | 5470 | Multi-staged random cluster sampling of households | 68\% | $\checkmark$ |  |
| WPR-B | China | 85 | Ham (2001c) | 1997 | 16000 | Multi-staged random cluster sampling of household of nine Provinces | Not available | $\checkmark \checkmark$ | $\checkmark$ |
|  | China | 85 | Li Xiang-ru et al. (2001) | 1996 | 8000 | Multi-staged random cluster sampling | Not available | $\checkmark$ |  |
|  | $\checkmark$, data included; X , data not included; L, leisure (discretionary) time; O, occupational; T, transport; D, domestic; M, multiple domains included in one/two question(s). |  |  |  |  |  |  |  |  |
| $D$ un | Data included means data were used as presented in publication and/or data were obtained in format for comparison; data not included means unavailable in comparable format. |  |  |  |  |  |  |  |  |
|  | Data used only in post hoc assessment of final predicted estimates of physical inactivity. |  |  |  |  |  |  |  |  |
| C |  Netherlands (1010), Portugal (1007), Spain (1000), Sweden (1001), United Kingdom (1490). |  |  |  |  |  |  |  |  |
| e Austria, Denmark, France, Germany, Italy, Netherlands, Sweden, Switzerland. |  |  |  |  |  |  |  |  |  |

alone. Like transport, domestic activities were sometimes included within the context of broader questions, particularly in instruments developed for use in developing countries. Details of the questionnaire, the domains and definition of inactivity for each data set included in the analyses are provided in Tables 10.4-10.6.

Below are brief summaries of the data found and included, described by geographical region.

## AFrica

Finding data on physical activity from countries in the African Region was most difficult.

Several studies were found for various sub-populations in South Africa (Levitt et al. 1993, 1999; Sparling et al. 1994; Steyn et al. 1985, 1991). But these were regional studies with modest sample sizes. In the absence of a single study with a nationally representative sample, published data from the larger, more recent risk factor study in a black population (known as the Black Risk Study, BRISK) (Steyn et al. 1991) were selected to represent the black South African population. Estimates of inactivity among the white South African population were imputed from American data after a qualitative assessment (Lambert 2001) of the similarities in urban lifestyles. Given national estimates from the United States include white and non-white populations, the extrapolation to white south Africans may slightly underestimate levels of activity. This approach adds greater uncertainty to the estimates for this region and this is addressed in later sections.

Data from two small studies, one in rural Ethiopia (Alemu and Lindtjorn 1995) and the other in the United Republic of Tanzania (Aspray et al. 2000), provided a useful post hoc validity check on our final estimates of discretionary-time inactivity. Both studies assessed physical activity across multiple domains that could not be disaggregated.

No data were found to represent those countries in the AFR-D subregion. Two studies reported data from Nigeria (Ezenwaka et al. 1997; Forrest et al. 2001) and one study submitted for publication from Cameroon (Sobngwi et al. 2002) were identified but all three studies reported physical activity using different summary scores and/or units. At the time of our analyses the results from reanalysis were not available.

## Americas

Data on discretionary-time physical inactivity and work-related activity from the United States and Canada were identified for the AMR-A subregion. Both countries have established surveillance systems collecting nationally representative data on patterns of physical activity (C. Craig, unpublished data, 2001; S. Ham, unpublished data, 2001a; Macera et al. 2001). Data on transport-related activity for the United States were
available from the National Personal Transportation Survey (Federal Highway Administration Research and Technical Support Center 1997) and data for Canada were found in a cross national comparison report (Pucher and Dijkstra 2000). No data were found for Cuba, the third country in this subregion.

For the AMR-B subregion national data on physical activity were found for Argentina, Brazil and Chile. In Brazil multiple data sets were found for different cities including Porto Alegre (Duncan et al. 1993) and São Paulo (Andrade et al. 2001). However data from the most recent, largest, most representative study, with sampling from 98 cities across Brazil was selected (Datafohla 1997). Data for Argentina were obtained from a recent national Gallup Poll that included one item on physical activity (Instituto Gallup De La Argentina 2001). Data for Chile were obtained from a recent published study conducted in a metropolitan area of Valparaiso. (Jadue et al. 1999) Two studies from Peru with data on work and discretionary-time activity were identified for AMRD (E. Jacoby, unpublished data, 2000). Only the data on discretionarytime activity were available for inclusion.

## Eastern Mediterranean

Only two studies were found reporting data for physical activity EMRB and EMR-D and neither study provided national estimates. One study from Saudi Arabia reported estimates for only discretionary-time inactivity in male adults (al-Refaee and al-Hazzaa 2001). While these data are limited in both domain and are for males only, the questions were deemed comparable and due to the lack of any alternative data they were included for EMR-B. The second study reported data on physical activity from a large sample of adults in a region of Egypt (the city of Cairo and surrounding villages) (Herman et al. 1995). The measure of inactivity assessed discretionary-time, transport and work-related activity in combination. Because it was not possible to disaggregate these components these data were used only as a post hoc check of the final predictive model.

## EUROPE

Although several individual European countries have collected data on various domains of physical activity in recent years (Hassmen et al. 2000; MacAuley et al. 1996, 1998) we selected a recent, large, multi-country study ( $n=15239$ ) assessing both discretionary-time and work-related physical activities (Institute of European Food Studies 2001). The study was selected for inclusion because it collected national representative samples from 15 countries and data were collected within a defined period of time using the same measurement instrument. The Physical and Nutrition European Union data set (known as PAN EU) provided $87 \%$ coverage of EUR-A; however these data were collected by commercial marketing companies in each country and there were no available data
on the response fraction. This concern increases uncertainty around these estimates but the data were deemed more comparable than using a larger number of country-specific estimates, collected using a variety of instruments and providing less coverage of the EUR subregions.

Data on transport-related activity are likely to exist for many European countries at either a national or regional level, often in departments of transportation or planning databases. However, these data remain difficult to locate and require translation. In lieu of obtaining these data directly we used a recent report that provided national estimates of walking and cycling for several European countries and Canada (Pucher and Dijkstra 2000).

No data were found for EUR-B. Data from the MONICA project that included two sites in Poland were considered, but were not nationally representative and were therefore excluded.

One study reported data from three countries in EUR-C, namely Estonia, Latvia and Lithuania (Pomerleau et al. 2000). Two sources were identified for data on physical activity in the Russian Federation. An unpublished research study investigated patterns of activity in an adult population in Moscow (Zabina et al. 2002) and a publication using data from the Russian Longitudinal Monitoring Survey (RLMS) (North Carolina Population Center 2001a). We were unable to complete analyses of the most recent RLMS data (1998/2000) and therefore selected the published estimates from survey six in 1995 and we were provided with additional unpublished analyses from this data set (Cockerham and Sneed 2001). Combined the two sources of data provide $63 \%$ coverage of EUR-C. No data were found on transport-related activity for EUR-C.

## South-East Asia

Our search found only a few studies with data on patterns of physical activity in India (Gupta et al. 1994; Singh et al. 1998) and one study with data from Bangladesh (Hypertension Study Group 2001). Other contacts through international agencies and electronic listserve identified several studies with unpublished data although none of these included nationally representative samples (Misra et al. 2001; K. Reddy, unpublished data, 2001). One study investigated an urban slum population (A. Misra, unpublished data, 2001) and another a population selected from an industrial workforce (K. Reddy, unpublished data, 2001). Both studies provided us with additional unpublished analyses, and we attempted to create a weighted (by urban and rural) national estimate. However, the results from these studies showed no clear pattern and we had low confidence in our estimate. Therefore, these data were not used in the final analyses. The small, published study with a population of older adults from rural and urban towns in Bangladesh and India was considered for SEAR-D. However, on close inspection these data were considered insufficient for inclusion.

## Western Pacific

Australia and New Zealand both have established health and population surveillance systems and from time to time conduct national surveys specifically aimed at assessing patterns of physical activity. Data from these specific surveys were selected for our analyses (Armstrong et al. 2000; Hillary Commission 1998). No national data on transportation activity were found for this region, thus we substituted data on cycling and walking for transport from a combination of two reports from Australia; one included a large survey with a representative sample of the population in Western Australia (Bull et al. 2000) and the other included data from a transportation project also undertaken in Western Australia (Travel Demand Management 1999).

Two studies published in English were found reporting data on dis-cretionary-time physical activity in Japan (Arai and Hisamichi 1998; Kono et al. 1999) and translation of data from another source was also provided (N. Murase, unpublished data, 2001). However, we chose to include data from the larger Japanese Lifestyle Monitoring Study (Iwai et al. 2000) even though only discretionary-time physical activity was assessed in a middle-aged adult population (40-60 years). A small study in Singapore with data from women only was also found but after considering the instrument and definition of inactivity it was excluded.

Subregion WPR-B is represented by data found from China. Nine published studies were identified (Bell 2001; Hong et al. 1994; Hu et al. 1997, 2002; Matthews et al. 2001; Paeratakul et al. 1998; Pan et al. 1997; Yu and Nissinen 2000; Yu et al. 2000) but excluded in preference for recent data from the China Health and Nutrition Survey (CHNS) (North Carolina Population Center 2001b). Micro data were available online (http://www.cpc.unc.edu/) and we conducted additional analyses on physical activity across different domains (Ham 2001c). Additional data from the State Sport General Administration of China were also found and these suggested notably higher estimates of discretionarytime activity compared with CHNS. After a careful review of the two questionnaires and input from experts familiar with China, the average of the two studies was used to estimate the prevalence of inactivity in China (Wang 2002). These data provided $85 \%$ coverage of WPR-B.

### 2.5 COMPARABILITY OF INSTRUMENTS ASSESSING PHYSICAL activity and estimates of exposure among INCLUDED STUDIES

Tables $10.4,10.5$ and 10.6 provide a summary of the data sets used to assess physical (in)activity in discretionary-time, work and transport domains, respectively. For each data source a brief description of the question and the definition of inactive is provided. Below is a summary on the comparability of the final selection of studies in each domain.
Table 10.4 Summary of questions used to assess discretionary-time physical (in)activity

| Subregion | Country | Reliability | Validation | Study title/instrument (reference) | Discretionary-time (in)activity |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  | Method of assessment | Definition of inactive |
| AFR-E | Ethiopia | Not available | Not available | Borana Health and Nutrition Study (Alemu and Lindtjorn 1995) | Seven-day recall via a series of close-ended questions. Participants were interviewed for 7 consecutive days every three months for one year. Activities coded as sleep (I MET), light (I.5 METs), moderate (4METs), hard (6METs), very hard (IOMETs) | Reported mean day EE in kcals/day and kcals/kg per day. (Reanalysis for this work defined inactive as no hard or very hard activities over a 7-day period) |
|  | South Africa | Not available | Not available | BRISK study (Steyn et al. 1991) | Question: After working hours do you get any regular (more than twice/week) exercise? If yes, is it light or strenuous? | No exercise outside of working hours |
| AMR-A | Canada | Not available | Not available | 1998 Physical Activity Monitor (C. Craig, unpublished data, 2001) | Prompted recall of participation in approximately 25 different activities over last 12 months. For each activity reported, number of times in last 12 months and average duration/occasion were recorded | Energy expenditure equal to or less than 0.5 KKD (kcals/kg per day) |
|  | USA | $\begin{aligned} & \text { Карра = } \\ & 0.57-0.77 \end{aligned}$ | Not available | National Physical Activity Survey 1999-2000 (Ham 2001a) | Question: In last month, have you participated in any physical activities or exercises such as brisk walking, bicycling, vacuuming, gardening, or anything else that causes small increases in breathing or heart rate? If yes, type, distance, usual minutes and frequency are recorded | No activities reported |


| AMR-B | Argentina Brazil | No No | No No | August Omnibus Wave-Gallup Poll (Instituto Gallup De La Argentina 2001) Brazil - 98 Cities Project (Datafohla 1997) | Question: Do you regularly practice some type of physical activity? If yes, what (choice of 8 activities or "other")? <br> Prompted recall of participation in 15 sports and activities (+ "other"). If yes, frequency during last month and pattern of participation over last 10 years was recorded | No practice of regular physical activity <br> No activities reported |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Chile | Pearson correlation for leisure index 0.74-0.83 | Pearson correlation $r=0.83$ (VO2 Peak) 0.44-0.56 <br> (EE from 3day diary) | Baecke Questionnaire of Habitual Physical Activity (Jadue et al. 1999) | Participation in sports assessed in two ways. One question asked Do you play sports? If yes, type, frequency, duration recorded. Second question asked During leisure time I play sports? Response is on 5 -point Likert scale (never-very often) | Persons who "never" do exercise (sports) regularly in their free (leisure) time |
| AMR-D | Peru | Not available | Not available | National Living Standards Measurement Survey (Cortez et al. 2002) | Question: In the last month, did you practice any sport such as soccer, volleyball, jogging or walking, among others? If yes, How many times during the last month did you participate in these activities? | No practice of physical activity in leisure time in last month |
| EMR-B | Saudi Arabia | Not available | Not available | Research project using original questions (al-Refaee and al-Hazzaa 2001) | Question: Do you currently do any type of physical activity in a typical week? If yes, Is it regular? How many days? For how long? What kind? | No current physical activity at all (no walking, jogging, swimming, gardening, yard work or any sporting activity) |
| EMR-D | Egypt | Not available | Not available | Research project using original survey (Herman et al. 1995) | One question assessed activity "outside the job" (including transportation to and from work, sporting activities, and other leisure time physical activity). Response was on a 4-point scale | No physical activity weekly |

Table I0.4 Summary of questions used to assess discretionary-time physical (in)activity (continued)

|  |  |  |  | Discretionary-time (in)activity |
| :--- | :--- | :--- | :--- | :--- |
| Subregion | Country | Reliability | Validation | Study title/instrument (reference) |

Australia

## WPR-A Japan

| WPR-A | Japan | One month <br> test-re-test: <br> light $r=$ | Tested <br> against VO2 <br> Pearson <br> correlation | Japan Lifestyle Monitoring <br> Study - (Minnesota <br> leisure-time questionnaire) <br> (lwai et al. 2000) |
| :--- | :--- | :--- | :--- | :--- |
|  |  | $\mathrm{r}=0.92$ | $\mathrm{r}=0.47$ |  |
| New Zealand | Not available | Not available | I998 Sport and Physical <br> Activity Survey (Hillary <br> Commission 1998) |  |
| WPR-B | China | Not available | Not available | China Health and Nutrition <br> Survey 2001 (Ham 2001c) |

Key: EE, energy expenditure; MET, metabolic equivalent. VO2, oxygen consumption.
a Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom.
Table 10.5 Summary of questions used to assess work-related physical (in)activity

|  |  |  |  |  | Occupational activity |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Subregion | Country | Reliability | Validation | Study title (reference) | Method of assessment | Definition of active |
| AFR-E | South Africa | Not available | Not available | BRISK study (Steyn et al. 1991) | Question: Does your work involve mostly: I) primarily sitting/standing; 2) a lot of walking; 3) hard physical (sweat) work? | Category 3: hard physical (sweat) work |
| AMR-A | Canada | Not available | Not available | 1998 Physical Activity Monitor (C. Craig, unpublished data, 2001) | One question assessed activities undertaken at home or at work like lifting boxes, carrying objects, climbing stairs or walking as part of your job or chores. Frequency was recorded. Response scale was: I) almost all the time; 2) almost three-quarters; 3) almost half of the time; 4) about a quarter and/or almost none of the time | Category I: almost all the time |
| AMR-A | USA | Not available | Not available | National Physical Activity Survey 1999-2000 (Ham 2001b; Macera et al. 2001) | Question: When you are at work which of the following best describes you: Mostly sitting or standing? Mostly walking? Mostly heavy labour? | Middle and upper category: mild physical exertion and heavy physical exertion |
| EUR-A | European Union ${ }^{\text {a }}$ | Not available | Not available | Pan-EU survey (Institute of European Food Studies 2001) | One question assessed typical day's activity (either at work, college, in the office or at home). Categories of responses: I) sitting down at work; 2) standing or walking around; 3) more physical work than any of the above. Approximate number of hours/day recorded | Category 3: more physical work than sitting down, or standing or walking around for 3-6 hours |


| EUR-C | Estonia, Latvia, Lithuania | Not available | Not available | Research project using original survey (Pomerleau et al. 2000) | Question not specified but responses coded as low; moderate (mixed sedentary and standing work); high (work requires a lot of walking, lifting, carrying-examples given include heavy housework); and very high (heavy manual work and example occupations given) | Categories: high (work requiring a lot of walking, lifting or carrying) or very high |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| WPR-A | Japan | Not available | Not available | Japan Lifestyle Monitoring Study (Iwai et al. 2000) | Frequency and average duration of various types of labour and other activities on the job, including household activities, for every two months within the last 12 months were queried. Responses were classified into 4 categories according to intensity: sedentary (I.5 METs), standing/walking (2.5 METs), moderately strenuous work (4.5 METs), strenuous work (7.5 METs). A total work activity score was calculated. Lists of activities for 5 major types of businesses were used as prompts | Strenuous work ${ }^{\text {b }}$ |
| WPR-B | China | Not available | Not available | China Health and Nutrition Survey 2001 (Ham 2001c) | Time spent in three categories of activities during a work day assessed. Categories: light or very light; moderate; heavy or very heavy. Examples provided for each category | Categories: heavy or very heavy |
|  | metabolic equivalen Belgium, Denma nce estimates for | Finland, France, <br> rk-related activit | ermany, Greece, were not calcula | land, Italy, Luxembourg, Nethe and were thus unavailable for | nds, Portugal, Spain, Sweden, United Kingdom. clusion. |  |

Table I0.6 Summary of questions used to assess transport-related physical activity

| Subregion | Country or area | Reliability | Validation | Study title (reference) |
| :--- | :--- | :--- | :--- | :--- |

## DISCRETIONARY-TIME (IN)ACTIVITY

The 17 studies providing data for 33 countries that were included in this report used 17 different measures to assess physical activity in discretionary time (Table 10.4). Seven studies assessed participation using an instrument comprising a list of specific activities (i.e. a prompted approach $n=6$ ) or by asking the respondent about "any activities" and recording the type (i.e. unprompted approach $n=1$ ). Both approaches assessed frequency and duration (as detailed in Table 10.1 and ascribed questionnaire format 1 and 2 , respectively).

Nine studies used either a single question or a small number of questions with either open or closed response format (Table 10.1, format 3). These instruments comprised the most diverse set of questions and rarely were the reliability or validity properties of the instruments reported. Only one study used an instrument that captured time spent doing categories of activities (format 5), where categories were defined by the intensity of activity (e.g. moderate-intensity and vigorous-intensity). Only one study used a detailed recall over the past seven days (format 4).

Despite these 17 questionnaires appearing quite different, closer inspection revealed considerable similarity in what each measure attempted to assess. Moreover, the definitions of inactive (level 1 exposure) were very comparable. Three quarters of the studies $(n=14)$ used "no", "none", "never" or "zero" as part of their quantification of inactive. In one study (C. Craig, unpublished data, 2001) inactive was defined on a continuous scale (as less than $3 \mathrm{kcals} / \mathrm{kg}$ per day). Three studies used the lowest category of activity as described in their question (Cockerham and Sneed 2001; Craig 2001; Pomerleau et al. 2000).

The degree of similarity across the definitions of inactive was sufficient for these data to be included in our analyses. However, it is recognized that this comparison is not perfect. Walking and gardening were included in some studies and not others; nor were these activities always defined the same way. Furthermore, the reference time frame and definition of "regular" used in questions were not always consistent. These differences remain limitations to the comparison of data on physical activity.

## WORK-RELATED (IN)ACTIVITY

Table 10.5 provides a summary of the questions used to assess workrelated physical (in)activity in each study for which we had at least some data for consideration. Three of the seven instruments were used in formats 1 or 2 and asked about three types of work activities, namely walking, standing and hard (or heavy) physical work/labour (e.g. EU, South Africa, the United States). Three studies used an instrument (format 3) assessing different categories of intensity (e.g. light, moderate and heavy) and this was deemed comparable to format 1 and 2 instru-
ments. The remaining instrument captured frequency of different types of work-related activity (Iwai et al. 2000).

While there were clear differences in the format of the instruments used to assess this domain, the more significant problem was that "nonwork" activities were also included in the wording of items and thus the physical activity prevalence estimates. For instance, Canada included "chores", and this may have been interpreted as household (domestic tasks) and Japan specifically cited "household activities". The wording and consequently broader scope of the questions severely limited the comparability of these measures of work-related physical activity-so much so that we concluded that only the questions used in the studies from the United States and China were sufficiently similar to allow comparison and thus only these data were used in subsequent analyses.

## TRANSPORT-RELATED (IN)ACTIVITY

Our search for data on transport-related physical activity, and more specifically patterns of cycling and walking to get to and from places, found very few national data sets. It is acknowledged that many countries collect these data as part of national or regional transportation planning and evaluation activities. However, it was beyond the scope of this project to complete an exhaustive search of these sources. Table 10.6 presents details of the six sources used in this study covering 15 countries.

Transportation data can be collected either by prompted recall during a telephone interview or from diaries completed daily over a specific time frame (usually 3-7 days). Trip origin and destination are usually recorded as well as trip mode and duration. This approach is common in developed countries and examples include the National Personal Transportation Survey in the United States (Federal Highway Administration Research and Technical Support Center 1997) and similar instruments used in several European countries (Pucher and Dijkstra 2000) and Australia (Travel Demand Management 1999). In contrast the Health and Nutrition Survey in China included very specific questions to capture the number of walk or cycle trips to a range of specific destinations.

Integrating measures of transport-related physical activity into the assessment of total physical (in)activity is at a very early stage. For this project we were limited to sources of data from three transportation surveys and one health survey. It is acknowledged that other studies previously mentioned and included under other domains may have captured all or some walking trips, either by using a distinct question or by specifying walking as an example of moderate-intensity activity (Armstrong et al. 2000). The same possibility exists for some cycling trips. However, in these cases it was not possible to disaggregate the activity into separate domains. Clearly further development of a standard set of questions for transport-related physical activity is required.

With the exception of China and the United States, all the studies used for this project presented data as "percent of trips" not as a proportion of people undertaking trips. In lieu of obtaining data in the latter format we analysed data from the United States to help derive an estimate from per cent of trips. This method is described in full detail in later sections.

## DOMESTIC-RELATED (IN)ACTIVITY

No data were obtained on physical (in)activity in the domestic domain. We found several studies that did include questions to assess this domain but the data were not reported. Like transport-related activity, some domestic-related activity may have been captured because of the wording of questions or the examples provided. For instance, "housework" could be specified as a moderate-intensity activity in addition to examples of sports and recreational pursuits. In this situation, it was usually impossible to disaggregate the data and, if used, this data could potentially lead to erroneous estimates.

Due to the lack of sufficient data in this domain it was not possible to estimate the magnitude of domestic physical activity or inactivity. It is recognized that this is an important domain and therefore a significant limitation to our final estimates. Moreover, it is highly likely that there is considerable variation in levels of domestic activity between developing and developed countries, within countries, between males and females and across ages. The magnitude and direction of these differences are not well known. Some data are available from studies investigating energy intake and expenditure but these studies usually investigate very specific populations and provide too few data to generalize.

## Combination measures of (in)activity across domains

Several countries used instruments that assessed physical activity across multiple domains. Often these questions presented a brief description and respondents chose a category (e.g. in Egypt) (Herman et al. 1995). In one case very detailed data were obtained via a seven-day recall instrument and the final analyses combined all activity across all domains (e.g. in Ethiopia) (Alemu and Lindtjorn 1995). Data from these approaches were excluded from our estimates in the specific domains. However, where appropriate these data were used post hoc to assess the fit of our final predicted estimates.

### 2.6 Treatment of data

## DEALING WITH DATA REPORTED FOR DIFFERENT AGE CATEGORIES

Much of the data obtained was not available using the same cut-point for age categories. The first and preferred solution involved contacting the original investigators and requesting data analysed by the relevant age groups. Where this was not possible or where data were no longer
accessible, an indirect method was used to compute age-specific estimates, by assuming uniform distribution of population and exposure in one-year increments of each age category.

Where data were available for only some of the age groups included in the lower and upper age categories used in this report (e.g. 15-29 years, 60-69 years, 70-79 years) the obtained prevalence estimate was applied to all years in the category. For example if data were available for only those aged $60-65$ years it was applied to our 60-69 year category, and also extrapolated to older age groups such as $70-79$ or $\geq 80$ years. The effect of this is to slightly underestimate inactivity in the older years within an age category (inactivity usually increases with age) and slightly overestimate the prevalence of inactivity within the youngest age category.

## DEALING with data from non-Nationally representative samples

Several data sets did not come from nationally representative samples. In these instances, descriptions of the study sample and information on the age, sex, ethnic/racial and rural/urban profile of the country were used to create adjusted weighted national estimates. For example, these steps were undertaken with data from Bangladesh and India, which were samples from older adults in urban and rural towns, and also for transport data from Australia that were weighted to form a national estimate.

Dealing with missing data by Sex, age and subregion
Table 10.7 provides a summary of the amount of data identified for each of the 14 subregions according to six age groups and both sexes. Of these 168 cells, $44 \%$ were missing data. For the 14 subregions, the missing cells ranged from $0 \%$ (AMR-A and WPR-B) to $100 \%$ (SEAR-B). Across subregions, missing data cells ranged from $32 \%$ for the four age groups spanning 15-69 years, and increased to $57 \%$ and $79 \%$ for the $70-79$ and $\geq 80$ age groups, respectively. These incomplete subregional by age and sex data cells presented two tasks:

- obtaining estimates for missing age and sex categories; and
- obtaining estimates for countries/subregions where no data source exists.

One approach to solving these two issues was to select the most comparable data set available and apply these values to missing cells. For example, data from EUR-C could be substituted for EUR-B, or WPR-B (which includes China) could be used to derive estimates for SEAR-B and SEAR-D. Similarly missing data for older adults in Europe could be derived from known data from the Americas. While plausible, this approach is weak because any limitations in the data for one subregion will be extended to the other. Moreover, it makes no attempt to better represent the likely differences that exist between countries.

Table 10.7 Summary of data available by subregion, sex and age ${ }^{\text {a }}$

| Subregion | Countries with data | Sex | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D |  | Male | 0 | 0 | 0 | 0 | 0 | 0 |
|  |  | Female | 0 | 0 | 0 | 0 | 0 | 0 |
| AFR-E | South Africa | Male | 14 | 18 | 18 | 16 | 1 | 1 |
|  | Ethiopia | Female | 14 | 18 | 18 | 17 | 1 | 1 |
| AMR-A | Canada | Male | 96 | 96 | 97 | 97 | 97 | 88 |
|  | USA | Female | 96 | 96 | 97 | 97 | 98 | 89 |
|  | Argentina | Male | 51 | 54 | 56 | 17 | 0 | 0 |
| AMR-B | Chile | Female | 51 | 53 | 56 | 18 | 0 | 0 |
|  | Brazil | Male | 36 | 39 | 40 | 0 | 0 | 0 |
| AMR-D | Peru | Female | 37 | 40 | 40 | 0 | 0 | 0 |
| EMR-B | Saudi Arabia | Male | 13 | 13 | 19 | 15 | 0 | 0 |
|  |  | Female | 0 | 0 | 0 | 0 | 0 | 0 |
| EMR-D | Egypt | Male | 21 | 22 | 22 | 21 | 22 | 0 |
|  |  | Female | 20 | 22 | 24 | 23 | 25 | 0 |
| EUR-A | European | Male | 91 | 92 | 91 | 93 | 78 | 20 |
|  | Union ${ }^{\text {b }}$ | Female | 91 | 92 | 91 | 92 | 78 | 24 |
| EUR-B | Estonia | Male | 0 | 0 | 0 | 0 | 0 | 0 |
|  | Latvia | Female | 0 | 0 | 0 | 0 | 0 | 0 |
| EUR-C | Lithuania | Male | 63 | 64 | 64 | 60 | 58 | 56 |
|  | Russian | Female | 62 | 64 | 64 | 61 | 60 | 61 |
|  | Federation |  |  |  |  |  |  |  |
| SEAR-B |  | Male | 0 | 0 | 0 | 0 | 0 | 0 |
|  |  | Female | 0 | 0 | 0 | 0 | 0 | 0 |
| SEAR-D | India | Male | 0 | 0 | 0 | 0 | 0 | 0 |
|  | Bangladesh | Female | 0 | 0 | 0 | 0 | 0 | 0 |
| WPR-A | Australia | Male | 15 | 96 | 98 | 97 | 10 | 0 |
|  | Japan |  |  |  |  |  |  |  |
|  | New Zealand | Female | 16 | 96 | 98 | 97 | 9 | 0 |
|  | Singapore | Male | 82 | 86 | 88 | 88 | 89 | 87 |
| WPR-B | China | Female | 82 | 86 | 87 | 87 | 88 | 88 |

${ }^{\text {a }}$ As \% of total population in each subregion.
b Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom.

An alternative approach was to develop a statistical model to predict national estimates of physical inactivity. Known correlates of inactivity at the individual level could be explored for their predictive value at the national level (e.g. level of education). In addition, a range of other parameters, possibly related to inactivity, such as indicators of economic development (e.g. per capita gross national product [GNP]), could be investigated. Using this "ecological" approach represents the first stage of the epidemiological approach to identifying and understanding the determinants of physical inactivity. Such a model was used to predict
values for countries with missing data and values for missing age/sex cells in countries with only some data. This approach has the advantage of using all available, comparable data on physical inactivity to develop a statistical relationship with other indicators for which there are data available worldwide. We also had to decide whether to create a single model to predict the sum total of inactivity (i.e. inactivity across all four domains as stated in the definition of exposure) or four separate models-one for each domain. The latter approach was selected because of the complexities of the behaviour and the likelihood that different parameters may explain levels of inactivity in different domains.

### 2.7 Estimating prevalence of level i exposure (inactive) USING REGRESSION ANALYSIS

Multivariate and linear regression analyses were used to create predictive models for discretionary-time, work-related and transport-related levels of inactivity. All data obtained from the worldwide search that met our inclusion criteria were considered for inclusion in this process. Only data that clearly represented a single domain were included in the domain-specific modelling exercise. Domestic physical inactivity could not be modelled due to insufficient data. Estimates from countries with measures of physical (in)activity that comprised multiple domains were used post hoc to assess the fit of predicted estimates of overall inactivity across all domains. Only one source of data per country was used.

From a practical stance it was necessary to treat all data within a domain as either values of exposed (namely \% inactive) or unexposed (\% active). In most cases we found it more convenient to work with a model predicting prevalence of doing some physical activity (unexposed) and in the final step to reverse the direction to inactivity (exposed). The exception was in the discretionary-time domain where the predictive model was developed for inactivity.

The modelling approach of level 1 exposure required six steps:

1. Create a predictive model to estimate discretionary-time physical inactivity, and subtract values from 100 to create estimates of physical activity.
2. Create a predictive model to estimate work-related physical activity.
3. Create a predictive model to estimate transport-related physical activity.
4. Sum the domain-specific estimates of physical activity and adjust for overlap across domains.
5. Scale estimates of physical activity and subtract values from 100 to compute final estimates of level 1 exposure (physically inactive).
6. Aggregate age by sex country-level estimates to create age by sex regional-level estimates.

Within each step estimates for each age (six categories) by sex (male/female) cell were modelled and predicted rather than working with and predicting only single point estimates to represent all persons from each country. Table 10.8 summarizes the potential parameters found from World Bank data (World Bank 1999) that were considered for inclusion in the models of (in)activity in each domain. Many of these potential parameters were highly associated including, for example, GNP and per cent of the population completing tertiary education (hereafter referred to as "\% tertiary education") (correlation matrix not reported). Criteria for parameters were the following:

1. Parameters must be available on a national level for most countries across most regions.
2. Parameters must come from a reliable (reputable) source.
3. Parameters must be from recent sources consistent with our exposure variable.

Within each domain, predicted estimates were limited to the countries for which data on the selected parameters were available from the World Bank ( $n=146$ ). After completion of the first five steps, the final task was to aggregate the age-sex-country specific values to create age by sex

Table 10.8 Summary list of potential parameters considered for modelling level I exposure

| Parameter categories | Potential parameter |
| :--- | :--- |
| Demographic | Country population <br> Regional population <br> Per cent urbanization <br> Typical population density experienced by an individual <br> GNP per capita <br> World Bank classification |
| Economic | Per cent completing tertiary education |
| Socioeconomic | Subregion |
| Geographical | Latitude of the country centroid |
| Climatalogical | Mean annual temperature |
| Employment | Per cent employed in agricultural sector |
|  | Per cent employed in manufacturing sector |
| Energy consumption/emissions | Per cent employed in service sector |
| Cars per thousand population |  |
| Data | Carbon dioxide emissions |

regional estimates. At this point known data replaced predicted values where available. Each of the above steps undertaken to estimate the prevalence of exposure level 1 (inactive) is described in detail below.

Step 1: Estimating discretionary-time physical activity
Data by age and sex from 32 countries provided the basis for estimating the prevalence of discretionary-time physical inactivity (level 1 exposure) for 12 age-sex categories in 146 countries. A mixed model for nested, repeated measures was used to identify a model with predicted values that best approximated the data. Twelve age-sex categories were treated as repeated measures nested within each country.

The general linear mixed model is

$$
\mathrm{Y}=\mathrm{X} \beta+\mathrm{Zu}+\mathrm{e}
$$

where: $Y=$ observed response variables
$\mathrm{X}=$ matrix of known values of covariates
$\beta=$ matrix of unknown fixed-effects parameters
$\mathrm{Z}=$ known design matrix of random effects
$\mathrm{u}=$ matrix of unknown random-effects parameters
$\mathrm{e}=$ error.

The model assumes that the error is independent and identically distributed (iid). A mixed model was chosen as most suitable for these data because we considered the possibility of correlation within countries (e.g. clustering of age by sex prevalences). Within-country estimates are likely to cluster because the level of economic development and culture is likely to determine if any data were available and relatively small differences in use of discretionary-time within a country; in contrast, the differences in measurement instruments and protocols would be greater between countries. In addition, using a mixed model allowed for the consideration of clustering of physical activity patterns by geographic region (our strata variable) and this was deemed desirable for the analysis. Mixed models are robust to two characteristics of our prevalence data, heterogeneous variances within clusters (e.g. countries and regions) and missing and unbalanced data. It is acknowledged that missing data may not be random. The effect of missing data is most likely to increase error in subregions with less coverage and reduce error in subregions with better coverage. We compensated for the limitation due to missing data in our calculations of age by sex by subregion estimates by increasing our $95 \%$ confidence limits (described in detail later).

A simple repeated (up to 12 age by sex cells per country) measures model is

$$
\mathrm{y}_{\mathrm{ij}}=\mu+\alpha_{\mathrm{ij}}+\mathrm{d}_{\mathrm{i}}+\mathrm{e}_{\mathrm{ij}}
$$

where $\mu$ and $\alpha_{\mathrm{ij}}$ are fixed-effects parameters, $\mathrm{d}_{\mathrm{i}} \sim \operatorname{iid} \mathrm{N}\left(0, \sigma^{2}\right)$ is the random-effect parameter, and $\mathrm{e}_{\mathrm{ij}} \sim \operatorname{iid} \mathrm{N}\left(0, \sigma^{2}\right)$.

Several tasks were undertaken before the multivariate modelling. Firstly, the skewed physical inactivity data available from 32 countries were transformed to obtain a more normal distribution. Various transformations were attempted but several countries (e.g. Chile, China, the Russian Federation) which had very high estimates of inactivity (ranging $85-96 \%$ ) prevented any transformation from creating a truly normal distribution. Nonetheless, the square root transformation created a more satisfactory variable with a distribution closest to normal (Figure 10.5).

Secondly, we created a variable regional strata to allow countries with data within a region to have a greater influence on the predicted values of other countries in the same region. Due to the overall lack of data across the 14 subregions, five new regional strata were developed to help our analyses. These strata were defined by considering geographical location, World Bank income classification, social, religious and cultural similarities and the availability of data on exposure. The five strata are shown in Table 10.9.

## Weighting

After some preliminary attempts at modelling discretionary-time physical activity and predicting age by sex by country estimates we introduced a weighting factor for each data set. This step was in addition to using the regional strata parameter, and weighted data to the world popula-

Figure 10.5 Transformed age $x$ sex estimates for discretionary-time physical inactivity data $(n=276)$ from 3I countries


Table I0.9 Country allocation to regional strata parameter for modelling level I exposure

| Regional strata created for model | Countries included |
| :--- | :--- |
| I Western industrialized | Western Europe (Austria, Belgium, Denmark, Finland, |
|  | France, Germany, Greece, Ireland, Israel, Italy, Luxembourg, |
|  | Netherlands, Norway, Portugal, Slovenia, Spain, Sweden, |
|  | Switzerland, United Kingdom); North America (Canada, |
|  | USA); Australia, New Zealand |
|  | Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, |
|  | Cuba, Dominican Republic, Ecuador, El Salvador, France |
|  | (Guiana), Guatemala, Guyana, Haiti, Honduras, Jamaica, |
|  | Mexico, Netherlands, Antilles, Nicaragua, Panama, Paraguay, |
|  | Peru, Puerto Rico, Suriname, Trinidad and Tobago, Uruguay, |
|  | Venezuela |
|  | Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, |
|  | Central African Republic, Chad, Congo, Democratic |
|  | Republic of the Congo, Eritrea, Ethiopia, Gabon, Gambia, |
|  | Ghana, Guinea, Guinea-Bissau, Côte d'lvoire, Kenya, |
|  | Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, |
|  | Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, |
|  | Senegal, Sierra Leone, Somalia, South Africa, Sudan, United |
|  | Republic of Tanzania, Togo, Uganda, Zambia, Zimbabwe |
|  | Cambodia, China, Democratic People's Republic of Korea, |
|  | Indonesia, Japan, Lao People's Democratic Republic, |
|  | Malaysia, Mongolia, Philippines, Republic of Korea, |
|  | Singapore, Sri Lanka, Thailand, Viet Nam |
|  | Eastern Europe (Albania, Belarus, Bosnia and Herzegovina, |
|  | Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, |
|  | Lithuania, Poland, Romania, Serbia and Montenegro, The |
|  | former Yugoslav Republic of Macedonia, Turkey); Central |
|  | and South Asia (Bangladesh, Bhutan, India, Nepal, Pakistan); |
|  | the Middle East (Afghanistan, Cyprus, Iran [lslamic Republic |
|  | of, Iraq, Jordan, Kuwait, Lebanon, Libyan Arab Jamahiriya, |
|  | Oman, Saudi Arabia, United Arab Emirates, Yemen); North |
|  | Africa (Algeria, Egypt, Morocco, Tunisia); and the Russian |
|  | Federation |

tion (aged $\geq 15$ years). A weighted model should produce similar parameter estimates regardless of the number of countries, or which specific countries, were included in the model. This was an important consideration because over half of the data were from China, Europe, North America and the Russian Federation. The intent was to avoid these data dominating the model and thus the final estimates.

In practice, known data from a country were weighted to the total population (by age and sex) of the regional stratum proportional to the country population in which the data were available. The sum total population was the total population (aged $\geq 15$ years) in the 146 countries for which the independent variable ( $\%$ tertiary education) was available.

Over half of all the available data on physical inactivity were from countries in the Western industrialized stratum and within this stratum these data represented $90 \%$ of the population. In the South-East Asia stratum, data were available from China, Japan and Singapore, and China represented $77 \%$ of the stratum's population. Using the weights improved the predicted prevalence estimates by allowing China to contribute to the model approximately 1.3 times its population, while limiting the influence of data from countries in the Western industrialized countries to slightly more than their population of each country ( 1.1 times).

The remaining strata had fewer countries with data and thus smaller proportions of the strata population were represented. In these cases, the countries with data determined the age by sex trends of inactivity while the covariate ( $\%$ tertiary education) influenced the intercept for each individual country. Individual weights are not shown because they varied for each country depending on what age by sex data were available and the age by sex population of the stratum.

## Preliminary models of discretionary-time physical inactivity

All of the parameters listed in Table 10.8 were explored in univariate and multivariate models. As mentioned many of the World Bank indicators were correlated and if used in combination would cause problems of collinearity, or unreliable parameter estimates resulting from correlated covariates. Using $\%$ tertiary education was found to explain almost equal variance as GNP and provided smaller residuals that were more normally distributed. Education was reasonably normally distributed. The single point estimate of education per country was applied equally to the 12 age by sex cells.

Figure 10.6 shows a plot of national estimates of discretionary-time physical inactivity by the indicator \% tertiary education for 31 countries. The figure shows that countries with lower tertiary education have a higher prevalence of discretionary-time inactivity.

In addition to exploring a range of demographic, economic and development indices we tested a model that included a variable rating the "data representativeness". As discussed there is insufficient evidence on which to base judgements about different types of instruments, thus we created a variable that differentiated national sources of data from nonnational (regional or metropolitan). China was coded as national because the sample was representative of 16 provinces. Including the data representativeness parameter did not, however, improve our model; thus it was removed.

## Final model of discretionary-time physical inactivity

The final model parameters were education (defined as \% tertiary education), age, sex, region, a three-way interaction termed age x sex x region and corresponding two-way terms. This parameter allowed trends for age to vary by sex and for each different region and it helped to

Figure I0.6 National estimates of discretionary-time physical inactivity by the indicator \% tertiary education ( $n=31$ countries)

improve the predictive power of our analysis. Available data suggest that the trend for inactivity across age can vary by sex and by region. Education allowed inactivity for each country to be based upon an external measure known to be highly associated with a number of risk factors, including discretionary-time physical inactivity, at the individual level. Other social and economic indicators were tested and some produced good results (e.g. particularly GNP per capita), but the best model fit and distribution of the residuals was obtained using the parameter education.

Applying the simple repeated measures model to estimate discretionary-time activity, we get

$$
y_{i j}=m+a_{i j}+d_{i j}+e_{i j}
$$

where $\quad y_{i j}$ is the response, prevalence estimates for age x sex x country categories;
m and $\mathrm{a}_{\mathrm{ij}}$ are fixed-effects parameters: intercept, age, sex, region, \% tertiary education, interaction parameters;
$\mathrm{d}_{\mathrm{i}}$ is the random-effect parameter, country; and
$\mathrm{e}_{\mathrm{ij}}$ is error.
The final model for predicted discretionary-time physical inactivity using data obtained from the World Bank is represented by:
square $\operatorname{root}($ DISCRETIONARY-TIME PHYSICAL
INACTIVITY $)_{\text {COUNTRY,AGE,SEX }}=\alpha_{i}+\beta_{1}(\mathrm{AGE})+\beta_{2}(\mathrm{SEX})$
$+\beta_{3}(\mathrm{REGION})+\beta_{4}(\mathrm{AGE} \times \mathrm{SEX})+\beta_{5}(\mathrm{AGE} \times \mathrm{REGION})$
$+\beta_{6}(\mathrm{SEX} \times \mathrm{REGION})+\beta_{7}(\mathrm{AGE} \times \mathrm{SEX} \times \mathrm{REGION})$
$+\beta_{8}(\%$ TERTIARY EDUCATION $)+\mathrm{d}_{\text {COUNTRY }}+\mathrm{e}_{\mathrm{COUNTRY,AGE,SEX}}$
where $\mathrm{d}_{\text {COUNTRY }} \sim$ iid $\mathrm{N}\left(0, \sigma_{\mathrm{d}}^{2}\right)$ is the random-effect intercept, and $\mathrm{e}_{\text {COUNTRY,AGE,SEX }} \sim$ iid $\mathrm{N}\left(0, \sigma_{\mathrm{d}}^{2}\right)$ is error.

Figure 10.7 shows a plot of the predicted estimates by actual data for 31 countries. The final model explained $52 \%$ of the within-country variance in the prevalence of physical inactivity. The remaining $48 \%$ of variance is attributed to variables not included in this model. Chile, China and the Russian Federation were identified as potential outliers and models were run including and excluding these data. The final model excluded only the Russian Federation and the predicted prevalence estimates fitted the data better by reducing the standard deviation of the residuals from 5.2 to 4.7 .

Modelling was done using SAS 8.1 with the restricted maximum likelihood (REML) method and unstructured covariance structure. The REML method is a better, more conservative estimation method than least squares. We modelled within-country data as repeated measures and "countries" as a random selection of all possible countries. Various

Figure 10.7 Plot of predicted estimates against actual data for discretionary-time physical inactivity $(n=276)$ from 31 countries

models were compared using education and GNP with manual inclusion/exclusion of possible outlying countries. The best-fit model was assessed by minimizing the standard deviation of the residuals calculated by transforming the predicted empirical best linear unbiased predictors (EBLUPs) as well as by minimizing the -2 residual log-likelihood, Akaike's information criterion (AIC), and Schwarz's information criterion (BIC).

The model gives each country its own random-effect intercept estimate which adjusted the overall model intercepts higher or lower by a small percentage. These random-effect intercepts help the model fit the data better than a fixed-effects model by allowing each country to have a different mean value. A shortcoming in using random-effect intercepts is that the random effects are not used to adjust the intercepts for the 114 countries with no data when calculating predicted values. Therefore, the predicted values may be biased if the random-effects estimates are not randomly distributed with a mean of zero. The random-effects estimates for our model were randomly distributed within strata. Figure 10.8 presents a histogram of the distribution of the model residuals by age and sex category estimates showing a normal distribution.

Figure 10.9 shows the residuals from the final model plotted by dis-cretionary-time physical inactivity. The plot shows an increasing trend toward positive residuals with higher estimates of physical inactivity; this is an expected effect of using skewed response data in a model that assumes normal distribution. The model underestimated high prevalence and overestimated low prevalence although the maximum absolute residual for any age-sex estimate was approximately $15 \%$.

Figure 10.8 Distribution of residuals from final model of discretionarytime physical inactivity $(n=276)$


Figure I0.9 Plot of residuals for predicted age- and sex-specific estimates of discretionary-time physical inactivity $(n=276)$ from 3I countries


Figure I0.10 Distribution of predicted national estimates of discretionary-time physical inactivity for all countries ( $n=146$ )


Figure 10.10 shows the distribution of predicted national estimates of discretionary-time physical inactivity from the final model for all countries (total $n=146$ ). The plot shows about one third of the countries had estimates of discretionary-time physical inactivity around $40 \%$. Just over one third of countries had predicted estimates ranging from 55 to 75 per cent.

Figure 10.II Predicted national estimates of discretionary-time physical activity by the indicator \% tertiary education ( $n=146$ )


The final task in this step was to convert the model output (discre-tionary-time physical inactivity) to estimates of physical activity. This was easily done by subtracting the values from 100. Figure 10.11 shows a plot of the predicted national estimates of discretionary-time activity for each country by the indicator $\%$ tertiary education ( $n=146$ ). Countries with data are more developed with higher levels of tertiary education (shown by squares). The predicted values (circles) show different levels of activity for the same values of education reflecting the influence of our regional strata parameter in the predictive model.

Step 2: Estimating work-Related physical activity
Although several sources of data on work-related physical (in)activity were found for a number of countries (see Table 10.5), after careful inspection and discussion with the authors, few data could be included. Specifically, we found several data sets that appeared to be limited to work-related physical activity that also included some domestic activities listed as "chores". This was true for data from Canada, Estonia, Japan, Latvia, Lithuania and the EU. Because this could inflate the prevalence estimates of work-related activity these data were excluded.

Several other countries addressed work-related activity (e.g. Egypt and Ethiopia) but these data could not be disaggregated from overall measures of physical activity. Therefore only data from China and the United States remained as distinct estimates of work-related physical activity for use in a predictive model. Fortunately these countries represent quite different societies and cultures. Moreover, because we had access to the
micro data sets, age-sex estimates for both China and the United States were computed. Only estimates for heavy physical activity at work (or its equivalent) were included.

Figure 10.12 shows the proportion of adults reporting heavy activity at work by age and sex for both China and the United States. Heavy physical activity at work declines over age and is less likely to be reported by females compared with males in both China and the United States.

A simple linear regression equation was developed using the economic development indicator of \% employed in agriculture (hereafter referred to as "\% agriculture"). This World Bank indicator was selected and used by making the following assumptions:

- the proportion of the population employed in agriculture would reflect the level of work-related activity undertaken in a country; and
- there is a linear relationship between the proportion of the population undertaking physical activity at work and the proportion of the total population employed in agriculture.

Figure 10.13 shows the prevalence of heavy work-related physical activity by the indicator \% agriculture for males and females in China and the United States. More men were active at work than women in both China and the United States but the difference was greater in the United States (absolute difference of $12 \%$ compared with $6 \%$ ).

The final model parameters used to predict prevalence of work-related physical activity were \% employed in agriculture ( $\% \mathrm{AGR}$ ), age, sex and

Figure I0.I2 Prevalence of heavy work-related physical activity by age and sex, for China and the United States


Figure 10.13 Proportion of males and females in China and the United States who undertake heavy physical activity at work, by the indicator \% agriculture


Source: China data obtained from the China Health and Nutrition Survey (Ham 2001c); United States data obtained from the NPAS (Ham 2001a) and \% agriculture indicator obtained from the World Bank (1999).
an interaction term "sex by agriculture". This interaction term allowed estimates for work-related activity to vary between sex and by age as suggested by data from China and the United States (see Figure 10.12). Prevalence estimates for age categories 15-59 years were assumed to be the same because there were no clear trends by age in the data from China and the United States. However, an adjustment was introduced for age categories typically associated with retirement. Estimates for work-related activity for age categories 60-69 years, 70-79 years and $\geq 80$ years were progressively lower than those for $<60$ years to account for the reduced likelihood of heavy physical work and increasing proportion of retired persons in each age group. We calculated the mean difference and reduced estimates by $13 \%$ and $19 \%$ for males and females aged $60-69$ years, respectively; and by $19 \%$ and $26 \%$, and $20 \%$ and $27 \%$ for males and females, aged $70-79$ and $\geq 80$ years, respectively (the reference estimate was $50-59$ years). The magnitude of these adjustments were calculated using the mean of the differences seen between the relevant age groups in China and the United States (Figure 10.12). Additional adjustments for sex differences are described below.

Estimates for females were also adjusted to have lower values than males as suggested by the data from China and the United States (Figure

Figure IO.14 Predicted national estimates of heavy work-related physical activity by the indicator \% agriculture $(n=146)$

10.12) but the magnitude of difference seems to vary (range 6-12\%). Thus, for the 146 countries we were working with, those countries with higher total persons employed in agriculture the difference between men and women was smallest (minimum difference set as China value namely $6 \%)$. In countries with fewer total persons employed in agriculture the difference was greatest (maximum difference set as United States value, namely $12 \%$ ).

The regression equation is represented by

$$
\begin{aligned}
& \% \text { heavy physical activity at work } \\
& \quad=\alpha_{\mathrm{i}}+\beta_{1}(\mathrm{AGE})+\beta_{2}(\mathrm{SEX})+\beta_{3}(\% \mathrm{AGR})+\beta_{4}(\% \mathrm{AGR} \times \mathrm{SEX})
\end{aligned}
$$

Figure 10.14 shows predicted national estimates of heavy physical activity at work by the indicator \% agriculture.

## Step 3: Estimating transport-related physical activity

There were few data available reporting the proportion of the population undertaking transport-related physical activity, specifically cycling and walking. However some data were found reporting the per cent of "total trips" that were undertaken by bike or by foot, although even these data were not readily available from national samples.

Data used for modelling transport-related activity came from 13 countries, namely Australia (Travel Demand Management 1999), China (North Carolina Population Center 2001b), the United States (Federal Highway Administration Research and Technical Support Center 1997) and Canada and nine European countries whose data were obtained in
a report comparing trends between North America and Europe (Pucher and Dijkstra 2000). Only the data from China and the United States were available with the variable "persons" rather than "trips" as the denominator. All trip data had to be converted using available information and some adjustments. We wanted to avoid assuming that a separate individual undertook each trip (e.g. 150 trips $=150$ persons) because this is highly unlikely. It is more likely that a smaller proportion of the population walks or cycles and that they account for the majority of the trips (e.g. 150 trips might equal 50 persons). To explore this further we conducted new analyses using the nationwide personal transportation survey (NPTS) database (Ham 2001b) and found that $5 \%$ of all trips were cycling and walking and that they were undertaken by just $10 \%$ of the sample population. This gives a ratio of trips to persons of 1:2. This ratio was adjusted for use with data from Australia $(1: 1.5)$ and European countries ( $1: 1.25$ ) based on the conservative assumption that more individuals undertake more trips in these countries compared with the United States.

Figure 10.15 shows the prevalence of active transport (walk and cycle trips) undertaken by males and females in China and the United States. There was little difference between men and women in both China and the United States but considerable difference between the two countries. However, one notable difference is car ownership. This parameter was explored as a potential predictor of activity in this domain.

Figure $\mathbf{1 0 . 1 5}$ Proportion of adults undertaking transport-related physical activity (cycling or walking) by age and sex, in China and the United States


A simple linear regression equation was developed using the economic development indicator of cars per thousand population (see Figure 10.16). The underlying assumption is that the prevalence of cycling and walking declines as car ownership increases.

The final regression equation is represented by

$$
\begin{aligned}
& \text { TRANSPORT-RELATED PHYSICAL ACTIVITY }=\alpha_{i}+\beta_{1}(\text { CARS }) \\
& =\text { TRANSPORT-RELATED PHYSICAL ACTIVITY } \\
& =73.69-(0.063 . \text { CARS })
\end{aligned}
$$

Figure 10.17 shows the predicted estimates (shown in circles) as well as actual country estimates (shown in squares) for transport-related physical activity data.

China had the lowest ownership of cars and the highest proportion of the population walking and cycling for transport. In contrast, the European countries had much higher car ownership and lower levels of active transport. Australia, Canada, New Zealand and the United States do not fit this model well. They have high estimates of car ownership, with values similar to European countries, but notably lower levels of active transport. This may be due to the degree to which these geographically large countries have developed public transportation infrastructure and the quality of these systems. It is also possible that the mean distance to work (or other destinations) is greater reflecting "sprawling" communities or particular land use patterns that may differ

Figure 10.16 Age-sex estimates of transport-related physical activity for 13 countries, by the indicator cars per thousand population ( $n=156$ )


Figure 10.17 Predicted estimates and actual country data for transportrelated physical activity (cycling and walking) by the indicator cars per thousand population ( $n=146$ countries)

between Australia, Europe and the United States. It is likely that a more sophisticated multivariate model, including average miles driven, would better describe and predict active transport. However in the absence of available data on these other possible parameters we limited our model to cars per thousand population.

Four countries had transport data by age and sex, but the pattern differed for each country (China, Germany, the Netherlands and the United States), therefore a model ignoring age and sex was developed. Australia and the United States were identified as outliers and excluded from the linear regression. In subsequent steps of the analyses the actual values for these countries, as well as from China, were used.

## Step 4: Summation of domain estimates and adjustment for overlap

Estimates of total physical activity were obtained by combining the estimates of physical activity in each of the sub-domains (with the exception of the domestic domain). A simple summation of the estimates of activity in three domains produced country-level estimates ranging from $41 \%$ to $178 \%$ (mean 120, median 122, SD 20). High estimates of activity in more than one domain produced values well over $100 \%$. This was anticipated because it is known that at least some individuals are active in more that one domain and a simple summation would fail to take this into account. However, the magnitude of the overlap between different domains, and for men and women, across age is not well established.

In the absence of published data we explored the pattern of activity in multiple domains using data from China and the United States (Ham

2001a, 2001b, 2001c). It was observed that $6.2 \%$ of adults in the United States were active in both work and discretionary domains, $7.1 \%$ were active in discretionary and transport domains, and $0.2 \%$ were active in both work and transport domains. These relationships are shown schematically in Figure 10.18(a).

Similar analyses were repeated with data from China and are shown in Figure 10.18(b). In China, 29\% of adults were active at work and in the transport domain; $5 \%$ were active in both the discretionary and transport domains but zero per cent were active in both work and discretionary domains. Therefore the pattern of overlap is quite different compared with the United States.

Finally, data on physical activity across multiple domains were found for Finland and considered in a similar way (Luoto et al. 1998). However, without the exact data on the overlap only a hypothesized relationship between domains is shown for Finland in Figure 10.18(c). Finbalt Health Monitor data indicate that over $90 \%$ of Finnish adults are active in their discretionary time, $50 \%$ are active in transport and $13 \%$ are active at work. Accepting these values as reasonable estimates, there must be a large overlap between domains and this is shown schematically in Figure 10.18(c).

Figure 10.18 illustrates that quite different patterns of activity across different combinations of domains are possible and they are likely to vary across different countries around the world, and may vary for men and women and across age. Without additional information on these relationships it was necessary to develop a standard adjustment to our raw sum total of activity (i.e. the sum of the estimates of discretionarytime, transport-related and work-related activity), and apply this to all countries.

Figure 10.18 Schematic representation of national data for three countries showing the proportion of activity in single and multiple domains
(a) United States

$\bigcirc$ Inactive
(b) China

o Discretionary domain

- Work domain
(c) Finland


O Transport domain

[^39]
## Adjustment of raw domain-specific estimates

Adjustment could be made to the estimates in each of the domains ( $n=$ 3) in an attempt to reduce the absolute magnitude of the prevalence estimate of total inactivity. It was also possible to consider an additional adjustment (upward) to accommodate the missing data on activity in the domestic domain. The latter was rejected because there were insufficient data to guide such an adjustment across age, sex or country. Undertaking three separate adjustments was also rejected in favour of a single adjustment and the estimate of transport-related activity was selected as our focus. Our estimate of transport-related activity was selected for the adjustment for overlapping domains for no better reason than it was judged to be the weakest data (both in quality and quantity).

With only data on physical activity across multiple domains from China and the United States (Figure 10.18 [a and b]) we were unable to identify a clear relationship, and therefore selected to reduce the raw estimate of transport-related physical activity by $60 \%$. This gross adjustment was further refined for each country by choosing to scale the magnitude of the adjustment to a range of $50-70 \%$ based on each country's GNP. The underlying assumption was that those countries with the lowest GNP were more likely to have more active transport (cycling and walking), and therefore a smaller adjustment was appropriate (i.e. $30 \%$ reduction in absolute value of the predicted estimate). In contrast, those countries with the highest GNP were likely to have less walking and cycling so they had a larger adjustment ( $50 \%$ reduction in absolute value of the predicted estimate).

These functions are represented below:

$$
\begin{aligned}
& \text { Adjusted TRANSPORT-RELATED PHYSICAL ACTIVITY } \\
& \quad=\alpha_{\mathrm{i}}+\beta_{1}(\mathrm{CARS})+\beta_{2}(\mathrm{GNP})+\beta_{3}(\mathrm{CARS})(\mathrm{GNP}) \\
& \text { Adjusted TRANSPORT-RELATED PHYSICAL ACTIVITY } \\
& =(73-0.06 \text { CARS })(85-3.35 \mathrm{GNP} / 100)
\end{aligned}
$$

Adjusted estimates for transport-related physical activity by the indicator cars per thousand population are shown in Figure 10.19 and this can be compared to the previous unadjusted data shown in Figure 10.17.

## Step 5: Scaling and computation of final estimates of exposure

The age- and sex-specific estimates for discretionary-time activity, workrelated activity and adjusted transport-related activity were summed to create a sum total of physical activity for each age by sex cell for each of the 146 countries. Using the adjusted transport scores, the sum totals had a range of $45-135$, a mean of $95(\mathrm{SD}=20.2)$ and median of 97 . The distribution of both the adjusted and unadjusted sum totals is shown in Figure 10.20. Despite the adjustment described in Step 4, the sum total for some age by sex categories exceeded $100 \%$. Given these values are

Figure IO.19 Adjusted estimates for transport-related physical activity (cycling and walking) by the indicator cars per thousand ( $n=147$ countries)

impossible some form of recalibration of at least the upper values was necessary. Scaling offered one way to address this final problem, which in itself is deemed to be primarily a result of insufficient data on levels of exposure within and between domains.

Once again, known data were used to guide this computation. Specifically, 12 age- and sex-specific estimates of total physical activity were available for both China and the United States $(n=24)$. These data were used to investigate what value to set as the mean to enable $z$-scores to be calculated and the data to be scaled. In addition, we set the maximum value for any age by sex cell for total activity as $98 \%$. This was derived from United States data on disability that suggests approximately $2 \%$ of the whole population may not be able to achieve any physical activity in any domain. Using a maximum value of $98 \%$ and exploring various values for the mean we identified the model of best fit for the 24 agesex category data for China and the United States. Predicted estimates of total physical activity from the best fit model are shown against actual data in Figure 10.21.

The final step was to scale the scores from Step 4 to create age x sex estimates within the range of $0-100$. We calculated the $z$-scores by setting the mean at $82 \%$ and maximum at $98 \%$ and these parameters determined the SD to be 4.33 . We chose to only scale estimates above the mean using the following formula:

If TOTAL SUM SCORE $>82$ then $82+z$-score $\times$
SD of desired distribution $(=4.33)$.

Figure 10.20 Distribution of estimates of total physical activity (a) before adjustment to transport, (b) after adjustment and (c) after standardizing scores above the mean


Predicted estimates that fell below $82 \%$ were not scaled. The final distribution of estimates of total activity is shown in Figure 10.20(c).

In addition to using data from China and the United States, available prevalence estimates of total physical activity (covering at least three domains) from Egypt and the United Republic of Tanzania were used to provide a check on the face validity of our final predicted estimates

Figure I0.2I Predicted and actual age $\times$ sex estimates of total physical activity for the USA $(n=12)$

derived from the adjusted (for activity in multiple domains i.e. Step 4) and now scaled (Step 5) model.

## Step 6: Aggregation of country estimates to create REGIONAL ESTIMATES

The penultimate step was to create prevalence estimates in terms of the exposure variable physical inactivity by subtracting physical activity scores from 100. In order to create age- and sex-specific estimates for 14 subregions, each age by sex by country prevalence estimate was weighted to the age by sex population of the country. The final estimates for each subregion were obtained by calculating the mean level of activity for each age category for males and females. Actual data were substituted for predicted estimates where available.

Estimates of exposure to physical inactivity by subregion, sex and age are shown in Table 10.10 and graphically in Figure 10.22(a). Figure 10.22(b) depicts physical activity by domain and subregion.

### 2.8 Estimating prevalence of level 2 exposure (INSUFFICIENTLY ACTIVE)

Consistent with our estimates of level 1 exposure, we attempted to consider all domains in our computation of level 2 exposure. We used the results of our literature search described earlier to identify those studies with data that matched or closely corresponded to our definition of level 2 exposure, those undertaking some physical activity but not sufficient amounts to meet the Centers for Disease Control and Prevention and the American College of Sports Medicine (CDC/ACSM) public health

Table 10.10 Exposure estimates by subregion, sex and age ${ }^{\text {a }}$

| Subregion | Sex | Exposure category | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | Inactive | 10 | 12 | 13 | 15 | 17 | 18 |
|  |  | Insufficient | 48 | 48 | 48 | 46 | 44 | 43 |
|  |  | Recommended | 42 | 40 | 40 | 39 | 39 | 39 |
|  | Female | Inactive | 12 | 11 | 12 | 15 | 18 | 19 |
|  |  | Insufficient | 45 | 52 | 51 | 48 | 46 | 45 |
|  |  | Recommended | 43 | 37 | 37 | 37 | 36 | 36 |
| AFR-E | Male | Inactive | 9 | 11 | 11 | 14 | 16 | 16 |
|  |  | Insufficient | 50 | 51 | 50 | 49 | 47 | 46 |
|  |  | Recommended | 41 | 38 | 38 | 37 | 38 | 38 |
|  | Female | Inactive | 11 | 10 | 10 | 13 | 15 | 15 |
|  |  | Insufficient | 47 | 56 | 55 | 51 | 50 | 50 |
|  |  | Recommended | 42 | 35 | 35 | 35 | 35 | 35 |
| AMR-A | Male | Inactive | 16 | 18 | 19 | 20 | 21 | 31 |
|  |  | Insufficient | 44 | 47 | 44 | 40 | 40 | 35 |
|  |  | Recommended | 40 | 35 | 36 | 40 | 39 | 34 |
|  | Female | Inactive | 21 | 22 | 20 | 26 | 30 | 40 |
|  |  | Insufficient | 36 | 41 | 40 | 38 | 36 | 31 |
|  |  | Recommended | 43 | 37 | 40 | 37 | 34 | 29 |
| AMR-B | Male | Inactive | 16 | 17 | 18 | 22 | 25 | 28 |
|  |  | Insufficient | 42 | 44 | 41 | 39 | 36 | 35 |
|  |  | Recommended | 42 | 39 | 40 | 39 | 39 | 37 |
|  | Female | Inactive | 22 | 26 | 27 | 36 | 39 | 41 |
|  |  | Insufficient | 33 | 32 | 31 | 30 | 30 | 29 |
|  |  | Recommended | 45 | 41 | 42 | 33 | 31 | 30 |
| AMR-D | Male | Inactive | 16 | 18 | 18 | 22 | 27 | 29 |
|  |  | Insufficient | 38 | 38 | 33 | 32 | 30 | 29 |
|  |  | Recommended | 46 | 45 | 49 | 46 | 43 | 42 |
|  | Female | Inactive | 21 | 25 | 29 | 39 | 45 | 47 |
|  |  | Insufficient | 28 | 26 | 24 | 24 | 22 | 22 |
|  |  | Recommended | 51 | 48 | 47 | 38 | 32 | 31 |
| EMR-B | Male | Inactive | 14 | 18 | 18 | 21 | 24 | 26 |
|  |  | Insufficient | 41 | 39 | 38 | 36 | 32 | 32 |
|  |  | Recommended | 44 | 43 | 43 | 42 | 44 | 42 |
|  | Female | Inactive | 18 | 20 | 20 | 24 | 30 | 32 |
|  |  | Insufficient | 36 | 36 | 35 | 32 | 31 | 30 |
|  |  | Recommended | 46 | 45 | 45 | 44 | 40 | 38 |
| EMR-D | Male | Inactive | 13 | 17 | 18 | 20 | 22 | 25 |
|  |  | Insufficient | 42 | 38 | 36 | 35 | 32 | 31 |
|  |  | Recommended | 45 | 45 | 46 | 45 | 46 | 44 |
|  | Female | Inactive | 17 | 19 | 19 | 22 | 28 | 30 |
|  |  | Insufficient | 36 | 36 | 35 | 32 | 31 | 30 |
|  |  | Recommended | 47 | 45 | 46 | 45 | 41 | 39 |
| EUR-A | Male | Inactive | 13 | 15 | 16 | 18 | 20 | 21 |
|  |  | Insufficient | 52 | 57 | 55 | 52 | 50 | 47 |
|  |  | Recommended | 35 | 29 | 30 | 30 | 30 | 32 |
|  | Female | Inactive | 17 | 18 | 18 | 22 | 24 | 28 |
|  |  | Insufficient | 47 | 51 | 51 | 45 | 45 | 42 |
|  |  | Recommended | 37 | 31 | 31 | 33 | 31 | 30 |

Table 10.10 Exposure estimates by subregion, sex and age ${ }^{\text {a }}$ (continued)

| Subregion | Sex | Exposure category | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EUR-B | Male | Inactive | 15 | 18 | 19 | 22 | 25 | 28 |
|  |  | Insufficient | 43 | 40 | 38 | 36 | 34 | 33 |
|  |  | Recommended | 42 | 42 | 43 | 42 | 41 | 39 |
|  | Female | Inactive | 18 | 20 | 21 | 26 | 31 | 32 |
|  |  | Insufficient | 37 | 37 | 36 | 33 | 32 | 32 |
|  |  | Recommended | 44 | 42 | 43 | 40 | 37 | 36 |
| EUR-C | Male | Inactive | 17 | 18 | 21 | 30 | 36 | 38 |
|  |  | Insufficient | 38 | 34 | 32 | 30 | 28 | 28 |
|  |  | Recommended | 45 | 48 | 47 | 39 | 35 | 34 |
|  | Female | Inactive | 20 | 26 | 27 | 38 | 39 | 40 |
|  |  | Insufficient | 32 | 31 | 30 | 27 | 27 | 26 |
|  |  | Recommended | 48 | 43 | 43 | 34 | 34 | 34 |
| SEAR-B | Male | Inactive | 13 | 15 | 15 | 15 | 14 | 14 |
|  |  | Insufficient | 43 | 43 | 43 | 47 | 52 | 52 |
|  |  | Recommended | 44 | 42 | 42 | 38 | 34 | 34 |
|  | Female | Inactive | 14 | 17 | 17 | 17 | 16 | 16 |
|  |  | Insufficient | 41 | 41 | 41 | 45 | 50 | 50 |
|  |  | Recommended | 44 | 42 | 42 | 38 | 34 | 34 |
| SEAR-D | Male | Inactive | 13 | 16 | 17 | 20 | 22 | 20 |
|  |  | Insufficient | 42 | 38 | 36 | 34 | 32 | 31 |
|  |  | Recommended | 45 | 46 | 47 | 46 | 47 | 49 |
|  | Female | Inactive | 17 | 18 | 19 | 22 | 24 | 26 |
|  |  | Insufficient | 36 | 35 | 34 | 32 | 30 | 30 |
|  |  | Recommended | 47 | 47 | 47 | 47 | 46 | 44 |
| WPR-A | Male | Inactive | 14 | 15 | 16 | 18 | 17 | 17 |
|  |  | Insufficient | 50 | 56 | 53 | 52 | 56 | 55 |
|  |  | Recommended | 35 | 29 | 30 | 30 | 27 | 28 |
|  | Female | Inactive | 16 | 19 | 18 | 20 | 17 | 17 |
|  |  | Insufficient | 48 | 49 | 50 | 49 | 55 | 54 |
|  |  | Recommended | 36 | 32 | 32 | 31 | 28 | 28 |
| WPR-B | Male | Inactive | 13 | 15 | 15 | 17 | 18 | 20 |
|  |  | Insufficient | 41 | 40 | 41 | 41 | 44 | 41 |
|  |  | Recommended | 46 | 44 | 45 | 41 | 38 | 38 |
|  | Female | Inactive | 15 | 16 | 17 | 20 | 20 | 19 |
|  |  | Insufficient | 40 | 39 | 38 | 38 | 41 | 38 |
|  |  | Recommended | 45 | 45 | 45 | 42 | 39 | 42 |

a As \% of total population.
recommendation of 150 minutes of moderate-intensity physical activity per week or equivalent. We found some data for the discretionary-time domain but no specific data on the level of insufficient activity in the transport- and occupation-related domains. Thus, with only a modest amount of empirical evidence we constructed a method to differentiate between those adults doing some activity but not meeting the recommended amount (i.e. level 2 exposure) and those adults meeting the current recommendations (level 3).

Figure 10.22 Prevalence of exposure by (a) subregion and (b) prevalence of physical activity by domain and subregion

$\square$ Level 1 (inactive) $\mathbb{Z}$ Level 2 (insufficiently active) $\square$ Level 3 (meet recommendations)
$\square$ Discretionary-time activity $\mathbb{Z}$ Work-related activity $■$ Transport-related activity

## Discretionary-time domain

Countries with some data on levels of insufficient activity in the discretionary-time domain were Australia, Canada, China, Estonia, Ethiopia, Finland, India, Latvia, Lithuania, the Russian Federation, the United Kingdom of Great Britain and Northern Ireland and the United States. Figure 10.23 presents the country-level estimates of \% insufficient

Figure 10.23 Prevalence of insufficiently active as a proportion of total adult population and insufficiently active as a proportion of adults who do at least some activity, in 12 countries $^{\text {a }}$

a Definitions of insufficiently active are similar but not identical in each country; comparisons should therefore be made with caution.
for these 12 countries. We emphasize that these data are not exactly comparable due to different instruments and calculations as well as slightly different definitions. With the exception of Ethiopia, these data represent levels of insufficient activity in the discretionary-time domain only.

Figure 10.23 shows that there was no clear pattern across this set of countries. For instance, Canada and Australia had similar prevalence estimates of insufficiently active, namely $23 \%$ and $25 \%$ respectively, but this represented $25 \%$ of the "active" Canadian population and $33 \%$ in Australia. In contrast, $45 \%$ of the adult population in the United States was classified as insufficient and this was $63 \%$ of the active population. Finland had a very high proportion of adults doing at least some activity ( $90 \%$ ) but of these, $63 \%$ did not meet the current recommendations. This was similar in Ethiopia. In the absence of any pattern in the proportion of adults classified as inactive with data ranging from $25-80 \%$, we assumed that $50 \%$ of adults who were doing some activity were not active enough to meet recommendations. This rule was applied to our calculations of insufficient for all countries, male and females and all ages except the lowest age category (15-29 years). Available data indicate that younger adults are more likely to meet recommended levels of activity; therefore for this age group we assumed only $40 \%$ of those active were insufficiently active.

## Transport domain

The data on physical activity in the transport domain are generally of poor quality. Some data from Europe suggest that approximately half of all walk and cycle trips are less than 3 km . Using this information, we estimated that $50 \%$ of adults classified as active in the transport domain would be classified as insufficient because at a moderate-intensity (walking pace) 3 km would take approximately 30 minutes). It is likely that trips may be longer in duration in developing countries; therefore we scaled our algorithm such that $60 \%$ of adults identified as active in countries defined as "high income" (World Bank 1999) were classified as insufficient compared with $50 \%$ and $40 \%$ of adults in "upper middle income" and the combined group of "lower middle" and "lower income", respectively. We applied no additional adjustment to account for likely age and sex differences.

## Work domain

Data on work-related physical activity are not often reported in terms of hours of different intensity activities. Our search found few data assessing this activity in this domain alone, and only data from China were reported in hours. It is suspected that there is a large degree of measurement error in self-reporting of heavy physical activity at work; therefore some proportion of those adults reporting heavy activity may be misclassified. We decided to assume that $30 \%$ of those identified as doing heavy physical activity at work in countries defined as "high income" (World Bank 1999) did not meet recommendations. In "upper middle", "lower middle" and "lower income" countries we adjusted this assumption to $15 \%$ thus allowing more work activity "to count" as recommended. We did no further adjustments based on age or sex. Moreover, we did not attempt to estimate the proportion of adults misclassified as inactive when they should have been classified as insufficient.

## Summing across domains and final adjustment of estimates

We used our age-sex-country estimates of level 1 exposure (inactive) within in each domain (discretionary, transport and work) to compute the prevalence of adults doing at least some activity for each age x sex x country category within each domain. This was simply $100-\%$ inactive. The algorithms described above were used to compute the prevalence of insufficient in each domain, which in turn were summed to derive an overall estimate for level 2 exposure.

We reviewed these data and noted the absence of any difference between males and females. In contrast, limited available data in the dis-cretionary-time domain suggest women are less likely to be undertaking the recommended level of activity (level 3, sufficient). In addition, there is evidence of a trend over age, with older adults more likely to be insuf-
ficient compared with younger adults who are more likely to be undertaking recommended levels. We therefore developed and imposed an age and sex adjustment which resulted in the best fit between our age-sexcountry estimates and available data. It was, however, not possible to get a good fit; but the results appear to reflect the regional variations even though certain individual countries may be poorly described. Computed estimates of level 2 exposure by subregion, sex and age are reported in Table 10.10.

### 2.9 Sources of uncertainty

Given the paucity of existing data in each of the domains of physical (in)activity and the complexity of the approach taken to predict estimates of exposure, there is a large degree of uncertainty around the final data. Both quantitative and qualitative approaches were taken to estimate the magnitude of this error and these steps are outlined below.

## Methods used to estimate uncertainty and compute standard error AND 95\% CONFIDENCE INTERVALS

## Level 1-Inactive

Subregions were categorized according to the level of estimate uncertainty by considering both the quantity and quality of available data within each subregion. Quantity of data was assessed using the proportion of the total subregional population represented by data. Subregions with fewer data from fewer countries were deemed to have much greater uncertainty than those subregions with more data from more countries. Quality of data was assessed by considering the source of data (national or non-national samples). The subregions were classified into four levels of uncertainty.

- AFR-D, EUR-B, SEAR-B, SEAR-D
- AFR-E, EMR-B, EMR-D
- AMR-B, AMR-D
- AMR-A, EUR-A, EUR-C, WPR-A, WPR-B


## highest uncertainty;

high uncertainty;
moderate-high uncertainty;
lowest uncertainty.

The predicted estimates within each domain (discretionary-time, work-related and transport-related), the missing data on domestic inactivity and the modelling procedures all represent sources of uncertainty. To account for these an adjustment factor was calculated for each age-sex-subregion estimate. The magnitude of the adjustment was proportional to the level of uncertainty surrounding the data and was a sliding scale inverse to the grading of data quality. The maximum adjustment was chosen to elicit a maximum coefficient of variation of approximately $30 \%$ for the least certain estimates. To achieve this the adjustment factor
was greatest for prevalence estimates in the subregions with the highest uncertainty ( $25 \%$ ) and smallest for those subregions with the lowest uncertainty $(10 \%)$. The adjustment factor for high uncertainty and mod-erate-high uncertainty were $20 \%$ and $15 \%$, respectively. The resulting uncertainty ranges for level 1 exposure are shown in Table 10.11.

The lack of data on older adults would add greater uncertainty to these estimates across all subregions. Furthermore, the omission of data (actual or predicted) on domestic-related activity may add differential levels of uncertainty to the final estimates of women compared with men and between subregions.

## Level 2-Insufficiently active

Overall there is more uncertainty around the estimates of level 2 exposure than for level 1 due to the paucity of data and the heterogeneity among the existing data. Therefore, larger rather then smaller confidence limits were used. We computed uncertainty ranges by calculating a $25 \%$ upper and lower margin of error around each age-sex-subregion predicted estimates (Table 10.11). Given a mean level of insufficient exercise of approximately $40 \%$ this produces a range of $\pm 10$ prevalence estimate points.

## 3. Estimating physical inactivity-disease RELATIONSHIPS

### 3.1 Selection of health outcomes

There is a large body of scientific evidence linking physical inactivity with a wide range of cardiovascular, musculoskeletal and mental health outcomes. All potential disease end-points were considered; however, for inclusion the following criteria had to be met:

1. the disease outcome is included in the Global Burden of Disease (GBD) disease list;
2. there was strong evidence for a causal association between physical inactivity and an increase in risk;
3. biologically plausible mechanisms exist to explain (at least partly) the association; and
4. sufficient information was available to allow quantification of hazard.

The disease end-points initially considered for inclusion were cardiovascular disease, specifically ischaemic heart disease and stroke, several site-specific cancers (colon, rectal, breast, prostate), type II diabetes, various musculoskeletal conditions (namely lower back pain, osteoarthritis, osteoporosis), falls and mental health outcomes (specifically depression). Literature pertaining to each of these was reviewed to
Table IO.II Range of uncertainty associated with estimates for level I and 2 exposure by subregion, sex and age

| Subregion | Sex | Exposure category | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | Inactive | 4.03-15.23 | 5.18-19.38 | 5.28-19.80 | 6.58-23.70 | 7.43-27.03 | 7.54-28.86 |
|  |  | Insufficient | 29.31-67.10 | 29.29-67.05 | 28.93-66.23 | 28.03-64.16 | 26.60-60.90 | 26.05-59.64 |
|  | Female | Inactive | 4.99-18.97 | 4.25-18.63 | 4.49-18.95 | 6.06-24.56 | 6.36-28.96 | 6.51-3I.63 |
|  |  | Insufficient | 27.39-62.70 | 31.50-72.12 | 31.10-71.21 | 29.09-66.59 | 28.17-64.49 | 27.37-62.67 |
| AFR-E | Male | Inactive | 4.51-\|2.5| | 6.00-16.00 | 6.15-16.33 | 7.70-19.82 | 8.83-22.29 | 9.18-22.96 |
|  |  | Insufficient | 30.48-69.79 | 31.00-70.98 | 30.65-70.17 | 29.76-68.13 | 28.40-65.03 | 27.98-64.07 |
|  | Female | Inactive | 5.68-15.52 | 5.04-14.36 | 5.32-14.94 | 7.37-19.59 | 8.34-21.22 | 8.73-21.97 |
|  |  | Insufficient | 28.79-65.91 | 33.78-77.34 | 33.17-75.94 | 31.21-71.46 | 30.51-69.85 | 30.20-69.14 |
| AMR-A | Male | Inactive | 11.93-19.95 | 12.68-23.60 | 13.12-25.70 | 14.65-25.07 | 13.91-28.25 | 20.09-41.77 |
|  |  | Insufficient | 26.57-60.83 | 28.46-65.16 | 26.91-61.61 | 24.45-55.99 | 24.08-55.13 | 21.50-49.22 |
|  | Female | Inactive | 14.03-27.45 | 13.65-30.29 | 11.90-28.62 | 14.87-36.69 | 18.66-41.50 | 24.15-55.43 |
|  |  | Insufficient | 22.01-50.39 | 25.08-57.43 | 24.44-55.96 | 22.89-52.40 | 21.64-49.53 | 18.73-42.89 |
| AMR-B | Male | Inactive | 10.53-21.07 | 11.56-23.24 | 11.93-24.93 | 12.90-31.20 | 14.02-35.98 | 15.92-39.72 |
|  |  | Insufficient | 25.75-58.95 | 26.5I-60.70 | 25.18-57.65 | 23.47-53.72 | 22.09-50.58 | 21.42-49.05 |
|  | Female | Inactive | 13.88-30.84 | 15.87-37.03 | 15.25-39.17 | 21.31-51.55 | 23.32-55.40 | 24.21-57.71 |
|  |  | Insufficient | 19.92-45.61 | 19.56-44.78 | 18.77-42.97 | 18.50-42.35 | 18.08-41.39 | 17.56-40.20 |
| AMR-D | Male | Inactive | 10.27-21.61 | 11.48-23.92 | 11.99-24.35 | 14.16-28.84 | 15.18-39.44 | 16.02-42.68 |
|  |  | Insufficient | 22.92-52.48 | 22.97-52.58 | 19.95-45.66 | 19.63-44.94 | 18.13-41.52 | 17.72-40.56 |
|  | Female | Inactive | 13.18-28.60 | 13.93-36.47 | 15.09-43.05 | 21.97-55.65 | 26.43-64.39 | 27.65-65.99 |
|  |  | Insufficient | 17.12-39.19 | 16.01-36.66 | 14.74-33.74 | 14.40-32.96 | 13.62-31.17 | 13.34-30.54 |
| EMR-B | Male | Inactive | 7.94-20.96 | 10.03-25.67 | 10.25-26.31 | 11.20-31.46 | 10.86-36.16 | 12.62-39.74 |
|  |  | Insufficient | 25.21-57.73 | 23.67-54.18 | 23.29-53.32 | 22.06-50.50 | 19.70-45.11 | 19.19-43.95 |
|  | Female | Inactive | 10.23-25.95 | 11.13-28.27 | 11.19-28.83 | 10.85-36.77 | 15.37-44.09 | 17.05-47.09 |
|  |  | Insufficient | 21.88-50.10 | 21.67-49.62 | 21.31-48.79 | 19.43-44.49 | 18.69-42.79 | 18.30-41.90 |
| EMR-D | Male | Inactive | 6.78-20.18 | 8.65-25.27 | 9.01-26.43 | 10.54-30.06 | 10.94-33.36 | 12.29-39.74 |
|  |  | Insufficient | 25.29-57.91 | 23.22-53.17 | 22.15-50.71 | 20.98-48.04 | 19.48-44.60 | 18.89-43.24 |
|  | Female | Inactive | 8.60-25.74 | 8.91-28.51 | 9.12-28.94 | 10.62-34.06 | 13.35-43.65 | 13.87-46.59 |
|  |  | Insufficient | 21.94-50.24 | 21.90-50.15 | 21.47-49.16 | 19.62-44.92 | 18.8\|-43.05 | 18.53-42.41 |

Table IO.II Range of uncertainty associated with estimates for level I and 2 exposure by subregion, sex and age (continued)

| Subregion | Sex | Exposure category | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EUR-A | Male | Inactive | 10.02-16.76 | 10.81-18.47 | 11.32-19.76 | 13.07-22.85 | 13.72-25.62 | 14.99-27.4\| |
|  |  | Insufficient | 31.56-72.25 | 34.57-79.14 | 33.33-76.31 | 31.82-72.85 | 30.61-70.08 | 28.68-65.65 |
|  | Female | Inactive | 12.19-21.05 | 12.97-22.65 | 13.02-23.28 | 15.29-29.63 | 16.08-31.66 | 18.39-36.95 |
|  |  | Insufficient | 28.39-65.00 | 31.17-71.37 | 30.84-70.62 | 27.25-62.39 | 27.19-62.26 | 25.82-59.11 |
| EUR-B | Male | Inactive | 6.69-21.85 | 8.40-27.00 | 8.92-28.38 | 10.11-31.75 | 10.84-38.98 | 11.87-43.05 |
|  |  | Insufficient | 26.17-59.92 | 24.49-56.07 | 23.30-53.36 | 22.09-50.57 | 20.59-47.15 | 20.16-46.16 |
|  | Female | Inactive | 8.58-27.62 | 9.47-30.11 | 9.80-30.92 | \|1.15-40.5 | | 12.82-48.68 | 13.36-50.64 |
|  |  | Insufficient | 22.77-52.14 | 22.73-52.04 | 22.15-50.72 | 20.32-46.52 | 19.59-44.85 | 19.32-44.24 |
| EUR-C | Male | Inactive | 11.50-21.54 | 13.5\|-22.49 | 14.88-27.98 | \|8.45-42.51 | 21.16-51.62 | 21.64-54.24 |
|  |  | Insufficient | 23.11-52.92 | 20.64-47.26 | 19.38-44.36 | 18.34-41.98 | 17.22-39.43 | 16.94-38.78 |
|  | Female | Inactive | 14.39-25.89 | 16.90-36.06 | 17.02-37.08 | 22.04-54.54 | 21.33-56.49 | 22.26-57.94 |
|  |  | Insufficient | 19.34-44.29 | 18.78-42.99 | 18.43-42.19 | 16.62-38.05 | 16.27-37.25 | 15.91-36.42 |
| SEAR-B | Male | Inactive | 5.73-20.03 | 6.82-23.70 | 6.83-23.71 | 7.02-23.90 | 6.40-21.78 | 6.59-22.31 |
|  |  | Insufficient | 26.18-59.93 | 25.96-59.44 | 25.95-59.41 | 28.47-65.18 | 31.73-72.65 | 31.57-72.29 |
|  | Female | Inactive | 6.41-22.57 | 7.60-26.52 | 7.53-26.29 | 7.73-26.43 | 7.09-24.25 | 7.28-24.72 |
|  |  | Insufficient | 25.22-57.74 | 24.63-56.38 | 24.85-56.89 | 27.39-62.71 | 30.70-70.28 | 30.56-69.96 |
| SEAR-D | Male | Inactive | 6.34-19.36 | 8.17-24.73 | 8.49-25.95 | 9.50-30.38 | 9.57-33.43 | 9.46-30.96 |
|  |  | Insufficient | 25.38-58.10 | 22.96-52.56 | 21.87-50.07 | 20.8I-47.64 | 19.33-44.24 | 18.88-43.23 |
|  | Female | Inactive | 8.11-24.91 | 9.10-27.58 | 9.23-28.03 | 10.11-33.21 | 10.94-36.94 | 11.50-41.28 |
|  |  | Insufficient | 21.94-50.23 | 21.26-48.67 | 20.84-47.72 | 19.17-43.89 | 18.38-42.08 | 18.02-41.25 |
| WPR-A | Male | Inactive | 10.31-18.25 | 11.15-19.07 | 12.05-20.37 | 13.65-21.97 | 12.86-20.80 | 12.61-21.35 |
|  |  | Insufficient | 30.60-70.06 | 34.00-77.84 | 32.46-74.33 | 31.79-72.79 | 33.91-77.65 | 33.70-77.16 |
|  | Female | Inactive | 11.72-20.60 | 13.40-24.08 | 13.36-23.00 | 15.30-23.84 | 12.98-21.36 | 12.49-22.29 |
|  |  | Insufficient | 28.93-66.24 | 29.80-68.22 | 30.59-70.03 | 30.06-68.81 | 33.44-76.55 | 33.13-75.85 |
| WPR-B | Male | Inactive | 10.07-16.44 | 11.91-18.93 | 11.53-18.07 | 13.22-21.62 | \|3.5|-23.27 | 14.49-26.19 |
|  |  | Insufficient | 24.93-57.07 | 24.39-55.83 | 24.65-56.44 | 25.13-57.53 | 26.55-60.79 | 25.14-57.56 |
|  | Female | Inactive | 11.28-18.34 | 12.76-19.90 | 13.08-20.58 | 15.09-25.27 | 14.88-25.98 | 14.42-24.56 |
|  |  | Insufficient | 24.30-55.64 | 23.64-54.13 | 23.05-52.77 | 23.19-53.09 | 24.83-56.85 | 23.14-52.99 |

Table IO.I2 Health outcomes considered in this analysis

| Health outcome | GBD <br> classification system outcome | Causal | Biological plausibility | Risk <br> data | Inclusion/Exclusion |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Ischaemic heart disease | Yes | Yes | Yes | Yes | Included |
| Stroke | Yes | Yes ${ }^{\text {a }}$ | Yes ${ }^{\text {a }}$ | Yes | Included ${ }^{\text {a }}$ |
| Type II diabetes | Yes | Yes | Yes | Yes | Included |
| Breast cancer | Yes | Yes | Yes | Yes | Included |
| Colon cancer | Yes | Yes | Yes | Yes | Included |
| Prostate cancer | Yes | Uncertain | Uncertain | Limited | Excluded |
| Rectal cancer | Yes | Uncertain | Uncertain | Limited | Excluded |
| Low back pain | Yes | Some | Some | Limited | Excluded |
| Osteoporosis | No | Some | Some | Limited | Excluded |
| Osteoarthritis | No | No | No | Limited | Excluded |
| Falls | No | Some | No | Limited | Excluded |
| Depression | Yes | No | Uncertain | Limited | Excluded |
| Obesity | No | Yes | Yes | No | Excluded |
| ${ }^{\text {a }}$ For ischaemic str | ke only. |  |  |  |  |

identify the level of evidence on causality. For several outcomes there is a considerable level of interest and a large number of studies, particularly in recent years, but insufficient support for the biological mechanism associated with physical inactivity. In these cases the disease outcomes were excluded but any future attempts at assessing attributable burden should update our review. Below is a brief summary of the evidence and proposed biological mechanisms associated with physical inactivity for each disease end-point considered. These are summarized in Table 10.12.

## CARDIovascular diseases

## Ischaemic heart disease

The strongest evidence for the benefits of physical activity pertains to the reduction of risk of mortality and morbidity from cardiovascular disease, particularly acute myocardial infarction and other forms of ischaemic heart disease (Berlin and Colditz 1990; Powell et al. 1987). These associations are generally strong and independent of the definition of physical activity used. Biologically plausible mechanisms for the effects of moderate and/or vigorous intensity physical activity on ischaemic heart disease have been identified through clinical and observational studies. The mechanisms or pathways include advantageous effects on athero-
sclerosis, lipid profile, ischaemia, blood pressure, thrombosis and fibrinolytic activity (Hardman and Stensel 2003).

A number of studies have shown physical activity directly and indirectly reduces the effects of excess cholesterol and other atherosclerotic agents (Durstine and Haskell 1994; Kramsch et al. 1981; Leon 1991). Participation in physical activity can improve total blood cholesterol levels (McMurray et al. 1998) and improve high density lipoprotein (HDL) subfraction profiles (Moore 1994). An increase in HDL is desirable because HDL transports cholesterol to the liver for elimination in the bile, and thus has an "anti-atherosclerotic" function. Physical activity has also been shown to increase the activity of lipoprotein lipase, which is involved in the removal of cholesterol from the blood (Stefanick and Wood 1994). There may be, however, a threshold for the relationship between physical activity and improvements in the HDL subfraction of cholesterol, with prolonged or intensive physical activity being more beneficial for HDL to cholesterol ratios (Kokkinos and Fernhall 1999).

Decreased risk of ischaemia may be due to positive adaptations in coronary circulation from structural adaptations following physical activity (Laughlin 1994; Tomanek 1994). Acute coronary events may be reduced by a reduction in thromboses by an increase in enzymatic (fibrinolytic) breakdown of blood clots and a decrease in platelet aggregation (Leon 1991).

Vigorous physical activity has been shown to decrease systolic and diastolic blood pressures (Arroll and Beaglehole 1992; Kelley and McClellan 1994; McMurray et al. 1998; Mensink et al. 1999). Moreover there is some evidence that participation in more moderateintensity activity may achieve similar or even greater effects than vigorous activity (Hagberg et al. 1989; Marceau et al. 1993; Matsusaki et al. 1992; Moreau et al. 2001). The proposed mechanisms by which blood pressure is lowered is via the immediate and temporary dilation of the peripheral blood vessels during physical activity and the ongoing effect of a reduction in sympathetic nervous system activity (Fagard and Tipton 1994).

## Stroke

A recent review of the dose-response relationship between physical activity and risk of stroke suggests current evidence remains equivocal (Kohl 2001). Only six studies from a total of 15 showed evidence of a dose-reponse relationship; eight did not and two studies suggested a "U" shaped distribution (Kohl 2001). While this presents an unclear picture, it is notable that many studies did not differentiate between the two different types of stroke: ischaemic and haemorrhagic, as we describe in detail below.

The biological mechanisms for the association between physical activity and ischaemic stroke and ischaemic heart disease are thought to be
similar, namely atherosclerosis and hypertensive disease (Kohl 2001). However, the biological pathway for haemorrhagic stroke is less clear. Given potentially different pathophysiological pathways, it follows that physical activity may be differentially related to one type (ischaemic) and not the other (haemorrhagic). In general, studies show a decrease in the risk of ischaemic stroke with increasing levels of physical activity (Ellekjaer et al. 2000; Wannamethee and Shaper 1999) but studies with separate risk by sub-type show no or smaller associations with varying levels of activity for haemorrhagic stroke (Hu et al. 2000).

Given the evidence is mixed for an association between overall stroke and physical inactivity and sufficient evidence on causality and burden exists only for ischaemic stroke, for this study, only ischaemic stroke was included.

## Type II Diabetes

A recent review (Ivy et al. 1999) showed that the benefits of physical activity in the prevention of type II diabetes are strongly supported by current research, especially among people already at risk (Kelley and Goodpaster 2001). In general, prospective observational studies show a lower incidence of type II diabetes in more active people compared with the least active in the population. Both moderate- and vigorous-intensity physical activity reduce the risk of type II diabetes in women (Hu et al. 1999), and the benefits appear to accrue also for males and in diverse populations (Folsom et al. 2000; Okada et al. 2000). Reduction in risk of type II diabetes appears to occur only from regular, sustained physical activity.

The biological mechanisms for this protection have not been clearly identified, but are likely to occur at systemic, tissue and cellular levels. Numerous studies and reviews have described either a short- and/or a long-term effect from participation in physical activity on glucose tolerance and carbohydrate metabolism. In general, physical activity increases sensitivity to insulin, improves glucose metabolism, reduces atherosclerosis risk, and reduces intra-abdominal fat distribution (Kelley and Goodpaster 2001). Physical inactivity appears to relate to increased risk of type II diabetes by two possible pathways, which are described in detail by Katzmarzyk et al. (1996). Briefly, one pathway involves the relationship between physical inactivity, a positive energy balance, and an increase in adiposity. The resulting insulin resistance leads to a reduction in plasma free fatty acid (FFA) clearance. Elevated blood FFA levels have detrimental effects on blood glucose, which result in increased pancreatic $\beta$ cell insulin secretion and hyperinsulinaemia in order to control blood glucose levels. The increased requirement for insulin causes $\beta$ cell impairment and reduced blood insulin levels, which increases the insulin resistant state, further reduces clearance of FFA from the blood, increases glucose levels and results in type II diabetes (Katzmarzyk et al. 1996).

A second mechanism suggests that a physically inactive lifestyle exposes a genetic predisposition in skeletal muscle, which can result in muscular insulin resistance. This results in increases in $\beta$ cell insulin secretion and hyperinsulinaemia to control blood glucose concentrations. One result of hyperinsulinaemia is a suppression of fatty acid oxidation and increases of triglyceride storage, and adipose cell hypertrophy. Adipose cells then become insulin resistant and there is a reduced ability to remove FFA from the blood. This results in an increase in glucose output from the liver and further development of muscle insulin resistance. Ultimately, the increased reliance on insulin to control blood glucose concentration results in $\beta$ cell impairment and development of type II diabetes.

The physiological adaptations that are responsible for the protective effects of physical activity subside within a short period of the cessation of physical activity, within two weeks of inactivity (Arciero et al. 1999; Dela et al. 1993; Rogers et al. 1990). Therefore physical activity must be undertaken regularly to provide benefits in terms of risk reduction.

The evidence for a causative association between physical inactivity and increased risk of type II diabetes was strong enough for inclusion.

## CANCERS

## Colon cancer

Numerous studies have shown the protective effect of physical activity on risk of colon cancer (Colditz et al. 1997; IARC 2002; Thune and Furberg 2001) and on the prevention of precancerous polyps in the large bowel (Neugut et al. 1996; Slattery et al. 1997). However, no definitive biological mechanisms have been identified to explain the relationship between physical inactivity and increased risk of colon cancer although several mechanisms have been proposed, which link physical activity and changes in physiologic measures.

One possible mechanism is the effect of prostaglandins on colon mucosal cell proliferation. Physical activity produces an increase in prostaglandin F2 alpha that increases intestinal motility, and a decrease in prostaglandin E2, which stimulates colon cell proliferation (Thor et al. 1985; Tutton and Barkla 1980). Further support for this possible mechanism comes from laboratory studies on rats and evidence in humans that aspirin and nonsteroidal anti-inflammatory drugs, also inhibitors of prostaglandin synthesis, reduce risk of colon cancer (Tomeo et al. 1999). A second mechanism relates physical inactivity and abdominal obesity and proposes that obesity may increase the risk of colon cancer via its influence on insulin-like growth factor-1 (IGF-1) and insulin-like growth factor binding proteins (IGFBP). Since insulin is a growth factor for colon mucosal cells this may promote cancerous cell production. Thus, it is possible that activity exerts its protective effect through reduced insulin levels, because physical activity has been shown
to reduce insulin levels in both obese individuals and those of healthy weight (Giovannucci et al. 1995; McKeown-Eyssen 1994; Tomeo et al. 1999). Insulin-like growth factor is associated with colon cancer risk (Giovannucci et al. 2000; Ma et al. 1999), and IGF is influenced by caloric intake and physical activity. Thus higher levels of physical activity may down-regulate IGF by increasing the production of the binding protein (IGFBP3). The nature of the evidence for an association between physical inactivity and colon cancer was sufficient to meet criteria for inclusion.

## Rectal cancer

There is less evidence for a significant causal association between physical inactivity and rectal (or colorectal) cancer. A recent review of studies in which separate risk estimates for colon and rectal cancers were provided found no association between physical activity and rectal cancer in $80 \%$ of studies included in their review (Thune and Furberg 2001). The authors suggest that the apparent protective effect of physical activity on colorectal cancer may be derived, in the main, from the association between physical activity and colon cancer (Thune and Furberg 2001). On the basis of this review rectal (or colorectal) cancer was not included in this study.

## Breast cancer

The majority of studies investigating the benefits of physical activity and breast cancer report a reduction in the risk of breast cancer among physically active women (Gammon et al. 1998; Latikka et al. 1998; Verloop et al. 2000). There is substantial evidence that discretionary-time and/or occupational physical activity is associated with approximately a $30 \%$ reduction in the risk of breast cancer in pre-, peri- and post-menopausal women (Thune and Furberg 2001). Increased number of lifetime ovulatory cycles and cyclic estrogen has been proposed as a risk factor for breast cancer (Henderson et al. 1985). The impact of regular physical activity on the secretion, metabolism, and excretion of the sex hormones estradiol and progesterone provides a biologically plausible causal mechanism for the reduction of risk for breast cancer among physically active women. However, elevated relative and absolute estrogen levels may increase the risk of breast cancer (McTiernan et al. 1996). The strength and nature of the evidence for an association between physical inactivity and risk of breast cancer meets criteria for inclusion.

## Prostate cancer

There is limited evidence showing vigorous activity may provide a protective effect against prostate cancer in men (Giovannucci et al. 1998) and other researchers have not found such a relationship (Liu et al. 2000). A recent review of 24 studies found that 14 suggested an inverse association of physical activity on prostate cancer, six showed no asso-
ciation, and an increased risk of prostate cancer was observed among the most physically active men in four other studies (Friedenreich and Thune 2001). Given these equivocal results prostate cancer was not included in this project.

## MUSCULOSKELETAL CONDITIONS

Participation in physical activity throughout the course of life can increase, maintain or reduce the decline of musculoskeletal health that generally occurs with ageing in sedentary people (Brill et al. 2000). Participation by older adults can help maintain strength and flexibility, resulting in an ability to continue to perform daily activities (Brill et al. 2000; Huang et al. 1998; Simonsick et al. 1993). Furthermore, participation can reduce the risk of falling and hip fractures in older adults. (Grisso et al. 1997; Lord 1995) Evidence for associations between physical inactivity and low back pain, osteoporosis and falls, and oesteoarthritis are examined below.

## Low back pain

Low back pain is a term applied to a group of conditions or symptoms where there is pain, muscle tension and/or stiffness in the lower back. Low back pain may or may not include sciatica and other "nerve" discomfort. Physical activity is associated with both increased and decreased risk for low back pain. Certain specific activities such as heavy physical work, lifting, twisting, pulling, pushing that generally occur in occupational or during some specific discretionary-time activities (e.g. certain sports, heavy yard work) can increase the risk of low back pain (Picavet and Schouten 2000; Vuori 2001). However, participation in certain types of exercise may also reduce risk of low back pain. Four potential mechanisms for the protective effects of specific types of physical activities were proposed by Suni (2000), namely: increased abdominal and back muscle strength; better mobility and flexibility of trunk; increased endurance of trunk muscles assisting in the maintenance of motor control and stability; and increased capacity for appropriate motor skills.

Although there is some evidence from randomized controlled trials to show that low back pain can be prevented by participation in specific exercises, the evidence on participation in general physical activity and reduction on risk of low back pain is neither consistent nor strong (Vuori 2001). The outcomes are often poorly defined in many studies and this is in part due to the problematic nature of the term "low back pain", which is used for a collection of conditions with various International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10) codes (M45-48). Further, the type of physical activity associated with a reduced risk of low back pain has yet to be fully elucidated. Since participation in specific types of activity can not be identified at a population level there is insufficient information on which to
quantify hazard. Low back pain therefore was not included as a health outcome in this project.

## Osteoporosis and falls

Osteoporosis is characterized by low bone mass and structural deterioration of bone tissue leading to bone fragility and increased risk of fractures. The development of osteoporosis has been shown to be associated with physical inactivity (Drinkwater 1994). The biological mechanism proposed for the benefits of physical activity is via the impact on bone density. Put simply, bone cells respond to mechanical loading by increasing bone mass through improvements between bone formation and bone resorption (Lanyon 1993) and this process is mediated by glucose-6phosphate, prostaglandins and nitric oxide (Pitsillides et al. 1995; Tang et al. 1995; Turner et al. 1995). Cross-sectional studies show that participation in weight-bearing physical activity is positively associated with bone density (Gutin and Kasper 1992). Undertaking weight-bearing activity is particularly important in the development of peak bone density for adolescents (Welten et al. 1994) and for middle-aged women (Zhang et al. 1992).

Although this mechanism is strongly supported by the American College of Sports Medicine (ACSM), their position statement on osteoporosis states the types of activity most effecting such change are still not clear (1995). A recent review came to similar conclusions (Vuori 2001).

Systematic reviews of the literature have identified the beneficial role of physical activity in reducing the risks of falls in the elderly (Gillespie and McMurdo 1998; Kujala et al. 2000). Participation in physical activity is likely to be beneficial through an increase in bone strength, muscle strength, balance and coordination (Gregg et al. 2000). Currently osteoporosis is not listed as an outcome in the GBD classification system for diseases and injuries, but rather is grouped with "other musculoskeletal conditions". Given we were unable to classify what proportion of other musculoskeletal conditions are attributable to osteoporosis, or quantify the proportion of osteoporosis attributable to physical inactivity, falls were excluded from this project. Excluding osteoporosis and falls is based upon application of the specific criteria of this report, and does not indicate that physical activity is not important in fall prevention, or that there is not evidence suggesting a causal relationship.

## Osteoarthritis

Physical activity appears to be beneficial for controlling the symptoms of osteoarthritis and maintaining the health of joints. However there is limited evidence that physical activity itself can prevent osteoarthritis.

Conversely, there is evidence that certain types of physical activity, such as sustained participation in training and competition in elite sports, can lead to injuries and may increase the risk of osteoarthritis (Kujala
et al. 1994, 1995; U.S. Department of Health and Human Services 1996). However, these studies are often based on extremely small samples and specific populations and this limits their generalization to whole populations. Moreover, participation in recreational running, as opposed to competitive athletics, over a long period was not shown to increase risk of osteoarthritis (Lane 1995). Due to a lack of evidence for a causal association, and the absence of clear biologically plausible mechanisms, osteoarthritis was excluded from this review.

## DEPRESSION

Observational studies demonstrate that participation in discretionarytime and/or occupational physical activity can reduce symptoms of depression and possibly stress and anxiety (Dunn et al. 2001; Glenister 1996; Hassmen et al. 2000; Paffenbarger et al. 1994). Physical activity may also confer other psychological and social benefits that impact on health. For example, participation by individuals can help build selfesteem (Sonstroem 1984), social skills among children (Evans and Roberts 1987), positive self-image among women (Maxwell and Tucker 1992) and improve quality of life among children and adults (Hassmen et al. 2000; Laforge et al. 1999; Morans and Mohai 1991). These benefits are probably due to a combination of participation in the activity itself and from the social and cultural aspects that can accompany physical activity.

However, interpretation of the evidence on how physical activity might improve mental health is difficult due to less than ideal methodology and inconsistencies in study designs. Often studies involve small samples and use different definitions and measures of mental health outcomes. Thus, although the literature supports a beneficial effect on relieving symptoms of depression and anxiety, and as a treatment modality, there is currently limited evidence that physical activity can reduce the risk of depression and no clear evidence for a causal association. Exclusion of depression does not indicate that physical activity is not considered important for mental health; it indicates rather that the current evidence does not provide sufficient information for compliance with the inclusion criteria for this project.

## Obesity

There is considerable interest in the role of physical activity on weight. The available literature comprises a large body of observational studies showing that habitual physical activity over a lifetime can attenuate the increase in weight normally associated with increasing age, and participation in appropriate amounts of activity can lead to weight maintenance, or even weight loss (Grundy et al. 1999). The latter is especially true if physical activity is combined with a restriction of dietary energy intake. There are several proposed biological mechanisms for the association between physical inactivity and obesity (Hill and Melanson
1999). The most simple being that obesity occurs when energy intake (dietary intake) exceeds total energy expenditure (including the contribution of physical activity).

In this book, obesity is assessed as a risk factor (see Chapter 8). Moreover, it is currently not listed as a condition in the GBD classification system, instead it is included within the group "endocrine disorders". For these reasons obesity did not meet our inclusion criteria. Like depression, osteoporosis and other excluded disease end-points, exclusion does not imply the lack of any identified association with physical activity nor any lack of importance. Indeed, for obesity there is an urgent need for research and public health interventions to control and reverse the epidemic. Many health outcomes mediated through obesity are also among those considered above.

### 3.2 Search strategies for data sources

Medline and manual searches were conducted to identify review articles, individual studies and published meta-analyses of studies on the relationship between physical activity and the selected disease outcomes. The search was limited to studies published in English from 1980 to 2001. Keywords used were physical activity or exercise, along with at least one of the relevant disease outcomes listed below.

## Coronary heart disease OR cardiovascular disease

Overall our search identified a total of seven quantitative or qualitative reviews on the relationship between physical activity and ischaemic heart disease. These were: a qualitative review by Powell et al. (1987); a quantitative replication of Powell et al. by Berlin and Colditz (1990); a qualitative review undertaken as part of the 1996 U.S. Surgeon General's report on physical activity and health (U.S. Department of Health and Human Services 1996); a quantitative meta-analysis of 12 cohort studies by Eaton in 1992; a review of the dose-response relationship by Kohl (2001); a review of physical fitness vs physical activity by Blair et al. (2001); and a recent quantitative meta-analysis of all studies included in the U.S. Surgeon General's report along with any subsequent papers by Williams (2001).

## Stroke OR cardiovascular disease

Studies included in the recent reviews by Kohl (2001), Blair et al. (2001) and the U.S. Surgeon General's report (U.S. Department of Health and Human Services 1996) were obtained. Medline and manual searches revealed only a small number of additional articles not included in these documents, and these were obtained as well.

## Colon cancer Or breast cancer

Recent reviews by Thune and Furberg (2001) and McTiernan et al. (1998) provided a comprehensive coverage of the literature. We found
and obtained only a limited number of additional articles with at least one of the exposure titles (physical activity or exercise) which had not been included in either of these reviews.

Non-Insulin-dependent diabetes Or NIDDM OR type II diabetes MELLITUS
A recent review by Kelley and Goodpaster (2001) provided a list of relevant papers. Our search revealed only one additional review and a small number of articles with at least one of the exposure titles (physical activity or exercise).

### 3.3 Criteria for study inclusion

The following criteria were established for inclusion of studies in the meta-analyses:

- timing of exposure (level of physical inactivity) preceded the health outcome;
- at least two categories of physical activity were included;
- the health outcome(s) of interest were defined separately (specifically, ischaemic heart disease could be separated from cardiovascular disease and ischaemic and heamorrhagic stroke were treated separately);
- the instrument used to measure physical activity was described;
- if the exposure variable was part of a larger intervention (e.g. effect of a combined diet and exercise programme), the physical inactivity component alone could be estimated;
- demographic information was provided on the study population;
- sample size was provided;
- loss during follow up was less than $20 \%$ (if applicable);
- relative risks were published or it was possible to calculate them; and
- confidence intervals or standard errors were published or could be computed.
In addition to the above criteria, if there were multiple publications concerning the same outcome from a single cohort or trial, only the most recent publication that satisfied the inclusion criteria was selected.


### 3.4 DESCRIPTION OF INCLUDED STUDIES, INCLUDING METHODOLOGICAL QUALITIES

## ISCHAEMIC HEART DISEASE

The search identified 43 papers on physical activity and ischaemic heart disease outcomes, representing 30 cohorts. After critically reviewing each of the papers against the inclusion criteria 20 papers were retained covering 18 separate cohorts. These studies are summarized in Table 10.13.

Six of the 23 excluded papers used fitness measures rather than a measure of physical activity (Blair et al. 1989; Ekelund et al. 1988; Hein et al. 1992; Lie et al. 1985; Mundal et al. 1987; Peters et al. 1983). Of the remainder, four presented total cardiovascular disease as the outcome (LaCroix et al. 1996; Sherman et al. 1994a, 1994b, 1999), nine were reports covering cohorts which were already included (Donahue et al. 1988; Johansson et al. 1988; Morris et al. 1980; Paffenbarger et al. 1984; Seccareccia and Menotti 1992; Shaper et al. 1991; Slattery and Jacobs 1988; Stampfer et al. 2000; Yano et al. 1984), in one case a relative risk could not be calculated (Pomrehn et al. 1982), and in three cases standard errors could not be determined (Garcia-Palmieri et al. 1982; Kannel et al. 1986; Lindsted et al. 1991).

Sample sizes were generally of the order of a several thousand people, with the median sample size around 8000 , but ranged from 636 to 99029 . Most studies adjusted for confounding factors notably age, sex and smoking. Many studies adjusted for the intermediate factors identified as being in the causal pathway, namely blood pressure and blood cholesterol.

Mortality was ascertained by linkage or review of state, municipal or national death records. Incident cases were identified through hospital discharge lists, linkage with health registers, self-report (with and without verification by hospital records), or by abstraction and examination of hospital records.

Twelve studies were conducted in western Europe and eight studies were from North America. No studies were found from Africa, Asia or the Eastern Mediterranean. Study populations were generally of middle socioeconomic status and mostly Caucasians, an exception being men of Japanese ancestry (Rodriguez et al. 1994).

Estimates of relative risk with and without an adjustment for blood pressure and cholesterol were extracted for meta-analyses where available (see Table 10.13). The total effect of physical inactivity, not the independent effect, is required to calculate the estimates of disease burden.

## ISCHAEMIC STROKE

Over 30 papers were considered for inclusion into our meta-analysis for ischaemic stroke. Only eight studies met our criteria, all of which had
Table 10.13 Summary of included studies relating to ischaemic heart disease

| Subregion | Study (year) | Name ${ }^{\text {a }}$ | $n$ | Cases | Sex | $\begin{gathered} \text { Age } \\ \text { (years) } \end{gathered}$ | Outcome | Follow-up (years) | Adjustment ${ }^{\text {b }}$ | $R R^{\text {c }}$ | $\begin{gathered} \text { Approximate } \\ \text { SE } \end{gathered}$ | 95\% Cl | PA measure ${ }^{\text {d }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AMR-A | Folsom et al.$(\mathrm{I} 997)^{*}$ | ARIC | 7852 | 97 | Female | 45-64 | IHD incidence ${ }^{e}$ | 4-7 | I,2,3,4,5,6 | 1.39 (1.64) | 0.34 | 0.71-2.70 | Sports index |
|  |  |  |  |  |  |  |  |  |  | 1.56 (1.79) | 0.33 | 0.82-2.98 | Leisure index |
|  |  |  | 6188 | 223 | Male |  |  |  |  | 1.20 (1.43) | 0.20 | 0.81-1.78 | Sports index |
|  |  |  |  |  |  |  |  |  |  | 1.12 (1.32) | 0.21 | 0.74-1.70 | Leisure index |
|  | Lee et al. (2000) | Harvard Alumni | 7307 | 482 | Male | Mean 66.1 | IHD incidence | 5 | 1,2,4,6 | 1.51 | 0.22 | 1.05-2.48 | DTPA + walking + stairs |
|  | Lee and Skerrett (2001)* | Women's Health | 39372 | 244 | Female | $\geq 45$ | IHD incidence | 4-7 | 1,2 | 1.82 | 0.20 | 1.23-2.69 | DTPA + walking + stairs |
|  | Leon et al. (1987) | MRFIT | 12138 | 225 | Male | 35-57 | IHD mortality | 6-8 | I,2,4,5 | 1.49 (1.56) | 0.16 | 1.09-2.04 | DTPA |
|  | Manson et al. (I999)* | US Nurses' Health | 72488 | 645 | Female | 40-65 | IHD mortality or AMI | 8 | I,2,3,4,5,6 | 1.52 (1.85) | 0.15 | 1.13-2.03 | DTPA + walking + stairs |
|  | Rodriguez et al. (1994) | Honolulu Heart Program | 8006 | 789 | Male | 45-68 | IHD incidence | 23 | I,2,3,4,5,6 | 1.05 (1.22) | 0.09 | 0.88-1.26 | All PA |


|  | $\begin{aligned} & \text { Sesso et al. } \\ & (2000)^{*} \end{aligned}$ | Harvard Alumni | 12516 | 2135 | Male | 39-88 | IHD incidence | 16 | I,2,3,4,6 | 1.23 (1.37) | 0.07 | 1.07-1.41 | Sports + walking + stairs |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Slattery et al. (1989) | US Railroad | 2548 |  | Male | 22-79 | IHD mortality | 17-20 | 1,2,4,5 | 1.28 (1.39) | 0.13 | 0.99-1.63 | DTPA |
| EUR-A | Bijnen et al. (1998) | Zutphen Elderly | 802 | 90 | Male | 64-84 | IHD mortality | 10 | 1,2 | 1.18 (1.18) | 0.26 | 0.71-1.96 | DTPA |
|  | Haapanen et al. $(1997)^{*}$ | $\ldots$ | 953 | 75 | Female | 35-63 | IHD incidence | 10 | 1,2 | 1.25 | 0.28 | 0.72-2.15 | DTPA |
|  |  |  | 754 | 108 | Male | 35-63 | IHD incidence | 10 |  | 1.98 | 0.25 | 1.18-3.15 | DTPA |
|  | Kaprio et al. $(2000)^{*}$ | Finnish Twins | 8177 | 723 | Male | 25-68 | IHD incidence | 2-20 | I,2,3,4,6 | 1.47 (1.92) | 0.16 | 1.07-2.01 | DTPA |
|  | Lakka et al. (I994)* | Kuopio | 1166 | 42 | Male | 42-60 | AMI | 2-8 | I,2,3,4,5,6 | 2.63 (2.94) | 0.51 | 0.97-7.15 | DTPA |
|  | Menotti and Seccareccia (1985) | Italian Railroad | 99029 | 614 | Male | 40-59 | AMI mortality | 5 | None | 1.97 | 0.12 | 1.56-2.49 | OPA |
|  | Morris et al. (1990) | British Civil Servants | 9376 | 474 | Male | 45-64 | IHD incidence | 10 | None | 1.32 | 0.15 | 0.98-1.77 | DTPA |
|  | Pekkanen et al. (1987) | Seven Countries | 636 | 106 | Male | 45-64 | IHD mortality | 20 | 1,2,4,5 | 1.30 (1.40) | 0.19 | 0.89-1.89 | All PA |
|  | Rosengren and Wilhelmsen (1997) | Goteborg | 7142 | 584 | Male | 47-55 | IHD mortality | 20 | 1,2,3,4,5,6 | 1.35 (1.45) | 0.14 | 1.03-1.78 | DTPA |

Table 10.13 Summary of included studies relating to ischaemic heart disease (continued)

| Subregion | Study (year) | Name ${ }^{\text {a }}$ | $n$ | Cases | Sex | Age (years) | Outcome | Follow-up (years) | Adjustment ${ }^{\text {b }}$ | $R R^{\text {c }}$ | Approximate SE | 95\% Cl | PA measure ${ }^{\text {d }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Salonen et al.(1982) | North Karelia | 3688 | 63 | Female | 30-59 | AMI | 6 | 1,2,3,4,5 | $\begin{aligned} & 1.50(1.70) \\ & 2.40(2.70) \end{aligned}$ | $\begin{aligned} & 0.31 \\ & 0.23 \end{aligned}$ | $\begin{aligned} & 0.82-2.75 \\ & 1.53-3.77 \end{aligned}$ | DTPA OPA |
|  |  |  | 3978 | 89 | Male | 30-59 | IHD mortality | 6 |  | $\begin{aligned} & 1.40 \text { (1.60) } \\ & 1.6 \text { (1.70) } \end{aligned}$ | $\begin{aligned} & 0.23 \\ & 0.19 \end{aligned}$ | $\begin{aligned} & 0.89-2.20 \\ & 1.10-2.32 \end{aligned}$ | DTPA OPA |
|  | Salonen et al. (1988) | North Karelia | 15088 | 102 | Male and female | 30-59 | IHD mortality | 6 | I,2,3,4,5,7 | 1.30 | 0.11 | 1.13-1.74 | All PA |
|  | Shaper et al. (1994) | British Regional | 5694 | 311 | Male | 40-59 | IHD incidence | 9.5 | I,2,3,5 | 2.50 | 0.22 | 1.64-3.85 | DTPA |
|  | Sobolski et al. (1987) | Belgian Fitness | 2106 | 62 | Male | 40-55 | IHD incidence | 5 | none | 0.76 | 0.35 | 0.38-1.51 | All PA |

Key: SE, standard error (of log relative risk); CI, confidence interval; PA, physical activity; IHD, ischaemic heart disease; AMI, acute myocardial infarction. a Name refers to the common name of the cohort study.
Adjusted for: $\mathrm{I}=$ age, $2=$ smoking, $3=\mathrm{BMI}$ or waist/hip ratio, $4=$ blood pressure, $5=$ cholesterol, $6=$ type II diabetes, $7=$ sex.
Relative risk estimate with adjustment as indicated in footnote b. Figure in brackets indicates relative risk estimate NOT adjusted for blood pressure and cholesterol and some other factors. DTPA refers to discretionary-time activity; Sports index refers to organized sporting activities; Leisure index refers to DTPA excluding organized sports; OPA refers to occupational-related activity.
Incidence includes both morbidity and mortality.

* Indicates data from study used to calculate summary relative risk for level 2 exposure (insufficiently active)
been published since 1990 (Table 10.14). The majority of studies were excluded because they did not separate the subtypes of stroke; instead they reported risk estimates for the combined outcome of ischaemic and heamorrhagic stroke. The studies included covered both males and females and the age range of $40-74$ years. Seven of the eight studies were from North America (Abbott et al. 1994; Evenson et al. 1999; Gillum et al. 1996; Hu et al. 2000; Kiely et al. 1994; Lee et al. 1999; Sacco et al. 1998) and one study was conducted in Europe (Sweden) (Harmsen et al. 1990). Median sample size was811 1473.

For stroke, cases were identified through hospital discharge lists, linkage with health registers, self-report (in some cases verified by hospital records) or by abstraction and examination of hospital records. Deaths were identified through linkage with national, state or municipal registries.

Details of each study and relative risk estimates are presented in Table 10.14.

## CANCERS

Over 100 papers published since 1980 were identified for the two sitespecific cancer types, namely, colon and breast cancer. For both types of cancer, the majority of cases were identified through hospital discharge lists, cancer registries or by direct hospital recruitment (for case-control studies). In a small number of studies, self-reported data were used, but in all but one study these data were verified by review of hospital records. Deaths were identified via linkage with national, municipal or state records.

Sample sizes varied greatly, with a median sample size of 1844 for breast cancer and 1472 for colon cancer. The studies of breast cancer included only females, a single study looking at male breast cancer was excluded. Colon cancer studies appeared to cover both sexes equally. The majority of studies controlled for various important confounding or intermediary factors including age, smoking, family history of cancer and BMI. In addition, parity, use of hormone replacement therapy, use of oral contraceptives and menopausal status were controlled for in studies on breast cancer, and alcohol consumption, caloric intake and dietary factors (such as fat and fibre intake) were considered in studies on colon cancer.

The majority of studies for both types of cancer were conducted in North America, western Europe and the Western Pacific (e.g. China and Japan) (Kato et al. 1990; Matthews et al. 2001). Details of the 43 studies on breast cancer and 30 studies on colon cancer are shown in Tables 10.15 and 10.16 .

## Type II DIABETES

Twenty-two recent studies (see Table 10.17) were identified addressing the relationship between physical inactivity and type II diabetes. A
Table IO.14 Summary of included studies relating to ischaemic stroke

| Subregion | Study (year) | $n$ | Cases | Sex | Age (years) | Outcome | Follow-up (years) | Adjustment ${ }^{\text {a }}$ | $R R^{\text {b }}$ | $\begin{gathered} \text { proxime } \\ \text { SE } \end{gathered}$ | 95\% Cl | PA measure ${ }^{\text {c }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AMR-A | Abbott et al. $(\mathrm{I} 994)^{*}$ | $\begin{aligned} & 1854 \\ & 1257 \end{aligned}$ | $\begin{aligned} & 110 \\ & 116 \end{aligned}$ | Male | 55-68 | Incidence ${ }^{\text {d }}$ | 22 | I,4,5,6,7 | $\begin{aligned} & 1.80(2.00) \\ & 1.20(3.7) \end{aligned}$ | $\begin{aligned} & 0.26 \\ & 0.21 \end{aligned}$ | $\begin{aligned} & 1.10-3.10 \\ & 0.80-1.80 \end{aligned}$ | All PA (non-smokers) <br> All PA (smokers) |
|  | Evenson et al. (1999) | 14575 | 189 | Male and female | 45-64 | Incidence | 7.2 | I,2,3,4,6,7 | $\begin{aligned} & 1.22 \\ & 1.11 \\ & 1.43 \end{aligned}$ | $\begin{aligned} & 0.23 \\ & 0.15 \\ & 0.21 \end{aligned}$ | $\begin{aligned} & 0.78-1.91 \\ & 0.83-1.49 \\ & 0.95-2.16 \end{aligned}$ | Sport PA <br> DTPA (non-sports) OPA |
|  | Gillum et al. $(1996)^{*}$ | $\begin{array}{r} 1285 \\ 1083 \\ 1473 \\ 1240 \\ 771 \end{array}$ | $\begin{array}{r} 60 \\ 186 \\ 48 \\ 179 \\ 94 \end{array}$ | Male <br> Female <br> Male and female | 45-64 <br> 65-74 <br> 45-64 <br> 65-74 <br> 45-74 | Incidence | 10-16 | I,2,3,4,5,6,8 | $\begin{aligned} & 1.10(1.10) \\ & 1.34(1.49) \\ & 2.89(3.05) \\ & 1.47(1.6) \\ & 1.43(1.43) \end{aligned}$ | $\begin{aligned} & 0.36 \\ & 0.20 \\ & 0.61 \\ & 0.26 \\ & 0.37 \end{aligned}$ | $\begin{aligned} & 0.54-2.23 \\ & 0.90-2.00 \\ & 0.87-9.55 \\ & 0.88-2.44 \\ & 0.70-2.94 \end{aligned}$ | DTPA <br> DTPA <br> DTPA <br> DTPA <br> DTPA |
|  | Hu et al. (2000)* | 72488 | 258 | Female | 40-65 | Incidence | 8 | I,2,3,4,5,6 | 1.20 (1.45) | 0.18 | 0.85-1.71 | All PA |
|  | Kiely et al. $(1994)^{*}$ | $\begin{aligned} & 1228 \\ & 1676 \end{aligned}$ | $\begin{aligned} & 107 \\ & 127 \end{aligned}$ | Male Female | 28-62 | Incidence | 18 | 1,2,3,4,5,6,8 | $\begin{aligned} & 1.89(2.27) \\ & 0.83(1.05) \end{aligned}$ | $\begin{aligned} & 0.23 \\ & 0.25 \end{aligned}$ | $\begin{aligned} & 1.20-2.96 \\ & 0.51-1.35 \end{aligned}$ | All PA <br> All PA |
|  | Lee et al. (1999)* | 21823 | 437 | Male | 40-84 | Incidence | 11.1 | I,2,3,4,5,6,7,8 | 1.05 | 0.13 | 0.82-1.36 | DTPA |
|  | Sacco et al. (1998) | $\begin{aligned} & 284^{e} \\ & 394^{e} \end{aligned}$ | $\begin{aligned} & 163 \\ & 206 \end{aligned}$ | Male Female | $\geq 40$ | Incidence |  | I,2,3,4,6,7,8 | $\begin{aligned} & 2.78 \\ & 2.94 \end{aligned}$ | $\begin{aligned} & 0.29 \\ & 0.35 \end{aligned}$ | $\begin{aligned} & 1.57-4.90 \\ & 1.48-5.84 \end{aligned}$ | DTPA <br> DTPA |
| EUR-A | Harmsen et al. (1990) | 7495 | 69 | Male | 47-55 | Mortality | 11.8 | None | 1.20 | 0.27 | 0.70-2.00 | All PA |
| Key: SE, standard error (of log relative risk); PA, physical activity. |  |  |  |  |  |  |  |  |  |  |  |  |
| Adjusted for: $\mathrm{I}=$ age, $2=$ smoking, $3=\mathrm{BMI}$ or waist/hip ratio, $4=$ blood pressure, $5=$ cholesterol, $6=$ type II diabetes, $7=$ sex, $8=$ history of cardiovascular disease.Relative risk estimate with adjustment as indicated in footnote a. Figure in brackets indicates relative risk estimate NOT adjusted for blood pressure and cholesterol and some other |  |  |  |  |  |  |  |  |  |  |  |  |
| Sports PA refers to organized sporting activities; DTPA refers to discretionary-time activity; OPA refers to occupational-related activity. |  |  |  |  |  |  |  |  |  |  |  |  |
| Incidence includes both morbidity and mortality. |  |  |  |  |  |  |  |  |  |  |  |  |
| Case-control study. "n" is number of controls. |  |  |  |  |  |  |  |  |  |  |  |  |
| Indicates data from study used to calculate summary relative risk for level 2 exposure (insufficiently active). |  |  |  |  |  |  |  |  |  |  |  |  |

Table 10.15 Summary of studies relating to breast cancer

| Subregion | Study (year) | n |  | Age (years) | Follow-up (years) | Adjustment ${ }^{\text {a }}$ | $R R^{\text {b }}$ | Approximate SE | 95\% CI | PA measure ${ }^{\text {c }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Cases | Controls |  |  |  |  |  |  |  |
| AMR-A | Albanes et al. (1989)* | 122 | $740{ }^{\text {e }}$ | 25-74 | 7-13 | 1 | 1.00 | 0.25 | 0.61-1.63 | DTPA |
|  |  |  |  |  |  |  | 1.11 | 0.31 | 0.60-2.04 | OPA |
|  | Bernstein et al. (1994) | 545 | 545 | $\leq 40$ |  | 3,4,5,7,8 | 1.25 | 0.20 | 0.84-1.85 | DTPA |
|  | Breslow et al. (2001) | 138 | 6160 | $\geq 50$ | 8-10 | Yes | 3.03 | 0.45 | 1.25-7.32 | DTPA |
|  | Calle et al. (1998) | 1780 | $563395^{\text {d }}$ | $\geq 29$ | 9 | Yes | 0.91 | 0.07 | 0.79-1.04 | OPA |
|  | Carpenter et al. (1999) | 1123 | 904 | 55-64 |  | 1,3,5,8 | 1.14 | 0.10 | 0.93-1.38 | DTPA |
|  | Cerhan et al. (1998)* | 46 | 1806 | 65-102 | 11 | Yes | 5.00 | 0.74 | 1.17-21.32 | DTPA |
|  | Chen et al. (1997) | 747 | 961 | 21-45 |  | 1,3,4,5,8 | 1.05 | 0.13 | 0.82-1.36 | DTPA |
|  | Coogan et al. (1997) | 4863 | 6783 | 17-74 |  | 1,3,4,5,8 | 1.16 | 0.06 | 1.03-1.31 | OPA |
|  | Coogan and Aschengrau (1999) | 233 | 670 | All ages |  | 1 | 1.25 | 0.30 | 0.69-2.25 | OPA (ever held job with at least moderate activity) |
|  | Dorgan et al. (1994) | 117 | $2307^{\text {d }}$ | $\geq 35$ | 28 | 1,4,5,9 | 0.77 | 0.31 | 0.42-1.41 | All PA |
|  | Fraser and Shavlik (1997) | 218 | $20341^{\text {d }}$ | $\geq 24$ | 6 | 1,3,4,5,6,8,9 | 1.27 | 0.14 | 0.97-1.67 | All PA |
|  | Friedenreich et al. (2001)* | 462 | 475 | $<80$ |  | 1,2,3,6,8 | 0.87 | 0.20 | 0.59-1.29 | All PA (pre-menopausal) |
|  |  | 771 | 762 | $<80$ |  |  | 1.33 | 0.15 | 0.99-1.79 | All PA (post-menopausal) |
|  | Frisch et al. (1987) | 69 | $5398{ }^{\text {d }}$ | 18-22 | 56 | 1,2,3,4,6,7,8 | 1.86 | 0.32 | 1.00-3.47 | DTPA (college athletics) |
|  | Gammon et al. (1998) | 1668 | 1505 | <45 |  | 3,9 | 0.98 | 0.10 | 0.81-1.19 | DTPA |

Table IO.I5 Summary of studies relating to breast cancer (continued)

| Subregion | Study (year) | n |  | Age (years) | Follow-up (years) | Adjustment ${ }^{\text {a }}$ | $R R^{\text {b }}$ | Approximate SE | 95\% Cl | PA measure ${ }^{\text {c }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Cases | Controls |  |  |  |  |  |  |  |
| Gilliland et al. (2001)* |  | 171 | 210 | 30-74 |  | I,4,5,6,7 | 2.70 | 0.36 | 1.33-5.47 | DTPA (post-menopausal Hispanic) |
|  |  | 228 | 224 |  |  | I,4,5,6,7 | 2.04 | 0.29 | 1.16-3.60 | DTPA (post-menopausal non-Hispanic) |
|  |  | 128 | 152 |  |  | 1,4,5,6 | 2.04 | 0.40 | 0.93-4.47 | DTPA (pre-menopausal Hispanic) |
|  |  | 115 | 183 |  |  | 1,4,5,6 | 0.69 | 0.39 | 0.32-1.49 | DTPA (pre-menopausal non-Hispanic) |
| Lee et al. (2001b)* |  | 411 | 39322 | $\geq 45$ | 4 | I,3,4,5,6,7,8,9 | 1.16 | 0.15 | 0.87-1.56 | EE of DTPA + stairs (all women) |
|  |  | 261 | 21482 |  |  | 1,3,4,5,6,7,8 | 1.28 | 0.19 | 0.88-1.86 | EE of DTPA + stairs (post-menopausal) |
|  | Marcus et al. (1999) | 790 | 864 | 20-74 |  | Not specified | 1.25 | 0.13 | 0.97-1.61 | DTPA (at age 12) |
|  | McTiernan et al. (1996)* | 537 | 492 | 50-64 |  | 1 | 1.67 | 0.21 | 1.10-2.52 | DTPA |
|  | Mittendorf et al. (1995) | 6888 | 9539 | 17-74 |  | Yes | 1.00 | 0.10 | 0.80-1.20 | DTPA (at age 14-22) |
|  | Moore et al. (2000)* | 1380 | $37105^{\text {d }}$ | 55-69 | 12 | 1,3,5,6,8,9 | 1.09 | 0.09 | 0.91-1.30 | DTPA |
|  | Rockhill et al. (1998) | 372 | $116671^{\text {d }}$ | 25-42 | 16 | Yes | 0.91 | 0.18 | 0.64-1.29 | DTPA in adolescence |
|  | Rockhill et al. (1999)* | 3137 | $121701^{\text {d }}$ | 30-55 | 16 | 1,3,4,5,6,8,9 | 1.12 | 0.05 | 1.02-1.24 | DTPA |
|  | Sesso et al. (1998) | 109 | $1566{ }^{\text {d }}$ | $\geq 40$ | 31 | 1,3 | 0.56 | 0.43 | 0.24-1.29 | DTPA (pre-menopausal) |
|  |  |  |  |  |  |  | 2.04 | 0.29 | 1.16-3.60 | DTPA (post-menopausal) |
|  | Shoff et al. (2000) | 4614 | 5817 | 20-74 |  | 3,4,5,8,9 | 1.08 | 0.08 | 0.92-1.26 | DTPA |
|  | Steenland et al. (1995) | 163 | $14407{ }^{\text {e }}$ | 25-74 | 13-17 | Yes | 1.11 | 0.30 | 0.62-2.00 | All PA |

$$
\begin{aligned}
& \text { Not specified } \\
& \text { I,3,4 } \\
& \text { I,3 } \\
& \text { I,2,3,4,6,7,8 } \\
& \text { I,2,3,4,7,8 } \\
& \text { I,3,6,7 } \\
& \text { I,4,5,8,9 } \\
& \text { I,3 } \\
& \text { Yes } \\
& \text { । } \\
& \text { I,3,4 } \\
& \text { I,2,3,4,8 } \\
& \text { I,2 }
\end{aligned}
$$

| 0.23-0.88 | DTPA |
| :--- | :--- |
| $0.41-3.02$ | DTPA (pre-menopausal) |
| $0.31-1.93$ | DTPA (post-menopausal) |
| $1.07-5.32$ | All PA |
|  |  |
| $1.21-2.26$ | DTPA (college athletics) |
| $1.58-23.59$ | DTPA (college athletics) |
| $0.64-2.16$ | All PA (pre-menopausal) |
| $0.81-2.45$ | All PA (post-menopausal) |
| $1.23-3.26$ | DTPA (at $30-39$ years) |
| $1.01-3.82$ | OPA (at 30-39 years) |
| $0.91-2.13$ | OPA (pre-menopausal) |
| $1.16-2.23$ | OPA (post-menopausal) |
| $1.11-1.41$ | DTPA recently |
| $1.07-1.30$ | OPA |
| $1.25-1.64$ | DTPA |
| $0.78-2.16$ | DTPA (pre-menopausal) |
| $0.72-1.40$ | DTPA (post-menopausal) |
| $0.75-1.99$ | OPA (pre-menopausal) |
| $0.81-1.64$ | OPA (post-menopausal) |
| $1.06-1.98$ | DTPA |
| $0.21-2.46$ | OPA |

$$
\wedge \quad \text { ㄴ ๓ } \quad \text { の } \ddagger
$$

|  | Sternfeld (1992) |
| :---: | :---: |
|  | Taioli et al. (1995) |
|  | Wyrwich and Wolinsky (2000) |
|  | Wyshak and Frisch (2000) |
| EUR-A | Fioretti et al. (1999) |
|  | Levi et al. (1999)* |
|  | Mezzetti et al. (1998)* |
|  | Moradi et al. (2000) |
|  | Moradi et al. (1999) |
|  | Pukkala et al. (1993) |
|  | Thune et al. (1997) |
|  | Verloop et al. (2000) |
| EUR-B | Dosemeci et al. (1993) |

Table IO.I5 Summary of studies relating to breast cancer (continued)

| Subregion | Study (year) | n |  |  | Follow-up (years) | Adjustment ${ }^{\text {a }}$ | $R R^{\text {b }}$ | Approximate SE | 95\% Cl | PA measure ${ }^{\text {c }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Cases | Controls | Age (years) |  |  |  |  |  |  |
| WPR-A | Friedenreich and Rohan (1995) | 110 | 110 | 20-74 |  | Yes | 1.67 | 0.35 | 0.84-3.31 | DTPA (pre-menopausal) |
|  |  | 258 | 258 |  |  |  | 1.37 | 0.26 | 0.82-2.28 | DTPA (post-menopausal) |
|  | Hirose et al. (1995) | 606 | 14864 | All ages |  | Not specified | 1.35 | 0.13 | 1.05-1.74 | DTPA (pre-menopausal) |
|  |  | 439 | 6170 |  |  |  | 1.39 | 0.15 | 1.04-1.87 | DTPA (post-menopausal) |
|  | Hu et al. (1997) | 87 | 202 | All ages |  | Yes | 0.99 | 0.32 | 0.53-1.85 | DTPA (pre-menopausal) |
|  |  | 67 | 159 |  |  |  | 1.89 | 0.53 | 0.67-5.33 | DTPA (post-menopausal) |
|  | Ueji et al. (1998) | 148 | 296 | 26-69 |  | 3,4,5,8 | 3.33 | 0.59 | 1.05-10.59 | DTPA (pre-menopausal) |
|  |  |  |  |  |  |  | 2.00 | 0.71 | 0.50-8.04 | DTPA (post-menopausal) |
|  |  |  |  |  |  |  | 1.67 | 0.62 | 0.49-5.62 | OPA (pre-menopausal) |
|  |  |  |  |  |  |  | 1.43 | 0.57 | 0.47-4.37 | OPA (post-menopausal) |
| WPR-B | Matthews et al. (2001) | 1459 | 1556 | 25-64 |  | 1,5,8,9 | 2.13 | 0.14 | 1.62-2.80 | DTPA (all ages) |
| Key: SE, standard error (of log relative risk); PA, physical activity; EE, energy expenditure. |  |  |  |  |  |  |  |  |  |  |
| a Adjusted for: I = age, $2=$ smoking, $3=\mathrm{BMI}$ or waist/hip ratio, $4=$ parity, $5=$ age at first birth, $6=$ use of hormone replacement therapy, history of breast cancer, $9=$ menopausal status. |  |  |  |  |  |  |  |  |  |  |
| b Relative risk estimate with adjustment as indicated in footnote a. |  |  |  |  |  |  |  |  |  |  |
| DTPA refers to discretionary-time activity; OPA refers to occupational-related activity. |  |  |  |  |  |  |  |  |  |  |
| Cohort/cross-sectional study - number of controls is total sample size. |  |  |  |  |  |  |  |  |  |  |
| Indicates data from study used to calculate summary relative risk for level 2 exposure (insufficiently active). |  |  |  |  |  |  |  |  |  |  |

Table I0.16 Summary of studies relating to colon cancer

Table 10.16 Summary of studies relating to colon cancer (continued)

| Subregion | Study (year) | n |  | Sex | $\begin{gathered} \text { Age } \\ \text { (years) } \end{gathered}$ | Follow-up (years) (cohort) | Adjustment ${ }^{\text {a }}$ | $R R^{\text {b }}$ | Approximate |  | PA measure ${ }^{\text {c }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Cases | Controls |  |  |  |  |  | SE | 95\% Cl |  |
|  | Thun et al. (1992) | 611 | 2073 | Male |  |  | I,3,6,8 | 1.67 | 0.37 | 0.81-3.44 | All PA |
|  |  | 539 | 2081 | Female |  |  |  | 1.11 | 0.41 | 0.50-2.48 | All PA |
|  | Vena et al. (1985) | 210 | 1431 | Male | 30-79 |  | I | 1.97 | 0.18 | 1.38-2.80 | OPA (time in sedentary job) |
|  | White et al. (1996)* | 251 | 233 | Male | 30-62 |  | I | 1.27 | 0.20 | 0.86-1.87 | All PA |
|  |  | 193 | 194 | Female |  |  |  | 1.49 | 0.32 | 0.80-2.79 | All PA |
|  | Whittemore et al. (1990) | 61 | 255 | Male | 20-79 |  |  | $\begin{aligned} & 1.59 \\ & 2.50 \end{aligned}$ | $\begin{aligned} & 0.20 \\ & 0.43 \end{aligned}$ | $\begin{aligned} & 1.07-2.35 \\ & 1.08-5.8 । \end{aligned}$ | DTPA OPA |
|  |  | 46 | 198 | Female |  |  |  | 2.00 | 0.26 | $1.20-3.33$ | DTPA |
|  |  |  |  |  |  |  |  | 1.20 | 0.51 | 0.44-3.27 | OPA |
| EUR-A | Clemmensen (1998) | 88 | $5248{ }^{\text {d }}$ | Male | 40-59 | 15 | Not specified | 2.00 | 0.25 | $1.23-3.26$ | All PA |
|  | Gerhardsson et al. (1986) | 5100 | $1055715^{\text {d }}$ | Male | 20-64 | 19 | I | 1.3 | 0.07 | 1.13-1.49 | OPA |
|  | Gerhardsson et al. (1988) | 121 | $16477^{\text {d }}$ | Male and female | 42-82 | 14 | 1,7 | 3.6 | 0.52 | 1.3-9.8 | All PA |
|  | Gerhardsson et al. | 163 | 624 | Male | 40-79 |  | 1,3,5,6 | 2.33 | 0.50 | 0.64-4.52 | All PA |
|  | (1990) | 189 | 624 | Female |  |  |  | 1.69 | 0.48 | 0.91-5.96 | All PA |
|  | Lynge and Thygesen |  | $2000000^{\text {d }}$ | Male | 20-64 | 11 | I | 1.38 | 0.16 | 1.01-1.89 | OPA |
|  | (1988) |  |  | Female |  |  |  | 1.73 | 0.24 | 1.06-2.68 | OPA |


|  | Tavani et al. (1999)* | 688 | 2073 | Male | 19-74 |  | 1,4,5 | 1.41 | 0.16 | 1.03-1.93 | $\begin{aligned} & \text { OPA at age } \\ & 30-39 \end{aligned}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 537 | 2081 | Female | 19-74 |  |  | 2.04 | 0.20 | 1.38-3.02 | $\begin{aligned} & \text { OPA at age } \\ & 30-39 \end{aligned}$ |
|  | Thune and Lund (1996) | $\begin{array}{r} 236 \\ 99 \end{array}$ | $\begin{aligned} & 53242^{\mathrm{d}} \\ & 28274^{\mathrm{d}} \end{aligned}$ | Male Female | 40-54 | 16.3 | 1,3 | $\begin{aligned} & 1.03 \\ & 1.59 \end{aligned}$ | $\begin{aligned} & 0.22 \\ & 0.25 \end{aligned}$ | $\begin{aligned} & 0.67-1.59 \\ & 0.97-2.59 \end{aligned}$ | All PA All PA |
| EUR-B | Dosemici et al. (1993) | 93 | 486 | Male | All ages |  | 1,2 | 1.67 | 0.09 | 1.40-1.99 | OPA |
| WPR-A | Fraser and Pearce (1993) | 1651 Population |  | Male | All ages |  | Not specified | 1.25 | 0.08 | 1.07-1.46 | OPA |
|  | Kato et al. (1990) | 1716 | 16600 | Male | $\geq 20$ |  | I,2,4,8 | 1.92 (proximal) 1.52 (distal) | $\begin{aligned} & 0.17 \\ & 0.12 \end{aligned}$ | $\begin{aligned} & 1.38-2.67 \\ & 1.19-1.94 \end{aligned}$ | $\begin{aligned} & \text { OPA } \\ & \text { OPA } \end{aligned}$ |
|  | Kato et al. (1990) | 132 | 528 | Male and female |  |  | Not specified | $\begin{aligned} & 1.67 \\ & 2.00 \end{aligned}$ | $\begin{aligned} & 0.28 \\ & 0.28 \end{aligned}$ | $\begin{aligned} & 0.96-2.89 \\ & 1.16-3.46 \end{aligned}$ | $\begin{aligned} & \text { DTPA } \\ & \text { OPA } \end{aligned}$ |
| WPR-B | Tang et al. (1999) | 71 | 71 | Female | 33-80 |  | 1,2,4,6 | 5.26 | 0.64 | 0.45-5.56 | DTPA |
|  |  | 92 | 92 | Male |  |  |  | 1.59 | 0.70 | 1.33-20.74 | DTPA |
|  | Whittemore et al. (1990) | 81 | 567 | Male | 20-79 |  | Not specified | $\begin{aligned} & 0.85 \\ & 1.41 \end{aligned}$ | $\begin{aligned} & 0.40 \\ & 0.46 \end{aligned}$ | $\begin{aligned} & 0.39-1.87 \\ & 0.57-3.47 \end{aligned}$ | DTPA OPA |
|  |  | 66 | 305 | Female |  |  |  | 2.50 | 0.47 | 1.00-6.28 | DTPA |
|  |  |  |  |  |  |  |  | 1.69 | 0.57 | 0.55-5.18 | OPA |

Key: SE, standard error (of log relative risk); PA, physical activity.
a Adjusted for: $\mathrm{I}=$ age, $2=$ smoking, $3=\mathrm{BMI}$ or waist/hip ratio, $4=$ alcohol consumption, $5=$ caloric intake, $6=$ other dietary factors (e.g. fat, fibre), $7=$ sex, $8=$ family history of colon cancer.
Relative risk estimate with adjustment as indicated in footnote a.
OPA refers to occupational-related activity; DTPA refers to discretionary-time activity.
Cohort/cross-sectional study - number of controls is total sample size.
Indicates data from study used to calculate summary relative risk for level 2 exposure (insufficiently active).
Table 10.17 Summary of studies relating to type II diabetes

| Subregion | Author (year) | n |  | Sex | Age (years) | Follow-up (years) (cohort) | Adjustment ${ }^{\text {a }}$ | $R R^{\text {b }}$ | Approximate SE | 95\% Cl | PA measure ${ }^{\text {c }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Cases | Controls |  |  |  |  |  |  |  |  |
| AFR-D | Dowse et al. (1991)* | 288 | $2362^{\text {d }}$ | Male | 25-74 |  | 1,3,6 | 4.31 | 0.24 | 1.07-2.69 | All PA |
|  |  | 314 | $2669^{\text {d }}$ | Female |  |  |  | 1.70 | 0.75 | 1.00-18.61 | All PA |
| AFR-E | Levitt et al. (1999) | 69 | $974{ }^{\text {d }}$ | Male and female | $\geq 15$ |  | I,3,6,7 | 1.75 | 0.25 | 1.07-2.86 | DTPA |
| AMR-A | Folsom et al. (2000)* | 1997 | $34257^{\text {d }}$ | Female | 55-69 | 12 | 1,2,3,6 | 1.27 | 0.06 | 1.13-1.43 | DTPA |
|  | Fulton-Kehoe et al. (2001)* | 167 | 1100 | Male and female | 20-74 |  | I,3,6,7 | 1.39 | 0.30 | 0.77-2.50 | All PA |
|  | Gurwitz et al. (1994) | 185 | $2737^{\text {d }}$ | Male and female | $\geq 65$ | 6 | 1,4,7 | 1.5 | 0.19 | 1.03-2.18 | DTPA |
|  | Helmrich et al. (1994) | 202 | $5990^{\text {d }}$ | Male |  | 14 | I,3,4,6 | 1.32 | 0.1 | 1.09-1.61 | DTPA |
|  | Hu et al. (1999)* | 1419 | $70102{ }^{\text {d }}$ | Female | 40-65 | 12 | 1,2,3,4,5,6 | 1.30 | 0.09 | 1.09-1.55 | DTPA |
|  | Hu et al. (2001)* | 1058 | $37918^{\text {d }}$ | Male | 40-75 | 10 | 1,2,3,6 | 1.45 | 0.12 | 1.15-1.83 | DTPA |
|  | James et al. (1998)* | 78 | $916{ }^{\text {d }}$ | Male and female | 30-55 | 5 | 1,3,7 | 2.86 | 0.54 | $1.00-8.16$ | DTPA |
|  | Kriska et al. (1993) | 131 | $353{ }^{\text {d }}$ | Male and female | 37-59 |  | 1,3,7 | 2.17 | 0.26 | 1.30-3.61 | DTPA before age 35 |
|  | Leonetti et al. (1989) | 78 | 79 | Male | Mean 61 |  | 1,3,6 | 1.69 | 0.22 | 1.10-2.60 | DTPA age 15-20 |
|  | Manson et al. (1991) | 1303 | $87253^{\text {d }}$ | Female | 34-59 | 8 | 1,3,6 | 1.20 | 0.06 | 1.07-1.35 | DTPA |
|  | Manson et al. (1992) | 285 | $21271^{\text {d }}$ | Male | 40-84 | 5 | 1,2,3,4,5 | 1.43 | 0.14 | 1.09-1.88 | DTPA |


| EUR-A | Baan et al. (1999)* | 69 49 | $\begin{aligned} & 503^{d} \\ & 5 I 3^{d} \end{aligned}$ | Male Female | 55-75 |  | I,2,3,6 | $\begin{aligned} & 1.39 \\ & 2.56 \end{aligned}$ | $\begin{aligned} & 0.38 \\ & 0.48 \end{aligned}$ | $\begin{aligned} & 0.66-2.93 \\ & 1.00-6.56 \end{aligned}$ | DTPA DTPA |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Haapanen et al. (1997)* | $\begin{aligned} & 64 \\ & 54 \end{aligned}$ | $\begin{aligned} & 891^{d} \\ & 973^{\mathrm{d}} \end{aligned}$ | Male Female | 35-63 | 10 | 1,2 | $\begin{aligned} & 1.54 \\ & 2.64 \end{aligned}$ | $\begin{aligned} & 0.31 \\ & 0.37 \end{aligned}$ | $\begin{aligned} & 0.83-2.84 \\ & 1.28-5.45 \end{aligned}$ | DTPA DTPA |
|  | Lynch et al. (1996) |  | $897{ }^{\text {d }}$ | Male | Middle aged |  | 1,3,6 | 2.27 | 0.35 | 1.14-4.51 | DTPA |
|  | Perry et al. (1995) | 194 | $7735^{\text {d }}$ | Male | 40-59 | 11-13 | 1,2,3,4,5 | 2.50 | 0.35 | 1.26-4.96 | DTPA |
|  | Tuomilehto et al. (2001) | 27 | $265^{\text {d }}$ | Male and female | 40-65 | Mean 3.3 | 3 | 3.33 | 0.5 | 1.25-8.87 | DTPA |
|  | Wannamethee et al. (2000)* | 196 | $5159{ }^{\text {d }}$ | Male | 40-59 | 16.8 | 1,2,3,4,5 | 1.61 | 0.29 | 0.91-2.85 | DTPA |
| WPR-A | Okada et al. (2000) | 444 | $6013{ }^{\text {d }}$ | Male | 35-60 | 6-16 | 1,2,3,4,6 | 1.33 | 0.11 | 1.07-1.65 | DTPA |
|  | Todoroki et al. (1994)* | 48 | $1113^{\text {d }}$ | Male | 49-56 |  | 2,3,6 | 2.00 | 0.46 | 0.81-4.93 | DTPA |
| WPR-B | Pan et al. (1997) | 2731 | $11028{ }^{\text {d }}$ | Male and female | 25-64 |  | 1,3,4,6,7 | 1.18 | 0.05 | 1.07-1.30 | OPA |
|  | Taylor et al. (1984) |  | $640^{\text {d }}$ | Male | $\geq 20$ |  | 1,3 | 2.50 | 0.44 | $1.06-5.92$ | All PA |
|  |  |  | $595{ }^{\text {d }}$ |  |  |  |  | 2.10 | 0.25 | 1.29-3.43 | All PA |

Key: SE, standard error (of log relative risk); PA, physical activity.
Adjusted for: $\mathrm{I}=$ age, $2=$ smoking, $3=\mathrm{BMI}$ or waist/hip ratio, $4=$ blood pressure, $5=$ cholesterol, $6=$ family history of type II diabetes, $7=$ sex.
Relative risk estimate with adjustment as indicated in footnote a.
DTPA refers to discretionary-time activity; OPA refers to occupational-related activity.
Cohort/cross-sectional study - number of controls is total sample size.
diverse range of populations have been studied from the regions of Africa (Mauritius, South Africa and the United Republic of Tanzania), Asia (China), the Western Pacific (Japan and Fiji) as well as North America and western Europe.

For type II diabetes, cases in all case-control studies were identified using oral glucose tolerance tests (OGTT). In the majority of cohort studies, identification was by OGTT or fasting plasma glucose tests at follow-up, or by self-report, validated by physician or medical record review. In only one case was unverified self-reported data used.

Sample sizes ranged widely with a median sample size of 1113. All studies adjusted for a variety of important confounding or intermediary factors including age, BMI, blood pressure, cholesterol and family history of diabetes.

### 3.5 Methods for meta-analysis

Our general method for meta-analyses was similar to that used by Berlin and Colditz (1990) and Eaton (1992) and combined estimates of the log relative risks using an inverse-variance weighting scheme (Berlin and Colditz 1990; Eaton 1992). This method gives relatively more importance to studies with a larger number of cases and produces a wider confidence interval for the pooled relative risk estimate than would be obtained by other methods (such as the Mantel-Haenszel method). Confidence intervals for the summary risks were derived using a pooled standard error.

For ischaemic heart disease and ischaemic stroke we conducted two different meta-analyses. Firstly, we pooled risk estimates derived from analyses which had adjusted for two intermediary variables (blood pressure and cholesterol) and secondly, we pooled estimates of risk without adjustment for these intermediary variables. The former (with adjustment) removes the effect due to other factors and identifies the independent risk due to physical inactivity. Estimates of risk without adjustment for other variables in the causal pathway indicate the total effect of exposure to inactivity. The first method partials out the contribution of inactivity to various disease end-points while the second is useful in calculating the total effect of removing exposure to inactivity (Greenland 1987). We report both results to enable comparison with results from previous meta-analyses, all of which used adjusted data, and to allow computation of avoidable burden of disease (which required the unadjusted analyses). For type II diabetes and breast and colon cancer, estimates of risk were computed using only adjusted data as noted in Tables 10.15-10.17.

The problems associated with measurement of behavioural risk factors are well known and to date few satisfactory solutions are available. Given the difficulties associated with assessing a complex exposure variable like physical inactivity and the heterogeneity among the instruments used across the epidemiological studies, one approach is to intro-
duce an adjustment at the analytical stage. Adjustment for measurement error within the meta-analysis has been previously applied to other risk factors (Bashir and Duffy 1997). We employed this technique in our analyses and present the results both adjusted for measurement error and unadjusted for measurement error.

Overall three separate analytical approaches were conducted to derive summary estimates of relative risk, as summarized below.

- Adjusted for intermediary variables: This method was undertaken for all disease outcomes and used pooled relative risk estimates from studies in which the analyses had controlled for the intermediate factors (blood cholesterol and blood pressure). These results have not incorporated the adjustment for measurement error and so can be compared to other published results that have controlled for intermediaries.
- Adjusted for intermediary variables AND adjusted for measurement error: This method is the same as above with the addition of a statistical adjustment for measurement error. This method was conducted for all disease outcomes and represents an extension of the above method.
- Unadjusted for intermediary variables AND adjusted for measurement error: This method pooled relative risk estimates from studies in which the analyses had not controlled for blood pressure and cholesterol. It was conducted for only ischaemic heart disease and ischaemic stroke and, for those studies without unadjusted risk estimates, the adjusted risk estimates were included.


## MEASUREMENT ERROR

The major sources of error arise from the variability in the measurement and/or definitions of the health outcome and the exposure variable (i.e. physical inactivity). A further source of variability may be due to the timing of the exposure. These will be discussed in turn below.

## Error due to variable measures of exposure

In general, measurement error is a product of the different ways in which physical inactivity is measured and/or categorized. Using different measures and definitions can result in quite different estimates of the level of exposure to physical inactivity. We identified three main sources of potential measurement error for exposure.

Firstly, the questionnaires used in the epidemiological studies included in our meta-analyses vary in their attempt to measure physical activity undertaken in different domains. The studies included for ischaemic heart disease mostly evaluated discretionary-time physical activity (Haapanen et al. 1997; Lakka et al. 1994; Leon et al. 1987; Rosengren and Wilhelmsen 1997; Shaper et al. 1994; Slattery et al. 1989), although in
some studies this was defined as "sport" and "leisure activities" (Folsom et al. 1997) and in other studies it was extended to include other activities such as walking and stair climbing (Lee and Skerrett 2001; Lee et al. 2000; Manson et al. 1999). In the Harvard Alumni study "sports" activity plus walking and stair climbing were assessed (Sesso et al. 2000). In several studies "all" physical activity was evaluated but the exact meaning and coverage of domains required a careful inspection of the specific instruments (Morris et al. 1990; Pekkanen et al. 1987; Rodriguez et al. 1994; Salonen et al. 1988; Sobolski et al. 1987). Only one study included in our meta-analysis provided an estimate of risk based on only occupational physical activity and one other study provided separate relative risk estimates for occupational and discretionary-time activity (Menotti and Seccareccia 1985; Salonen et al. 1982). It is notable that the majority of studies published post 1980 focus more on discretionarytime activity whereas literature published between 1958-1980 is dominated by studies of work-related activity.

Pooling results from studies with measures of exposure across discretionary time, sports, walking, stair climbing and "all" activity could have a marked affect on the overall estimates of risk. One solution, previously used by Berlin and Colditz (1990), is to separately pool results from studies assessing different domains (e.g. occupation only, discretionary time only). However, in this study the pooled summary relative risk estimates were applied to prevalence estimates of exposure in which multiple domains had been considered. We therefore considered the inclusion of studies with diverse definitions of activity, possibly across more than one domain as acceptable. Nonetheless, this source of heterogeneity among the literature is noted, as is the desirability for much greater comparability of measures of exposure between studies.

A second source of measurement error could arise from the different instruments used to assess exposure in the various domains. Five of the 19 included cohort studies used the Minnesota Leisure Time Physical Activity questionnaire and three used the Harvard Alumni instrument, but the remaining 11 studies used different and sometimes unspecified survey tools. Different questionnaires have different properties and concomitant variation in the level of validity and reliability. The instruments used in the studies included in our meta-analyses assessed physical activity over differing referent time periods (e.g. the previous year, last week or usual week) and used various response formats (Folsom et al. 1997; Lakka et al. 1994; Lee and Paffenbarger 2000). To address this concern we included an adjustment for measurement error in our meta-analyses using information on the reliability of each instrument. The rationale and methodology are described below.

A third potential source of error relating to the exposure variable may result from the different ways in which physical activity data are categorized. For instance, different studies have used different methods of categorization (e.g. tertiles, quartiles, quintiles, sextiles) (Folsom et al.

1997; Lee and Paffenbarger 2000; Leon et al. 1987; Rodriguez et al. 1994; Rosengren and Wilhelmsen 1997; Sesso et al. 2000; Shaper et al. 1994), but not all studies provide a full description of the level of activity in each category to facilitate between-study comparison. In addition to the variability between studies there is also likely to be within-category variability (Blair and Jackson 2001). Indeed, Berlin and Colditz (1990) discussed this problem and concluded that this could explain why some studies failed to show a protective affect due to physical activity.

We attempted to reduce the error from pooling results from different studies using different categories by identifying from each study the categories of activity most consistent with our definitions of exposure. We selected the "sedentary" or lowest category from each study as the "most exposed" group and this was deemed most comparable to our level 1 exposure (inactive). Similarly, the categories of activity within each study were reviewed to identify the most comparable group(s) for our level 2 exposure (insufficient activity). Finally, and most importantly, we extracted from each study the estimate of relative risk pertaining to a referent group equivalent to our level 3 exposure, namely, a level of activity equal to at least 150 minutes of moderate-intensity activity per week (at least 2.5 hours of activity per week or an energy expenditure of at least $4000 \mathrm{~kJ} /$ week). Some studies did not clearly define this group and in these cases we carefully reviewed each category within the study and made a subjective assessment. In this regard, our meta-analysis differs from previous reviews which have chosen to assess the relative risk of inactivity using a referent group defined as the "most active" or "vigorous activity" or "high fitness".

## Error due to measures of disease end-points

Another possible source of measurement error arises from different definitions or classifications of the disease end-points, particularly when disease subtypes exist within a health outcome. For example, studies with ischaemic heart disease as an outcome may include mortality and/or morbidity from all types of ischaemic heart disease (including angina pectoris), only myocardial infarction or myocardial infarction and sudden death combined. However, the etiology of these conditions may differ, as may the contribution of physical inactivity.

In assessing ischaemic heart disease, Berlin and Colditz (1990) undertook separate meta-analyses for acute myocardial infarction, ischaemic heart disease incidence and for ischaemic heart disease mortality. They found no difference in the relative risk estimates from the analysis of each outcome separately, although the strength of association (defined by the range of confidence intervals) was greater for mortality than morbidity. Eaton (1992) examined the differences between pooled estimates from studies using ischaemic heart disease mortality and those using clinical ischaemic heart disease as health outcomes and also found no dif-
ference in the size or the strength of the association. Therefore, given no evidence to the contrary, we combined ischaemic heart disease mortality and morbidity in our meta-analyses.

There appears to be no association between physical activity and hemorrhagic stroke. As a result in studies assessing stroke, misclassification of subtypes may have led to null associations between physical activity level and disease outcome (Kohl 2001). Therefore, for inclusion in our meta-analyses, only studies that clearly differentiated and reported results for ischaemic stoke were included.

For type II diabetes, colon cancer and breast cancer there is a paucity of information on the level of any misclassification and the effect on relative risk estimates. However, any misclassification of colon cancer (i.e. colon, colorectal and rectal cancers) would, if random, most likely underestimate the relationship between physical activity and risk for colon cancer. The positive effect of physical activity would be diluted because of null associations included due to misclassification.

## Error due to timing and duration of exposure

Epidemiological data may also be inconsistent due to the timing of the exposure, in this case physical inactivity. However it is simpler to consider the timing of physical activity and ask when during the life course is activity required for risk reduction? For ischaemic heart disease and type II diabetes the activity is required to be regular and recent for risk reduction. However, for other conditions (e.g. breast cancer, colon cancer) activity during childhood or adolescence may be important to reduce risk.

Among the studies included in the meta-analyses a range of followup periods can be seen-for example, from 2 to 23 years for ischaemic heart disease, from 8 to 22 years for ischaemic stroke and from 4 to 56 years for breast cancer. Measurement error associated with assessing long-term physical activity patterns by a questionnaire administered only at baseline is likely to be important. It was, however, not possible to account for this in our meta-analyses but it is highly recommended that this issue be explored in future research.

## Adjustment for measurement error

We considered the issue of measurement error associated with self reported recall of physical (in)activity and concluded that such error could lead to an underestimate of the association between physical activity and disease end-points. Thus, an adjustment for measurement error was made to each estimate of relative risk extracted from the individual studies based on the reliability of the instrument used to measure exposure. Specifically, the adjustment involved multiplying the beta coefficients (log relative risks) by the inverse of the study-specific test-retest correlation coefficient as previously used with other risk factors (Bashir
and Duffy 1997), although we found no prior evidence of its use with physical inactivity.

The test-retest coefficient for each instrument was obtained in the first instance from the article as cited by the authors of the epidemiological study. When this was not available we searched the published literature including a compendium of instruments which contained an excellent summary of test-retest reliability data for many instruments (Pereira et al. 1997). If no information was found for a specific instrument but a sufficiently detailed description was available, we assigned a test-retest coefficient based on the correspondence between characteristics of the instrument as described by the authors and other known instruments. In those cases where little or no information on the instrument was provided the average score of known instruments was applied. This process was undertaken for all studies included for each of the disease outcomes. Table 10.18 reports the test-retest coefficient for the instruments used in studies addressing ischaemic heart disease.

## Heterogeneity

Although the fixed-effects model gives relatively more importance to bigger studies than smaller studies in the summary relative risk estimate, the effect of any heterogeneity not accounted for by study size remains. Despite our study inclusion criteria heterogeneity among the studies was significant for all conditions, except for ischaemic heart disease, as we describe below. No attempt to identify the sources of heterogeneity was undertaken. However, it is probable that differences in study design such as combining studies with different types of physical activity, different measurement instruments, different follow-up time, and differences in disease outcomes contributed to the heterogeneity (Berlin and Colditz 1990). Tests of heterogeneity and bias were conducted to assess the quality of the data as well as to reveal any evidence of heterogeneity and/or publication bias. The extent of bias was assessed by regressing the standard errors to the relative risk estimates and testing whether the regression coefficient was equal to zero (Sterne and Egger 2001).

Heterogeneity was assessed using funnel plots by drawing $95 \%$ confidence limits around the summary risk estimate for various values of the standard error. In the absence of heterogeneity, $95 \%$ of the point estimates should lie within these limits.

### 3.6 Attenuation for age

The relative risk associated with physical activity and some disease endpoints has been shown to be lower for older age groups compared with younger age groups (Sesso et al. 2000). This is consistent with an age attenuation seen in other intermediate risk factors such as systolic blood pressure and cholesterol (MacMahon et al. 1990). It is desirable for any calculation of attributable burden to include an attempt to better represent the differential risk across age. However, overall there were too few

Table IO.I8 Test-re-test reliability for instruments used in studies on ischaemic heart disease

| Authors | Year | Study name | Instrument | Test-re-test reliability |
| :---: | :---: | :---: | :---: | :---: |
| Sesso et al. | 2000 | Harvard Alumni | Harvard Alumni | 0.72 |
| Morris et al. | 1990 | British Civil Servants | (No specific name) | 0.75 |
| Rodriguez et al. | 1994 | Honolulu | Framingham | 0.45 |
| Menotti and Seccareccia | 1985 | Seven Counties Italian Railroad | (No specific name) | 0.70 |
| Rosengren and Wilhelmsen | 1997 | Goteborg Primary <br> Prevention | (No specific name) | 0.75 |
| Lee and Paffenbarger | 2000 | Harvard Alumni | Harvard Alumni | 0.72 |
| Lee et al. | 2001 a | US Women's Health | Harvard Alumni | 0.72 |
| Leon et al. | 1987 | MRFIT | Minnesota LTPA | 0.92 |
| Lakka et al. | 1994 | Kuopio | Minnesota LTPA | 0.92 |
| Sobolski et al. | 1987 | Belgian Fitness | Minnesota LTPA | 0.92 |
| Shaper et al. | 1994 | British Regional | Minnesota LTPA | 0.92 |
| Folsom et al. | 1997 | ARIC | Baecke Physical Activity | 0.93 |
| Salonen et al. | 19881982 | North Karelia (two cohorts) | (No specific name) | 0.75 |
| Slattery et al. | 1989 | US Railroad | Minnesota LTPA (Early version) | 0.82 |
| Pekkanen et al. | 1987 | Finish Cohort | (No specific name) | 0.75 |
| Haapanen et al. | 1997 | Finish Cohort | (No specific name) | 0.80 |
| Manson et al. | 1999 | US Nurses' Health | (No specific name) | 0.79 |
| Kaprio et al. | 2000 | Finnish Twins | (No specific name) | 0.75 |
| Bijnen et al. | 1998 | Zutphen Elderly | Zutphen Elderly | 0.93 |

data on physical inactivity by age to compute age-specific estimates of relative risks.

In lieu of this, the summary relative risk estimates from our metaanalyses were applied to the age categories 30-44 years, 45-59 years and 60-69 years on the basis that this age range was well represented by the studies included in our analyses. In addition, the summary estimate was applied to the 15-29-year age group for all disease end-points except breast cancer. In this case there was sufficient evidence to support differential relative risk estimates for pre-/peri- and postmenopausal women and we classified these groups as $15-44$ years, and $\geq 45$ years, respectively.

Reviewing data from two studies on ischaemic heart disease in which older age cohorts were examined (Bijnen et al. 1998; Sesso et al. 2000)
we found older adults had around half the excess of the average risk estimate from all other studies combined. This formed the basis of age attenuation estimates. Specifically, $25 \%$ of the excess risk was used to estimate risk for the $70-79$-year age group and $50 \%$ of the excess risk was used for the age group $\geq 80$ years. These algorithms were used to compute age attenuated estimates for all other disease end-points. It is possible that there is no attenuation across age for increased risk of cancer, employing the age attenuation described may have added to the likely underestimation of burden.

### 3.7 Extrapolation of hazard estimates

In general, there was a paucity of data on the effects of physical activity and reduction in risk of disease among populations not of European decent. For ischaemic heart disease and ischaemic stroke, the majority of studies were conducted with men of primarily European descent. For colon and breast cancer the population base was broader but still mostly of North American and European origin. Studies on type II diabetes included data from a much wider range of regions and the reduction in risk was in the same direction across different population groups. Like previous reviews (Powell and Blair 1994), we found no established reasons to suggest that the association between physical inactivity and chronic diseases would differ across diverse populations. Thus, in the absence of evidence to the contrary, we assume that the relative risk estimates for all conditions can be applied across all populations.

### 3.8 Estimated hazards

Where possible, separate relative risks for males and females were used to derive the summary estimates. If studies provided specific risk estimates for various age groups and/or for different domains of activity in addition to an overall summary risk, the estimate across all age groups and/or across all domains was extracted for our analyses (e.g. Morris et al. 1990). If no summary score was presented, the age/domain specific estimates were included as one study in the meta-analysis (e.g. Gillum et al. 1996).

## IsCHAEMIC HEART DISEASE

Table 10.19 reports the summary risk estimates from meta-analyses undertaken with and without an adjustment for measurement error. Funnel plots showed little evidence of bias but some heterogeneity in these meta-analyses (Figure 10.24). The regression test showed no evidence of bias ( $P=0.34$ ).

## IsCHAEMIC STROKE

The analyses described above for ischaemic heart disease were repeated for ischaemic stroke and the results are reported in Table 10.20. Funnel plots (Figure 10.25) showed some evidence of heterogeneity and large

Table 10.19 Summary relative risk estimates for ischaemic heart disease ${ }^{\text {a }}$ for level I (inactive) and level 2 (insufficiently active) exposure, by age and sex

|  | NO adjustment for measurement error ${ }^{\text {b }}$ |  | WITH adjustment for measurement error ${ }^{\text {b }}$ |  |
| :---: | :---: | :---: | :---: | :---: |
| Level I (inactive) |  |  |  |  |
| Age group (years) | Males | Females | Males | Females |
| 0-4 | NA | NA | NA | NA |
| 5-14 | NA | NA | NA | NA |
| 15-29 | 1.47 (1.39-1.56) | 1.47 (1.39-1.56) | 1.71 (1.58-I.85) | 1.71 (1.58-I.85) |
| 30-44 | 1.47 (1.39-1.56) | 1.47 (1.39-1.56) | 1.71 (1.58-1.85) | 1.71 (1.58-1.85) |
| 45-59 | 1.47 (1.39-1.56) | 1.47 (1.39-1.56) | 1.71 (1.58-1.85) | 1.71 (1.58-1.85) |
| 60-69 | 1.47 (1.39-1.56) | 1.47 (1.39-1.56) | 1.71 (1.58-1.85) | 1.71 (1.58-1.85) |
| 70-79 | 1.34 (1.26-1.42) | 1.34 (1.26-1.42) | 1.50 (1.38-1.61) | 1.50 (1.38-1.61) |
| $\geq 80$ | 1.21 (1.14-1.29) | 1.21 (1.14-1.29) | 1.30 (1.2I-I.4I) | 1.30 (1.21-1.4I) |

Level 2 (insufficiently active)

| Age group (years) | Males | Females | Males | Females |
| :---: | :---: | :---: | :---: | :---: |
| 0-4 | NA | NA | NA | NA |
| 5-14 | NA | NA | NA | NA |
| 15-29 | 1.31 (1.2I-I.4I) | 1.31 (1.21-I.41) | 1.44 (1.28-1.62) | 1.44 (1.28-1.62) |
| 30-44 | 1.31 (1.2I-I.41) | 1.31 (1.2I-I.4I) | 1.44 (1.28-1.62) | 1.44 (1.28-1.62) |
| 45-59 | 1.31 (1.21-I.41) | 1.31 (1.2I-I.4I) | 1.44 (1.28-1.62) | 1.44 (1.28-1.62) |
| 60-69 | 1.31 (1.2I-I.41) | 1.31 (1.2I-I.41) | 1.44 (1.28-1.62) | 1.44 (1.28-1.62) |
| 70-79 | 1.22 (1.13-1.32) | 1.22 (1.13-1.32) | 1.31 (1.17-1.48) | 1.31 (1.17-1.48) |
| $\geq 80$ | 1.14 (1.06-1.24) | 1.14 (1.06-1.24) | 1.20 (1.07-1.35) | 1.20 (1.07-1.35) |

NA Not applicable.
a Incidence and mortality.
b Summary risk estimates computed using estimates adjusted for confounding variables (e.g. age, sex) but NOT adjusted for blood pressure and cholesterol. If these were unavailable from any study the available overall adjusted relative risk estimate was used.
bias in these analyses (regression $P<0.001$ ). We were unable to account for this bias, however publication bias (i.e. preferential publication and increased citation of studies with significant results) and the restriction to studies published in English may have contributed.

## BREAST CANCER

Summary risk estimates were computed for pre- and peri-menopausal women combined and for post-menopausal women. These estimates

Figure 10.24 Funnel plot of studies used in ischaemic heart disease metaanalysis for ischaemic heart disease, level I exposure (inactive)


Figure 10.25 Funnel plot of studies used in ischaemic stroke meta-analysis for ischaemic heart disease, level I exposure (inactive)


Table 10.20 Summary relative risk estimates for ischaemic stroke ${ }^{\mathrm{a}}$ for level I (inactive) and level 2 (insufficiently active) exposure, by age and sex

|  | NO adjustment for measurement error ${ }^{\text {b }}$ |  | WITH adjustment for measurement error ${ }^{\text {b }}$ |  |
| :---: | :---: | :---: | :---: | :---: |
| Level I (inactive) |  |  |  |  |
| Age group (years) | Males | Females | Males | Females |
| 0-4 | NA | NA | NA | NA |
| 5-14 | NA | NA | NA | NA |
| 15-29 | 1.39 (1.24-1.56) | 1.39 (1.24-1.56) | 1.53 (1.31-1.79) | 1.53 (1.31-1.79) |
| 30-44 | 1.39 (1.24-1.56) | 1.39 (1.24-1.56) | 1.53 (1.31-I.79) | 1.53 (1.31-1.79) |
| 45-59 | 1.39 (1.24-1.56) | 1.39 (1.24-1.56) | 1.53 (1.31-I.79) | 1.53 (1.31-1.79) |
| 60-69 | 1.39 (1.24-1.56) | 1.39 (1.24-1.56) | 1.53 (1.31-I.79) | 1.53 (1.31-1.79) |
| 70-79 | 1.28 (1.14-1.44) | 1.28 (1.14-1.44) | 1.38 (1.18-1.60) | 1.38 (1.18-1.60) |
| $\geq 80$ | 1.18 (1.05-1.33) | 1.18 (1.05-1.33) | 1.24 (1.06-1.45) | 1.24 (1.06-1.45) |

Level 2 (insufficiently active)

| Age group (years) | Males | Females | Males | Females |
| :--- | :---: | :---: | :---: | :---: |
| $0-4$ | NA | NA | NA | NA |
| $5-14$ | NA | NA | NA | NA |
| $15-29$ | $0.97(0.87-1.15)$ | $0.97(0.87-1.15)$ | $1.10(0.89-1.37)$ | $1.10(0.89-1.37)$ |
| $30-44$ | $0.97(0.87-1.15)$ | $0.97(0.87-1.15)$ | $1.10(0.89-1.37)$ | $1.10(0.89-1.37)$ |
| $45-59$ | $0.97(0.87-1.15)$ | $0.97(0.87-1.15)$ | $1.10(0.89-1.37)$ | $1.10(0.89-1.37)$ |
| $60-69$ | $0.97(0.87-1.15)$ | $0.97(0.87-1.15)$ | $1.10(0.89-1.37)$ | $1.10(0.89-1.37)$ |
| $70-79$ | $0.97(0.87-1.15)$ | $0.97(0.87-1.15)$ | $1.08(0.87-1.33)$ | $1.08(0.87-1.33)$ |
| $\geq 80$ | $0.97(0.87-1.15)$ | $0.97(0.87-1.15)$ | $1.05(0.85-1.30)$ | $1.05(0.85-1.30)$ |

NA Not applicable.
a Incidence and mortality.
b Summary risk estimates computed using estimates adjusted for confounding variables (e.g. age, sex) but NOT adjusted for blood pressure and cholesterol. If these were unavailable from any study the available overall adjusted relative risk estimate was used.
were applied to the 15-29- and 30-44-year age groups and those groups of women aged $\geq 45$ years, respectively. The results shown in Table 10.21 indicate a somewhat stronger association between level 1 exposure and breast cancer for post-menopausal women (1.34) compared to pre- and peri-menopausal women (1.25). This relationship is also evident for level 2 exposure. Funnel plot (Figure 10.26) revealed some evidence of bias and heterogeneity. The regression test showed significant bias among studies of post-menopausal women $(P=0.002)$ but not among pre- and peri-menopausal women ( $P=0.23$ ).

Table I0.2I Summary relative risk estimates for breast cancer ${ }^{a}$ for level I (inactive) and level 2 (insufficiently active) exposure, by age and sex

|  | Adjusted for confounding variables but NO adjustment for measurement error |  | Adjusted for confounding variables WITH adjustment for measurement error |  |
| :---: | :---: | :---: | :---: | :---: |
| Level I exposure (inactive) |  |  |  |  |
| Age group (years) | Males | Females | Males | Females |
| 0-4 | NA | NA | NA | NA |
| 5-14 | NA | NA | NA | NA |
| 15-29 | NA | 1.13 (1.04-1.22) | NA | 1.25 (1.20-1.30) |
| 30-44 | NA | 1.13 (1.04-1.22) | NA | 1.25 (1.20-1.30) |
| 45-59 | NA | 1.13 (1.04-1.22) | NA | 1.34 (1.29-1.39) |
| 60-69 | NA | 1.13 (1.04-1.22) | NA | 1.34 (1.29-1.39) |
| 70-79 | NA | 1.09 (1.01-1.18) | NA | 1.25 (1.21-1.30) |
| $\geq 80$ | NA | I. 06 (0.98-I.15) | NA | 1.16 (1.1I-I.20) |
| Level 2 exposure (insufficiently active) |  |  |  |  |
| Age group (years) | Males | Females | Males | Females |
| 0-4 | NA | NA | NA | NA |
| 5-14 | NA | NA | NA | NA |
| 15-29 | NA | 1.09 (1.03-1.16) | NA | 1.13 (1.04-1.22) |
| 30-44 | NA | 1.09 (1.03-1.16) | NA | 1.13 (1.04-1.22) |
| 45-59 | NA | 1.09 (1.03-1.16) | NA | 1.13 (1.04-1.22) |
| 60-69 | NA | 1.09 (1.03-1.16) | NA | 1.13 (1.04-1.22) |
| 70-79 | NA | 1.07 (I.01-I.l3) | NA | 1.09 (1.01-I.18) |
| $\geq 80$ | NA | 1.04 (0.98-1.1 I) | NA | 1.06 (0.98-I.15) |
| NA Not applicable. <br> a Incidence and mortality. |  |  |  |  |

## COLON CANCER

Table 10.22 reports the results of risk associated with colon cancer for men and women. Separate analyses by sex were conducted but showed a non-significant difference, thus pooled estimates are provided (data not shown). Funnel plots showed some evidence of bias but little heterogeneity (see Figure 10.27), however the regression test revealed a significant bias $(P<0.001)$. This bias would appear to be attributable to the inclusion of three studies which each had relatively large estimates of risk and small standard errors (Gerhardsson et al. 1988; Longnecker et al. 1995; Tang et al. 1999). Excluding these studies resulted in a regression test p-value of 0.30 and had no effect on the summary risk estimates.

Figure 10.26 Funnel plot of studies used in breast cancer meta-analysis for ischaemic heart disease, level I exposure (inactive)


Figure 10.27 Funnel plot of studies used in colon cancer meta-analysis for ischaemic heart disease, level I exposure (inactive)


Type II DIABETES
The risk of type II diabetes associated with level 1 exposure was 1.45 and 1.31 , with and without the adjustment for measurement error, respectively (see Table 10.23). Level 2 exposure was associated with a relative risk of 1.24 (with measurement adjustment) and 1.17 (without

Table 10.22 Summary relative risk estimates for colon cancer ${ }^{\text {a }}$ for level I (inactive) and level 2 (insufficiently active) exposure, by age and sex

|  | Adjusted for confounding variables but NO adjustment <br> for measurement error |  | Adjusted for confounding variables WITH adjustment for measurement error |  |
| :---: | :---: | :---: | :---: | :---: |
| Level I exposure (inactive) |  |  |  |  |
| Age group (years) | Males | Females | Males | Females |
| 0-4 | NA | NA | NA | NA |
| 5-14 | NA | NA | NA | NA |
| 15-29 | 1.43 (1.38-1.49) | 1.43 (1.38-1.49) | 1.68 (1.55-1.82) | 1.68 (1.55-1.82) |
| 30-44 | 1.43 (1.38-1.49) | 1.43 (1.38-1.49) | 1.68 (1.55-1.82) | 1.68 (1.55-1.82) |
| 45-59 | 1.43 (1.38-1.49) | 1.43 (1.38-1.49) | 1.68 (1.55-1.82) | 1.68 (1.55-1.82) |
| 60-69 | 1.43 (1.38-1.49) | 1.43 (1.38-1.49) | 1.68 (1.55-1.82) | 1.68 (1.55-1.82) |
| 70-79 | 1.31 (1.26-1.36) | 1.31 (1.26-1.36) | 1.48 (1.36-1.60) | 1.48 (1.36-1.60) |
| $\geq 80$ | 1.20 (1.15-1.24) | 1.20 (1.15-1.24) | 1.30 (1.20-1.40) | 1.30 (1.20-1.40) |
| Level 2 exposure (insufficiently active) |  |  |  |  |
| Age group (years) | Males | Females | Males | Females |
| 0-4 | NA | NA | NA | NA |
| 5-14 | NA | NA | NA | NA |
| 15-29 | 1.11 (1.03-1.20) | 1.11 (1.03-1.20) | 1.18 (1.05-1.33) | 1.18 (1.05-1.33) |
| 30-44 | 1.11 (1.03-1.20) | 1.11 (1.03-1.20) | 1.18 (1.05-1.33) | 1.18 (1.05-1.33) |
| 45-59 | 1.11 (1.03-1.20) | 1.11 (1.03-1.20) | 1.18 (1.05-1.33) | 1.18 (1.05-1.33) |
| 60-69 | 1.11 (1.03-1.20) | 1.11 (1.03-1.20) | 1.18 (1.05-1.33) | 1.18 (1.05-1.33) |
| 70-79 | 1.08 (1.00-1.17) | 1.08 (1.00-1.17) | 1.13 (1.01-1.27) | 1.13 (1.01-1.27) |
| $\geq 80$ | 1.05 (0.97-1.14) | 1.05 (0.97-1.14) | 1.08 (0.97-1.22) | 1.08 (0.97-1.22) |
| NA Not applicable. <br> a Incidence and mortality. |  |  |  |  |

measurement adjustment). Age attenuation was applied to the age groups of $70-79$ years and $\geq 80$ years.

The funnel plot showed no evidence of heterogeneity (Figure 10.28) but strong evidence of bias for these analyses (regression test $P<0.001$ ). We were unable to account for this bias, however it is possible that publication bias may have contributed.

## Discussion of relative risk estimates

Overall the results from our set of meta-analyses are similar to previous reports. For instance, the summary estimate of the independent effect of inactivity (1.38) is similar to previous findings of Eaton (1992) and the

Table 10.23 Summary relative risk estimates for type II diabetes for level I (inactive) and level 2 (insufficiently active) exposure, by age and sex

|  | Adjusted for confounding variables but NO adjustment for measurement error |  | Adjusted for confounding variables WITH adjustment for measurement error |  |
| :---: | :---: | :---: | :---: | :---: |
| Level I exposure (inactive) |  |  |  |  |
| Age group (years) | Males | Females | Males | Females |
| 0-4 | NA | NA | NA | NA |
| 5-14 | NA | NA | NA | NA |
| 15-29 | 1.31 (1.24-1.39) | 1.31 (1.24-1.39) | 1.45 (1.37-1.54) | 1.45 (1.37-I.54) |
| 30-44 | 1.31 (1.24-1.39) | 1.31 (1.24-1.39) | 1.45 (1.37-1.54) | 1.45 (1.37-1.54) |
| 45-59 | 1.31 (1.24-1.39) | 1.31 (1.24-1.39) | 1.45 (1.37-1.54) | 1.45 (1.37-1.54) |
| 60-69 | 1.31 (1.24-1.39) | 1.31 (1.24-1.39) | 1.45 (1.37-1.54) | 1.45 (1.37-1.54) |
| 70-79 | 1.22 (1.15-1.30) | 1.22 (1.15-1.30) | 1.32 (1.25-1.40) | 1.32 (1.25-1.40) |
| $\geq 80$ | 1.14 (1.08-1.21) | 1.14 (1.08-I.2I) | 1.20 (1.14-1.28) | 1.20 (1.14-1.28) |

Level 2 exposure (insufficiently active)

| Age group (years) | Males | Females | Males | Females |
| :---: | :---: | :---: | :---: | :---: |
| 0-4 | NA | NA | NA | NA |
| 5-14 | NA | NA | NA | NA |
| 15-29 | 1.17 (1.08-1.27) | 1.17 (1.08-1.27) | 1.24 (1.10-1.39) | 1.24 (1.10-1.39) |
| 30-44 | 1.17 (1.08-1.27) | 1.17 (1.08-1.27) | 1.24 (1.10-1.39) | 1.24 (1.10-1.39) |
| 45-59 | 1.17 (1.08-1.27) | 1.17 (1.08-1.27) | 1.24 (1.10-1.39) | 1.24 (1.10-1.39) |
| 60-69 | 1.17 (1.08-1.27) | 1.17 (1.08-1.27) | 1.24 (1.10-1.39) | 1.24 (1.10-1.39) |
| 70-79 | 1.12 (1.04-1.22) | 1.12 (1.04-1.22) | 1.18 (1.04-I.32) | 1.18 (1.04-1.32) |
| $\geq 80$ | 1.08 (1.00-1.17) | 1.08 (1.00-1.17) | I.II (0.99-I.25) | 1.11 (0.99-I.25) |
| NA Not applicable |  |  |  |  |

recent findings reported by Williams (2001). Eaton (1992) found an association between physical inactivity and ischaemic heart disease of 1.37 (1.27-1.48) from a set of 14 studies that assessed discretionary time and/or well-defined occupational physical activity, with ischaemic heart disease mortality or morbidity as disease end-points. More recently, Williams (2001) pooled results from 16 cohorts with measures of dis-cretionary-time physical activity or "all physical activity" and ischaemic heart disease end-points. The author reported pooled relative risk estimates against percentiles of activity and showed relative risks of between $0.65-0.75$ for the 70th percentile and above of activity compared with the referent group of zero activity. Inverting his result to allow comparison with this study produced a relative risk of about 1.4.

Figure 10.28 Funnel plot of studies used in type II diabetes meta-analysis for ischaemic heart disease, level I exposure (inactive)


Our results are also consistent with earlier reviews undertaken by Powell et al. (1987) and Berlin and Colditz (1990) when the differences in protocols are reviewed. Powell et al. (1987) pooled 47 studies with diverse methodological properties and included studies with exposure of either fitness or physical activity. Their qualitative summary estimate of relative risk was 1.9 but the authors noted the magnitude of this association ranged from 1.5-2.4. Berlin and Colditz (1990) replicated the work of Powell et al. with the addition of any subsequent published work, and reported several estimates of relative risk specific to occupational physical activity and non-occupational activity, and for separate cardiovascular disease outcomes. Also, they undertook separate analyses for studies with two and with three levels of exposure. Using only those studies included by Powell et al., they computed a summary estimate of risk for discretionary-time inactivity and ischaemic heart disease (morbidity) of 1.6 with a confidence interval of 1.3-1.8 (see published paper Table 4, analysis B) (Berlin and Colditz 1990). This did not change substantially when additional studies were added ( $\mathrm{RR}=1.5,95 \% \mathrm{CI}$ 1.4-1.7, see published paper Table 5, analysis B).

Berlin and Colditz refined their analyses by separating studies based on epidemiological criteria and classified studies as "satisfactory" and "unsatisfactory". Including only the former studies, relative risk estimates ranged from $1.3(0.7-2.6)$ to $1.9(1.0-3.6)$ for ischaemic heart disease mortality (see published paper Table 6). Thus, comparing our
results with the intricate set of analyses of Berlin and Colditz and other previous work of Powell et al. (1987), and Eaton (1992) revealed our results are similar and well within the range of published confidence limits.

We found a weaker relationship for the effect of physical activity on ischaemic stroke than on ischaemic heart disease, the relative risk estimate for the inactive group being 1.53. The existence of a dose-response relationship between physical activity and stroke is not yet established (Kohl 2001) and there is no evidence of a dose-response effect seen in this study, with the relative risk of the insufficiently active group being no different from 1 .

The relative risk of physical inactivity associated with developing type II diabetes was 1.45 (1.37-1.54) and this was applied to men and women. Previous studies have reported a differential protective effect across sex, however we found only modest evidence for level 1 exposure and no evidence for level 2 exposure. Any apparent difference may have been due to the large degree of heterogeneity among studies. A dose-response relationship across levels of exposure was observed and this is consistent with a recent review of the benefits of physical activity in preventing or delaying the development of diabetes (Kelley and Goodpaster 2001).

The relative risk associated with colon cancer was 1.68 (1.55-1.82) for inactive adults compared with those who are reaching recommended levels of activity. The insufficiently active group had a lower relative risk of 1.18 (1.05-1.33) compared to the referent group. These results indicate a dose-response relationship across exposure and this is consistent with previous research (Thune and Furberg 2001). No previous quantitative meta-analyses have been undertaken on physical inactivity and colon cancer, although the recent review by Thune and Furberg (2001) concluded that the majority of studies showed an independent protective effect ranging in magnitude of between $10 \%$ and $70 \%$ (2001).

We found the relative risks associated with breast cancer were 1.25 (1.20-1.30) for women aged 15-44 years and 1.34 (1.29-1.39) for women aged 45-69 years. These results are consistent with a previously published qualitative review that showed a $30 \%$ reduction in risk of breast cancer among pre-, peri- and post-menopausal women, with a graded dose- response seen in around half of the studies reviewed (Thune and Furberg 2001). Evidence of a dose-response effect was observed with the relative risk of 1.13 (1.04-1.22) associated with the insufficiently active group.

In summary, our estimates of risk of ischaemic heart disease are consistent with previous research, given a careful inspection of the study inclusion criteria and treatment of the data. There are less data by which to compare our results on ischaemic stroke and no previous quantitative reviews addressing type II diabetes and breast and colon cancers. Across
all health outcomes, except ischaemic stroke, a dose-response relationship was observed, with the most inactive (level 1 exposure) associated with having the greatest risk. There is considerable heterogeneity across studies aimed at answering apparently similar questions. This is a notable limitation to conducting meta-analyses. This field of research would benefit from improved measures of exposure; and greater consistency in the reporting of results between studies would advance this field.

### 3.9 RISK REVERSIBILITY AND TEMPORAL ASPECTS OF THE RISK FACTOR-DISEASE RELATIONSHIP

There is little direct evidence on risk reversibility associated with increases in levels of physical activity. This is primarily due to the lack of randomized controlled trials and the fact that a long period between exposure to physical inactivity and disease outcomes is required for assessment. Further, the few data that are available mostly relate to risk reduction in all-cause mortality and ischaemic heart disease and predominantly among white men from middle to upper socioeconomic groups. Despite these limitations, we reviewed the available literature and attempted to estimate the magnitude and time lag associated with change in level of activity and change in risk. Our conclusions for each health outcome are summarized below.

## Ischaemic heart disease and ischaemic stroke

Increase in levels of physical activity can reduce the risk of ischaemic heart disease (Paffenbarger et al. 1994). Among Harvard alumni, physically inactive men who started a moderate-vigorous physical activity routine reduced risk of ischaemic heart disease mortality by $41 \%$ compared with men who remained inactive (Paffenbarger et al. 1993). This is consistent with findings on the benefits of changes in cardiorespiratory fitness, which although not included in this review may help inform the potential nature of risk reversibility due to physical inactivity (Blair et al. 1995). Blair et al. (1995) reported a reduction in risk of ischaemic heart disease due to improvements in cardiorespiratory fitness. They found that men classified as unfit at their initial examination and fit at their subsequent examination had a $44 \%$ reduction in risk of ischaemic heart disease mortality than did men who were unfit at both examinations (Blair et al. 1995). Furthermore, these associations were independent of other potentially confounding factors (e.g. smoking, BMI, systolic blood pressure, cholesterol) (Blair et al. 1995, 1996). Moreover, there is evidence to suggest that recent participation in physical activity, rather than activity performed in the past, is required for reduction in risk of ischaemic heart disease (Paffenbarger et al. 1978).

Although there is limited empirical support, there are indications that reversibility for ischaemic heart disease could be complete, and could occur within a relatively short period of time following reversal of exposure, perhaps as short as several months or up to two years. This assump-
tion is based on the available evidence pertaining to changes in physical inactivity, indirect evidence from studies using fitness parameters, and known acute effects of physical activity on elements of the physiological pathways which reduce cardiovascular risk.

There is little evidence to guide the estimates of risk reversibility for ischaemic stroke following changes in exposure to physical inactivity. As the biological pathways are considered to be the same as those for ischaemic heart disease, it is possible that there may be a similar relationship for a reduction in risk.

## BREAST CANCER AND COLON CANCER

There are no data to help quantify what degree of risk reduction may occur over what time frame for breast cancer. The paucity of data on physical activity and risk reduction required us to use indicative information on which to base such estimates. Studies on the cessation of postmenopausal hormone replacement therapy and breast cancer rates provide indirect estimates, as the biological pathways (i.e. through hormone levels, as described in section 1) are considered to be similar to that assumed for physical activity's impact on breast cancer.

Major reviews of the epidemiological evidence suggest complete reversibility. That is, women who stop postmenopausal hormone replacement therapy generally revert to the same breast cancer rates as women who had never had hormone replacement therapy, over a period of two to five years (Collaborative Group on Hormonal Factors in Breast Cancer 1997; La Vecchia et al. 2001). There is, however, debate on the biological plausibility and mechanisms of this observed reduction in risk (Bieber and Barnes 2001).

Again there are few data to indicate what the magnitude of risk reduction may be over time for colon cancer. Lee et al. (1991) suggested that although a lifetime or at least consistent participation during adulthood in physical activity is required for greatest protection from risk of colon cancer, some reduction ( $13 \%$ over $11-15$ years) is seen in men who become active after being inactive.

## Type II Diabetes

There are two sources of information to help with risk reversibility for type II diabetes. Data from the Nurses' Health Study show that women who increased their physical activity levels over six years experienced $29 \%$ lower rates of type II diabetes than those who remained inactive over the same period $(R R=0.71)$ (Hu et al. 1999). However, women who were active at the beginning and remained active over the six years still had lower rates of type II diabetes than those who became active ( $\mathrm{RR}=0.59$ ). Recently, the Diabetes Prevention Program Research Group (2002) showed that, among people who had impaired glucose tolerance, a lifestyle intervention combining moderate amounts of physical activity
with a controlled diet resulted in a $58 \%$ reduction in the incidence of type II diabetes over 2 to 5 years. Since this risk reduction resulted from a combined lifestyle intervention with a population already at increased risk, the effect of only changes in physical activity in the general population is assumed to be less. These estimates suggest a risk reversibility of $25 \%$ over five years for sedentary adults becoming sufficiently active and $35 \%$ over four years for insufficiently active adults becoming sufficiently active.

## 4. Burden of disease attributable to PHYSICAL INACTIVITY

The fraction of burden (mortality and morbidity combined) attributable to physical inactivity for each disease end-point is shown in Figure 10.29. Globally physical inactivity contributed to an estimated $22 \%$ of ischaemic heart disease, $11 \%$ of ischaemic stroke, $14 \%$ of type II diabetes, $16 \%$ of colon cancer and $10 \%$ of breast cancer. The results show small differences between males and females, due in part to differences in the level of exposure and to the different distribution of events between the sexes. There were small, non-significant differences in the attributable fractions across subregions (data not shown). For ischaemic heart disease the subregional differences ranged from $21 \%$ (SEAR-D, EMR-D and WPR-B) to $23 \%$ (AMR-B, EUR-C and WPR-A). A slightly wider range was seen for ischaemic stroke $9 \%$ (AFR-D) to $14 \%$ (EUR-C).

The number of deaths and number of DALYs were computed for each disease outcome (Figures 10.30 and 10.31, respectively). Over one

Figure 10.29 Attributable burden of disease due to physical inactivity for ischaemic heart disease, ischaemic stroke, type II diabetes, colon cancer and breast cancer


Figure I 0.30 Attributable mortality due to physical inactivity for ischaemic heart disease, ischaemic stroke, type II diabetes, colon cancer and breast cancer


Figure I 0.31 Attributable DALYs due to physical inactivity for ischaemic heart disease, ischaemic stroke, type II diabetes, colon cancer and breast cancer

million deaths due to ischaemic heart disease were attributable to physical inactivity ( 730000 males and 634000 females). Another 308000 deaths from ischaemic stroke were also attributable to inactive and insufficiently active lifestyles. Combined with deaths from type II diabetes (116000), colon cancer (90000) and breast cancer (45000) a total of 1.92 million deaths could be prevented from increasing levels of physical activity (Figure 10.30). Similarly inactive lifestyles contributed to a loss of 19 million DALYs worldwide.

## 5. Discussion

The attributable fraction for ischaemic heart disease morbidity and mortality ( $22 \%$ ) is slightly lower than previous estimates that range from $23 \%$ (Hahn et al. 1990) to $35 \%$ (Powell and Blair 1994). This variation is explained by differences in the estimates used for prevalence and relative risk. Previous attempts have used rounded estimates of risk (1.9) (Powell and Blair 1994). We estimated relative risks that are consistent but marginally lower than previous reports because we only accepted data from studies reporting an exposure of physical activity not physical fitness; risk of exposure was computed relative to the current public health recommendations not a referent group of "high level of activity" or "vigorously active"; an adjustment for attenuation of risk over age was included; and, perhaps most importantly, we used a trichotomous not continuous measure of exposure. We acknowledge that no attempt was made to account for the additional benefit associated with being highly active and/or engaging in activities of vigorous intensity.

Furthermore, past estimates of exposure have been based on only one domain, namely discretionary time. Single domain estimates will overestimate exposure by not including activity undertaken in other domains. Another important difference with previous estimates is in the number of categorical levels used in the analyses. We created a trichotomous measure; in contrast previous work had used four levels of exposure (Powell and Blair 1994). Estimates of level 2 exposure should be interpreted with great caution. Available data were the most difficult to compare and we acknowledge our assumptions will inevitably include some misclassification. Our global estimate of insufficiently active was $41 \%$, although data show that in different countries it can range from $20 \%$ to $70 \%$. For both level 1 and level 2 there is greatest uncertainty around prevalence estimates for Africa, Asia and the Eastern Mediterranean and across all subregions for adults aged $>60$ years. There is a corresponding high level of uncertainty around the estimates of disease burden for these subregions and age groups. A difference of $\pm 10 \%$ in prevalence estimates would make a considerable difference in the estimated magnitude of the disease burden.

Given the noted differences between our input variables and those used in previous reports, our global estimates of attributable fraction appear consistent, but we suggest that they too are likely to underestimate the true burden due to inactive lifestyles. Moreover, a number of the limitations combine to make it more difficult to compare with other continuously measured risk factors such as blood pressure and cholesterol.

In summary, the multiple model of exposure provided a useful platform for these analyses of attributable burden but recent, comparable country-level data on inactivity would be far more preferable. This would require, however, having more countries systematically collect
data on the distribution of physical inactivity. Well developed, culturally appropriate measurement instruments should be used to collect comparable data. While there are examples of progress in this area (Craig et al. 2003), with few exceptions, instruments to assess exposure across the diverse cultures of Africa, Asia and the Eastern Mediterranean have yet to be well tested or widely used.

Emerging evidence is pointing towards the important role of physical activity in avoiding outcomes (e.g. falls, poor mental health) as well as other disease end-points (e.g. osteoarthritis and osteoporosis) that did not meet our inclusion criteria. Future replications of this work should review the evidence to ensure a complete a picture as possible is obtained on the burden attributable to inactivity. Across these outcomes as well as those included, the exact nature of the dose-response relationship across all disease end-points is not well defined. Moreover, the heterogeneity across studies is notable and represents a limitation to conducting meta-analyses in this field. Researchers are encouraged to pursue this research agenda systematically to maximize the expedient pooling of data and furthering of our knowledge.

Physically inactive lifestyles account for $3.3 \%$ of deaths and morbidity worldwide. Successful promotion of more active lifestyles would prevent at least 2 million premature deaths and almost 20 million DALYs worldwide. Success in prevention requires greater commitment through policy development and resource allocation at the national and local level. Diverse challenges face developed and developing countries as they consider the current and future rapid changes in patterns of activity by domain.

## 6. Projected estimates of future exposure

Calculation of the avoidable burden of disease requires predicated estimates of exposure in the future. Estimates are usually based on trends and patterns seen over past years adjusted based on assumptions and models of factors likely to influence exposure in the future. We looked for data on patterns of activity over time for predicting level 1 and level 2 exposure in 2010, 2020 and 2030. Our search found few trend data and those available were mostly from developed countries and had similar limitations regarding their comparability as previously discussed.

### 6.1 Methods

In the absence of data on trends to guide our calculations of future exposure we considered the factors that might influence inactivity in the next 30 years. Considering each domain separately and in combination, we proposed that economic development was a central factor to changes in the level of physical activity undertaken in the work place, at home, through transportation and in leisure time. As economies develop, new mechanical, computer and biological technologies will reduce the

Figure 10.32 Predicted ${ }^{\text {a }}$ estimates of physical activity for 145 countries in 2000, by domain and GNP per capita

a Actual physical activity estimates for countries were used when available.
demand for physical tasks in the work place as well as drive the shift in employment from agriculture to manufacturing and service industries. Both factors will result in lower levels of physical activity in the occupational domain. Furthermore, as economies develop and individuals have higher incomes, there will be a shift in travel behaviour, specifically walking and cycling will decrease in favour of modes of personal transportation (motor vehicle) and to a lesser extent public transit. Discretionary time is likely to increase as economies develop and there is a higher level of wealth. While there is more time for discretionary physical activity there also will be more alternative options, many of which will require low levels of physical activity, for example, computer-based entertainment and television. Figure 10.32 shows our predicted estimates for physical activity in 2000 by domain and GNP.

We also considered whether the changes in physical activity within each domain over time could have a lag effect, such that increases in available leisure time would result some time after the shifts in the work place and changes in transportation. However, we had insufficient data to model future change in each domain separately or to explore our hypothetical time lag. Instead we used the net effect or the summary estimate of physical inactivity and GNP (shown in Figure 10.33) to model the association based on current (year 2000) levels. We then used dif-

Figure I0.33 Per cent physically inactive (level I exposure) for 145 countries by GNP per capita

ferentiation to compute the change in inactivity associated with a change in GNP over time and predicted estimates of GNP to solve the equation and compute estimates for physical inactivity in 2010, 2020 and 2030. This method is described in detail below.

## METHODS FOR THE PROJECTION OF FUTURE ESTIMATES OF LEVEL 1 EXPOSURE

Prevalence of inactivity was related to GNP per capita for each country and prevalence of malnutrition (at the regional level rather than country level) as below. Malnutrition (as measured by childhood underweight) was included in the model along with GNP because preliminary regression analyses using per cent childhood mortality improved the model fit substantially for developing countries. However, future predictions for child mortality were not available therefore malnutrition was used as a proxy measure. Because country level estimates of malnutrition also did not exist, predicted estimates at the regional level were used (see chapter $2)$.

$$
\begin{aligned}
& \log \left(\% \text { Inactivity }_{2000}\right) \\
& \quad=\mathrm{B}_{1}(\mathrm{GNP} \text { per capita } 2000)+\mathrm{B}_{2}(\% \text { malnutrition } 2000)
\end{aligned}
$$

Linear regression was used to estimate $B_{1}$ and $B_{2}$. The relationship between GNP per capita in 2000 and the prevalence of inactivity for

Table 10.24 Mean (range) GNP per capita and \% malnourished for years 2000 and 2030

| World Bank income <br> classification | GNP per capita (US\$ IO00) |  |  | \% children malnourished |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 2000 | 2030 |  | 2000 | 2030 |
|  | $3.6(0.5-14.9)$ | $8.2(0.9-53.8)$ |  | $18.8(2.3-45.9)$ | $18.4(0.0-50.7)$ |
| High income | $22.7(12.3-33.6)$ | $59.6(31.9-105.9)$ |  | $3.1(2.3-8.1)$ | $0.0(0.0-0.5)$ |

2000 showed that inactivity increased with GNP per capita when GNP was less than US\$ 5000 and decreased when GNP was greater than US\$ 5000 (Figure 10.33). Therefore, we divided countries into two groups using World Bank income classification: a low/middle-income group, and high-income group (including Organisation for Economic Co-operation and Development [OECD] and non-OECD countries). The two groups were modelled separately to give two sets of parameter estimates for $B_{1}$ and $\mathrm{B}_{2}$. Inactivity ( $\%$ ) was log-transformed and all variables were centred to reduce collinearity and to set the regression intercepts to zero by using mean \% inactivity equal to zero. Change in physical inactivity over time was computed using predicted changes in GNP per capita (from World Bank estimates) and predicted change in malnutrition (World Bank 1999; chapter 2).

The assumption that the relationship seen in year 2000 would continue to characterize countries in year 2030 is examined in Table 10.24. Mean GNP increases in both groups between year 2000 and 2030 with a considerable increase in the ranges with only a small overlap between groups. Malnutrition, predicted at the regional level and aggregated to create a group estimate showed only a very small decrease over time in the low/middle-income group reflecting the worsening conditions predicted for some regions (e.g. Africa). For high-income countries malnutrition is predicted to approach zero by 2030 (chapter 2).

## Methods for the projection of future estimates of level 2 exposure

Age by sex estimates of level 2 exposure (insufficiently active) were computed by holding constant our predicted values for year 2000 and applying these uniformly for 2010, 2020 and 2030. We chose this simple approach based on evidence that changes in the population distribution of physical (in)activity occur between levels of both inactive and insufficiently active, as well as between insufficiently active and sufficient (level 3 unexposed). We had no data on which to estimate the magnitude of the shift between level 2 and level 3. Thus, our predicted values of level 2 exposure are based on the assumption that the changes calculated for level 1 (predicted based on changes in economic development as described above) are equal to the change from level 2 to level 3 -the net result of which is that level 2 estimates remain unchanged over time.

This approach is clearly limited but without more data from different populations on both current trends and patterns over time it was the only feasible approach.

### 6.2 Results

The parameter estimates for the model used to estimate future level 1 exposure show the different relationship between physical inactivity estimates in 2000 and GNP per capita and \% malnutrition for each group (see Figure 10.33). Final age by sex estimates for 14 subregions for 2010, 2020 and 2030 are shown in Tables $10.25,10.26$ and 10.27 , respectively. These data indicate an increasing physical inactivity in all subregions, for males and females and across all age groups. The magnitude of increase in level 1 exposure was of the order of $3-4 \%$. We did not model any change in level 2 exposure. Recent evidence suggests these data are likely to grossly underestimate change in exposure. Between 1999 and 2002, Western Australia experienced a $4 \%$ decline in the proportion of the population meeting recommended levels of physical activity (McCormack et al. 2003). These data may be more indicative of the possible magnitude of negative change to expect in developed market economies in coming years if no action is undertaken. Taken in combination with the changing patterns of lifestyle people will experience in developing economies, the future burden of disease attributable to inactivity may be considerable.
Table I0.25 Projected estimates (\%) of physical inactivity in year 2010

| Subregion | Exposure level | Males |  |  |  |  |  | Females |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Inactive | 11 | 13 | 14 | 16 | 18 | 19 | 13 | 12 | 13 | 16 | 19 | 20 |
|  | Insufficient | 48 | 48 | 48 | 46 | 44 | 43 | 45 | 52 | 51 | 48 | 46 | 45 |
|  | Recommended | 41 | 39 | 38 | 38 | 38 | 38 | 42 | 36 | 36 | 36 | 35 | 35 |
| AFR-E | Inactive | 9 | 12 | 12 | 15 | 16 | 17 | 12 | 11 | 11 | 14 | 16 | 16 |
|  | Insufficient | 50 | 51 | 50 | 49 | 47 | 46 | 47 | 56 | 55 | 51 | 50 | 50 |
|  | Recommended | 41 | 37 | 38 | 36 | 37 | 37 | 41 | 33 | 34 | 35 | 34 | 34 |
| AMR-A | Inactive | 17 | 19 | 20 | 21 | 22 | 32 | 22 | 23 | 21 | 27 | 31 | 41 |
|  | Insufficient | 44 | 47 | 44 | 40 | 40 | 35 | 36 | 41 | 40 | 38 | 36 | 31 |
|  | Recommended | 39 | 34 | 36 | 39 | 38 | 33 | 42 | 36 | 39 | 35 | 33 | 28 |
| AMR-B | Inactive | 17 | 18 | 20 | 23 | 26 | 29 | 23 | 28 | 28 | 38 | 40 | 42 |
|  | Insufficient | 42 | 44 | 41 | 39 | 36 | 35 | 33 | 32 | 31 | 30 | 30 | 29 |
|  | Recommended | 41 | 38 | 39 | 38 | 38 | 36 | 44 | 40 | 41 | 32 | 30 | 29 |
| AMR-D | Inactive | 17 | 19 | 19 | 23 | 28 | 30 | 22 | 26 | 30 | 40 | 46 | 48 |
|  | Insufficient | 38 | 38 | 33 | 32 | 30 | 29 | 28 | 26 | 24 | 24 | 22 | 22 |
|  | Recommended | 45 | 43 | 48 | 45 | 42 | 41 | 50 | 48 | 46 | 36 | 32 | 30 |
| EMR-B | Inactive | 16 | 19 | 19 | 22 | 25 | 27 | 19 | 21 | 21 | 25 | 31 | 33 |
|  | Insufficient | 41 | 39 | 38 | 36 | 32 | 32 | 36 | 36 | 35 | 32 | 31 | 30 |
|  | Recommended | 43 | 42 | 43 | 42 | 43 | 41 | 45 | 43 | 44 | 43 | 38 | 37 |
| EMR-D | Inactive | 15 | 18 | 19 | 21 | 23 | 26 | 18 | 20 | 20 | 23 | 30 | 31 |
|  | Insufficient | 42 | 38 | 36 | 35 | 32 | 31 | 36 | 36 | 35 | 32 | 31 | 30 |
|  | Recommended | 43 | 44 | 45 | 44 | 45 | 43 | 46 | 44 | 45 | 45 | 39 | 39 |

Table I0.25 Projected estimates (\%) of physical inactivity in year 2010 (continued)

| Subregion | Exposure level | Males |  |  |  |  |  | Females |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EUR-A | Inactive | 14 | 15 | 16 | 19 | 21 | 22 | 17 | 19 | 19 | 23 | 25 | 29 |
|  | Insufficient | 52 | 57 | 55 | 52 | 50 | 47 | 47 | 51 | 51 | 45 | 45 | 42 |
|  | Recommended | 34 | 28 | 29 | 29 | 29 | 31 | 36 | 30 | 30 | 32 | 30 | 29 |
| EUR-B | Inactive | 15 | 19 | 20 | 22 | 26 | 28 | 19 | 40 | 38 | 36 | 34 | 33 |
|  | Insufficient | 43 | 40 | 38 | 36 | 34 | 33 | 37 | 37 | 36 | 33 | 32 | 32 |
|  | Recommended | 42 | 41 | 42 | 42 | 40 | 39 | 44 | 23 | 26 | 31 | 34 | 35 |
| EUR-C | Inactive | 18 | 19 | 23 | 32 | 37 | 39 | 21 | 28 | 28 | 39 | 40 | 41 |
|  | Insufficient | 38 | 34 | 32 | 30 | 28 | 28 | 32 | 31 | 30 | 27 | 27 | 26 |
|  | Recommended | 44 | 47 | 45 | 38 | 35 | 33 | 47 | 41 | 42 | 34 | 33 | 33 |
| SEAR-B | Inactive | 14 | 16 | 16 | 17 | 15 | 16 | 16 | 18 | 18 | 18 | 17 | 17 |
|  | Insufficient | 43 | 43 | 43 | 47 | 52 | 52 | 41 | 41 | 41 | 45 | 50 | 50 |
|  | Recommended | 43 | 41 | 41 | 36 | 33 | 32 | 43 | 41 | 41 | 37 | 33 | 33 |
| SEAR-D | Inactive | 14 | 18 | 18 | 21 | 23 | 21 | 18 | 20 | 20 | 23 | 25 | 28 |
|  | Insufficient | 42 | 38 | 36 | 34 | 32 | 31 | 36 | 35 | 34 | 32 | 30 | 30 |
|  | Recommended | 44 | 44 | 46 | 45 | 45 | 48 | 46 | 45 | 46 | 45 | 45 | 42 |
| WPR-A | Inactive | 15 | 16 | 17 | 19 | 18 | 18 | 17 | 19 | 19 | 20 | 18 | 18 |
|  | Insufficient | 50 | 56 | 53 | 52 | 56 | 55 | 48 | 49 | 50 | 49 | 55 | 54 |
|  | Recommended | 35 | 28 | 30 | 29 | 26 | 27 | 35 | 32 | 31 | 31 | 27 | 28 |
| WPR-B | Inactive | 14 | 17 | 16 | 19 | 20 | 21 | 16 | 17 | 18 | 21 | 22 | 21 |
|  | Insufficient | 41 | 40 | 41 | 41 | 44 | 41 | 40 | 39 | 38 | 38 | 41 | 38 |
|  | Recommended | 45 | 43 | 43 | 40 | 36 | 38 | 44 | 44 | 44 | 41 | 37 | 41 |

Table IO.26 Projected estimates (\%) of physical inactivity in year 2020

| Subregion | Exposure level | Males |  |  |  |  |  | Females |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Inactive | 11 | 13 | 13 | 16 | 18 | 19 | 13 | 12 | 13 | 16 | 19 | 20 |
|  | Insufficient | 48 | 48 | 48 | 46 | 44 | 43 | 45 | 52 | 51 | 48 | 46 | 45 |
|  | Recommended | 41 | 39 | 39 | 38 | 38 | 38 | 42 | 36 | 36 | 36 | 35 | 35 |
| AFR-E | Inactive | 9 | 12 | 12 | 15 | 16 | 17 | 11 | 11 | 11 | 14 | 16 | 16 |
|  | Insufficient | 50 | 51 | 50 | 49 | 47 | 46 | 47 | 56 | 55 | 51 | 50 | 50 |
|  | Recommended | 41 | 37 | 38 | 36 | 37 | 37 | 42 | 33 | 34 | 35 | 34 | 34 |
| AMR-A | Inactive | 17 | 19 | 20 | 21 | 22 | 32 | 21 | 23 | 21 | 27 | 31 | 41 |
|  | Insufficient | 44 | 47 | 44 | 40 | 40 | 35 | 36 | 41 | 40 | 38 | 36 | 31 |
|  | Recommended | 39 | 34 | 36 | 39 | 38 | 33 | 43 | 36 | 39 | 35 | 33 | 28 |
| AMR-B | Inactive | 17 | 19 | 20 | 23 | 26 | 29 | 24 | 28 | 28 | 38 | 41 | 42 |
|  | Insufficient | 42 | 44 | 41 | 39 | 36 | 35 | 33 | 32 | 31 | 30 | 30 | 29 |
|  | Recommended | 41 | 37 | 39 | 38 | 38 | 36 | 43 | 40 | 41 | 32 | 29 | 29 |
| AMR-D | Inactive | 17 | 19 | 19 | 23 | 28 | 31 | 22 | 26 | 30 | 40 | 47 | 48 |
|  | Insufficient | 38 | 38 | 33 | 32 | 30 | 29 | 28 | 26 | 24 | 24 | 22 | 22 |
|  | Recommended | 45 | 43 | 48 | 45 | 42 | 40 | 50 | 48 | 46 | 36 | 31 | 30 |
| EMR-B | Inactive | 16 | 19 | 19 | 22 | 25 | 27 | 19 | 21 | 21 | 25 | 31 | 33 |
|  | Insufficient | 41 | 39 | 38 | 36 | 32 | 32 | 36 | 36 | 35 | 32 | 31 | 30 |
|  | Recommended | 43 | 42 | 43 | 42 | 43 | 41 | 45 | 43 | 44 | 43 | 38 | 37 |
| EMR-D | Inactive | 15 | 18 | 19 | 21 | 23 | 26 | 18 | 20 | 20 | 24 | 30 | 31 |
|  | Insufficient | 42 | 38 | 36 | 35 | 32 | 31 | 36 | 36 | 35 | 32 | 31 | 30 |
|  | Recommended | 43 | 44 | 45 | 44 | 45 | 43 | 46 | 44 | 45 | 44 | 39 | 39 |

Table I0.26 Projected estimates (\%) of physical inactivity in year 2020 (continued)

| Subregion | Exposure level | Males |  |  |  |  |  | Females |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EUR-A | Inactive | 14 | 15 | 16 | 19 | 20 | 22 | 17 | 19 | 19 | 23 | 25 | 28 |
|  | Insufficient | 52 | 57 | 55 | 52 | 50 | 47 | 47 | 51 | 51 | 45 | 45 | 42 |
|  | Recommended | 34 | 28 | 29 | 29 | 30 | 31 | 36 | 30 | 30 | 32 | 30 | 30 |
| EUR-B | Inactive | 15 | 19 | 20 | 22 | 26 | 28 | 19 | 21 | 21 | 27 | 32 | 33 |
|  | Insufficient | 43 | 40 | 38 | 36 | 34 | 33 | 37 | 37 | 36 | 33 | 32 | 32 |
|  | Recommended | 42 | 41 | 42 | 42 | 40 | 39 | 44 | 42 | 43 | 40 | 36 | 35 |
| EUR-C | Inactive | 18 | 19 | 23 | 32 | 38 | 39 | 14 | 17 | 17 | 17 | 16 | 16 |
|  | Insufficient | 38 | 34 | 32 | 30 | 28 | 28 | 32 | 31 | 30 | 27 | 27 | 26 |
|  | Recommended | 44 | 47 | 45 | 38 | 34 | 33 | 54 | 52 | 53 | 56 | 57 | 58 |
| SEAR-B | Inactive | 14 | 17 | 17 | 17 | 16 | 16 | 16 | 19 | 18 | 19 | 17 | 17 |
|  | Insufficient | 43 | 43 | 43 | 47 | 52 | 52 | 41 | 41 | 41 | 45 | 50 | 50 |
|  | Recommended | 43 | 40 | 40 | 36 | 32 | 32 | 43 | 40 | 41 | 36 | 33 | 33 |
| SEAR-D | Inactive | 14 | 18 | 19 | 21 | 23 | 22 | 18 | 20 | 20 | 23 | 25 | 28 |
|  | Insufficient | 42 | 38 | 36 | 34 | 32 | 31 | 36 | 35 | 34 | 32 | 30 | 30 |
|  | Recommended | 44 | 44 | 45 | 45 | 45 | 47 | 46 | 45 | 46 | 45 | 45 | 42 |
| WPR-A | Inactive | 15 | 16 | 17 | 18 | 17 | 17 | 17 | 19 | 19 | 20 | 18 | 18 |
|  | Insufficient | 50 | 56 | 53 | 52 | 56 | 55 | 48 | 49 | 50 | 49 | 55 | 54 |
|  | Recommended | 35 | 28 | 30 | 30 | 27 | 28 | 35 | 32 | 31 | 31 | 27 | 28 |
| WPR-B | Inactive | 15 | 17 | 16 | 19 | 20 | 22 | 16 | 18 | 18 | 22 | 22 | 21 |
|  | Insufficient | 41 | 40 | 41 | 41 | 44 | 41 | 40 | 39 | 38 | 38 | 41 | 38 |
|  | Recommended | 44 | 43 | 43 | 40 | 36 | 37 | 44 | 43 | 44 | 40 | 37 | 41 |

Table I0.27 Projected estimates (\%) of physical inactivity in year 2030

| Subregion | Exposure level | Males |  |  |  |  |  | Females |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Inactive | 11 | 13 | 13 | 16 | 18 | 19 | 13 | 12 | 13 | 16 | 19 | 20 |
|  | Insufficient | 48 | 48 | 48 | 46 | 44 | 43 | 45 | 52 | 51 | 48 | 46 | 45 |
|  | Recommended | 41 | 39 | 39 | 38 | 38 | 38 | 42 | 36 | 36 | 36 | 35 | 35 |
| AFR-E | Inactive | 9 | 12 | 12 | 15 | 16 | 17 | 11 | 10 | 11 | 14 | 16 | 16 |
|  | Insufficient | 50 | 51 | 50 | 49 | 47 | 46 | 47 | 56 | 55 | 51 | 50 | 50 |
|  | Recommended | 41 | 37 | 38 | 36 | 37 | 37 | 42 | 34 | 34 | 35 | 34 | 34 |
| AMR-A | Inactive | 17 | 19 | 20 | 20 | 22 | 32 | 21 | 23 | 21 | 26 | 31 | 40 |
|  | Insufficient | 44 | 47 | 44 | 40 | 40 | 35 | 36 | 41 | 40 | 38 | 36 | 31 |
|  | Recommended | 39 | 34 | 36 | 40 | 38 | 33 | 43 | 36 | 39 | 36 | 33 | 29 |
| AMR-B | Inactive | 17 | 19 | 20 | 23 | 26 | 29 | 24 | 28 | 29 | 38 | 41 | 42 |
|  | Insufficient | 42 | 44 | 41 | 39 | 36 | 35 | 33 | 32 | 31 | 30 | 30 | 29 |
|  | Recommended | 41 | 37 | 39 | 38 | 38 | 36 | 43 | 40 | 40 | 32 | 29 | 29 |
| AMR-D | Inactive | 17 | 19 | 19 | 23 | 29 | 31 | 22 | 26 | 30 | 40 | 47 | 48 |
|  | Insufficient | 38 | 38 | 33 | 32 | 30 | 29 | 28 | 26 | 24 | 24 | 22 | 22 |
|  | Recommended | 45 | 43 | 48 | 45 | 41 | 40 | 50 | 48 | 46 | 36 | 31 | 30 |
| EMR-B | Inactive | 16 | 19 | 20 | 23 | 25 | 27 | 19 | 21 | 21 | 25 | 31 | 33 |
|  | Insufficient | 41 | 39 | 38 | 36 | 32 | 32 | 36 | 36 | 35 | 32 | 31 | 30 |
|  | Recommended | 43 | 42 | 42 | 41 | 43 | 41 | 45 | 43 | 44 | 43 | 38 | 37 |
| EMR-D | Inactive | 15 | 18 | 19 | 22 | 23 | 26 | 18 | 20 | 20 | 24 | 30 | 32 |
|  | Insufficient | 42 | 38 | 36 | 35 | 32 | 31 | 36 | 36 | 35 | 32 | 31 | 30 |
|  | Recommended | 43 | 44 | 45 | 43 | 45 | 43 | 46 | 44 | 45 | 44 | 39 | 38 |

Table I0.27 Projected estimates (\%) of physical inactivity in year 2030 (continued)

| Subregion | Exposure level | Males |  |  |  |  |  | Females |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EUR-A | Inactive | 14 | 15 | 16 | 19 | 20 | 22 | 17 | 18 | 19 | 23 | 24 | 28 |
|  | Insufficient | 52 | 57 | 55 | 52 | 50 | 47 | 47 | 51 | 51 | 45 | 45 | 42 |
|  | Recommended | 34 | 28 | 29 | 29 | 30 | 31 | 36 | 31 | 30 | 32 | 31 | 30 |
| EUR-B | Inactive | 15 | 19 | 20 | 22 | 26 | 29 | 19 | 21 | 22 | 27 | 32 | 33 |
|  | Insufficient | 43 | 40 | 38 | 36 | 34 | 33 | 37 | 37 | 36 | 33 | 32 | 32 |
|  | Recommended | 42 | 41 | 42 | 42 | 40 | 38 | 44 | 42 | 42 | 40 | 36 | 35 |
| EUR-C | Inactive | 18 | 19 | 23 | 32 | 38 | 39 | 21 | 28 | 28 | 40 | 40 | 41 |
|  | Insufficient | 38 | 34 | 32 | 30 | 28 | 28 | 32 | 31 | 30 | 27 | 27 | 26 |
|  | Recommended | 44 | 47 | 45 | 38 | 34 | 33 | 47 | 41 | 42 | 33 | 33 | 33 |
| SEAR-B | Inactive | 15 | 17 | 17 | 17 | 16 | 16 | 16 | 19 | 19 | 19 | 18 | 18 |
|  | Insufficient | 43 | 43 | 43 | 47 | 52 | 52 | 41 | 41 | 41 | 45 | 50 | 50 |
|  | Recommended | 42 | 40 | 40 | 36 | 32 | 32 | 43 | 40 | 40 | 36 | 32 | 32 |
| SEAR-D | Inactive | 14 | 18 | 19 | 22 | 23 | 22 | 18 | 20 | 20 | 23 | 26 | 28 |
|  | Insufficient | 42 | 38 | 36 | 34 | 32 | 31 | 36 | 35 | 34 | 32 | 30 | 30 |
|  | Recommended | 44 | 44 | 45 | 44 | 45 | 47 | 46 | 45 | 46 | 45 | 44 | 42 |
| WPR-A | Inactive | 15 | 15 | 16 | 18 | 17 | 17 | 16 | 19 | 18 | 20 | 17 | 18 |
|  | Insufficient | 50 | 56 | 53 | 52 | 56 | 55 | 48 | 49 | 50 | 49 | 55 | 54 |
|  | Recommended | 35 | 29 | 31 | 30 | 27 | 28 | 36 | 32 | 32 | 31 | 28 | 28 |
| WPR-B | Inactive | 15 | 17 | 16 | 19 | 20 | 22 | 16 | 18 | 18 | 22 | 22 | 21 |
|  | Insufficient | 41 | 40 | 41 | 41 | 44 | 41 | 40 | 39 | 38 | 38 | 41 | 38 |
|  | Recommended | 44 | 43 | 43 | 40 | 36 | 37 | 44 | 43 | 44 | 40 | 37 | 41 |

## Acknowledgements

The authors wish to extend their sincere thanks to all the colleagues and researchers worldwide who assisted in either providing existing data, re-analysing data and/or providing contacts. The generous contributions of the following people helped us complete this project: Erinaldo Adrande, Ahmed al-Hashemi, Hazzaa al-Hazzaa, Jawad al-Lawati, Wahid al-Kharus, Terry Aspray, Adrian Bauman, Angèle Beaulieu, Pascal Bovet, Steve Blair, Clareann Bunker, Nick Cavill, L. Cavalli-Sforza, Beatriz Champagne-Marcet, Dr Chandran, Constanza Cilley, William Cockerham, Cora Craig, Catherine Cross, Constanza Cilley, Maximilian de Courten, Kathy Douglas, Michael Engelgau, Ted Forbes, Gauden Galea, Bill Haskill, Louise Hayes, Francisco Holway, Frank Hu, Toni Hunt, Enrique Jacoby, Konrad Jamrozik, Andrea Kriska, Darwin Labarthe, Vicky Lambert, I-Min Lee, N Mallick, Beatriz Champagne Marcet, Miguel Martinez-Gonzalez, Sandra Matsudo, Victor Matsudo, Willem Van Mechelen, Abdulrahman Musaiger, Norio Murase, Anoop Misra, Pekka Oja, Andrzej Pajak, Rimma Potemkina, Joceline Pomerleau, Barry Popkin, Antonio Prista, Jasem Ramadan, Srinath Reddy, Vikram Singh, Christine Snead, Eugene Sobngwi, Soeharsono Soemantri, Michael Sjöström, Elena Subirats, Roger Vaughan, Guijing Wang, Dr Yang, C. Yajnik, Helena Zabina.

In addition, the authors wish to extend particular thanks to Barbara Ainsworth, David Buchner and Carl Caspersen for their internal review and to the anonymous external reviewers who provided valuable comments on the methods and assumptions. Finally, the authors thank colleagues leading the CRA project Majid Ezzati, Alan Lopez, Anthony Rodgers and Steve Vander Hoorn for their support and technical assistance.

## Note

1 See preface for an explanation of this term.

## References

Abbott RD, Rodriguez BL, Burchfiel CM, Curb JD (1994) Physical activity in older middle-aged men and reduced risk of stroke: The Honolulu Heart Program. American Journal of Epidemiology, 139:881-893.
Ainsworth BE, Haskell WL, Whitt MC et al. (2000) Compendium of physical activities: an update of activity codes and MET intensities. Medicine and Science in Sports and Exercise, 32:S498-504.
al-Refaee SA, al-Hazzaa HM (2001) Physical activity profile of Saudi adult males: implications for health. Saudi Medical Journal, 22:784-789.

Albanes D, Blair A, Taylor P (1989) Physical activity and risk of cancer in the NHANES I population. American Journal of Public Health, 79:744-750.

Alemu T, Lindtjorn B (1995) Physical activity, illness and nutritional status among adults in a rural Ethiopian community. International Journal of Epidemiology, 24:977-983.

American College of Sports Medicine (1995) ACSM position stand on osteoporosis and exercise. Medicine and Science in Sports and Exercise, 27:i-vii.
Andrade E, Matsudo V, Andrade D, Araujo T, Oliveira L, Braggion GRV (2001) Level of physical activity in adults, including and excluding walking, according to the knowledge of Agita Sâo Paulo Program. American College of Sports Medicine Scientific Conference, St Louis, MO.
Arai Y, Hisamichi S (1998) Self-reported exercise frequency and personality: a population-based study in Japan. Perceptual and Motor Skills, 87:13711375.

Arciero PJ, Vukovich MD, Holloszy JO, Racette SB, Kohrt WM (1999) Comparison of short-term diet and exercise on insulin action in individuals with abnormal glucose tolerance. Journal of Applied Physiology, 86: 1930-1935.

Armstrong TP, Bauman A, Davies J (2000) Physical activity patterns of Australian adults-results of the 1999 National Physical Activity survey. Australian Institute of Health and Welfare, Canberra.
Arroll B, Beaglehole R (1992) Does physical activity lower blood pressure? A critical review of the clinical trials. Journal of Clinical Epidemiology, 45:419-428.

Aspray TJ, Mugusi F, Rashid S et al. (2000) Rural and urban differences in diabetes prevalence in Tanzania: the role of obesity, physical inactivity and urban living. Transactions of the Royal Society of Tropical Medicine and Hygiene, 94:637-644.

Baan CA, Stolk RP, Grobbee DE, Witteman JC, Feskens EJ (1999) Physical activity in elderly subjects with impaired glucose tolerance and newly diagnosed diabetes mellitus. American Journal of Epidemiology, 149:219-227.
Ballard-Barbash R, Schatzkin A, Taylor P, Kahle L (1990) Association of change in body mass with breast cancer. Cancer Research, 50:2152-2155.
Bashir SA, Duffy SW (1997) The correction of risk estimates for measurement error. Annals of Epidemiology, 7:154-164.
Bell ACG (2001) Weight gain and its predictors in Chinese adults. International Journal of Obesity and Related Metabolic Disorders, 25:1079-1086.

Berlin JA, Colditz G (1990) A meta-analysis of physical activity in the prevention of coronary heart disease. American Journal of Epidemiology, 132: 612-628.

Bernstein L, Henderson BE, Hanisch R, Sullivan-Halley J, Ross RK (1994) Physical exercise and reduced risk of breast cancer in young women. Journal of the National Cancer Institute, 86:1403-1408.
Bieber E, Barnes R (2001) Breast cancer and HRT—what are the data? International Journal of Fertility and Women's Medicine, 46:73-78.

Bijnen F, Caspersen C, Feskens E, Saris W, Mosterd W, Kromhout D (1998) Physical activity and 10 -year mortality from cardiovascular diseases and all causes. The Zutphen Elderly Study. Archives of Internal Medicine, 158:1499-1505.
Blair S, Kampert J, Kohl H et al. (1996) Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. Journal of American Medical Association, 276:205-210.
Blair S, Kohl H, Barlow C, Paffenbarger RS Jr, Gibbons L, Macera C (1995) Changes in physical fitness and all-cause mortality. Journal of American Medical Association, 273:1093-1098.

Blair SN, Cheng Y, Holder JS (2001) Is physical activity or physical fitness more important in defining health benefits? Medicine and Science in Sports and Exercise, 33:S379-399.

Blair SN, Jackson AS (2001) Physical fitness and activity as separate heart disease risk factors: a meta-analysis. Medicine and Science in Sports and Exercise, 33:762-764.

Blair SN, Kohl HW, Gordon NF, Paffenbarger RS Jr (1992) How much physical activity is good for health? Annual Review of Public Health, 13:99-126.
Blair SN, Kohl HW, Paffenbarger RS Jr, Clark DG, Cooper KH, Gibbons LW (1989) Physical fitness and all-cause mortality. A prospective study of healthy men and women. Journal of American Medical Association, 262:2395-2401.
Bouchard C (2001) Physical activity and health: introduction to the dose-response symposium. Medicine and Science in Sports and Exercise, 33:347-350.

Breslow RA, Ballard-Barbash R, Munoz K, Graubard BI (2001) Long-term recreational physical activity and breast cancer in the National Health and Nutrition Examination Survey I Epidemiologic Follow-up study. Cancer Epidemiology, Biomarkers and Prevention, 10:805-808.
Brill P, Macera C, Davis D, Blair S, Gordon N (2000) Muscular strength and physical function. Medicine and Science in Sports and Exercise, 32:412416.

Brownson R, Chang J, Davis J, Smith C (1991) Physical activity on the job and cancer in Missouri. American Journal of Public Health, 81:639-642.
Brownson RC, Zahm SH, Chang JC, Blair A (1989) Occupational risk of colon cancer. An analysis by anatomic subsite. American Journal of Epidemiology, 130:675-687.

Bull FC, Milligan R, Rosenberg M, MacGowan H (2000) Physical activity levels of Western Australian adults, 1999. The Health Department of Western Australia and the Ministry of Sport and Recreation, Perth, Western Australia. (Available at http://www.ausport.gov.au/fulltext/2001/wa/PActivity.pdf).
Calle EE, Murphy TK, Rodriguez C, Thun MJ, Heath CW Jr (1998) Occupation and breast cancer mortality in a prospective cohort of US women. American Journal of Epidemiology, 148:191-197.
Carpenter CL, Ross RK, Paganini-Hill A, Bernstein L (1999) Lifetime exercise activity and breast cancer risk among post-menopausal women. British Journal of Cancer, 80:1852-1858.

Caspersen CJ, Stephens T (1994) The demography of physical activity. In: Physical activity, fitness and health. Bouchard C, Shepard R, Stephens T, eds. Human Kinetics, Champaign, IL.
CDC (1998) Behavioral risk factor surveillance system questionnaire. Centers for Disease Control and Prevention. Available at http://www.cdc.gov/ brfss/questionnaires/pdf-ques/98brfss.pdf.
Cerhan JR, Chiu BC-H, Wallace RB et al. (1998) Physical activity, physical function, and the risk of breast cancer in a prospective study among elderly women. Journals of Gerontology, 53A:M351-356.
Chen CL, White E, Malone KE, Daling JR (1997) Leisure-time physical activity in relation to breast cancer among young women (Washington, United States). Cancer Causes and Control 8:77-84.

Chockalingam A (2000) Let's not make the same mistakes. The Lancet, 356:S9.
Clemmensen IH (1998) Physical activity and cancer. [Dissertation.] University of Copenhagen, Denmark.

Cockerham WC, Sneed C (2001) Reanalysis of data from the Russian Longitudinal Monitoring Study Survey 6. Department of Sociology, University of Alabama at Birmingham, AL (unpublished). Data set available from the University of North Carolina http://www.cpc.unc.edu/rlms/.
Colditz G, Cannuscio C, Frazier A (1997) Physical activity and reduced risk of colon cancer: implications for prevention. Cancer Causes and Control, 8: 649-667.

Collaborative Group on Hormonal Factors in Breast Cancer (1997) Breast cancer and hormone replacement therapy: collaborative re-analysis of data from 51 epidemiological studies of 52705 women with breast cancer and 108411 women without breast cancer. The Lancet, 350:1047-1059.

Coogan PF, Newcomb PA, Clapp RW, Trentham-Dietz A, Baron JA, Longnecker MP (1997) Physical activity in usual occupation and risk of breast cancer (United States). Cancer Causes and Control, 8:626-631.
Coogan PF, Aschengrau A (1999) Occupational physical activity and breast cancer risk in the upper Cape Cod cancer incidence study. American Journal of Industrial Medicine, 36:279-285.
Cortez R, Jacoby ER, Alarcrón G (2002) Determinantes de la obesidad y el sobrepeso en el Perú [Determinants of obesity and overweight in Peru]. In: Salud, equidad y pobreza en el Perú: teoría y nuevas evidencias [Health, equity and poverty in Perú: theory and new evidence]. Cortez R, ed. Centro de Investigación de la Universidad del Pácifico, Lima.

Craig C. (2001) Data from the 1998 physical activity monitor. Unpublished document available at http://www.cflri.ca.

Craig CL, Marshall AL, Sjöström M et al. (2003) International physical activity questionnaire: 12 -country reliability and validity. Medicine and Science in Sports and Exercise, 35:1381-1395.
Datafohla (1997) Pesquisa de prática de atividades físicas no Brasil. [Research on physical activity practices in Brazil]. Relatório Datafolha, Brasil. Available at http://www1.folha.uol.com.br/folha/datafolha.

Dela F, Handberg A, Mikines K, Vinten J, Galbo H (1993) GLUT 4 and insulin receptor binding and kinase activity in trained human muscle. Journal Physiology London, 469:615-624.

Department of Transport Local Government and the Regions (2002) UK National Travel survey. The Stationery Office, London.
Diabetes Prevention Program Research Group (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. The New England Journal of Medicine, 346:393-403.
Dishman RK, Buckworth J (1996) Increasing physical activity: a quantitative synthesis. Medicine and Science in Sports and Exercise, 28:706-719.
Donahue RP, Abbott RD, Reed DM, Yano K (1988) Physical activity and coronary heart disease in middle-aged and elderly men: the Honolulu Heart Program. American Journal of Public Health, 78:683-685.

Dorgan JF, Brown C, Barrett M et al. (1994) Physical activity and risk of breast cancer in the Framingham Heart study. American Journal of Epidemiology, 139:662-669.

Dosemeci M, Hayes RB, Vetter R et al. (1993) Occupational physical activity, socioeconomic status, and risks of 15 cancer sites in Turkey. Cancer Causes and Control, 4:313-321.
Dowse GK, Zimmet PZ, Gareeboo H et al. (1991) Abdominal obesity and physical inactivity as risk factors for NIDDM and impaired glucose tolerance in Indian, Creole, and Chinese Mauritians. Diabetes Care, 14:271-282.
Drinkwater B (1994) 1994 C. H. McCloy research lecture: does physical activity play a role in preventing osteoporosis. Research Quarterly for Exercise and Sport, 65:197-206.

Du S, Lu B, Zhai F, Popkin BM (2002) The nutrition transition in China: a new stage of the Chinese diet. In: Nutrition transition: diet and disease in the developing world. Caballero B, Popkin BM, eds. Academic Press, San Diego, CA.

Duncan BB, Schmidt MI, Achutti AC, Polanczyk CA, Benia LR, Maia AA (1993) Socioeconomic distribution of noncommunicable disease risk factors in urban Brazil: the case of Porto Alegre. Bulletin of the Pan American Health Organization, 27:337-349.
Dunn AL, Trivedi MH, O’Neal HA (2001) Physical activity dose-response effects on outcomes of depression and anxiety. Medicine and Science in Sports and Exercise, 33:S587-597.

Durstine J, Haskell W (1994) Effects of exercise training on plasma lipids and lipoproteins. Exercise and Sport Sciences Reviews, 22:477-521.

Eaton CB (1992) Relation of physical activity and cardiovascular fitness to coronary heart disease, Part I: a meta-analysis of the independent relation of physical activity and coronary heart disease. Journal of the American Board Family Practice, 5:31-42.
Ekelund LG, Haskell WL, Johnson JL, Whaley FS, Criqui MH, Sheps DS (1988) Physical fitness as a predictor of cardiovascular mortality in asymptomatic North American men. The Lipid Research Clinics Mortality Follow-up study. New England Journal of Medicine, 319:1379-1384.

Ellekjaer H, Holmen J, Ellekjaer E, Vatten L (2000) Physical activity and stroke mortality in women. Ten-year follow-up of the Nord-Trondelag health survey, 1984-1986. Stroke, 31:14-18.

Evans J, Roberts G (1987) Physical competence and the development of children's peer relations. Quest, 39:23-25.
Evenson K, Rosamond W, Cai J et al. (1999) Physical activity and ischaemic stroke risk. The Atherosclerosis Risk in Communities study. Stroke, 30:1333-1339.
Ezenwaka CE, Akanji AO, Akanji BO, Unwin NC, Adejuwon CA (1997) The prevalence of insulin resistance and other cardiovascular disease risk factors in healthy elderly southwestern Nigerians. Atherosclerosis, 128:201-211.
Fagard R, Tipton C (1994) Physical activity, fitness and hypertension. In: Physical activity, fitness and bealth. Bouchard C, Shephard R, Stephens T, eds. Human Kinetics, Champaign, IL.

Federal Highway Administration Research and Technical Support Center (1997) Nationwide Personal Transportation survey. U.S. Department of Transportation, Washington, DC. Available at http://www-cta.orn/gov/npts.
Fioretti F, Tavani A, Bosetti C et al. (1999) Risk factors for breast cancer in nulliparous women. British Journal of Cancer, 79:1923-1928.
Folsom AR, Arnett DK, Hutchinson RG, Liao F, Clegg LX, Cooper LS (1997) Physical activity and incidence of coronary heart disease in middle-aged women and men. Medicine and Science in Sports and Exercise, 29:901-909.
Folsom AR, Caspersen CJ, Taylor HL et al. (1985) Leisure time physical activity and its relationship to coronary risk factors in a population-based sample. The Minnesota Heart Survey. American Journal of Epidemiology, 121:570-579.

Folsom AR, Kushi LH, Hong CP (2000) Physical activity and incident diabetes mellitus in postmenopausal women. American Journal of Public Health, 90:134-138.

Forrest KY, Bunker CH, Kriska AM, Ukolif A, Huston SL, Markovic N (2001) Physical activity and cardiovascular risk factors in a developing population. Medicine and Science in Sports and Exercise, 33:1598-1604.
Fraser G, Pearce N (1993) Occupational physical activity and risk of cancer of the colon and rectum in New Zealand males. Cancer Causes and Control, 4:45-50.

Fraser G, Shavlik D (1997) Risk factors, lifetime risk, and age at onset of breast cancer. Annals of Epidemiology, 7:375-382.

Friedenreich CM, Courneya KS, Bryant HE (2001) Relation between intensity of physical activity and breast cancer risk reduction. Medicine and Science in Sports and Exercise, 33:1538-1545.
Friedenreich CM, Rohan TE (1995) Physical activity and risk of breast cancer. European Journal of Cancer Prevention, 4:145-151.
Friedenreich CM, Thune I (2001) A review of physical activity and prostate cancer risk. Cancer Causes and Control, 12:461-475.

Frisch RE, Wyshak G, Witschi J, Albright N, Albright T, Schiff I (1987) Lower lifetime occurrence of breast cancer and cancers of the reproductive system among former college athletes. International Journal of Fertility, 32:217-225.

Fulton-Kehoe D, Hamman RF, Baxter J, Marshall J (2001) A case-control study of physical activity and non-insulin dependent diabetes mellitus (NIDDM). The San Luis Valley Diabetes study. Annals of Epidemiology, 11:320-327.
Gammon MD, Schoenberg JB, Britton JA et al. (1998) Recreational physical activity and breast cancer risk among women under age 45 years. American Journal of Epidemiology, 147:273-280.
Garabrant GH, Peters JM, Mack TM, Bernstein L (1984) Job activity and colon cancer risk. American Journal of Epidemiology, 119:1005-1014.
Garcia-Palmieri MR, Costas RJ, Cruz-Vidal M, Sorlie PD, Havlik RJ (1982) Increased physical activity: a protective factor against heart attacks in Puerto Rico. American Journal of Cardiology, 50:749-755.

Gerhardsson M, Floderus B, Norell SE (1988) Physical activity and colon cancer risk. International Journal of Epidemiology, 17:743-746.

Gerhardsson M, Norell SE, Kiviranta H, Pedersen NL, Ahlbom A (1986) Sedentary jobs and colon cancer. American Journal of Epidemiology, 123: 775-780.

Gerhardsson de Verdier M, Steineck G, Hagman U, Rieger A, Norell SE (1990) Physical activity and colon cancer: a case-referent study in Stockholm. International Journal of Cancer, 46:985-989.
Gillespie ND, McMurdo M (1998) Falls in old age: inevitable or preventable? Scottish Medical Journal, 43:101-103.
Gilliland F, Yu-Fen L, Baumgartner K, Crumley D, Samet J (2001) Physical activity and breast cancer risk in Hispanic and non-Hispanic white women. American Journal of Epidemiology, 154:442-450.

Gillum R, Mussolino M, Ingram D (1996) Physical activity and stroke incidence in women and men. American Journal of Epidemiology, 143:860-869.
Giovannucci E, Ascherio A, Rimm EB, Colditz G, Stampfer MJ, Willett WC (1995) Physical activity, obesity, and risk for colon cancer and adenoma in men. Annals of Internal Medicine, 122:327-334.
Giovannucci E, Leitzmann M, Spiegelman D et al. (1998) A prospective study of physical activity and prostate cancer in male health professionals. Cancer Research, 58:5117-5122.
Giovannucci E, Pollak M, Platz E et al. (2000) A prospective study of plasma insulin-like growth factor-I and binding protein-3 and risk of colorectal neoplasia in women. Cancer Epidemiology, Biomarkers and Prevention, 9:345-349.

Glenister D (1996) Exercise and mental health: a review. Journal of the Royal Society of Health, 116:7-13.
Gregg E, Pereira M, Caspersen CJ (2000) Physical activity, falls, and fractures among older adults: a review of the epidemiologic evidence. Journal of the American Geriatrics Society, 48:883-893.

Greenland S. (1987) Quantitative methods in the review of epidemiologic literature. Epidemiologic Reviews, 9:1-30.

Grisso JA, Kelsey JL, O’Brien LA et al. (1997) Risk factors for hip fracture in men. hip fracture study group. American Journal of Epidemiology, 145: 786-793.

Grundy SM, Blackburn G, Higgins M, Lauer R, Perri MG, Ryan D (1999) Physical activity in the prevention and treatment of obesity and its comorbidities: evidence report of independent panel to assess the role of physical activity in the treatment of obesity and its comorbidities. Medicine and Science in Sports and Exercise, 31:1493-1500.

Gupta R, Gupta HG, Keswani P, Sharma A, Gupta VP, Gupta KD (1994) Coronary heart disease and coronary risk factor prevalence in rural Rajasthan. Journal of Applied Physiology International, 42:24-26.
Gurwitz JH, Field TS, Glynn RJ et al. (1994) Risk factors for non-insulindependent diabetes mellitus requiring treatment in the elderly. Journal of the American Geriatrics Society, 42:1235-1240.
Gutin B, Kasper M (1992) Can vigorous exercise play a role in osteoporosis prevention? A review. Osteoporosis International, 2:55-69.
Haapanen N, Miilunpalo S, Vuori I, Oja P, Pasanen M (1997) Association of leisure time physical activity with the risk of coronary heart disease, hypertension and diabetes in middle-aged men and women. International Journal of Epidemiology, 26:739-747.
Hagberg J, Montain S, Martin WI, Ehsani A (1989) Effect of exercise training in 60- to 69-year-old persons with essential hypertension. American Journal of Cardiology, 64:348-353.

Hahn R, Teutsch S, Rothberg R, Marks J (1990) Excess deaths from nine chronic diseases in the United States, 1986. Journal of American Medical Association, 264:2654-2659.

Ham S (2001a) Unpublished analysis of National Physical Activity Survey (NPAS) 1999-2000 data. Available from author at Epidemiology Section, Division of Nutrition and Physical Activity, Centers for Disease Control and Prevention, Atlanta, GA.

Ham S (2001b) Unpublished analysis of the 1995 U.S. National Personal Transportation Survey data. Details available from author at Epidemiology Section, Division of Nutrition and Physical Activity, Centers for Disease Control and Prevention, Atlanta, GA. Data set available at http://wwwcta.ornl.gov/npts/1995/doc/index.shtml.
Ham S (2001c) Unpublished analysis of the China Health and Nutrition survey. Details available from author at Epidemiology Section, Division of Nutrition and Physical Activity, Centers for Disease Control and Prevention, Atlanta, GA. Data set available at http://www.cpc.unc.edu/.

Hardman A, Stensel D (2003) Physical activity and health: the evidence explained. Routledge, London.

Harmsen P, Rosengren A, Tsipogianni A, Wilhelmsen L (1990) Risk factors for stroke in middle-aged men in Goteborg, Sweden. Stroke, 21:223-229.

Hassmen P, Koivula N, Uutela A (2000) Physical exercise and psychological wellbeing: a population study in Finland. Preventive Medicine, 30:17-25.

Hein HO, Suadicani P, Gyntelberg F (1992) Physical fitness or physical activity as a predictor of ischaemic heart disease? A 17-year follow-up in the Copenhagen Male study. Journal of Internal Medicine, 232:471-479.
Helmrich SP, Ragland DR, Paffenbarger RS Jr (1994) Prevention of non-insulindependent diabetes mellitus with physical activity. Medicine and Science in Sports and Exercise, 26:824-830.
Henderson B, Ross R, Judd H, Krailo M, Pike M (1985) Do regular ovulatory cycles increase breast cancer risk? Cancer, 56:1206-1208.
Herman WH, Ali MA, Aubert RE et al. (1995) Diabetes mellitus in Egypt: risk factors and prevalence. Diabetic Medicine, 12:1126-1131.
Hill J, Melanson E (1999) Overview of the determinants of overweight and obesity: current evidence and research issues. Medicine and Science in Sports and Exercise, 33:S515-521.

Hillary Commission (1998) Fitness and leisure report: a discussion paper by the Physical Activity Taskforce. Hillary Commission for Sport, Wellington, New Zealand. Accessed in 2001 from: www.hillarysport.org.nz. Details now available at http://www.sparc.org.nz.
Hillsdon M, Cavill N, Nanchahal K, Diamond A, White IR (2001) National level promotion of physical activity: results from England's ACTIVE for LIFE Campaign. Journal of Epidemiology and Community Health, 55: 755-761.

Hillsdon M, Thorogood M, Anstiss T, Morris J (1995) Randomised controlled trials of physical activity promotion in free living populations: a review. Journal of Epidemiology and Community Health, 49:448-453.

Hirose K, Tajima K, Hamajima N et al. (1995) A large-scale, hospital-based casecontrol study of risk factors of breast cancer according to menopausal status. Japanese Journal of Cancer Research, 86:146-154.
Hong Y, Bots ML, Pan X et al. (1994) Physical activity and cardiovascular risk factors in rural Shanghai, China. International Journal of Epidemiology, 23:1154-1158.

Hu FB, Leitzmann MF, Stampfer MJ, Colditz GA, Willett WA, Rimm EB (2001) Physical activity and television watching in relation to risk for type 2 diabetes mellitus in men. Archives of Internal Medicine, 161:1542-1548.

Hu FB, Sigal RJ, Rich-Edwards JW et al. (1999) Walking compared with vigorous physical activity and risk of type 2 diabetes in women: a prospective study. Journal of American Medical Association, 282:1433-1439.

Hu FB, Stampfer MJ, Colditz G et al. (2000) Physical activity and risk of stroke in women. Journal of American Medical Association, 283:2961-2967.
Hu G, Pekkarinen H, Hanninen O, Yu Z, Guo Z, Tian H (2002) Commuting, leisure-time physical activity, and cardiovascular risk factors in China. Medicine and Science in Sports and Exercise, 34:234-238.
Hu YH, Nagata C, Shimizu H, Kaneda N, Kashiki Y (1997) Association of body mass index, physical activity, and reproductive histories with breast cancer: a
case-control study in Gifu, Japan. Breast Cancer Research and Treatment, 43:65-72.

Huang Y, Macera C, Blair S, Brill P, Kohl H, Kronenfeld J (1998) Physical fitness, physical activity, and functional limitation in adults aged 40 and older. Medicine and Science in Sports and Exercise, 30:1430-1435.
Hypertension Study Group (2001) Prevalence, awareness, treatment and control of hypertension among elderly in Bangladesh and India: a Multicentre study. Bulletin of the World Health Organization, 79:490-500.
Institute of European Food Studies (2001) A Pan-EU survey on consumer attitudes to physical activity, body-weight and health-a report by the Institute of European Food Studies. Trinity College, Dublin.

Instituto Gallup De La Argentina (2001) Results from a national survey of physical activity in Argentine. Unpublished document available at http://www.gallup-international.com/scripts/member.asp?id=2.

IARC (2002) Weight control and physical activity. IARC Handbook of Cancer Prevention (Vol. 6). International Agency for Research on Cancer Press, Lyon.
Ivy J, Zderic T, Fogt D (1999) Prevention and treatment of non-insulindependent diabetes mellitus. In: Exercise and sport sciences reviews. Holloszy J, ed. Lippincott Williams and Wilkins, Philadelphia, PA.
Iwai N, Yoshiike N, Saitoh S, Nose T, Kushiro T, Tanaka H (2000) Leisure-time physical activity and related lifestyle characteristics among middle-aged Japanese. Japan Lifestyle Monitoring Study Group. Journal of Epidemiology, 10:226-233.

Jacobs DR, Ainsworth BE, Hartman TJ, Leon AS (1993) A simultaneous evaluation of 10 commonly used physical activity questionnaires. Medicine and Science in Sports and Exercise, 25:81-91.
Jadue L, Vega J, Escobar MC et al. (1999) Risk factors for non communicable diseases: methods and global results of the CARMEN program basal survey. Revista Medica de Chile, 127:1004-1013.
James SA, Jamjoum L, Raghunathan TE, Strogatz DS, Furth ED, Khazanie PG (1998) Physical activity and NIDDM in African-Americans. The Pitt County study. Diabetes Care, 21:555-562.
Johansson S, Rosengren A, Tsipogianni A, Ulvenstam G, Wiklund I, Wilhelmsen L (1988) Physical inactivity as a risk factor for primary and secondary coronary events in Goteborg, Sweden. European Heart Journal, 9:S8-19.

Kannel WB, Belanger A, D’Agostino R, Israel I (1986) Physical activity and physical demand on the job and risk of cardiovascular disease and death: the Framingham study. American Heart Journal, 112:820-825.

Kaprio J, Kujala UM, Koskenvuo M, Sarna S (2000) Physical activity and other risk factors in male twin-pairs discordant for coronary heart disease. Atherosclerosis, 150:193-200.
Kato I, Tominaga S, Ikari A (1990) A case-control study of male colorectal cancer in Aichi Prefecture, Japan: with special reference to occupational activity level, drinking habits and family history. Japanese Journal of Cancer Research, 81:115-121.

Katzmarzyk PT, Leonard WR, Stephen MA, Berti PR, Ross AG (1996) Differences between observed and predicted energy costs at rest and during exercise in three subsistence-level populations. American Journal of Physical Anthropology, 99:537-545.
Kelley D, Goodpaster B (2001) Effects of exercise on glucose homeostasis in type 2 diabetes mellitus. Medicine and Science in Sports and Exercise, 33: S495-501.
Kelley G, McClellan P (1994) Antihypertensive effects of aerobic exercise: a brief meta-analytic review of randomized controlled trials. American Journal of Hypertension, 7:115-119.

Kesaniemi YA, Danforth EJ, Jensen MD, Kopelman PG, Lefebvre P, Reeder BA (2001) Dose-response issues concerning physical activity and health: an evidence-based symposium. Medicine and Science in Sports and Exercise, 33:S351-358.

Kiely DK, Wolf PA, Cupples LA, Beiser AS, Kannel WB (1994) Physical activity and stroke risk-the Framingham study. American Journal of Epidemiology, 140:608-620.

Kohl HW (2001) Physical activity and cardiovascular disease: evidence for a dose response. Medicine and Science in Sports and Exercise, 33:S472-483.
Kokkinos P, Fernhall B (1999) Physical activity and high density lipoprotein cholesterol levels: what is the relationship? Sports Medicine, 28:307-314.
Kono S, Handa K, Hayabuchi H et al. (1999) Obesity, weight gain and risk of colon adenomas in Japanese men. Japanese Journal of Cancer Research, 90:805-811.

Kramsch D, Aspen A, Abramowitz B, Kreimendahl T, Hood WJ (1981) Reduction of coronary atherosclerosis by moderate conditioning exercise in monkeys on an atherogenic diet. New England Journal of Medicine, 305: 1483-1489.

Kriska AM, Caspersen CJ (1997) Introduction to a collection of physical activity questionnaires. Medicine and Science in Sports and Exercise, 29:S59.

Kriska AM, LaPorte RE, Pettitt DJ et al. (1993) The association of physical activity with obesity, fat distribution and glucose intolerance in Pima Indians. Diabetologia, 36:863-869.

Kujala U, Kaprio J, Kannus P, Sarna S, Koshenvuo M (2000) Physical activity and osteoporotic hip fracture risk in men. Archives of Internal Medicine, 160:705-708.

Kujala U, Kaprio J, Sarna S (1994) Osteoarthritis of weight bearing joints of lower limbs in former elite male athletes. British Medical Journal, 308: 231-234.

Kujala UM, Kettunen J, Paananen H et al. (1995) Knee osteoarthritis in former runners, soccer players, weight lifters, and shooters. Arthritis and Rheumatology, 38:539-46I.

La Vecchia C, Brinton L, McTiernan A (2001) Menopause, hormone replacement therapy and cancer. Maturitas, 39:97-115.

LaCroix AZ, Leveille SG, Hecht JA, Grothaus LC, Wagner EH (1996) Does walking decrease the risk of cardiovascular disease hospitalizations and death in older adults? Journal of American Geriatrics Society, 44:113-120.

Laforge RG, Rossi JS, Prochaska JO, Velicer WF, Levesque DA, McHorney CA (1999) Stage of regular exercise and health-related quality of life. Preventive Medicine, 28:349-360.
Lakka TA, Venäläinen JM, Rauramaa R, Salonen R, Tuomilehto J, Salonen JT (1994) Relation of leisure-time physical activity and cardiorespiratory fitness to the risk of acute myocardial infarction in men. New England Journal of Medicine, 330:1549-1554.

Lane N (1995) Exercise: a cause of osteoarthritis. Journal of Rheumatology, 43:S3-6.

Lanyon L (1993) Osteocytes, strain detection, bone modelling and remodelling. Calcified Tissue International, 53:S102-107.

Latikka P, Pukkala E, Vihko V (1998) Relationship between the risk of breast cancer and physical activity. Sports Medicine, 26:133-143.

Laughlin M (1994) Effects of exercise training on coronary circulation: introduction. Medicine and Science in Sports and Exercise, 26:1226-1229
Lee I-M, Skerrett P-J (2001) Physical activity and all-cause mortality: what is the dose-response relation? Medicine and Science in Sports and Exercise, 33:S459-471.

Lee I-M, Hennekens C, Berger K, Buring J, Manson J (1999) Exercise and risk of stroke in male physicians. Stroke, 30:1-6.
Lee I-M, Manson JE, Ajani U, Paffenbarger RS Jr, Hennekens CH, Buring JE (1997) Physical activity and risk of colon cancer: the Physicians' Health study (United States). Cancer Causes and Control, 8:568-574.

Lee I-M, Paffenbarger RS Jr (2000) Associations of light, moderate and vigorous intensity physical activity with longevity. American Journal of Epidemiology, 151:293-299.

Lee I-M, Paffenbarger RS Jr, Hsieh C (1991) Physical activity and risk of developing colorectal cancer among college alumni. Journal of the National Cancer Institute, 83:1324-1329.
Lee I-M, Rexrode KM, Cook NR, Manson JE, Buring JE (2001a) Physical activity and coronary heart disease in women: is "no pain, no gain" passé? Journal of the American Medical Association, 285:1447-1454.
Lee I-M, Rexrode KM, Cook NR, Hennekens CH, Buring JE (2001b) Physical activity and breast cancer risk: the Women's Health study. Cancer Causes Control, 12:137-145.
Lee I-M, Sesso HD, Paffenbarger RS Jr (2000) Physical activity and coronary heart disease risk in men: does the duration of exercise episodes predict risk? Circulation, 102:981-986.

Leenders NY, Sherman WM, Nagaraja HN, Kien CL (2001) Evaluation of methods to assess physical activity in free-living conditions. Medicine and Science in Sports and Exercise, 33:1233-1240.

Leon AS, Connett J, Jacobs D, Rauramaa R (1987) Leisure-time physical activity levels and risk of coronary heart disease and death. The Multiple Risk Factor Intervention Trial. Journal of American Medical Association, 258:2388-2395.

Leon A (1991) Effects of exercise conditioning on physiological precursors of coronary heart disease. Journal of Cardiopulmonary Rehabilitation, 11:46-57.

Leonetti DL, Fujimoto WY, Wahl PW (1989) Early-life background and the development of non-insulin-dependent diabetes mellitus. American Journal of Physical Anthropology, 79:345-355.

Levi F, Pasche C, Lucchini F, La Vecchia C (1999) Occupational and leisure time physical activity and the risk of breast cancer. European Journal of Cancer, 35:775-778.

Levitt NS, Katzenellenbogen JM, Bradshaw D, Hoffman MN, Bonnici F (1993) The prevalence and identification of risk factors for NIDDM in urban Africans in Cape Town, South Africa. Diabetes Care, 16:601-607.
Levitt NS, Steyn K, Lambert EV et al. (1999) Modifiable risk factors for type 2 diabetes mellitus in a peri-urban community in South Africa. Diabetic Medicine, 16:946-950.
Lie H, Mundal R, Erikssen J (1985) Coronary risk factors and incidence of coronary death in relation to physical fitness. Seven-year follow-up study of middle-aged and elderly men. European Heart Journal, 6:147-157.

Lindsted KD, Tonstad S, Kuzma JW (1991) Self-report of physical activity and patterns of mortality in Seventh-Day Adventist men. Journal of Clinical Epidemiology, 44:355-364.

Liu S, Lee I, Linson P, Ajani U, Buring JE, Hennekens C (2000) A prospective study of physical activity and risk of prostate cancer in US physicians. International Journal of Epidemiology, 29:29-35.
Li Xiang-ru et al. (2001) Status quo and the strategies of the implementation of the national fitness program in urban communities in China [in Chinese]. Sport Science, 21:1-6.
Longnecker MP, Gerhardsson le Verdier M, Frumkin H, Carpenter CL (1995) A case-control study of physical activity in relation to risk of cancer of the right colon and rectum in men. International Journal of Epidemiology, 24:4250.

Lord S (1995) The effect of a 12 month exercise trial on balance, strength and falls in older women. Journal of American Geriatrics Society, 43: 1198-1206.

Luoto R, Prattala R, Uutela A, Puska P (1998) Impact of unhealthy behaviours on cardiovascular mortality in Finland, 1978-1993. Preventive Medicine, 27:93-100.

Lynch J, Helmrich SP, Lakka TA et al. (1996) Moderately intense physical activities and high levels of cardiorespiratory fitness reduce the risk of non-insulindependent diabetes mellitus in middle-aged men. Archives of Internal Medicine, 156:1307-1314.

Lynge E, Thygesen L (1988) Use of surveillance systems for occupational cancer: data from the Danish National system. International Journal of Epidemiology, 17:493-500.

Ma J, Pollak M, Giovannucci E (1999) Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-1 and IGFbinding protein-3. Journal of the National Cancer Institute, 91:620-625.
MacAuley D, Mc C, Stott G et al. (1998) Levels of physical activity, physical fitness and their relationship in the Northern Ireland Health and Activity Survey. International Journal of Sports Medicine, 19:503-511.
MacAuley D, McCrum EE, Stott G et al. (1996) Physical activity, physical fitness, blood pressure, and fibrinogen in the Northern Ireland health and activity survey. Journal of Epidemiology and Community Health, 50:258-263.
Macera C, Ham S, Jones D, Kimsey D, Ainsworth BE, Neff L (2001) Limitations on the use of a single screening question to measure sedentary behaviour. American Journal of Public Health, 91:2010-2012.
Macera CA, Powell KE (2001) Population attributable risk: implications of physical activity dose. Medicine and Science in Sports and Exercise, 33: S635-639.

MacMahon S, Peto R, Cutler J et al. (1990) Blood pressure, stroke and coronary heart disease. Part I, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. The Lancet, 335:765-774.

Manson J, Hu F, Rich-Edwards J et al. (1999) A prospective study of walking as compared with vigorous exercise in the prevention of coronary heart disease in women. New England Journal of Medicine, 341:650-658.
Manson JE, Nathan DM, Krolewski AS, Stampfer MJ, Willett WC, Hennekens CH (1992) A prospective study of exercise and incidence of diabetes among US male physicians. Journal of the American Medical Association, 268:63-67.
Manson JE, Rimm EB, Stampfer MJ et al. (1991) Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. The Lancet, 338: 774-778.

Marceau M, Kouame N, Lacourciere Y, Cleroux J (1993) Effects of different training intensities on 24-hour blood pressure in hypertensive subjects. Circulation, 88:2803-2811.

Marcus PM, Newcomb PA, Storer BE (1994) Early adulthood physical activity and colon cancer risk among Wisconsin women. Cancer Epidemiology, Biomarkers and Prevention, 3:641-644.

Marcus PM, Newman B, Moorman PG et al. (1999) Physical activity at age 12 and adult breast cancer risk (United States). Cancer Causes and Control, 10:293-302.
Markowitz S, Morabia A, Garibaldi K, Wynder E (1992) Effect of occupational and recreational activity on the risk of colorectal cancer among males: a casecontrol study. International Journal of Epidemiology, 21:1057-1062.
Martinez ME, Giovannucci E, Spiegelman D, Hunter DJ, Willett WC, Colditz GA (1997) Leisure-time physical activity, body size, and colon cancer in
women. Nurses' Health Study Research group. Journal of the National Cancer Institute, 89:948-955.

Martinez-Gonzalez MA, Martinez JA, Hu FB, Gibney MJ, Kearney J (1999) Physical inactivity, sedentary lifestyle and obesity in the European Union. International Journal of Obesity and Related Metabolic Disorders, 23: 1192-1201.

Matsusaki M, Ikeda M, Tashiro E, Koga M, Miura S, Ideishi M (1992) Influence of workload on the antihypertensive effect of exercise. Clinical and Experimental Pharmacology and Physiology, 19:471-479.
Matthews CE, Shu XO, Jin F et al. (2001) Lifetime physical activity and breast cancer risk in the Shanghai Breast Cancer Study. British Journal of Cancer, 84:994-1001.

Maxwell K, Tucker L (1992) Effects of weight training on the emotional wellbeing and body image of females: predictors of greatest benefit. American Journal of Health Promotion, 6:338-344.
McCormack G, Milligan R, Giles-Corti B, Clarkson JP (2003) Physical activity levels in Western Australian Adults 2002: results from the adult physical activity survey and pedometer study. Western Australian Government, Perth, Western Australia.
McKeown-Eyssen G (1994) Epidemiology of colorectal cancer revisited: are serum triglycerides and/or plasma glucose associated with risk? Cancer Epidemiology, Biomarkers and Prevention, 3:687-695.

McMurray RG, Ainsworth BE, Harrell JS, Griggs TR, Williams OD (1998) Is physical activity or aerobic power more influential on reducing cardiovascular disease risk factors? Medicine and Science in Sports and Exercise, 30:1521-1529.

McTiernan A, Stanford JL, Weiss NS, Daling JR, Voigt LF (1996) Occurrence of breast cancer in relation to recreational exercise in women age 50-64 years. Epidemiology, 7:598-604.
McTiernan A, Ulrich C, Slate S, Potter J (1998) Physical activity and cancer etiology: associations and mechanisms. Cancer Causes and Control, 9: 487-509.

Menotti A, Seccareccia F (1985) Physical activity at work and job responsibility as risk factors for fatal coronary heart disease and other causes of death. Journal of Epidemiology and Community Health, 39:325-329.
Mensink GBM, Ziese T, Kok FJ (1999) Benefits of leisure-time physical activity on the cardiovascular risk profile at older age. International Journal of Epidemiology, 28:659-666.
Mezzetti M, La-Vecchia C, Decarli A, Boyle P, Talamini R, Franceschi S (1998) Population attributable risk for breast cancer: diet, nutrition, and physical exercise. Journal of the National Cancer Institute, 90:389-394.
Misra A, Sharma R, Pandey RM, Khanna N (2001) Adverse profile of dietary nutrients, anthropometry and lipids in urban slum dwellers of northern India. European Journal of Clinical Nutrition, 9:727-734.

Mittendorf R, Longnecker MP, Newcomb PA et al. (1995) Strenuous physical activity in young adulthood and risk of breast cancer (United States). Cancer Causes and Control, 6:347-353.

Moore DB, Folsom AR, Mink PJ, Hong CP, Anderson KE, Kushi LH (2000) Physical activity and incidence of postmenopausal breast cancer. Epidemiology, 11:292-296.
Moore S (1994) Physical activity, fitness and atherosclerosis. In: Physical activity, fitness and health: international proceedings and consensus statement. Bouchard C, Shepherd R, Stephens T, eds. Human Kinetics, Champaign, IL.
Moradi T, Adami HO, Bergstrom R et al. (1999) Occupational physical activity and risk for breast cancer in a nationwide cohort study in Sweden. Cancer Causes and Control, 10:423-430.

Moradi T, Nyren O, Zack M, Magnusson C, Persson I, Adami HO (2000) Breast cancer risk and lifetime leisure-time and occupational physical activity (Sweden). Cancer Causes and Control, 11:523-531.

Morans R, Mohai P (1991) Leisure resources, recreation activity and the quality of life. In: Benefits of leisure. Driver B, Brown P, Peterson G, eds. Venture Publishing, Philadelphia, PA.
Moreau K, Degarmo R, Langley J et al. (2001) Increasing daily walking lowers blood pressure in postmenopausal women. Medicine and Science in Sports and Exercise, 33:1825-1831.

Morris JN, Clayton DG, Everitt MG, Semmence AM, Burgess EH (1990) Exercise in leisure time: coronary attack and death rates. British Heart Journal, 63:325-334.

Morris JN, Everitt MG, Pollard R, Chave SP, Semmence AM (1980) Vigorous exercise in leisure-time: protection against coronary heart disease. The Lancet, 2:1207-1210.

Mundal R, Erikssen J, Rodahl K (1987) Assessment of physical activity by questionnaire and personal interview with particular reference to fitness and coronary mortality. European Journal of Applied Physiology and Occupational Physiology, 56:245-252.
Murray CJL, Lopez AD, eds. (1996a) Quantifying the burden of disease and injury attributable to ten major risk factors. In: The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, vol. 1. Murray CJL, Lopez AD, eds. Harvard School of Public Health on behalf of WHO, Cambridge, MA.
Murray CJL, Lopez AD, eds. (1996b) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, vol. 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Neugut A, Terry M, Hocking G, Mosca L, Garbowski G, Forde K (1996) Leisure and occupational physical activity and risk of colorectal adenomatous polyps. International Journal of Cancer, 68:744-748.

North Carolina Population Center (2001a) Russian Longitudinal Monitoring survey. Unpublished document available at http://www.cpc.unc.edu/rlms/.

North Carolina Population Center (2001b) China Health and Nutrition survey. Unpublished document available at http://www.cpc.unc.edu/.

Okada K, Hayashi T, Tsumura K, Suematsu C, Endo G, Fujii S (2000) Leisuretime physical activity at weekends and the risk of type 2 diabetes mellitus in Japanese men: the Osaka Health Survey. Diabetic Medicine, 17:53-58.
Paeratakul S, Popkin BM, Keyou G, Adair LS, Stevens J (1998) Changes in diet and physical activity affect the body mass index of Chinese adults. International Journal of Obesity and Related Metabolic Disorders, 22:424-431.
Paffenbarger RS Jr, Hyde RT, Wing AL, Steinmetz CH (1984) A natural history of athleticism and cardiovascular health. Journal of American Medical Association, 252:491-495.

Paffenbarger RS Jr, Kampert JB, Lee I-M, Hyde RT, Leung RW, Wing AL (1994) Changes in physical activity and other lifeway patterns influencing longevity. Medicine and Science in Sports and Exercise, 26:857-865.

Paffenbarger RS Jr, Hyde R, Wing A, Jung D, Kampert J (1993) The association of changes in physical-activity level and other lifestyle characteristics with mortality among men. New England Journal of Medicine, 328:538-545.
Paffenbarger RS Jr, Wing AL, Hyde RT (1978) Physical activity as an index of heart attack risk in college alumni. American Journal of Epidemiology, 108:161-175.

Pan XR, Yang WY, Li GW, Liu J (1997) Prevalence of diabetes and its risk factors in China, 1994. National Diabetes Prevention and Control Cooperative Group. Diabetes Care, 20:1664-1669.

Pate RR, Pratt M, Blair SN, Macera C, Bouchard C, Buchner D et al. (1995) Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. Journal of American Medical Association, 273:402-407.
Pekkanen J, Nissinen A, Marti B, Tuomilehto J, Punsar S, Karvonen MJ (1987) Reduction of premature mortality by high physical activity: a 20 -year followup of middle-aged Finnish men. The Lancet, 1:1473-1477.
Pereira MA, FitzerGerald SJ, Gregg EW et al. (1997) A collection of physical activity questionnaires for health-related research. Medicine and Science in Sports and Exercise, 29:S1-205.

Perry IJ, Wannamethee SG, Walker MK, Thomson AG, Whincup PH, Shaper AG (1995) Prospective study of risk factors for development of non-insulin dependent diabetes in middle aged British men. British Medical Journal, 310:560-564.

Peters RK, Cady LD, Bischoff DP, Bernstein L, Pike MC (1983) Physical fitness and subsequent myocardial infarction in healthy workers. Journal of American Medical Association, 249:3052-3056.
Picavet H, Schouten J (2000) Physical load in daily life and low back problems in the general population-The MORGEN study. Preventive Medicine, 31:506-512.

Pitsillides A, Rawlenson S, Suswillo R, Bourrin S, Zaman G, Lanyon L (1995) Mechanical strain-induced NO production by bone cells: a possible role in adaptive bone (re)modelling. Journal of the Federation of American Societies for Experimental Biology, 9:1614-1622.
Pomerleau J, McKee M, Robertson A et al. (2000) Physical inactivity in the Baltic countries. Preventive Medicine, 31:665-672.
Pomrehn PR, Wallace RB, Burmeister LF (1982) Ischaemic heart disease mortality in Iowa farmers. The influence of life-style. Journal of American Medical Association, 248:1073-1076.
Powell K, Blair S (1994) The public health burdens of sedentary living habits: theoretical but realistic estimates. Medicine and Science in Sports and Exercise, 26:851-856.

Powell KE, Thompson PD, Caspersen CJ, Kendrick JS (1987) Physical activity and the incidence of coronary heart disease. Annual Review of Public Health, 8:253-287.

Pratt M, Macera CA, Blanton C (1999) Levels of physical activity and inactivity in children and adults in the United States: current evidence and research issues. Medicine and Science in Sports and Exercise, 31:S526-533.
Pucher J, Dijkstra L (2000) Making walking and cycling safer: lessons from Europe. Transportation Quarterly, 54:2-33.
Pukkala E, Poskiparta M, Apter D, Vihko V (1993) Life-long physical activity and cancer risk among Finnish female teachers. European Journal of Cancer Prevention, 2:369-376.

Reddy KS, Yusuf S (1998) Emerging epidemic of cardiovascular disease in developing countries. Circulation, 97:596-601.

Rockhill B, Willett WC, Hunter DJ, Manson JE, Hankinson SE, Colditz GA (1999) A prospective study of recreational physical activity and breast cancer risk. Archives of Internal Medicine, 159:2290-2296.
Rockhill B, Willett WC, Hunter DJ et al. (1998) Physical activity and breast cancer risk in a cohort of young women. Journal of the National Cancer Institute, 90:1155-1160.
Rodriguez BL, Curb JD, Burchfiel CM et al. (1994) Physical activity and 23-year incidence of coronary heart disease morbidity and mortality among middle-aged men. The Honolulu Heart Program. Circulation, 89:2540-2544.
Rogers MA, King DS, Hagber JM, Ehsani AA, Holloszy JO (1990) Effect of 10 days of physical inactivity on glucose tolerance in master athletes. Journal of Applied Physiology, 68:1833-1837.

Rosengren A, Wilhelmsen L (1997) Physical activity protects against coronary death and deaths from all causes in middle-aged men. Evidence from a 20year follow-up of the primary prevention study in Goteborg. Annals of Epidemiology, 7:69-75.
Sacco R, Gan R, Boden-Albala B et al. (1998) Leisure-time physical activity and ischaemic stroke risk. The Northern Manhattan Stroke Study. Stroke, 29:380-387.

Sallis J, Owen N (1999) Physical activity and behavioural medicine. Sage Publications, Thousand Oaks, CA.

Sallis JF, Saelens BE (2000) Assessment of physical activity by self-report: status, limitations, and future directions. Research Quarterly for Exercise and Sport, 71:S1-14.

Salonen JT, Puska P, Tuomilehto J (1982) Physical activity and risk of myocardial infarction, cerebral stroke and death: a longitudinal study in eastern Finland. American Journal of Epidemiology, 115:526-537.
Salonen JT, Slater JS, Tuomilehto J, Rauramaa R (1988) Leisure time and occupational physical activity: risk of death from ischaemic heart disease. American Journal of Epidemiology, 127:87-94.

Seccareccia F, Menotti A (1992) Physical activity, physical fitness and mortality in a sample of middle aged men followed-up 25 years. Journal of Medicine and Physical Fitness, 32:206-213.

Sesso HD, Paffenbarger RS Jr, Lee I-M (1998) Physical activity and breast cancer risk in the College Alumni Health Study (United States). Cancer Causes and Control, 9:433-439.
Sesso HD, Paffenbarger RS Jr, Lee I-M (2000) Physical activity and coronary heart disease in men: The Harvard Alumni Health study. Circulation, 102:975-980.

Severson R, Nomura A, Grove J, Stemmermann G (1989) A prospective analysis of physical activity and cancer. American Journal of Epidemiology, 130: 522-529.

Shaper A, Wannamethee G, Walker M (1994) Physical activity, hypertension and risk of heart attack in men without evidence of ischaemic heart disease. Journal of Human Hypertension, 8:3-10.

Shaper AG, Wannamethee G, Weatherall R (1991) Physical activity and ischaemic heart disease in middle-aged British men. British Heart Journal, 66:384-394.

Sherman SE, D'Agostino RB, Cobb JL, Kannel WB (1994a) Does exercise reduce mortality rates in the elderly? Experience from the Framingham Heart study. American Heart Journal, 128:965-972.
Sherman SE, D’Agostino RB, Cobb JL, Kannel WB (1994b) Physical activity and mortality in women in the Framingham Heart Study. American Heart Journal, 128:879-884.

Sherman SE, D'Agostino RB, Silbershatz H, Kannel WB (1999) Comparison of past versus recent physical activity in the prevention of premature death and coronary artery disease. American Heart Journal, 138:900-907.

Shoff S, Newcomb P, Trentham-Dietz A et al. (2000) Early-life physical activity and postmenopausal breast cancer: effect of body size and weight change. Cancer, Epidemiology, Biomarkers and Prevention, 9:591-595.
Simonsick EM, Lafferty ME, Phillips CL et al. (1993) Risk due to inactivity in physically capable older adults. American Journal of Public Health, 83:1443-1450.

Singh RB, Bajaj S, Niaz MA, Rastogi SS, Moshiri M (1998) Prevalence of type 2 diabetes mellitus and risk of hypertension and coronary artery disease in rural and urban population with low rates of obesity. International Journal of Cardiology, 66:65-72.
Slattery ML, Edwards SL, Boucher KM, Anderson K, Caan BJ (1999) Lifestyle and colon cancer: an assessment of factors associated with risk. American Journal of Epidemiology, 150:869-877.
Slattery ML, Jacobs D (1988) Physical fitness and cardiovascular disease mortality. The US Railroad study. American Journal of Epidemiology, 127:571-580.

Slattery ML, Jacobs D, Nichaman MZ (1989) Leisure time physical activity and coronary heart disease death. The US Railroad study. Circulation, 79:304-311.

Slattery ML, Potter J, Caan B et al. (1997) Energy balance and colon cancer-beyond physical activity. Cancer Research, 57:75-80.
Slattery ML, Schumacher MC, Smith KR, West DW, Abd-Elghany N (1988) Physical activity, diet, and risk of colon cancer in Utah. American Journal of Epidemiology, 128:989-999.
Sobngwi E, Mbanya JN, Unwin et al. (2002) Physical activity and its relationship with obesity, hypertension and diabetes in urban and rural Cameroon. International Journal of Obesity and Related Metabolic Disorders, 26: 1009-1016.

Sobolski J, Kornitzer M, De Backer G et al. (1987) Protection against ischaemic heart disease in the Belgian Physical Fitness Study: physical fitness rather than physical activity? American Journal of Epidemiology, 125:601-610.

Sonstroem R (1984) Exercise and self-esteem. Sport Science Review, 12:123-155.

Sparling PB, Noakes TD, Steyn K et al. (1994) Level of physical activity and CHD risk factors in black South African men. Medicine and Science in Sports and Exercise, 26:896-902.
Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC (2000) Primary prevention of coronary heart disease in women through diet and lifestyle. The New England Journal of Medicine, 343:16-22.
Steenland K, Nowlin S, Palu S (1995) Cancer incidence in the National Health and Nutrition Survey I. Follow-up data: diabetes, cholesterol, pulse and physical activity. Cancer Epidemiology, Biomarkers and Prevention, 4: 807-811.

Stefanick M, Wood PD (1994) Physical activity, lipid and lipoprotein metabolism, and lipid transport. In: Physical activity, fitness and health: International proceedings and consensus statement. Bouchard C, Shephard R, Stephens T, eds. Human Kinetics, Champaign, IL.
Sterne JA, Egger M (2001) Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. Journal of Clinical Epidemiology, 54: 1046-1055.

Sternfeld B (1992) Cancer and the protective effect of physical activity: the epidemiological evidence. Medicine and Science in Sports and Exercise, 24:1195-1209.

Steyn K, Jooste PL, Bourne L et al. (1991) Risk factors for coronary heart disease in the black population of the Cape Peninsula. The Black Risk study. South African Medical Journal, 79:480-485.
Steyn K, Jooste PL, Langenhoven ML et al. (1985) Coronary risk factors in the coloured population of the Cape Peninsula. South African Medical Journal, 67:619-625.

Suni J (2000) Health-related fitness test battery for middle-aged adults with emphasis on musculoskeletal and motor tests. [Dissertation.] University of Jyväskylä, Finland.
Taioli E, Barone J, Wynder E (1995) A case-control study on breast cancer and body mass. The American Health Foundation-Division of Epidemiology. European Journal of Cancer, 31A:723-728.

Tang L, Raab-Cullen M, Yee J, Jee WD (1995) Prostaglandin E2 increases skeletal response. Journal of Bone and Mineral Research, 10:S256.
Tang R, Wang JY, Lo SK, Hsieh LL (1999) Physical activity, water intake and risk of colorectal cancer in Taiwan: a hospital-based case-control study. International Journal of Cancer, 82:484-489.
Tavani A, Braga C, La Vecchia C et al. (1999) Physical activity and risk of cancers of the colon and rectum: an Italian case-control study. British Journal of Cancer, 79:1912-1916.

Taylor R, Ram P, Zimmet P, Raper LR, Ringrose H (1984) Physical activity and prevalence of diabetes in Melanesian and Indian men in Fiji. Diabetologia, 27:578-582.

Thor P, Konturek J, Konturek S, Anderson J (1985) Role of prostaglandins in control of intestinal motility. American Journal of Physiology, 248: G353-359.

Thun MJ, Calle EE, Namboodiri MM et al. (1992) Risk factors for fatal colon cancer in a large prospective study. Journal of the National Cancer Institute, 84:1491-1500.

Thune I, Brenn T, Lund E, Gaard M (1997) Physical activity and the risk of breast cancer. New England Journal of Medicine, 336:1269-1275.
Thune I, Furberg A-S (2001) Physical activity and cancer risk: dose-response and cancer, all sites and site-specific. Medicine and Science in Sports and Exercise, 33:S530-550.

Thune I, Lund E (1996) Physical activity and risk of colorectal cancer in men and women. British Journal of Cancer, 73:1134-1140.

Todoroki I, Shinchi K, Kono S, Imanishi K (1994) Lifestyle and glucose tolerance: a cross-sectional study of Japanese men. Annals of Epidemiology, 4:363-368.

Tomanek R (1994) Exercise-induced coronary angiogenesis: a review. Medicine and Science in Sports and Exercise, 26:1245-1251.

Tomeo CA, Colditz G, Willett WC et al. (1999) Harvard Report on Cancer Prevention. Volume 3: prevention of colon cancer in the United States. Cancer Causes Control, 10:167-180.

Travel Demand Management (1999) TravelSmart 2010: a 10-year plan. Department of Transport, Perth, Western Australia. Available at http:// www.dpi.wa.gov.au/travelsmart/index.html.
Tuomilehto J, Lindstrom J, Eriksson JG et al. (2001) Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. New England Journal of Medicine, 344:1343-1350.
Turner C, Owan I, Takano Y, Madali S, Murrell G (1995) Nitric oxide plays a role in bone mechanotransduction. Journal of Bone and Mineral Research, 10:S235.

Tutton P, Barkla D (1980) Influence of prostaglandin analogues on epithelial cell proliferation and xenograft growth. British Journal of Cancer, 41:47-51.

Ueji M, Ueno E, Osei-Hyiaman D, Takahashi H, Kano K (1998) Physical activity and the risk of breast cancer: a case-control study of Japanese women. Journal of Epidemiology, 8:116-122.
U.S. Department of Health and Human Services (1996) Physical activity and health: a report of the Surgeon General. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, The Presidents' Council on Physical Fitness and Sports, Atlanta, GA.
U.S. Department of Health and Human Services (2001) Increasing physical activity: a report on recommendations of the task force on community preventive services. Available at http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5018a1.htm.
Vena JE, Graham S, Zielezny M, Swanson MK, Barnes RE, Nolan J (1985) Lifetime occupational exercise and colon cancer. American Journal of Epidemiology, 122:357-365.
Verloop J, Rookus MA, van der Kooy K, van Leeuwen FE (2000) Physical activity and breast cancer risk in women aged 20-54 years. Journal of the National Cancer Institute, 92:128-135.

Vuori I (2001) Dose-response of physical activity and low back pain, osteoarthritis, and osteoporosis. Medicine and Science in Sports and Exercise, 33:S551-586.

Wang G (2002) personal correspondence on data on physical activity in Chinaresults published by the Ministry of Sport [in Chinese]. Division of Nutrition and Physical Activity, Centers for Disease Control and Prevention, Atlanta, GA.

Wannamethee S, Shaper A (1999) Physical activity and the prevention of stroke. Journal of Cardiovascular Risk, 6:213-216.
Wannamethee SG, Shaper AG, Alberti KG (2000) Physical activity, metabolic factors, and the incidence of coronary heart disease and type 2 diabetes. Archives of Internal Medicine, 160:2108-2116.

Wei M, Gibbons LW, Mitchell TL, Kampert JB, Lee CD, Blair SN (1999) The association between cardiorespiratory fitness and impaired fasting glucose and type 2 diabetes mellitus in men. Annals of Internal Medicine, 130:89-96.

Welten D, Kemper H, Post G et al. (1994) Weight-bearing activity during youth is a more important factor for peak bone mass than calcium intake. Journal of Bone and Mineral Research, 9:1089-1096.
White E, Jacobs EJ, Daling JR (1996) Physical activity in relation to colon cancer in middle-aged men and women. American Journal of Epidemiology, 144:42-50.

Whittemore A, Wu-Williams A, Lee M et al. (1990) Diet, physical activity, and colorectal cancer among Chinese in North America and China. Journal of the National Cancer Institute, 82:915-126.

Williams P-T (2001) Physical fitness and activity as separate heart disease risk factors: a meta-analysis. Medicine and Science in Sports and Exercise, 33: 754-761.

World Bank (1999) World development indicators 1999. World Bank, Development Data Group, Washington, DC.
Wyrwich KW, Wolinsky FD (2000) Physical activity, disability, and the risk of hospitalization for breast cancer among older women. Journals of Gerontology, 55:M418-421.
Wyshak G, Frisch R (2000) Breast cancer among former college athletes compared to non-athletes: a 15-year follow-up. British Journal of Cancer, 82:726-730.

Yano K, Reed DM, McGee DL (1984) Ten-year incidence of coronary heart disease in the Honolulu Heart Program. Relationship to biologic and lifestyle characteristics. American Journal of Epidemiology, 119:653-666.

Yu Z, Nissinen A (2000) Associations between socioeconomic status and, cardiovascular risk factors in an urban population in China. Bulletin of the World Health Organization, 78:1296-1305.
Yu Z, Nissinen A, Vartiainen E, Song G, Guo Z, Tian H (2000) Changes in cardiovascular risk factors in different socioeconomic groups: seven year trends in a Chinese urban population. Journal of Epidemiology and Community Health, 54:692-696.

Zabina H, Schmid T, Potemkina R, Glasunov I, Deev A (2002) Evaluation of physical activity levels in Russia based on the international physical activity questionnaire (IPAQ). American College of Sports Medicine, 49th annual meeting, May 2002, St. Louis, MO.

Zhang J, Feldblum P, Fortney J (1992) Moderate physical activity and bone density among perimenopausal women. American Journal of Public Health, 82:736-738.

## Chapter II

# Smoking and oral tobacco use 

Majid Ezzati and Alan D. Lopez

## Summary

Smoking has been causally associated with increased mortality from several diseases. This chapter provides global and regional estimates of premature mortality and disease burden in 2000 caused by tobacco use, including an analysis of uncertainty. It also describes a method for estimating the future burden of disease that could be avoided through smoking cessation or prevention.

Comparable data, especially age-specific data, on the prevalence of smoking are often unavailable or inaccurate. More importantly, current prevalence of smoking is a poor proxy for cumulative hazards of smoking, which depend on factors such as age at which smoking began, duration of smoking, the number of cigarettes smoked per day, cigarette characteristics such as tar and nicotine content or filter type, and smoking behaviour such as degree of inhalation. We used the smoking impact ratio (SIR) as a marker for accumulated smoking risk. SIR uses lung cancer mortality in excess of never-smokers as a biological marker for accumulated hazards of smoking. Lung cancer mortality data were from the Global Burden of Disease (GBD) mortality database. Neversmoker lung cancer mortality rates were estimated based on the household use of coal in unvented stoves for each subregion. ${ }^{1}$ Age-sex-specific SIR was divided into three categories: zero, medium ( $0<S I R \leq 0.5$ ), and high ( $0.5<\operatorname{SIR} \leq 1.0$ ).

Estimates of mortality and disease burden due to smoking were made for lung cancer, upper aerodigestive cancer, all other cancers, chronic obstructive pulmonary disease (COPD), other respiratory diseases, cardiovascular diseases and selected other medical causes. All diseases for which no plausible physio-biological causal mechanism is currently known were excluded from the analysis of the category "other medical causes", as were the impacts of maternal smoking during pregnancy. No estimates were made for non-medical causes (injuries). Estimates were
limited to ages 30 years and above. The American Cancer Society Cancer Prevention Study, Phase II (CPS-II), with follow-up for the years 1982-88, was the reference population and provided data on exposuredisease relationships. For China, the exposure-disease relationship was estimated from the retrospective proportional mortality analysis of Liu et al. (1998). Relative risks were corrected for confounding and extrapolation to other regions using conservative correction factors for groups of diseases. For non-fatal outcomes, we assumed that the same attributable fraction as mortality applied to all cancers and COPD (i.e. smoking changes mortality by changing incidence) and used only one half of this value for other health outcomes (i.e. assumed smoking changes mortality by changing incidence and severity/case fatality).

Estimates of reduction in risk after smoking cessation were obtained from the re-analysis of CPS-II, with follow-up for the years 1982-98, for lung cancer, COPD and cardiovascular diseases. To obtain the estimates of exposure under a "business-as-usual" scenario, we assigned male and female populations of countries to one of the ten main or transitional stages of a descriptive model of the smoking epidemic based on available prevalence data. This model was used to divide per capita tobacco consumption data into age-sex-specific estimates. Historical consumption and mortality trends were used to project lung cancer mortality and SIR under the business-as-usual scenario.

Quantitative analysis of uncertainty was conducted for five input parameters: population lung cancer mortality, never-smoker lung cancer mortality, reference population smoker and never-smoker lung cancer mortality, relative risks and relative risk correction factors. Uncertainty in population lung cancer was based on a critical review of the quality of country mortality reporting. Uncertainty in never-smoker mortality rates was based on the likely presence of other risk factors for lung cancer, and uncertainty in reference population lung cancer mortality and in relative risk estimates was estimated from the statistical uncertainty in CPS-II data and data from the Chinese retrospective proportional mortality analysis of Liu et al. (1998). Uncertainty in the correction factor was based on the potential and measured confounding of smoking hazard estimates due to other risk factors. Three other sources of uncertainty were discussed qualitatively: uncertainty associated with the use of SIR as the exposure variable, uncertainty associated with exposure to environmental tobacco smoke (ETS), and uncertainty in the estimates of non-fatal conditions.

In the year 2000, there were an estimated 4.83 ( $95 \%$ CI $3.94-5.93$ ) million deaths in the world attributable to smoking. Of these, 2.41 ( $95 \%$ CI 1.80-3.15) million deaths were in developing countries and 2.43 ( $95 \%$ CI $2.13-2.78$ ) million deaths in industrialized countries. There were 3.84 million global smoking-attributable deaths among men ( 2.02 million in developing countries and 1.81 million in industrialized countries) and 1.00 million among women ( 0.38 million in developing
countries and 0.61 million in industrialized countries). The leading causes of death due to smoking were cardiovascular diseases with 1.69 million deaths, COPD with 0.97 million deaths and lung cancer with 0.85 million deaths. In addition to those cases shared with smokers, there were an estimated 60000 deaths from oral tobacco use in SEAR-D.

Smoking and oral tobacco use accounted for $4.1 \%$ of healthy life years lost in the world in 2000. Given the demographic and epidemiological transitions and current smoking patterns in the developing world, the health loss from smoking will grow even larger unless effective interventions and policies that reduce smoking among males and prevent increases among females in developing countries are implemented.

## 1. Introduction

Tobacco is cultivated in many regions of the world and can be legally purchased in all countries. The dried leaf of the plant nicotiana tabacum is used for smoking, chewing or as snuff. Smoking has been causally associated with substantially increased risk of premature mortality from lung cancer as well as other medical causes (Doll et al. 1994; Liu et al. 1998; U.S. Department of Health and Human Services 1989; Zaridze and Peto 1986). As a result, in populations where smoking has been common for many decades, tobacco use accounts for a considerable proportion of premature mortality, as illustrated by estimates of smokingattributable deaths in industrialized countries (Peto et al. 1992).

There have been large increases in smoking in developing countries, especially among males, over the last part of the twentieth century (Corrao et al. 2000; WHO 1997). The first estimates of the health consequences of smoking in China and India have also shown substantially increased risk of mortality and disease among smokers (Dikshit and Kanhere 2000; Gupta and Mehta 2000; Liu et al. 1998; Niu et al. 1998; Gajalakshmi et al. 2003). This chapter provides estimates of premature mortality and morbidity caused by smoking in all regions of the world, with complete description of methods and data sources, including quantitative estimates of uncertainty. It also describes a method for estimating the future burden of disease due to smoking that could be avoided through cessation or prevention.

It is well-known that the accumulated hazards of smoking depend on factors such as the age at which smoking began, the number of cigarettes smoked per day, cigarette characteristics such as tar and nicotine content or filter type, and smoking behaviour such as degree of inhalation (Fletcher and Peto 1977; Liu et al. 1998; Peto 1986). Current prevalence of smoking alone is therefore an insufficient indicator of accumulated risk from smoking. Although such pattern variables have been studied in a few industrialized countries (Nicolaides-Bouman et al. 1993), few data are available elsewhere. This lack of knowledge about important parameters of smoking patterns and history, coupled with the fact that
smoking is currently increasing in many developing countries, motivates using an exposure variable that better describes accumulated risk in populations with varying smoking histories.

Peto et al. (1992) used data on absolute lung cancer mortality to obtain the proportions of mortality from lung cancer as well as various other diseases attributable to smoking. We extended this method for indirect estimation of excess mortality due to smoking to all regions of the world and also included analysis of morbidity and uncertainty. The large retrospective proportional mortality analysis in China (Liu et al. 1998) and newer studies from India provided a means for calibrating and verifying the method for developing countries using direct estimates from the world's largest countries.

### 1.1 Evidence for causality

Smoking has been one of the most extensively studied human health risks, with detailed epidemiological research dating back to the 1930s (Doll and Hill 1950; Levin et al. 1950; Mills and Porter 1950; Muller 1939; Schairer and Schoniger 1943; Schrek et al. 1950; Wassink 1948; Wynder and Graham 1950) (see Table 1 in Doll 1986 for a summary of early studies). The sample size and details of data and methods have grown in subsequent studies in a number of countries, leading to a growing list of more than 60000 publications on the hazards of smoking (Lopez 1999). The evidence for the causal relationship between smoking and all-cause and cause-specific mortality, as well as the mechanisms of disease causation have been extensively reviewed (Doll 1986, 1998b; U.S. Department of Health and Human Services 1989). Added to these are a number of recent studies, especially from developing countries such as China and India (Doll et al. 1994; Gajalakshmi et al. 2003; Gupta and Mehta 2000; Lam et al. 2001; Liu et al. 1998; Niu et al. 1998). We provide a brief summary of this evidence.

Randomized controlled trials of smoking and health, with randomization of exposure to cigarette smoking, would be impossible for ethical and logistical reasons. Therefore, research on the health impacts of smoking has been based on observational epidemiology and statistical adjustment. The first evidence for the causal relationship between smoking and mortality was from case-control studies. After the initial evidence from case-control studies, numerous prospective cohort studies were initiated to establish the relationship between smoking and mortality. These included studies in the United Kingdom of Great Britain and Northern Ireland (Male British Doctors) (Doll and Peto 1976; Doll et al. 1994), the United States of America (Males in 25 States, Males in 9 States, U.S. Veterans, California Occupations, the American Cancer Society Cancer Prevention Study Phases I and II) (Dunn et al. 1960; Garfinkel 1985; Hammond 1966; Hammond and Horn 1958; Kahn 1966; Peto et al. 1992; Thun et al. 1997a, 1997b), Japan (Hirayama 1977), Canada (Canadian Veterans) (Best et al. 1961), Sweden (Cederlof et al. 1975),
and more recently India (Gupta and Mehta 2000) and a number of other countries. Evidence for the relationship between smoking and total mortality or cause-specific mortality has been consistent among the studies. The causes of mortality from smoking include lung cancer and cancer of various other sites, cardiovascular diseases, COPD and other respiratory diseases, gastric ulcer and a number of other causes.

In addition to establishing the basic causal relationship between smoking and mortality, the results from all these studies exhibit the features and criteria that have been considered important in establishing causality:

1. consistently large risks relative to life-long non-smokers-more than a 20 -fold increase in risk for some diseases;
2. existence of a dose-response relationship that shows increasing risk with increasing number of cigarettes smoked and duration of smoking (Doll 1986; Peto 1986; Thun et al. 1997b; U.S. Department of Health and Human Services 1989);
3. consistency of mortality (in particular for lung cancer) with smoking characteristics and cigarette type (such as tar and nicotine yield) across sexes and populations (Doll 1986);
4. risk reduction with smoking cessation, and a decreasing risk gradient with increasing time since smoking cessation (Doll 1986; U.S. Department of Health and Human Services 1989); and
5. biological plausibility. Cigarette smoke contains a number of known human carcinogens and other toxics (Hecht 2003). The toxic (in particular carcinogenic) properties of cigarette smoke have also been established in studies with laboratory animals (U.S. Department of Health and Human Services 1989). More recently, the pathophysiological mechanisms for some non-cancer effects such as COPD and cardiovascular diseases have also been studied and established (see U.S. Department of Health and Human Services 1989 for a summary).

## 2. Methods and data

### 2.1 Exposure variable

## LUNG CANCER MORTALITY AS AN INDICATOR OF ACCUMULATED

 SMOKING HAZARDPeto et al. (1992) observed that the level of lung cancer mortality compared with never-smokers is an indicator of the accumulated hazard of smoking and the "maturity" of the smoking epidemic in a population. The relationship between cumulative smoking and lung cancer from various populations confirm this relationship (Peto 1986; Yamaguchi
et al. 2000). Based on this observation, the smoking impact ratio (SIR) is defined as population lung cancer mortality in excess of never-smokers, relative to excess lung cancer mortality for a known reference group of smokers. Formally, the ratio in Equation 1 measures the absolute excess lung cancer mortality due to smoking in the study population, relative to the absolute excess lung cancer mortality in life-long smokers of the reference population.

$$
\begin{equation*}
S I R=\frac{C_{L C}-N_{L C}}{S_{L C}^{*}-N_{L C}^{*}} \tag{1}
\end{equation*}
$$

$C_{L C}$ : (age-sex-specific) lung cancer mortality rate in the study population (e.g. country of analysis)
$N_{L C}$ : (age-sex-specific) lung cancer mortality rate of never-smokers in the same population
$S_{\tilde{L C}}^{*}$ and $N_{\tilde{L} C}^{*}$ : (age-sex-specific) lung cancer mortality rates for smokers and never-smokers in a reference population

Liu et al. (1998) found that in China, the relative risk of mortality from lung cancer as a result of smoking is approximately constant in different cities and villages whose non-smoker lung cancer mortality rates varied by a factor of 10 (see Figure 4 in Liu et al. 1998). A constant relative risk means that smoking results in a larger absolute excess mortality (i.e. the numerator of Equation 1) where never-smoker lung cancer mortality is higher (and smaller absolute excess mortality where neversmoker lung cancer mortality is lower). Therefore, to be converted to an indictor of the maturity of the smoking epidemic, the numerator and denominator of Equation 1 need to be normalized with the respective never-smoker lung cancer mortality rates. We defined the backgroundadjusted SIR by the following relationship:

$$
\begin{equation*}
S I R=\frac{C_{L C}-N_{L C}}{S_{L C}^{*}-N_{L C}^{*}} \times \frac{N_{L C}^{*}}{N_{L C}} \tag{2}
\end{equation*}
$$

where $C_{L C}, N_{L C}, S_{\tilde{L} C}^{*}$, and $N_{\tilde{L} C}$ are defined as above.
Following Peto et al. (1992), we used the CPS-II study population (described below) as the reference population. This is because, among the numerous studies of smoking and cause-specific mortality (U.S. Department of Health and Human Services 1989), CPS-II is one of the very few with follow-up conducted when the smoking epidemic was at its highest levels, especially for men. Therefore, the vast majority of (male) CPS-II current-smokers had been lifelong cigarette smokers. Further, the estimates of increased risk of mortality among smokers were available for both men and women and in smaller age groups than in other studies of smoking and mortality, such as the Male British Doctors cohort (Doll et al. 1994).

It is straightforward to show that SIR equals the proportion of refer-ence-population (i.e. CPS-II) smokers in a mix of smokers and neversmokers, which has the same lung cancer mortality rate as the study population (Peto et al. 1992). ${ }^{2}$ This provides a convenient interpretation of SIR: using excess lung cancer mortality over never-smokers, SIR captures the accumulated hazards of smoking by converting the smokers in the study population into equivalents of smokers in the reference population where hazards for other diseases have been measured (Peto et al. 1992).

SIR values were calculated for individual countries and then averaged (population-weighted) across the 14 reporting subregions by age group and sex. The age groups used in the analysis were $0-4,5-14,15-29$, $30-44,45-59,60-69,70-79$, and $\geq 80$, which are the reporting age groups used in the GBD study. No deaths before the age of 30 years were attributed to smoking. Peto et al. (1992) used 35 as the lowest age for considering the impacts of smoking. Because of the GBD age grouping, we also included the 30-34 age group. The number of deaths attributed to smoking among people between 30 and 34 years of age is likely to be very small.

SIR values larger than 1.0 were set to 1.0 . This occurred in the case of males in the 30-44 age group in 17 European countries and one Western Pacific island, and in the 45-59 age group in three countries in eastern Europe. Relatively low lung cancer mortality in younger ages can lead to unstable SIR values. This is particularly the case if the neversmoker rates are estimated with error, which is more likely in younger ages when lung cancer is relatively rare. Further, although a SIR larger than 1.0 may seem to imply that a population which consists of some smokers and some never-smokers had higher lung cancer mortality than CPS-II life-long smokers, factors such as the type and number of cigarettes or the age at which smoking began can result in such a pattern, especially where prevalence of smoking is high. The age of smoking initiation is particularly important for SIR values in earlier ages such as those affected in this analysis. For example, historical lung cancer mortality data show SIR larger than 1.0 among British males aged $<60$ years in some years between 1950 and 1970 and American males between 1968 and 1976. We nonetheless set the SIR for these groups to 1.0 to avoid any potential overestimation of risk.

## EXPOSURE CATEGORIES

$S I R$ is a measure of exposure in the population, vs the individual. In fact, since $S I R$ is based on the lung cancer mortality rate, it cannot be defined at the level of an individual. It is nonetheless possible to divide a population into subgroups with different SIRs subject to the constraint that the overall population SIR should remain unchanged.

Suppose that a population with a specific SIR is divided into three subgroups with $S I R_{1}, S I R_{2}$ and $S I R_{3}$. Then

$$
\begin{equation*}
S I R=a_{1} S_{1 R_{1}}+a_{2} S_{2}+a_{3} S_{2} \tag{3}
\end{equation*}
$$

and

$$
\begin{equation*}
a_{1}+a_{2}+a_{3}=1.0 \tag{4}
\end{equation*}
$$

where $a_{1}, a_{2}$, and $a_{3}$ are the fractions of the population in the three subgroups. The above system of equations is under-determined for any division of more than two subgroups. If one considers never-smokers as one of the subgroups (subgroup 1) then, by definition of the smoking impact ratio, $S I R_{1}=0$ and

$$
\begin{equation*}
S I R=a_{2} S_{2} R_{2}+a_{3} S_{3} \tag{5}
\end{equation*}
$$

If $p$ denotes the fraction of the population who have been smokers at some stage of their lives: $a_{1}=1-p$ and $a_{2}+a_{3}=p$.

After some manipulation, it can be shown that

$$
\begin{equation*}
a_{2}=\frac{S I R-p S I R_{3}}{S I R_{2}-S_{3}} \tag{6a}
\end{equation*}
$$

and

$$
\begin{equation*}
a_{3}=\frac{S I R-p S I R_{2}}{S I R_{3}-S_{2}} \tag{6b}
\end{equation*}
$$

But prevalence data ( $p$ ) are often unavailable or inaccurate, because of both differences in the definition of "ever-smoker" (someone who has smoked at least one cigarette in their life) and data collection complexity. In this work, the additional requirement for age-specific information makes the existing prevalence data-which are often defined for youth and adults only-even less useful. Therefore, in dividing the population into subgroups, the calculation of the fraction of the population in each subgroup is subject to the constraint that the overall $S I R$ remains unchanged.

The specific categories that were used correspond to $S I R_{1}=0$, $0<S I R_{1} \leq 0.5$, and $0.5<S I R_{3} \leq 1.0$. For the last two categories, the midpoints ( 0.25 and 0.75 ) were used in estimating the fraction of population in the two categories ( $a_{2}$ and $a_{3}$ ), once again subject to the constraint that the overall SIR be equal to the original population SIR. The analysis was conducted to divide each age-sex-subregion $S I R$ value into three categories.

## ThEORETICAL-MINIMUM-RISK EXPOSURE DISTRIBUTION

Because the harmful effects of smoking outweigh the potential health benefits for a small number of diseases (such as parkinsonism) by many
orders of magnitude (Doll 1986, 1998a), the dose-response relationships for smoking and overall mortality and burden of disease are monotonically increasing. Therefore, the minimum risk from smoking would occur in a population in which no one had ever smoked. Further, although such a population may not be achievable in practice, there are no physical limits to reduction of smoking. Therefore, the theoretical minimum exposure distribution was chosen as a population with a SIR of zero.

## Population and lung cancer mortality statistics

The age-sex-specific population estimates for the 191 WHO Member States were from the United Nations Population Division, and mortality statistics from WHO's GBD database. GBD mortality statistics are divided by country, sex, age (five-year age groups up to 85 and then $\geq 85$ ), and more than 150 causes of death. The sources of mortality data include vital registration and sample registration, population laboratories, and epidemiological studies. Details of mortality data and cause-of-death analysis methods are described elsewhere (Mathers et al. 2002) and are summarized below where relevant.

The reliability of the $S I R$ is determined by the reliability of lung cancer mortality estimates. In countries with good vital registration and medical certification of deaths (approximately 75 countries), lung cancer mortality is diagnosed with a high degree of accuracy. For example, microscopic confirmation of diagnosis against the cause reported on death certificates has suggested a $95 \%$ or higher confirmation rate in these settings (Percy and Muir 1989; Percy et al. 1990). In approximately another 50 countries, vital registration of mortality is incomplete and medical certification of the cause much less reliable. Standard demographic techniques (Bennett and Horiuchi 1984; Hill 1981; Preston et al. 2000) are used in the GBD project to correct all-cause death rates by age for these populations, and lung cancer rates are adjusted accordingly. Finally, for countries without vital registration, overall age-specific death rates were first determined using model life-tables (Lopez et al. 2002). Total cancer death rates are then estimated based on regional information about proportionate cancer mortality. Within this death rate, the distribution by site is based on regional incidence patterns from cancer registries reporting to the International Agency for Research on Cancer, IARC (Parkin et al. 1992, 1997). This indirect procedure is likely to entail considerable uncertainty, as we describe below.

### 2.2 RISK FACTOR-DISEASE RELATIONSHIPS

American Cancer Society Cancer Prevention Study (ACS CPS-II)

The American Cancer Society's Cancer Prevention Study, phase II (CPSII) is a prospective study of smoking and death in more than one million Americans aged $>30$ years when they completed a questionnaire in 1982,
with the latest follow-up in 1998. A complete description of the study is provided elsewhere and summarized below (Garfinkel 1985; Peto et al. 1992; Thun et al. 1995, 1997a, 1997b, 2000). In 1992, when the first six-year (1982-1988) results were obtained, mortality follow-up was virtually complete for the first two years, and about 98-99\% complete for the next four. Because some conditions that cause death in the first two years may have affected smoking habits at entry (e.g. those diagnosed with lung cancer may have stopped smoking because of their disease or related symptoms), analysis was restricted to years 3-6 inclusive (1984-1988) (Peto et al. 1992). The analysis related deaths (subdivided by cause, sex and five-year age groups at the time of death) to personyears (with accounting for incompleteness) for those who in 1982 had never smoked regularly, and for those who were then current cigarette smokers (Peto et al. 1992). Most of the CPS-II current-smokers were lifelong cigarette smokers with a mean consumption of about 20 cigarettes per day. Relative risks for cause-specific mortality among smokers in the CPS-II population are provided in Table 11.1.

## Retrospective proportional mortality study in China

In a retrospective study of one million deaths in 24 urban centres and 74 rural areas of China, Liu et al. (1998) used proportional mortality analysis to obtain relative risks for three groups of diseases (neoplastic, respiratory and cardiovascular diseases). As explained in Liu et al. (1998), proportional mortality analysis cannot estimate excess mortality for these causes of death in the reference group. Therefore, in this study (Liu et al. 1998) the attributable fraction for causes other than the above three groups was considered to be zero. At the same time, since a few of the deaths in the reference group were also due to smoking, this method would underestimate (as zero) the proportion of mortality due to the causes in the reference group. The relative risks and attributable fractions of cause-specific mortality from Liu et al. (1998) are summarized in Table 11.2.

## DISEASE OUTCOMES AND HAZARD SIZE

Lung cancer mortality attributable to smoking was obtained as the difference between population lung cancer mortality and that of neversmokers (i.e. the numerator of Equation 1). For all other diseases, the relative risks from CPS-II (Table 11.1) were used. To obtain mortality from other causes, a "mixture" of CPS-II smokers and non-smokers was taken to give a SIR equal to that of the study population (as described above, the proportion of smokers in this mixture equals the SIR of the study population). This mixture was then used together with the causespecific relative risks from CPS-II (Table 11.1) to estimate population attributable fractions (PAF) in the study population for different diseases.

Table II.I Selected relative risks for cause-specific mortality for years 3-6 inclusive of ACS CPS-II prospective study of one million American adults

| Cause (ICD-9 code) | Male | Female |
| :--- | ---: | ---: |
| Lung cancer (162) | 24.22 | 12.50 |
| Upper aerodigestive cancer (mouth, oropharynx or oesophagus) <br> (I40-150 and 16I) | 7.87 | 6.95 |
| Other cancer (rest of 140-209) | 1.69 | 1.20 |
| Chronic obstructive pulmonary disease (490-492, 496) | 13.82 | 14.21 |
| Other respiratory diseases (460-466, 480-487, 38I-382) |  |  |
| $35-59$ | 3.05 | 2.69 |
| $60-64$ | 2.31 | 2.68 |
| $65-69$ | 2.09 | 2.52 |
| $70-74$ | 2.00 | 2.00 |
| $\geq 75$ | 1.54 | 1.44 |

Infectious and parasitic diseases (00I-I39, 320-323, 6|4-6|6 with the exception of those above), maternal and perinatal conditions (630-676, 760-779), neuro-psychiatric conditions (290-319, 324-359), cirrhosis of the liver (57I), congenital anomalies (740-759), and non-medical causes (injuries) (E800-999) ${ }^{\text {c }}$

| Other medical causes (rest of $000-799$ ) |  |  |
| :--- | :--- | :--- |
| $35-59$ | 3.05 | 2.69 |
| $60-64$ | 2.31 | 2.68 |
| $65-69$ | 2.09 | 2.52 |
| $70-74$ | 2.00 | 2.00 |
| $\geq 75$ | 1.54 | 1.44 |

NA Not applicable.
a For lung cancer, the population attributable fraction was obtained by direct subtraction of
population lung cancer mortality from that of non-smokers.
This category also includes tuberculosis ( $010-018,137$ ) for which there are separate estimates of
relative risk in China.

The use of relative risk models in estimating tobacco-attributable mortality has been criticized in the past (Lee 1996). While both additive and multiplicative risk models have been used in epidemiology (Moolgavkar and Venzon 1987), the use of a relative risk approach for common risk factors for common diseases is standard in epidemiological literature based on its ability to capture the "risk magnification" role of most risk factors. In the particular case of smoking, Liu et al. (1998) found that in China, the relative risks for mortality from lung cancer and other major diseases are approximately constant in different cities and

Table II.2 Selected relative risks and attributable fractions of mortality from the retrospective mortality analysis of one million deaths in China

| Cause (ICD-9 code) | Ages 35-69 years |  | Ages $\geq 70$ years |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Weighted mean relative risks (SE) | \% deaths attributable to smoking | Weighted mean relative risks (SE) | \% deaths attributable to smoking |
| Male |  |  |  |  |
| Malignant neoplasm (140-208) | 1.51 (0.02) | 24.4 | 1.39 (0.03) | 18.7 |
| Lung cancer (162) | 2.72 (0.05) | 52.3 | 2.47 (0.07) | 46.6 |
| Oesophageal cancer (150) | 1.61 (0.04) | 27.9 | - | 18.2 |
| Stomach cancer (151) | 1.35 (0.03) | 18.1 | - | 9.1 |
| Liver cancer (155) | 1.40 (0.03) | 20.2 | - | 14.7 |
| Mouth, pharynx, larynx, pancreas or bladder (140-9, I6I, I57, I88) | 1.51 (0.05) | 24.6 | - | 19.1 |
| Other malignant neoplasm | 1.24 (0.03) | 13.1 | - | - |
| Respiratory | 1.31 (0.02) | 17.2 | 1.54 (0.02) | 24.6 |
| Chronic obstructive pulmonary disease (ICD 490-492, 496, 416-4I7) | 1.43 (0.03) | 22.6 | 1.63 (0.03) | 27.4 |
| Respiratory tuberculosis (0ІІ, 012, 018) | 1.20 (0.04) | 11.3 | - | 24.5 |
| Other respiratory (rest of 460-519) | 1.07 (0.05) | 4.2 | - | - |
| Cardiovascular (390-415, 4I8-459) | 1.15 (0.02) | 8.5 | 1.06 (0.02) | 3.4 |
| Stroke (430-439) | 1.17 (0.02) | 10.0 | - | 4.2 |
| Ischaemic heart disease (4I0-4I4) | 1.28 (0.03) | 14.7 | - | 2.8 |
| Other cardiovascular diseases <br> (all cardiovascular except 430-439, $4\|0-4\| 4,4\|6-4\| 7)$ | 0.94 (0.03) | NA | - | - |
| Other causes (reference group) | 1.00 | NA | 1.00 | NA |

Female

| Malignant Neoplasm (140-208) | 1.37 (0.04) | 4.0 | 1.37 (0.03) | 4.7 |
| :---: | :---: | :---: | :---: | :---: |
| Lung cancer (162) | 2.64 (0.08) | 19.4 | 2.50 (0.09) | 20.1 |
| Oesophageal cancer (150) | 1.34 (0.08) | 2.8 | - | 4.3 |
| Stomach cancer (151) | 1.17 (0.06) | 1.7 | - | 2.3 |
| Liver cancer (I55) | 1.22 (0.06) | 2.4 | - | 4.2 |
| Mouth, pharynx, larynx, pancreas, or bladder (140-9, 161, 157, 188) | 1.53 (0.09) | 6.4 | - | 9.1 |
| Other malignant neoplasm | 1.04 (0.04) | 0.5 | - | - |
| Respiratory | 1.61 (0.05) | 7.5 | 1.59 (0.03) | 7.4 |
| Chronic obstructive pulmonary disease (ICD 490-492, 496, 416-4I7) | 1.72 (0.05) | 9.3 | 1.70 (0.03) | 8.6 |
| Respiratory tuberculosis (011, 012, 018) | 1.29 (0.08) | 2.8 | - | 6.7 |
| Other respiratory (rest of 460-519) | 1.14 (0.09) | 1.5 | - | - |

Table II.2 Selected relative risks and attributable fractions of mortality from the retrospective mortality analysis of one million deaths in China (continued)

| Cause (ICD-9 code) | Ages 35-69 years |  | Ages $\geq 70$ years |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Weighted mean relative risks (SE) | \% deaths attributable to smoking | Weighted mean relative risks (SE) | \% deaths attributable to smoking |
| Cardiovascular (390-4I5, 4I8-459) | 1.01 (0.03) | 0.2 | 1.02 (0.02) | 0.2 |
| Stroke (430-439) | 0.97 (0.03) | NA | - | - |
| Ischaemic heart disease (4I0-4I4) | 1.30 (0.05) | 4.1 | - | 2.0 |
| Other cardiovascular diseases <br> (all cardiovascular except 439-439, $4\|0-4\| 4,4\|6-4\| 7)$ | 0.94 (0.05) | NA | - | - |
| Other causes (reference group) | 1.00 | NA | 1.00 | NA |

- Values not reported in Liu et al. (1998).

NA Not applicable.
Source: Liu et al. (1998), Tables I and 5.
villages where background (non-smoker) mortality rates for the same disease varied significantly. This finding also has been confirmed in studies which stratified on serum cholesterol for cardiovascular diseases (Jee et al. 1999).

The exception to the use of CPS-II relative risks was tobaccoattributable mortality in China, for which direct estimates of the fraction of cause-specific mortality due to smoking from 1990 were available. For China, attributable fractions for diseases other than lung cancer were obtained using relative risks from Liu et al. (1998). Since smoking has been increasing in China over the past few decades, we used SIR estimates to capture the impact of this trend. The relative risks for each smoker from Liu et al. (1998) were converted to relative risks for each unit of SIR (i.e. equivalent to life-long CPS-II smoker) by backcalculation in 1990 and used with 2000 SIR estimates. The upper aerodigestive cancer category was constructed from oesophageal cancer and cancers of five minor sites (mouth, pharynx, larynx, pancreas and bladder) by using weights based on the number of deaths in each disease category. Relative risks per unit of SIR for China (which correspond to the same level of accumulated smoking hazard as Table 11.1 for CPS-II life-long smokers), are provided in Table 11.3.

Although smokers are found to have increased mortality due to causes other than those in Table 11.1, none of the deaths from non-medical causes (injuries) and the following medical causes (for which there are no known pathways of causality) were attributed to smoking: infectious and parasitic diseases (with the exception of those in the category "other respiratory disease" for which there were estimates of hazard), maternal

Table II. 3 Relative risks for each unit of SIR ${ }^{a}$ for China obtained by back-calculation from 1990 estimates of attributable mortality from Table II. 2

|  | Male |  |  | Female $^{\mathrm{b}}$ |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Cause | $35-69$ years | $\geq 70$ years |  | $35-69$ years | $\geq 70$ years |
| Lung cancer | 20.97 | 35.76 |  | 14.19 | 15.06 |
| Upper aerodigestive cancer | 7.71 | 9.90 |  | 2.78 | 3.89 |
| Other cancer | 4.39 | 5.12 |  | 1.56 | 2.30 |
| Chronic obstructive pulmonary disease | 6.32 | 16.03 |  | 6.62 | 6.26 |
| Tuberculosis | 3.32 | 7.32 |  | 2.58 | 2.49 |
| Other respiratory diseases | 1.80 | 3.14 |  | 1.83 | 1.79 |
| Cardiovascular diseases | 2.69 | 2.40 |  | 1.11 | 1.11 |
| Other causes (reference category) |  | 1.00 | 1.00 |  | 1.00 |

${ }^{\text {a }}$ Equivalent to a CPS-II life-long smoker in Table II.I.
b Relative risks for females are likely to be unstable due to low prevalence of female smoking in China.
c Since the proportional mortality method used by Liu et al. (I998) does not allow obtaining attributable mortality for causes in the reference group, the relative risks were estimated as one. At the same time, since some deaths in this category are also due to smoking the relative risks from CPS-II were used for these causes.
and perinatal conditions (see below for discussion), neuro-psychiatric conditions, cirrhosis of the liver and congenital anomalies.

There is increasing evidence of an association between smoking and tuberculosis (Dhillon et al. 2000; Gajalakshmi et al. 2003; Lam et al. 2001; Liu et al. 1998). The role of air pollutants in increased risk of infectious pulmonary disease has also been suspected to be through weakening of lung function and defence mechanisms (Thomas and Zelikoff 1999). At the same time tuberculosis is a highly communicable disease whose transmission dynamics are affected by a number of factors. Further, progress from infection to disease and mortality is highly dependent on control and treatment mechanisms. Because of this crucial role of cofactors in tuberculosis incidence and mortality, and their concentration in specific sectors of society, we did not extrapolate relative risks for this disease from one setting to another. We considered tuberculosis as a separate cause of mortality due to smoking only in China where published direct estimates on the relationship were available. We grouped tuberculosis with the category "other respiratory diseases", which have a lower relative risk, in all other countries. Not including tuberculosis as a separate cause of death for the rest of the world is a conservative assumption. This is particularly true in the case of India where recent evidence indicates risks larger than those in China (Dhillon et al. 2000; Gajalakshmi et al. 2003).

Many deaths from burns and other injuries due to fires are also attributable to smoking. For example, pooled studies from Australia, the

United Kingdom and the United States show an attributable fraction of 0.23 for fire-related injuries due to smoking (English et al. 1995). The number of fire-related deaths and injuries are highly dependent on local circumstances including population density, housing, the availability of emergency services, etc. A relative risk approach would, therefore, be unsuitable for their analysis, which is better conducted using injury registries. Given the large burden of injuries due to fires, excluding nonmedical causes undoubtedly results in an underestimation of the health impacts of tobacco.

There has also been growing evidence of the impact of maternal smoking on maternal and child health. Specific outcomes include still births and neonatal deaths, perinatal mortality, low birth weight, primary and secondary infertility, ectopic pregnancy and spontaneous abortion and a number of other child and maternal conditions (U.S. Department of Health and Human Services 1989, 2001). Despite the evidence for the relationship, there has been little quantification of the impact of maternal exposure on child disease and mortality, especially in settings with high background levels of child and maternal mortality. Further, these estimates would require estimating what fraction of female smokers continues to smoke during pregnancy. Not including quantitative estimates of the health hazards for maternal and child health also underestimates the burden of disease due to smoking.

Before using the relative risks from CPS-II, we reduced the excess risk attributed to smoking using constant correction factors as described below. As explained by Peto et al. (1992), a constant correction factor, although arbitrary, avoids overestimating mortality due to confounding in the ACS CPS-II relative risk estimates (which were adjusted for age and sex only) as well as extrapolation of relative risk values from this population to other populations, where exposure to other risk factors that could modify the effects of smoking in a non-multiplicative way may be different. The issue of the confounding of the CPS-II risk estimates used by Peto et al. (1992) and the arbitrary nature of the correction factor were used to challenge the indirect estimates of mortality due to this risk factor (Lee 1996; Sterling et al. 1993).

In studies other than CPS-II, the overall impact of confounding due to diet on the estimates of excess risk of mortality from smoking has been found to be considerably less than half of the excess risk for cardiovascular diseases (such as those studies reviewed as a part of the metaanalysis of ETS and ischaemic heart disease [IHD] by Law et al. 1997 or in analysis of multiple cardiovascular disease risk factors in Japan Hirayama 1990). Alcohol consumption and smoking are correlated in many populations, including in the United States where the ACS CPS-II, which was used to obtain relative risks, was conducted. Alcohol may affect cardiovascular disease (in particular IHD) in both beneficial and harmful ways depending on the patterns of consumption (Britton and McKee 2000; Puddey et al. 1999; chapter 12 in this book). Alcohol
was associated with a reduction in cardiovascular disease mortality risk in CPS-II (Thun et al. 1997c). But the overall death rates were lowest among men and women reporting about one drink daily and increased with heavier drinking, particularly among adults aged $<60$ years with lower risk of cardiovascular disease (Thun et al. 1997c). It may also have been the case that CPS-II smokers had more harmful drinking patterns, which would confound the relative risk estimates for cardiovascular disease. At the same time, given that the beneficial impacts of alcohol in this cohort persisted even at higher consumption levels, any possible confounding effect on estimates of smoking hazard size is likely to be considerably less than one half of the excess risk. In the Nurses' Health Study, Kawachi et al. (1997) found that the relative risk for cardiovascular diseases as a result of smoking after adjustment for multiple covariates (3.74) was larger than the unadjusted relative risk (3.47); the change was not statistically significant. The reduction in cardiovascular disease excess risk due to adjustment for age, community, history of hypertension and diabetes, consumption of alcohol, systolic blood pressure, and frequency of walking in a study of the relationship between smoking and mortality in three United States communities was $20 \%$ for men (2.0 to 1.8). Adjustment resulted in an increase in female relative risk (from 1.7 to 1.8) (LaCroix et al. 1991); neither change was statistically significant.

Among cancers, upper aerodigestive cancers are the diseases for which confounding due to alcohol consumption may be most important. There is evidence from studies in different parts of the world on the interaction (excess over multiplicative) between alcohol and tobacco for various upper aerodigestive cancers (Flanders and Rothman 1982; Kinjo et al. 1998). As described by Flanders and Rothman (1982), this interactive effect as well as the correlation between exposure to these two risk factors, would imply larger risks from smoking than would be the case in the absence of interaction and correlation. Therefore, reduction of risk by one half is likely to not only account for potential confounding but also result in conservative estimates.

In response to the criticism about the lack of empirical evidence for confounding correction, CPS-II data were re-analysed together with adjustment for potential confounders (Malarcher et al. 2000; Thun et al. 2000). Malarcher et al. (2000) found that adjusting for multiple covariates (age, education, alcohol use, hypertension status and diabetes status) only marginally changed the fraction of mortality attributable to smoking in the United States, as shown in Table 11.4. As seen in the table, except for cerebrovascular disease (CVD) among men, adjustment for confounding had no or little effect, or even resulted in a slight increase, on the smoking-attributable mortality.

In a more detailed re-analysis of CPS-II data, Thun et al. (2000) adjusted for age, race, education, marital status, occupation ("blue collar" worker) and total weekly consumption of citrus fruits and vegetables in estimat-

Table II.4 Effect of adjustment for confounding on the fraction of cause-specific mortality attributable to smoking

|  | Male |  |  | Female |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
|  | Age-adjusted | Fully-adjusted |  | Age-adjusted | Fully-adjusted |
| Lung cancer | $0.91(0.89-0.92)$ | $0.89(0.87-0.91)$ |  | $0.71(0.68-0.74)$ | $0.77(0.71-0.80)$ |
| COPD | $0.85(0.82-0.88)$ | $0.88(0.83-0.92)$ |  | $0.70(0.65-0.74)$ | $0.68(0.61-0.76)$ |
| IHD | $0.23(0.20-0.26)$ | $0.24(0.21-0.27)$ |  | $0.08(0.06-0.10)$ | $0.10(0.06-0.13)$ |
| CVD | $0.16(0.09-0.22)$ | $0.10(0.02-0.19)$ |  | $0.10(0.07-0.12)$ | $0.09(0.05-0.12)$ |

Source: Malarcher et al. (2000).
ing the relative risk of mortality due to a range of neoplasms, cardiovascular diseases and respiratory diseases. In addition, the analysis for cardiovascular diseases adjusted for current aspirin use or alcohol consumption, body mass index (BMI), physical activity at work or leisure and weekly consumption of fatty foods. For lung cancer and COPD, the analysis also adjusted for occupational exposure to asbestos. Similar to the analysis of Malarcher et al. (2000), with the exception of stroke among men, whose relative risk declined from 2.9 ( $95 \%$ CI 2.3-3.7) to $2.4(95 \%$ CI 1.8-3.0) for the 35-64-year age group and from $1.8(95 \%$ CI 1.6-2.2) to 1.5 ( $95 \%$ CI 1.2-1.8) for those aged $>64$ years, excess risks increased, stayed unchanged or decreased by small amounts. As in the case of Malarcher et al. (2000), the largest decrease was in the case of stroke for males ( $17 \%$ decrease in excess risk for the 35-64-year age group and $38 \%$ for those aged $\geq 65$ years). Overall, Thun et al. (2000) concluded that adjustment for confounding reduced their estimates of mortality attributable to smoking in the United States by approximately $1 \%$.

Finally, the Chinese retrospective study provides some evidence that background disease rates do not change the proportion of mortality attributable to smoking (Figures 4 and 5 in Liu et al. 1998). In extrapolating hazard size from industrialized to developing countries, one would have to deal with availability of health services (such as medical care for cardiovascular disease patients) if extrapolation were based on absolute risk (i.e. number of deaths) because the number of fatal cases would vary based on available treatment. If extrapolation is made based on risk of mortality for smokers relative to non-smokers, however, availability of health services would not bias the estimates if smokers and non-smokers have similar access to these services.

Based on this evidence of the robustness of CPS-II relative risk estimates to adjustment for confounding and extrapolation across settings with different background mortality rates, we modified the choice of correction factor from that used by Peto et al. (1992). We used a correction factor of $30 \%$, approximately equal to the largest reduction in excess risk due to confounding seen in the re-analysis of CPS-II data, to reduce
the excess risk for all cause-specific relative risks (see also discussion of uncertainty). This choice continues to be conservative to account for any potential residual confounding or any other potential sources of overestimation arising from extrapolation across regions. For the category, "other medical causes", where the extent of confounding is still not known, we continued to attribute only one half of the excess mortality estimated by CPS-II.

The excess risk for China was reduced by $5 \%$. Confounding due to other risk factors is likely to have been negligible in China because of the more homogenous level of exposure to other risk factors (e.g. diet, alcohol, and indoor air pollution from coal) in the population compared to the CPS-II population (R. Peto, personal communication, 2001). The proportional mortality method used by Liu et al. (1998) also is not affected by confounding due to any risk factor that increases mortality in the study and reference disease categories (such as effects of alcohol on cancer and liver cirrhosis). Finally, since the relative risks are used for China only, effect modification due to other risk factors is not a concern. The $5 \%$ reduction of excess risk was used to account for the potential small level of residual confounding.

## ASSESSMENT OF MORBIDITY

Virtually all causes of mortality (diseases) that are affected by smoking are chronic diseases due to long-term exposure. Therefore, a reasonable assumption would be that smoking increases mortality by increasing disease incidence, rather than modifying case fatality among those who would already be affected by the disease. Since the non-fatal component of disability-adjusted life years (DALYs) (i.e. years lived with disability or YLD) is based on incidence, this assumption would imply that the relative risks for mortality and incidence are equal. Among the causes of mortality considered, the most obvious exception to this assumption may be asthma and respiratory infections, where smoking may affect incidence, case fatality or both. For these diseases we assumed that one half of the morbidity obtained from the above approach is due to smoking. To provide conservative estimates, we also considered that for those diseases where medical interventions may reduce case fatality (such as some cardiovascular diseases) and those for which smokers have a smaller excess risk compared to non-smokers (such as the "other medical conditions" category), the increased risk of mortality may be partially due to increased case fatality. Therefore, for these diseases also, we assumed that one half of the above attributable fraction was due to smoking. In summary, the attributable fraction of morbidity due to smoking was assumed to be the same as mortality for all cancers and COPD, and one half of the latter for all other causes.

### 2.3 Choice of never-smoker lung cancer mortality in the STUDY POPULATION

In Equation 2, $C_{L C}$ is from the GBD mortality database and $S_{\tilde{L C}}^{*}$ and $N_{\tilde{L} C}^{*}$ directly from CPS-II. The only parameter to be estimated indirectly is $N_{L C}$, since direct estimates of never-smoker lung cancer mortality are known for very few countries. Figure 11.1 shows non-smoker lung cancer mortality for the United States (from CPS-II) and China (from Liu et al. 1998). The detailed data in Liu et al. (1998) further allow dividing the Chinese rates into urban and rural areas, and the latter into coastal and inland rural.

As seen in Figure 11.1, age-specific non-smoker lung cancer mortality is considerably higher in China than in the United States for both males and females. In China itself, there are also marked differences between different parts of the country with urban areas having the highest non-smoker lung cancer mortality rates and inland rural areas the lowest (also seen in Figure 4 in Liu et al. 1998). The difference between the non-smoker lung cancer mortality rates for the United States and China, and between urban and rural regions of China is explained by the Chinese patterns of household energy use over the past few decades. Coal is a common household fuel in China, often burned in stoves and buildings without adequate ventilation (Du et al. 1996; Liu et al. 1998; Smith et al. 1993; Wang et al. 1996). Exposure to coal smoke and cooking fumes has been associated with increased lung cancer incidence in China (Du et al. 1996; He et al. 1991; Wang et al. 1996). In inland rural regions of China, where incomes are the lowest, biomass (including crop residues and wood) has been the dominant household fuel compared with coastal villages and cities where coal has been more commonly used.

The relationship between biomass fuels and lung cancer has been absent or considerably smaller than that between coal and lung cancer (Bruce et al. 2000; Ko et al. 1997; Smith and Liu 1993; Sobue 1990) as seen also in chapter 18. Although urban air pollution has been linked to increased lung cancer mortality in some studies, the size of the risk is considerably smaller than the effects of smoking or direct exposure to coal smoke (Doll 1978; Jedrychowski et al. 1990; Nyberg et al. 2000; Pope et al. 2002; Vena 1982). Further, the impact of ambient air pollution on lung cancer has been found to be smaller among non-smokers than among smokers, with one study finding increased risk of lung cancer as a result of urban air pollution among smokers only (Jedrychowski et al. 1990; Vena 1982). For example, exposure to high levels of urban air pollution, even in regions where coal was used extensively, was found to increase the risk of mortality from lung cancer by approximately $14 \%$ among non-smokers (vs $40 \%$ among smokers) (Jedrychowski et al. 1990). Coupled with the fact that only small fractions of national populations live in the most polluted urban areas,

Figure II.I Non-smoker mortality from lung cancer in different populations


Note: Chinese rates correspond to 1990. Note that the scales on the female and male charts are different to maintain the resolution for females.

Source: Ezzati and Lopez (2003).
overall population lung cancer mortality is not expected to be greatly affected by urban air pollution. For example, the lung cancer mortality rate among non-smoking American women remained constant at approximately 12 per 100000 between 1960 and 1986 despite changes in exposure to urban air pollution (U.S. Department of Health and Human Services 1989).

Based on the pattern of background lung cancer mortality rates, and the underlying risk factors for increased lung cancer mortality in China and its various regions, background (never-smoker) lung cancer rates for the different subregions were based on the estimated use of coal for domestic energy in unvented stores as provided in chapter 18. We used Chinese non-smoker rates for China, a weighted average of Chinese and CPS-II non-smoker for SEAR-D where coal is also used for household fuel (with weights for Chinese rates equal to the prevalence of coal use), and CPS-II non-smoker rates for the remaining countries of the world where domestic coal use in unvented stoves is negligible. The remaining risk factors with potential effects on lung cancer mortality (ambient air pollution, occupational hazards, indoor air pollution from radon or biomass smoke, etc.) affect all populations in varying degrees. The net impacts of these other risk factors are considered as sources of uncertainty in extrapolating never-smoker lung cancer mortality from one setting to another.

### 2.4 Risk reversibility

Estimating the burden of disease that might be avoidable as a result of exposure removal requires knowledge of risk reversibility (the decline in risk after exposure is removed) for those who are current smokers. ${ }^{3}$ For those who never begin to smoke, of course, the entire disease burden as a result of smoking is avoided. The benefits of smoking cessation for reduction of mortality risk have been reported in a number of studies (Best et al. 1961; Doll and Hill 1956; Doll and Peto 1976; Doll et al. 1994; Dunn et al. 1960; Hammond 1966; Kahn 1966; U.S. Department of Health and Human Services 1990c). Many studies have also considered the reduction in risk with time since cessation. Most of these studies have focused on the reduction in the risk of lung cancer and cardiovascular causes since cessation. Although the estimates of risk reduction time have varied among different studies, possibly due to different smoking histories and other differences (such as age at which cessation occurred) among the study populations, the benefits of cessation have been consistently demonstrated.

To capture the history and accumulated hazards of smoking, we continued to use smoking impact ratio (SIR) as the exposure variable in considering risk reversibility. We first estimated risk reversibility for lung cancer to obtain estimates of reduction in SIR values after cessation. For diseases other than lung cancer, we estimated the reduction in disease risk per unit of SIR-our exposure variable-using the studies on risk
reversibility. Together with the SIR estimates after cessation, these provided estimates of avoidable mortality for causes other than lung cancer, which also took into consideration the full history of smoking before cessation.

## LUNG CANCER

The reduction in lung cancer risk with smoking cessation has been demonstrated in a number of studies (Alderson et al. 1985; Higgins et al. 1988; Kahn 1966; Lubin et al. 1984; Pathak et al. 1986; Peto and Doll 1984; Peto et al. 2000; Sobue et al. 1991; U.S. Department of Health and Human Services 1990a). Peto et al. (2000) provided an analysis of risk reduction for those who would have been life-long smokers but stopped smoking at various ages using two case-control studies for lung cancer, and national lung cancer mortality statistics in 1950 and 1990 (see Figure 3 in Peto et al. 2000). Table 11.5 shows the reduction in risk, relative to those who continue to smoke, as a function of time since cessation. The results are presented for all ages to increase statistical stability (S. Darby, personal communication, 2001). The large difference between male and female estimates of reversibility in this study is because female estimates were based on relatively few cases of exsmokers with lung cancer (S. Darby, personal communication, 2001). Also, because the smoking epidemic for females lagged the male epidemic in the United Kingdom, as elsewhere, the comparison of 1950 and 1990 lung cancer mortality rates was likely to have included few females who had stopped smoking after the peak of the epidemic.

These estimates of decline in risk are consistent with those from a number of prospective cohort studies (summarized in Table 3 in U.S. Department of Health and Human Services 1990a) including Male British Doctors, United States Veterans, and American Cancer Society CPS-I and CPS-II. When all subjects are considered, these prospective

Table II. 5 Decline in the risk of lung cancer mortality with time since smoking cessation

| Smoking status | Male | Female |
| :--- | :---: | :---: |
| Current smoker | 1.00 | 1.00 |
| Years since cessation among former smokers |  |  |
| $0-9$ | 0.66 | 0.69 |
| $10-19$ | 0.42 | 0.21 |
| $20-29$ | 0.18 | 0.05 |
| $\geq 30$ | 0.08 | - |
| - |  |  |
| Source: Peto et al. (2000). |  |  |

studies show a consistent increase in relative risk in the first five years from the date of cessation (U.S. Department of Health and Human Services 1990a). This can be attributed to selection bias, as those who die in the first few years after cessation are likely to have stopped smoking because they had been diagnosed with a disease. This is confirmed in the separate analysis of those with and without history of chronic disease in CPS-II. In this analysis, those without a history of chronic disease have a consistent declining trend in lung cancer mortality risk compared to all respondents (Figure 11.2). To estimate lung cancer risk reversibility with cessation, we used a re-analysis of CPS-II data (1998 follow-up) excluding those subjects with a history of chronic disease (lung cancer, COPD, cardiovascular diseases) (Figure 11.2).

Despite some remaining differences, the male and female reversibility estimates in Figure 11.2 are much more similar than those in Table 11.5. With better estimates of reversibility among females, we took sexspecific estimates of reversibility to be consistent with the sex-specific smoker and non-smoker lung cancer mortality rates used in the definition of SIR. For the last two cessation periods ( $30-34$ and $\geq 35$ years)

Figure II. 2 Relative risk of lung cancer among former smokers (compared with lifelong non-smokers) with time since cessation


[^40]female estimates show a slight increase in risk compared to the 25-29 year cessation period. This anomaly, which is not statistically significant, is probably due to a small number of events in these groups, itself caused by the more recent maturity of the female smoking epidemic in the United States compared to men. Too few women would have stopped smoking so long ago to have had more than 35 years of cessation and a large number of lung cancer events. For these two cessation periods, we simply assumed that the declining trend observed between the 20-24 and 25-29 year periods would continue.

Avoidable lung cancer mortality can be estimated from the difference between lung cancer risk as it would have been without cessation and the risk that results from cessation. If, as above, we denote smoker and non-smoker lung cancer mortality rates in the reference cohort as $S_{\tilde{L} C}^{*}$ and $N_{\tilde{L} \text { c }}^{*}$, and lung cancer mortality rates for those who stopped smoking $x$ years ago as $S_{\bar{L} C, x}^{*}$, then for this cohort of former smokers $S_{\bar{L} C, x}^{*}-N_{\tilde{L} C}^{*}$ is the mortality attributable to past smoking, and $S_{\tilde{L} C}-S_{\tilde{L} C, x}$ is the mortality avoidable because they stopped smoking $x$ years ago. If $R R$ and $R R_{x}$ are the relative risks for life-long smokers without and with cessation respectively, then $S_{\tilde{L} C}^{*}=R R \times N_{\tilde{L} C}^{*}$ and $S_{\tilde{L} C, x}^{*}=R R_{x} \times N_{\tilde{L} C}^{*}$. Attributable and avoidable mortality $x$ years after cessation is then $N_{\tilde{L} C}^{*} \times\left(R R_{x}-1\right)$ and $N_{\tilde{L} C}^{*} \times\left(R R-R R_{x}\right)$. Of course, for those who never begin to smoke, $S_{\overline{L C}, x}^{*}=N_{\tilde{L} C}^{*}$ and no mortality is attributed to this risk factor and $S_{\tilde{L} C, x}-N_{\tilde{L} C}^{*}$ (i.e. the whole excess mortality due to smoking) is avoidable. In Table 11.5, for example, for males who stop smoking, $34 \%, 58 \%$ and $82 \%$ of lung cancer mortality would be avoided within 10 , 20 and 30 years from cessation, respectively, compared to what would have happened had they continued to smoke. The remaining $66 \%$, $42 \%$ and $18 \%$ of mortality in these time periods is still attributable to their past smoking, despite cessation.

So far, by using CPS-II estimates of risk reversibility, we have implicitly addressed risk reversibility among those who have smoked since youth and stop smoking at a given age (because ACS smokers were lifelong smokers, including at cessation time). We argued earlier, however, that such information is rare and unreliable, and that the only measure of accumulated history of smoking is the SIR. We also showed earlier that for each population, SIR is the equivalent prevalence of life-long smokers. Therefore, to obtain estimates of avoidable burden in each cohort, a fraction of whom have smoked for some period, the estimates of attributable and avoidable mortality $x$ years after cessation are given by $\operatorname{SIR} \times\left(S_{\tilde{L} C, x}-N_{\tilde{L} C}^{*}\right)=\operatorname{SIR} \times N_{\tilde{L} C} \times\left(R R_{x}-1\right)$ and $\operatorname{SIR} \times\left(S_{\tilde{L} C}^{*}-S_{\tilde{L} C, x}\right)$ $=\operatorname{SIR} \times N_{\tilde{L} C}^{*} \times\left(R R-R R_{x}\right)$ respectively where, SIR is measured at the time of cessation (say in 2000 for estimates of avoidable burden as a result of cessation in 2000). We emphasize that this estimation, like all other uses of $S I R$, is based on the assumption that the equivalence of SIR with life-long smokers holds in estimating the benefits of cessation.

Finally, to apply this information to populations with different background (non-smoker) mortality rates, we used the results of Liu et al. (1998) who found that in China the relative risk for mortality from lung cancer was approximately constant in different cities where the nonsmoker lung cancer mortality varies by a factor of 10 . Therefore for populations whose background mortality is $N_{L C}$, the attributable burden due to past smoking $x$ years after cessation is given by $\operatorname{SIR} \times N_{L C} \times$ $\left(R R_{x}-1\right)$ and the avoidable burden as a result of cessation by $S I R \times N_{L C}$ $\times\left(R R-R R_{x}\right)$.

## SIR AFTER CESSATION

To estimate the decline in risk for other diseases, while accounting for the pre-cessation history of smoking, we needed to also estimate the value of $S I R$ after cessation. Once again noting that $S I R$ is the equivalent of life-long smokers in the population, the lung cancer mortality rate in the population as a whole, $C_{L C}$, can be re-written as:

$$
C_{L C}=S I R \times R R \times N_{L C}+(1-S I R) \times N_{L C}
$$

If the same number of life-long smokers continue to smoke, $S I R$ will remain more or less constant for the cohort and in each time period their new SIR will be given by the same relationship, with $R R$ referring to the age-specific relative risk. On the other hand, if the smokers stop smoking, after $x$ years their relative risk will decrease to $R R_{x}$ a value less than $R R$ at any time after cessation (Table 11.5). In this case, the new (postcessation) lung cancer mortality rate, $C_{L C, x}$, is given by:

$$
C_{L C, x}=S I R \times R R_{x} \times N_{L C}+(1-S I R) \times N_{L C}
$$

Replacing this value for $C_{L C, x}$ in Equation 2 to obtain the postcessation smoking impact ratio, $S I R_{x}$, gives:

$$
S I R_{x}=\frac{S I R \times R R_{x} \times N_{L C}+(1-S I R) \times N_{L C}-N_{L C}}{S_{L C}^{*}-N_{L C}^{*}} \times \frac{N_{L C}^{*}}{N_{L C}}
$$

with some algebraic simplification and noting that $S_{\tilde{L} C, x}^{*}=R R \times N_{\tilde{L} C}^{*}$,

$$
\operatorname{SIR}_{x}=\operatorname{SIR} \frac{R R_{x} \times N_{L C}-N_{L C}}{R R \times S_{L C}^{*}-N_{L C}^{*}} \times \frac{N_{L C}^{*}}{N_{L C}}
$$

or

$$
\operatorname{SIR}_{x}=S I R \times \frac{R R_{x}-1}{R R-1}
$$

In other words, smoking cessation after $x$ years results in scaling down the population $S I R$ values relative to what they would be if smoking continued, with the scaling factor equal to the ratio of excess lung cancer risk after cessation to excess lung cancer risk if cessation did not occur. ${ }^{4}$

## DISEASES OTHER THAN LUNG CANCER

The decline of relative risk for cardiovascular diseases as a function of time since cessation has been estimated in a number of studies, resulting in a range of estimates for the rate of decline (Cook et al. 1986; Dobson et al. 1991; Kawachi et al. 1997; LaCroix et al. 1991; Rogot and Murray 1980; Rosenberg et al. 1985, 1990; U.S. Department of Health and Human Services 1990c). Kawachi et al. (1997) also considered the decline in the risk of other causes of mortality, including all cancers and external causes.

To convert the estimates of relative risk reduction from the time of cessation for diseases other than lung cancer to the estimates of relative risk per unit of SIR we used the following notation: Let $R R$ and $R R^{\prime}$ be the age-specific risks of lung cancer and disease $A$ respectively for lifelong smokers in the absence of cessation; $R R_{x}$ and $R R_{x}^{\prime}$ the age-specific relative risks at some specific time, $x$, after cessation; and $R R_{S I R}^{\prime}$ and $R R_{S I R, x}^{\prime}$ the relative risks of disease $A$ per unit of SIR without and with cessation respectively. In all cases, $R R_{S I R}^{\prime}=R R^{\prime}$ since in the absence of cessation, the relative risks for life-long smokers (from CPS-II for example) would apply. It is $R R_{S I R, x}^{\prime}$ that we are interested in, which, with the decline in SIR as outlined above, is now the risk of disease $A$ "per unit of former life-long smoker". Four scenarios are possible.

1. If the rate of decline in relative risk for cause $A$ is the same as the rate of decline in relative risk for lung cancer, then the relative risk per unit of SIR would stay constant at the pre-cessation level (except for age effects). This is because, in this case, all the benefits of cessation for both disease $A$ and lung cancer (which move together) would be captured in the decline of $\operatorname{SIR}$ (to $S I R_{x}$ ). Therefore, in this case $R R_{S I R, x}^{\prime}=R R^{\prime}$.
2. If there had been no decline in lung cancer risk (i.e. $R R_{x}=R R$ and $S I R_{x}=S I R$ ) but there was a decline in the risk of disease $A$ (i.e. $R R_{x}^{\prime}<R R^{\prime}$ ), then the relative risk per unit of SIR would decline at the same rate as the decline in relative risk for disease $A$ after cessation. This is because, in this case, none of the benefits of cessation would be captured by decline in $S I R$, and therefore they need to be included in the estimates of relative risk per unit of SIR. In this case $R R_{S I R, x}^{\prime}=R R_{x}^{\prime}$.
3. If disease $A$ is an acute condition, such as an injury or death (or possibly an acute respiratory infection) due to smoking-caused fires then cessation would result in complete removal of the risk for $A$, and $R R_{S I R, x}^{\prime}=1$.
4. If cessation results in lowering of the risk for both lung cancer and disease $A$, but at different rates, then the relative risk of disease $A$ per unit of SIR after cessation, is given by the following relationship:

$$
R R_{S I R, x}^{\prime}=1+\left(R R_{x}^{\prime}-1\right) \times \frac{R R_{x}^{\prime}-1}{R R^{\prime}-1} \times \frac{R R-1}{R R_{x}-1}
$$

The product of the terms $\frac{R R_{x}^{\prime}-1}{R R^{\prime}-1}$ and $\frac{R R-1}{R R_{x}-1}$ captures the relative drop in excess risk for disease $A$ (first term) compared to lung cancer (second term). If the risk of disease $A$ declines faster than lung cancer, then the relative risk per unit of SIR will decline to capture the additional benefits of cessation for disease $A$, compared to lung cancer. On the other hand, if the risk of disease $A$ declines more slowly than lung cancer, then the relative risk per unit of SIR will rise to capture the reduced benefits of cessation for disease $A$, compared to lung cancer. ${ }^{5}$ Scenarios $1-3$ above are special cases of this general relationship.

The estimates of $R R$ and $R R_{x}$ (for lung cancer) can be obtained from Figure 11.2 for different times since cessation. To estimate $R R^{\prime}$ and $R R_{x}^{\prime}$ for COPD and cardiovascular diseases, we used CPS-II data. In the analysis of risk reversibility for cardiovascular diseases for the first six-year follow up (1988), the estimates for those who had quit for less than one year showed inconsistent patterns (Table 11.6) even though subjects with cancer, heart disease and stroke were excluded at baseline (U.S. Department of Health and Human Services 1990b). This has been attributed to both high incidence of smoking resumption and the possible inclusion of subjects who stopped smoking as a result of symptoms from undiagnosed illness (U.S. Department of Health and Human Services 1990b).

Table II. 6 Decline in the risk of cardiovascular disease mortality with time since smoking cessation

|  | $R R$ (Male) |  |  | $R R$ (Female) |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Smoking status | $\leq 20$ cig/day | $>20 \mathrm{cig} / \mathrm{day}$ |  | $\leq 20 \mathrm{cig} / \mathrm{day}$ | $>20 \mathrm{cig} / \mathrm{day}$ |
| Current smoker | 1.93 | 2.02 |  | 1.76 | 2.27 |
| Years since cessation among former smokers |  |  |  |  |  |
| $<1$ | 1.43 | 2.56 |  | 2.13 | 1.41 |
| $1-2$ | 1.61 | 1.57 |  | 0.87 | 1.16 |
| $3-5$ | 1.49 | 1.41 |  | 1.31 | 0.96 |
| $6-10$ | 1.28 | 1.63 |  | 0.74 | 1.88 |
| $11-15$ | 0.99 | 1.16 |  | 1.2 | 1.37 |
| $\geq 16$ | 0.88 | 1.09 |  | 1.17 | 1.12 |

Source: U.S. Department of Health and Human Services (1990b).

The rate of decline in relative risk and the unstable behaviour in the early years since cessation were nonetheless consistent with other studies (see Table 2 in U.S. Department of Health and Human Services 1990b and the estimates [Kawachi et al. 1997] from the Nurses' Health Study).

Re-analysis of ACS CPS-II data (1998 follow up) with more stringent exclusion of subjects with a history of chronic disease (lung cancer, COPD, vascular diseases) shows a more consistent pattern of risk reversibility, especially if the first cessation period is considered as 2 years, which reduces the likelihood of high smoking resumption rates. Figure 11.3 shows the estimates of relative risk for COPD and cardiovascular disease with time since cessation for males and females using this re-analysis. To apply risk reversibility with cessation to the CPS-II and Chinese baseline relative risk estimates in Tables 11.1 and 11.3, we estimated the relative decline in relative risks after $x$ years of cessation (i.e. the ratio of relative risk $x$ years after cessation to current smokers in Figure 11.3) and applied this to CPS-II or Chinese relative risks (i.e. $R R^{\prime}$ ) to obtain $R R_{x}^{\prime}$.

For cardiovascular disease among men, former smokers with long cessation periods have lower risk than never-smokers (statistically not significant). This pattern, however, is implausible and is likely to be due to unobserved confounders that have affected the behaviour of former smokers (e.g. the same factor may have motivated cessation and change in diet or physical activity) and hence, lowered their risk. To account for this, we re-adjusted the values for the 25-29 and 30-34-year cessation periods based on the trends between the 20-24 and $\geq 35$ periods. For females, the estimates for cessation beyond 10 years were re-adjusted based on the male trends to asymptotically reach 1 for the longest cessation periods. The re-adjusted values are shown as dotted lines in Figure 11.3.

Given the potentially similar biological mechanisms of carcinogenesis, we assumed that all other cancers had the same rate of decline in relative risk as lung cancer. This is also seen in estimates of relative risk since smoking cessation for cancers including and excluding lung cancer analysed by Kawachi et al. (1997). Therefore, as described in scenario 1 above, the relative risk per unit of SIR remains constant for these diseases and the benefits of cessation in terms of avoidable mortality are fully captured by the decline in SIR. For all other disease categories, we used scenario 4 to estimate the risk per unit of SIR after cessation. We assumed that these diseases, for which reversibility data were not available, followed the same time pattern of decline as cardiovascular diseases, which decline more rapidly than lung cancer.

### 2.5 Analysis of uncertainty

In one taxonomy, uncertainty in quantitative risk assessment can be divided into parameter uncertainty and model uncertainty (Finkel 1990; National Research Council 1994). Parameter uncertainty includes the

Figure II. 3 Relative risk of chronic obstructive pulmonary disease (COPD) and cardiovascular diseases among former smokers


Note: Zero years represent current smokers. The estimates at 40 years represent cessation of more than 35 years in the subjects. Data are from ACS CPS-II 1998 follow up.

Source: American Cancer Society, unpublished data, 2003.
uncertainty quantifiable using random-variable methods such as that in a risk factor distribution, the magnitude of the risk factor-disease relationship, and burden of disease estimates. Model uncertainty is defined as uncertainty due to gaps in scientific theory (Finkel 1990; National Research Council 1994). In other words, model uncertainty occurs in those aspects of the analysis which are not currently quantifiable using a random-variable statistical methodology. In risk assessment, model uncertainty, broadly defined, also includes cases where exposure distributions or exposure-response relationships for populations are not known and are extrapolated from others. The uncertainty reported here is the $95 \%$ range of the combined distribution. In the following sections we describe the sources of uncertainty and the approach for quantifying parameter uncertainty.

## Parameter uncertainty

Uncertainty for each of the variables and parameters in the analysis was estimated separately. These include population lung cancer mortality and never-smoker lung cancer mortality for each subregion, lung cancer mortality for smokers and never-smokers in the reference population, and relative risks. The first four parameters are those needed for the estimation of SIR which, when combined with the relative risk of mortality, gives the fraction of cause-specific mortality due to smoking. Estimates of uncertainty for individual parameters were combined in a simulation using Latin hypercube sampling to obtain overall uncertainty for SIR and relative risk values, and eventually the fraction of mortality or morbidity due to smoking. The parameter-specific uncertainty estimates were obtained as follows:

1. Population lung cancer mortality: The GBD estimates of mortality do not currently include complete analysis of uncertainty. To obtain uncertainty estimates for population lung cancer mortality, we assigned each country into one of four uncertainty categories based on the quality of available mortality data, as we described earlier. Lung cancer mortality information for the four country categories were assigned uncertainty ranges equal to $10 \%, 20 \%, 40 \%$ and $80 \%$ of the best-estimate, with a triangular distribution. Since for countries with good vital registration, $95 \%$ or higher confirmation rates of the cause reported on death certificates have been found (Percy et al. 1990; Percy and Muir 1989), 10\% is more than double the observed uncertainty and a conservative estimate of true uncertainty. Eighty per cent uncertainty for the least certain countries, i.e. those with no established mortality reporting, implies that we have allowed nearly all of lung cancer mortality to be due to misclassification. Given that zero lung cancer mortality is implausible, higher levels of uncertainty could occur only on the upper side of these estimates, making this a conservative assumption for lung cancer mortality and SIR. Countries with less complete and/or less reliable data were

Table II. 7 Uncertainty (as \% of best-estimate) assigned to reported or estimated country level lung cancer mortality

| Uncertainty | Country $^{\text {a }}$ |
| :--- | :--- |
| I0\% | AMR-A - all; AMR-B - Chile, Costa Rica, Dominica, Uruguay; EMR-B - |
|  | Bahrain; EUR-A - all except Croatia; EUR-B - Poland, Slovakia; EUR-C - |
|  | Estonia, Hungary, Latvia, Lithuania, Russian Federation; WPR-A - all except |
|  | Brunei Darussalam |

a By subregion.
assumed to have $20 \%$ or $40 \%$ uncertainty around the best-estimate of lung cancer mortality. Table 11.7 shows the countries assigned to each uncertainty category.
2. Never-smoker lung cancer mortality: For China, estimates of the uncertainty for never-smoker lung cancer mortality rates from the proportional mortality study were used (Liu et al. 1998). For all other countries, where non-smoker lung cancer mortality rates were assumed to be those of CPS-II, we assumed an uncertainty of $15 \%$ around CPS-II estimates plus the statistical uncertainty of the CPS-II estimates themselves. Fifteen per cent is approximately equivalent to the whole (non-smoker) population being exposed to the highest levels of air pollution, such as that in the centre of the industrial city of Cracow, Poland, which resulted in a $14 \%$ increase in lung cancer mortality over the period 1980-1985 (Jedrychowski et al. 1990) or the whole (never-smoker) population being exposed to more than $10 \mu \mathrm{~g} / \mathrm{m}^{3}$ of $\mathrm{PM}_{2.5}$ (particulates below 2.5 microns in diameter) (Pope et al. 2002). Because the net difference between accumulated exposure to additional lung cancer risk factors (radon, urban air pollution, biomass smoke) in any two countries is less than the whole population, $15 \%$ is a relatively large uncertainty range for never-smoker lung cancer mortality.
3. Reference population smoker and never-smoker lung cancer mortality: We used statistical uncertainty of the CPS-II population since the reference population is constant in all regions of the world and the excess mortality of the reference population of smokers is simply a normalizing factor.
4. Relative risk: Uncertainties in relative risks were obtained directly from the CPS-II and Chinese mortality studies as well as the decline in risk due to cessation.
5. Confounding and extrapolation of relative risk and correction factor: To avoid overestimation of risk as a result of confounding in CPS-II relative risks, as well as from the extrapolation of relative risk to other regions, we have reduced the relative risk values by $30 \%$ to $50 \%$ for various diseases. To account for uncertainty in the level of the correction factor, we assumed that the $30 \%$ correction factor could vary between $10 \%$ and $50 \%$ with a triangular distribution. The lower end of this range corresponds to the typical amount of reduction in excess risk seen after adjustment for covariates in the re-analysis of CPS-II data (Thun et al. 2000) and seen in other studies such as the magnitude of confounding due to diet found by Law et al. (1997). The upper end is the conservative correction used by Peto et al. (1992) before the level of confounding was known and larger than those in the CPS-II re-analysis (Thun et al. 2000). The $50 \%$ correction factor used for the "other medical causes" group was assumed to be highly uncertain and in the range of $10 \%$ to $90 \%$ with a triangular distribution. We also assumed that the $5 \%$ correction factor for estimates of relative risk for China were in the range of $0-10 \%$ with a triangular distribution.

## Model uncertainty

Although lung cancer mortality provides a biological marker for accumulated exposure to the hazards of smoking in general, the question remains whether it equally represents the accumulated hazards for each of the diseases considered in this analysis. Two factors may reduce the accuracy of excess lung cancer and SIR as markers of accumulated hazard:

1. If the cigarette smoke characteristics that cause lung cancer are not fully correlated with those that cause the other hazard: While the carcinogens in cigarette smoke are the cause of cancer, the concentration or size distribution of respirable particles or other pollutants may be a better indicator of some of the other health effects. The validity of the SIR method compared to direct estimates in the United States was estimated by Peto et al. (1992). Comparisons with other (largely direct) methods for other countries also show consistent results (Bronnum-Hansen and Juel 2000; Valkonen and van Poppel 1997).
2. If the time-to-hazard (or hazard accumulation function as described in chapter 1) is different for lung cancer and other diseases caused by smoking: Lung cancer is caused by exposure and hazard accumulated over an extended period of time spent smoking. This property may be shared by some other outcomes of smoking, such as other cancers and COPD. Some of the other health effects of smoking, such as acute respiratory diseases, may occur after immediate exposure, and others, such as cardiovascular disease, are determined by a period of exposure that may be somewhat shorter than that of lung cancer. With these differences, when smoking is on the rise, an indicator based on excess lung cancer-which generally occurs later-would underestimate the impacts of those diseases that occur earlier. On the other hand some time after smoking begins to decline, an indicator based on excess lung cancer would overestimate the impacts of those diseases, where risk declines faster with cessation than lung cancer. ${ }^{6}$ Since smoking has been increasing in most regions of the world over the past two or three decades, the net effect of this would be an underestimation of current global mortality due to cardiovascular and some other diseases in our analysis.

In addition to the uncertainty in the use of SIR as exposure variable, the following other sources of uncertainty have not been quantified:
3. The impact of environmental tobacco smoke on SIR estimates: The carcinogenic effects of cancer agents are likely to apply at low levels without a threshold (Peto 1978). Therefore exposure to environmental tobacco smoke (ETS) is a likely cause of lung cancer. This relationship, as well as the impact of ETS on cardiovascular disease, has also been established in epidemiological studies (Australia National Health and Medical Research Council 1997; Environmental Protection Agency 1992; Hackshaw et al. 1997; Law et al. 1997; UK Department of Health Scientific Committee on Tobacco and Health 1998). Although we did not conduct a separate analysis of the burden of disease due to ETS, because of its relationship to lung cancer, ETS exposure affects SIR estimates.

Of the four variables in the $S I R$ relationship ( $C_{L C}, N_{L C}, S_{\tilde{L} C}^{*}$, and $N_{\tilde{L} C}^{*}$ in Equation 2), the effects of ETS are smallest on life-long smokers ( $S_{\tilde{L} C}^{*}$ ) who are almost completely affected by direct smoking. The lung cancer mortality of the population as a whole $\left(C_{L C}\right)$ is also affected more by direct smoking but nonetheless captures the effects of both direct and indirect exposure to tobacco smoke. The effects of ETS are largest on never-smokers ( $N_{L C}$ and $N_{L C C}^{*}$ ) who would otherwise not be exposed to tobacco smoke. Therefore, both $N_{L C}$ and $N_{\tilde{L} C}^{*}$ are larger in the presence of ETS than they would be in its absence. Since $C_{L C}$ is almost always smaller than $S_{\tilde{L} C}^{*}$, increasing non-smoker lung cancer rates will result in
a larger relative reduction (through subtraction) in the numerator of Equation 1 compared to the denominator, and therefore an underestimation of SIR values.
4. Estimates of non-fatal health outcomes: To obtain relative risks for non-fatal effects of these diseases, we have assumed that smoking increases mortality by increasing disease incidence (rather than modifying case fatality) for cancers and COPD. We have assumed a potential change in either the incidence or the severity/case fatality for all other causes. Given that most of the diseases affected by smoking are chronic diseases as a result of chronic exposure, this is a conservative assumption and requires further investigation using epidemiological studies on the relationship between smoking and disease incidence.
5. Future exposure and risk reversibility: Estimates of avoidable burden are possibly the most uncertain component of a risk assessment exercise because of the number of assumptions that, by definition, are needed for estimates of future burden. These include estimates of the "business-as-usual" trends of smoking and lung cancer and estimates of risk reversibility (reduction in relative risk) for those current smokers who stop smoking.

The future estimates of lung cancer mortality and SIR (see below) are based on a number of assumptions. These include the descriptive model of the tobacco epidemic and its parameters, the estimates of total tobacco consumption, and the statistical model used for estimating lung cancer based on consumption. As we discussed under risk reversibility, the estimates of decline in the relative risk for lung cancer, and more importantly for other diseases, after smoking cessation are derived from a limited number of studies and extrapolated to people of different ages, sexes and smoking histories. At the same time it may be possible that the benefits of cessation are more dependent on the specific history of smoking such as duration of smoking and age at cessation, than the current accumulated risk (which is captured by $S I R$ ) as also seen in the comparison between male and female reductions in lung cancer risk in Table 11.5. In this case, the use of $S I R$ as the indicator of pre-cessation history would create an additional source of uncertainty. At the same time, after a few decades, the potential health benefits of smoking reduction occur among those who would be prevented from smoking altogether. In this case, the longer-term estimates of avoidable burden would be less dependent on the specifics of risk reversibility and depend only on the estimates of risk among life-long smokers, which are known with much greater certainty.

### 2.6 ADdItIonal oral cancer mortality due to oral tobacco USE

Oral tobacco use, in the form of chewing of betel-quid with tobacco, is common in many parts of south Asia, and in particular in the Indian sub-continent (Bhonsle et al. 1992). Although oral tobacco use has been associated with increased risk of all-cause mortality (Gupta and Mehta 2000), we focus here on the risk of oral cancer which is the most-widely known and studied outcome of this risk behaviour (Gupta et al. 1982; International Agency for Research on Cancer (IARC) 1984; U.S. Department of Health and Human Services 1986), and a leading form of cancer mortality in the region (Parkin et al. 1997). Smoking in India is much more common among men than women whereas the prevalence of tobacco chewing is of comparable magnitude, although still higher among men (Corrao et al. 2000; Gupta 1996; WHO 1997). Given the correlation between smoking and oral tobacco use, many cases of oral cancer are likely to be affected by both habits. To report the total burden of disease due to both forms of tobacco use, we made estimates of only those cases of oral cancer that are caused by tobacco chewing in addition to smoking. The estimates were applied to SEAR-D only. Oral tobacco use is common throughout the region, whose mortality estimates are nonetheless dominated by those from India.

We obtained estimates of the fraction of total chewers who do not smoke $\left(p_{c}\right)$ and the fraction of total smokers who do not chew $\left(p_{s}\right)$ using a survey of tobacco habits in various ethnic and religious groups in India (Gupta 1996). The estimates of oral cancer for smokers who do not chew $\left(p_{s}\right)$ are included in those from the application of the SIR method. Those who both smoke and chew ( $p_{s c}$ as a fraction of total smokers), have a higher risk than smokers-only, by a factor $R R_{\text {(chewing+smoking)/smoking }}$, because of their additional chewing habit. We used $R R_{\text {(chewingtsmoking)/smoking }}$ to increase the CPS-II relative risk estimates for the upper-aerodigestive cancer category for this group, ${ }^{7}$ and applied these increased risk estimates to oral cancer mortality. The difference between these new estimates of oral cancer (using increased relative risk) and those using the CPS-II risk estimate, are the additional deaths due to oral tobacco use over and above those accounted for by smoking, among those who both chew and smoke. Finally, for those who chew only $\left(p_{c}\right)$, we used the corresponding relative risk, $R R_{\text {(chewing and not smoking) }}$, to obtain estimates of oral cancer mortality, which are independent from those obtained using the SIR method. For this group, we directly used the estimated prevalence of chewing. Although oral cancer as a result of tobacco chewing is also dependent on accumulated exposure, the relative risk estimates from recent literature are directly applicable because they were used in the same time period and region where epidemiological studies took place.

For males, we used the estimate that $20 \%$ of all smokers also chew (WHO 1997), which is close to the largest overlap of any ethnic or reli-
gious group found in a study of socio-demographic characteristics of tobacco users (Gupta 1996). Assuming a $40 \%$ prevalence of smoking and a $65 \%$ prevalence of total tobacco use, this implies that $32 \%$ of adult males smoke only, $8 \%$ of adult males both smoke and chew, and another $25 \%$ chew only, consistent with other estimates (WHO 1997). For females, we used a $33 \%$ prevalence of total tobacco use and a $3 \%$ prevalence of smoking (Corrao et al. 2000; WHO 1997). We also assumed that all female smokers also chew tobacco (Gupta 1996), hence $30 \%$ of adult females are chewers-only.

A review of studies prior to 1982 that estimated the relative risk of oral cancer due to tobacco chewing is provided by Gupta et al. (1982). We did not use these estimates since the definitions of cancer sites or control for covariates were not consistent with more recent studies. All these studies nonetheless show statistically significant increased risk of oral cancer due to tobacco chewing. A number of recent studies have estimated the relative risks for oral cancer (or one of its sub-types) based on stratification by smoking and tobacco chewing, all finding a significant increase in the risk of oral cancer among chewers regardless of their smoking habits (Balaram et al. 2002; Dikshit and Kanhere 2000; Rao et al. 1994; Sankaranarayanan et al. 1989a, 1989b, 1990). A description of recent studies and their estimates of relative risk (or odds ratio) of oral cancer due to chewing tobacco among smokers and non-smokers is provided in Table 11.8. Those studies that provided separate analysis according to the frequency of chewing also consistently show an increasing dose-response relationship. Note that in some of the studies, those who both chew and smoke tobacco had lower risk than those who used oral tobacco only. This may be because smokers chewed less tobacco than those whose only habit was tobacco chewing. We used a relative risk of 3.64 for chewing only and 1.7 for those who both chew and smoke relative to smokers. These values, which are lower than almost all other studies, are from non-drinkers in the study of Rao et al. (1994) to avoid confounding due to alcohol consumption, an important risk factor for oral cancer.

## 3. Results

### 3.1 Exposure (SIR) estimates

Figure 11.4 shows SIR estimates for females and males in developing and industrialized countries. We have also shown in the figure legend the best estimate of adult smoking prevalence for each subregion. As discussed above, prevalence estimates are uncertain and based on data from a limited number of countries. At the same time, as discussed by Jha et al. (2002), despite uncertainties in country-level and age-specific estimates, they provide a reasonable indication of current smoking status among adults in each subregion. The subregion WPR-B includes China.
Table II.8 Summary of studies on the relationship between tobacco chewing, smoking and oral cancer ${ }^{a}$

| Study location (reference) | Cancer site (ICD code) | Number of cases and control choice | Covariate adjustment | Effect size |
| :---: | :---: | :---: | :---: | :---: |
| Kerala (Sankaranarayanan et al. 1989a) | Oral tongue and the floor of the mouth (ICD 14I.I-I4I.4 and 144) | 228 cases; 453 hospital-based controls matched for age, sex and religion | Age | $\begin{aligned} & R R_{\text {chewing }}=6.13 ; R R_{\text {smoking }}=4.98 ; \\ & R R_{\text {chewing }} \text { +smoking } \\ & =7.2^{\mathrm{b}} \end{aligned}$ |
| Kerala (Sankaranarayanan et al. 1989b) | Gingiva (ICD 143.0 and I43.1) | 187 cases; 895 hospital-based controls with respiratory, intestinal, and genito-urinary infections, excluding malignancy in sites other than head and neck | Age | $\begin{aligned} & R R_{\text {chewing }}=11.76 ; R R_{\text {smokking }}=4.21 \text {; } \\ & R R_{\text {chewing+smoking }}=16.48^{\mathrm{b}} \end{aligned}$ |
| Kerala (Sankaranarayanan et al. 1990) | Buccal and labial mucosa (ICD 145.0, 145.I, and 145.6-140.3 and 140.4) | 414 cases; 895 hospital-based controls with non-malignant conditions | Age, religion | $\begin{aligned} & R R_{\text {chewing }}=14.28 ; R R_{\text {smoking }}=4.21 \text {; } \\ & R R_{\text {chewingtsmoking }}=21.46^{\text {b }} \end{aligned}$ |
| Bangalore (Nandakumar et al. 1990) | Oral cavity (140-14\| and 143-145 excluding 141.0) | 348 cases; 348 hospital-based controls with non-malignant conditions matched for age, sex and residence | Age, sex, religion and diet | $\begin{aligned} & R R_{\text {chewing }}=10.2 ; R R_{\text {smoking }}=3.5 ; \\ & R R_{\text {chewing }+ \text { smoking }}=9.2^{b} \end{aligned}$ |

Table II. 8 Summary of studies on the relationship between tobacco chewing, smoking and oral cancer ${ }^{\text {a }}$ (continued)

| Study location (reference) | Cancer site (ICD code) | Number of cases and control choice | Covariate adjustment | Effect size |
| :---: | :---: | :---: | :---: | :---: |
| Bombay (Rao et al. 1994) | Oral (excluding 141.0 and 145.3) | 713 male cases; 635 male hospital-based controls without cancer, benign tumour and infectious disease | Age and area of residence | $\begin{aligned} & \text { Non-drinkers: } R R_{\text {chewing }}=3.64 ; \\ & R R_{\text {smoking }}=1.69 ; R R_{\text {chewing+smoking }}=2.93 \\ & \text { Drinkers: } R R_{\text {chewing }}=4.32 ; R R_{\text {smoking }}=2.44 ; \\ & R R_{\text {chewingtsmoking }}=8.88 \end{aligned}$ |
| Bhopal (Dikshit and Kanhere 2000) | Oral cavity (140.0-144.9, 145.0-145.2, 145.5-145.9) | 148 male cases; 260 controls randomly selected from a survey of tobacco habits | Age | $\begin{aligned} & O R_{\text {chewing }}=10.6 ; O R_{\text {smoking }}(<10 \mathrm{cig} / \text { day })=1.0 ; \\ & O R_{\text {smoking }}(\geq 20 \text { cig } / \text { day })=4.9 ; O R_{\text {chewing }+ \text { smoking }} \\ & (<20 \text { cig } / \text { day })=8.4 ; O R_{\text {chewingtsmoking }} \\ & (\geq 20 \text { cig } / \text { day })=16.3 \end{aligned}$ |
| Southern India (Balaram et al. 2002) | Oral | 591 cases; hospital-based controls matched by medical centre, age and sex | Age, sex, medical centre, education and alcohol | $\begin{aligned} & O R_{\text {chewing }}=9.19 ; O R_{\text {smoking }} \\ & (<20 \mathrm{cig} / \text { day })=1.78 ; O R_{\text {smoking }} \\ & (\geq 20 \mathrm{cig} / \text { day })=3.69 ; O R_{\text {chewing+smoking }} \\ & (<20 \mathrm{cig} / \text { day })=8.86 ; O R_{\text {chewingtsmoking }} \\ & (\geq 20 \mathrm{cig} / \text { day })=6.69^{\mathrm{b}} \end{aligned}$ |

All studies were case-control and took place in India.
b The effect estimates are for males only since none or few of the female cases smoked.

Figure II.4 Estimates of smoking impact ratio (SIR) by age, sex and subregion

(b) Developing countries (male) II


Figure II.4 Estimates of smoking impact ratio (SIR) by age, sex and subregion (continued)

(d) Developing countries (female) II


Figure II.4 Estimates of smoking impact ratio (SIR) by age, sex and subregion (continued)

(f) Industrialized countries (female)


[^41]Hence the aggregate estimates for this subregion are dominated by the Chinese data. This division allows the characteristics of the other countries in the subregion to also be considered, in particular given the role of higher background lung cancer mortality in China. Table 11.9 provides the SIR distribution, divided as defined earlier.

A number of important features of the global smoking epidemic can be seen in the different panels of Figure 11.4:

1. The largest accumulated risk of smoking for males aged $<70$ years in the countries of eastern Europe and the former Soviet Union (EUR-B and EUR-C). For older males (aged $\geq 70$ years), North America and western European countries have the largest accumulated risk. These two results are consistent with the very high current and recent prevalence of smoking among eastern European men which has been sustained for several decades, and the longer history of smoking among North American and western European men.

The accumulated hazards of smoking among men were lowest in AMR-D, AFR-D and AFR-E. In these subregions, the rise in smoking has been a recent phenomenon. The results for China (WPR-B) are particularly important and instructive. The background-adjusted SIR values for China were still fairly low, despite high lung cancer mortality in this country. At the time of the Liu et al. (1998) study (in 1990) the relative risk of lung cancer for a Chinese smoker was less than 3.0 because of the more recent start of the epidemic in this country. Over the next few decades, increasing accumulated exposure to smoking may well result in rising lung cancer mortality in China to levels comparable to populations with life-long smokers (such as those in North America and western Europe where the relative risks for lung cancer are approximately 20). Together with the finding of Liu et al. (1998) that smoking acts to "amplify" the high background rates of lung cancer, this increase in accumulated hazard will result in enormous lung cancer (and other) mortality in China.

Examining prevalence and SIR patterns simultaneously also emphasizes the importance of considering accumulated risks. For example, the current prevalence of smoking among adult men in AMR-A was equal to AFR-D and lower than all other subregions in the developing world. At the same time, the SIR values for AMR-A males were larger than those of most developing subregions because of the histories of smoking. In AMR-A, smoking has been declining and current prevalence would underestimate the current impacts of smoking. In developing countries, on the other hand, smoking has been rising in recent decades with current prevalence being comparable to AMR-A, but the accumulated hazards were still lower.
2. For women, AMR-A had the single highest accumulated risk from smoking. Although women in many countries in western Europe
Table II. 9 Prevalence of exposure by subregion, age and sex

| Subregion | Exposure variable ${ }^{\text {a }}$ | Prevalence of exposure (\%) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 years ${ }^{\text {b }}$ |  | 5-14 years ${ }^{\text {b }}$ |  | 15-29 years ${ }^{\text {b }}$ |  | 30-44 years |  | 45-59 years |  | 60-69 years |  | 70-79 years |  | $\geq 80$ years |  |
|  |  | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female |
| AFR-D | I | NA | NA | NA | NA | NA | NA | 75 | 96 | 75 | 96 | 75 | 96 | 75 | 96 | 75 | 96 |
|  | 2 | NA | NA | NA | NA | NA | NA | 25 | 4 | 23 | 3 | 23 | 4 | 25 | 4 | 25 | 4 |
|  | 3 | NA | NA | NA | NA | NA | NA | 0 | 0 | 2 | I | 2 | I | 0 | 0 | 0 | 0 |
| AFR-E | I | NA | NA | NA | NA | NA | NA | 55 | 88 | 55 | 88 | 55 | 88 | 55 | 88 | 55 | 88 |
|  | 2 | NA | NA | NA | NA | NA | NA | 40 | 7 | 43 | 8 | 45 | 10 | 45 | 12 | 45 | 11 |
|  | 3 | NA | NA | NA | NA | NA | NA | 5 | 5 | 2 | 4 | 0 | 2 | 0 | 0 | 0 | 0 |
| AMR-A | 1 | NA | NA | NA | NA | NA | NA | 56 | 66 | 55 | 59 | 50 | 45 | 55 | 39 | 72 | 78 |
|  | 2 | NA | NA | NA | NA | NA | NA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|  | 3 | NA | NA | NA | NA | NA | NA | 44 | 34 | 45 | 41 | 50 | 55 | 45 | 61 | 28 | 22 |
| AMR-B | I | NA | NA | NA | NA | NA | NA | 60 | 75 | 60 | 75 | 60 | 75 | 60 | 75 | 60 | 75 |
|  | 2 | NA | NA | NA | NA | NA | NA | 4 | 20 | 15 | 21 | 18 | 20 | 24 | 22 | 25 | 17 |
|  | 3 | NA | NA | NA | NA | NA | NA | 35 | 5 | 25 | 4 | 22 | 5 | 16 | 3 | 15 | 8 |
| AMR-D | 1 | NA | NA | NA | NA | NA | NA | 61 | 84 | 61 | 84 | 61 | 84 | 61 | 84 | 61 | 84 |
|  | 2 | NA | NA | NA | NA | NA | NA | 39 | 16 | 39 | 16 | 39 | 16 | 39 | 16 | 39 | 16 |
|  | 3 | NA | NA | NA | NA | NA | NA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EMR-B | 1 | NA | NA | NA | NA | NA | NA | 66 | 77 | 66 | 94 | 66 | 94 | 66 | 94 | 66 | 94 |
|  | 2 | NA | NA | NA | NA | NA | NA | 8 | 0 | 13 | 0 | 8 | 0 | 12 | 0 | 11 | 3 |
|  | 3 | NA | NA | NA | NA | NA | NA | 27 | 23 | 22 | 6 | 26 | 6 | 23 | 6 | 24 | 3 |
| EMR-D | 1 | NA | NA | NA | NA | NA | NA | 63 | 92 | 63 | 92 | 63 | 92 | 63 | 92 | 63 | 100 |
|  | 2 | NA | NA | NA | NA | NA | NA | 0 | 0 | 25 | 7 | 29 | 7 | 33 | 8 | 33 | 0 |
|  | 3 | NA | NA | NA | NA | NA | NA | 37 | 8 | 13 | 1 | 8 | 1 | 4 | 0 | 5 | 0 |
| EUR-A | 1 | NA | NA | NA | NA | NA | NA | 37 | 74 | 45 | 74 | 52 | 74 | 63 | 74 | 63 | 74 |
|  | 2 | NA | NA | NA | NA | NA | NA | 0 | 0 | 0 | 5 | 0 | 13 | 0 | 12 | 0 | 14 |
|  | 3 | NA | NA | NA | NA | NA | NA | 63 | 26 | 55 | 20 | 48 | 13 | 37 | 14 | 37 | 12 |

Table II. 9 Prevalence of exposure by subregion, age and sex (continued)

| Subregion | Exposure variable ${ }^{a}$ | Prevalence of exposure (\%) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 years ${ }^{\text {b }}$ |  | 5-14 years ${ }^{\text {b }}$ |  | 15-29 years ${ }^{\text {b }}$ |  | 30-44 years |  | 45-59 years |  | 60-69 years |  | 70-79 years |  | $\geq 80$ years |  |
|  |  | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female |
| EUR-B | 1 | NA | NA | NA | NA | NA | NA | 12 | 74 | 25 | 74 | 55 | 74 | 55 | 74 | 55 | 74 |
|  | 2 | NA | NA | NA | NA | NA | NA | 0 | 2 | 0 | 12 | 0 | 18 | 5 | 21 | 30 | 2 |
|  | 3 | NA | NA | NA | NA | NA | NA | 88 | 23 | 75 | 14 | 45 | 8 | 40 | 4 | 15 | 0 |
| EUR-C | I | NA | NA | NA | NA | NA | NA | 2 | 85 | 7 | 85 | 42 | 85 | 42 | 85 | 42 | 85 |
|  | 2 | NA | NA | NA | NA | NA | NA | 0 | 6 | 0 | 7 | 0 | 10 | 0 | 7 | 46 | 15 |
|  | 3 | NA | NA | NA | NA | NA | NA | 98 | 9 | 93 | 8 | 58 | 5 | 58 | 9 | 12 | 0 |
| SEAR-B | 1 | NA | NA | NA | NA | NA | NA | 40 | 97 | 40 | 97 | 40 | 97 | 40 | 97 | 40 | 100 |
|  | 2 | NA | NA | NA | NA | NA | NA | 0 | 3 | 24 | 0 | 35 | 0 | 43 | 2 | 43 | 0 |
|  | 3 | NA | NA | NA | NA | NA | NA | 60 | 0 | 36 | 3 | 25 | 3 | 17 | 0 | 17 | 0 |
| SEAR-D | 1 | NA | NA | NA | NA | NA | NA | 55 | 91 | 55 | 91 | 55 | 91 | 55 | 91 | 55 | 91 |
|  | 2 | NA | NA | NA | NA | NA | NA | 6 | 9 | 19 | 6 | 24 | 6 | 33 | 9 | 35 | 9 |
|  | 3 | NA | NA | NA | NA | NA | NA | 39 | 0 | 26 | 3 | 21 | 4 | 12 | 0 | 10 | 0 |
| WPR-A | 1 | NA | NA | NA | NA | NA | NA | 51 | 85 | 51 | 85 | 51 | 85 | 51 | 84 | 51 | 85 |
|  | 2 | NA | NA | NA | NA | NA | NA | 28 | 8 | 37 | 6 | 25 | 4 | 4 | 0 | 0 | 0 |
|  | 3 | NA | NA | NA | NA | NA | NA | 20 | 7 | 12 | 9 | 23 | 9 | 45 | 16 | 49 | 15 |
| WPR-B <br> (excluding <br> China) | 1 | NA | NA | NA | NA | NA | NA | 31 | 90 | 31 | 90 | 31 | 90 | 31 | 90 | 31 | 90 |
|  | 2 | NA | NA | NA | NA | NA | NA | 0 | 0 | 26 | 2 | 27 | 0 | 34 | 0 | 42 | 0 |
|  | 3 | NA | NA | NA | NA | NA | NA | 69 | 10 | 43 | 8 | 42 | 10 | 35 | 9 | 27 | 10 |
| China | 1 | NA | NA | NA | NA | NA | NA | 37 | 96 | 37 | 96 | 37 | 96 | 37 | 96 | 37 | 96 |
|  | 2 | NA | NA | NA | NA | NA | NA | 63 | I | 63 | 2 | 63 | 1 | 63 | I | 63 | 3 |
|  | 3 | NA | NA | NA | NA | NA | NA | 0 | 2 | 0 | 1 | 0 | 3 | 0 | 2 | 0 | 1 |

[^42]NA Not applicable.
(EUR-A) have smoked for a long time (in particular in the United Kingdom), smoking is a more recent phenomenon in the southern parts of the continent, and therefore the overall SIR is still lower than in AMRA. SIR values for women were consistently low in developing countries except for younger and middle-aged women in AMR-B (Latin America and the Caribbean) and young adult women in EMR-B. Once again, comparing SIR values for females in AMR-A, whose current prevalence of smoking is $22 \%$ (reflecting recent declines in female smoking in North America), with those for males in many subregions of the developing world, who have higher current prevalence but lower SIR, illustrates the inadequacy of current prevalence as a marker for smoking risk. ${ }^{8}$
3. Age patterns of SIR estimates for the different subregions also provide information about the state of the smoking epidemic. For each age-sex group, SIR is excess lung cancer, relative to the same age-sex group of American smokers in the 1980s. Among men in industrialized countries, the SIR values were relatively constant across ages in AMR-A, but decline with age in other industrialized regions, with EUR-A being closest to the constant pattern. These inter-subregional and intra-subregional age patterns imply that, when compared to the same age group of American smokers in the 1980s, age patterns of smoking have been relatively constant in North America whereas smoking has been more concentrated among younger and middle-age men in Europe and the Western Pacific, especially in the former Soviet Union (EUR-C). In the developing countries of Latin America and the Caribbean, the Eastern Mediterranean, and sub-Saharan Africa, smoking also seems to have had fairly constant effects across ages, when each is compared to the same age group of American smokers in the 1980s. In Asia, on the other hand, smoking had a greater impact on younger and middle-aged male cohorts than at older ages when compared to the same age group of American smokers in the 1980s.

Overall, the age patterns of SIR were less variant among women in both developing and industrialized countries compared to men. In AMRB, EMR-B, EMR-D, and the EUR subregions there is more smoking among younger and middle-age women than at older ages, when each is compared to the same age group of female American smokers in the 1980s. In Asia, female smoking seems to peak among the middle-aged cohorts, which may reflect social factors that would prevent smoking among many young women. In North America, the distribution was more towards older age groups when compared to the same age group of female American smokers in the 1980s.

### 3.2 Mortality and disease burden due to smoking

Tables 11.10 and 11.11 provide the estimated number of smoking-attributable deaths and DALYs for males and females in developing and industrialized countries. Table 11.12 divides the estimates of global mortality
due to smoking into broad causes and age groups. Although the results were estimated for the eight age groups described in Table 11.9 (assumed to be zero for the first three age groups $<30$ years), they are reported in two age groups ( $30-69$ and $\geq 70$ years) to be comparable with previous estimates of mortality due to smoking, such as those in Peto et al. (1992). The distribution of mortality and DALYs by broad disease groups is given in Figures 11.5 and 11.6.

The 4.83 ( $95 \%$ CI $3.94-5.93$ ) million deaths due to smoking accounted for $12 \%$ of total global adult (aged $\geq 30$ years) mortality. ${ }^{9}$ The shares of adult male and female total mortality due to smoking were $18 \%$ and $5 \%$, respectively. Of these deaths, 2.69 million were among those aged 30-69 years, resulting in a larger number of life years lost to premature mortality, and 2.14 million among those aged $>69$ years. As seen in a comparison of Tables 11.10 and 11.11, although developing and industrialized countries accounted for virtually equal numbers of global mortality, the burden of disease associated with this risk factor

Table II.IO Mortality (in millions) due to smoking in developing and industrialized countries, 2000

|  | Male | Female | Total |
| :--- | :---: | :---: | :---: |
| Developing $^{a}$ | $2.02(1.56-2.50)^{c}$ | $0.38(0.25-0.65)^{c}$ | $2.41(1.80-3.15)^{c}$ |
| Industrialized $^{\text {b }}$ | $1.81(1.62-2.02)^{c}$ | $0.61(0.52-0.75)^{c}$ | $2.43(2.13-2.78)^{c}$ |
| Total | $3.84(3.17-4.53)^{c}$ | $1.00(0.76-1.40)^{c}$ | $4.83(3.94-5.93)^{c}$ |

a Developing countries include those in AFR, AMR-B, AMR-D, EMR, SEAR and WPR-B subregions.
b Industrialized countries include those in AMR-A, EUR and WPR-A.
c Numbers in parentheses are $95 \%$ confidence intervals.

Table II.II Loss of healthy life years (in thousands of DALYs) due to tobacco-caused mortality and morbidity in developing and industrialized countries, 2000

|  | Male | Female | Total |
| :--- | :---: | :---: | :---: |
| Developing $^{a}$ | $28015(4.4 \%)^{c}$ | $4962(0.8 \%)^{c}$ | $32977(2.7 \%)^{c}$ |
| Industrialized $^{b}$ | $20162(17 \%)^{c}$ | $5942(6.2 \%)^{c}$ | $26104(12 \%)^{c}$ |
| Total | $48177(6.3 \%)^{c}$ | $10904(1.6 \%)^{c}$ | $59081(4.1 \%)^{c}$ |

[^43]Table II.I2 Global mortality due to smoking by cause, sex and age, 2000

| Cause ${ }^{\text {a }}$ | Male |  |  |  | Female |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | No. of deaths (000s) | Fraction of total mortality (\%) | No. of deaths (000s) | Fraction of total mortality (\%) | No. of deaths (000s) | Fraction of total mortality (\%) | No. of deaths (000s) | Fraction of total mortality (\%) |
|  | 30-69 years |  | $\geq 70$ years |  | 30-69 years |  | $\geq 70$ years |  |
| Lung cancer | 398 | 77 | 294 | 82 | 77 | 44 | 79 | 54 |
| Upper aerodigestive cancer ${ }^{\text {b }}$ | 152 | 46 | 66 | 42 | 17 | 12 | 15 | 13 |
| Other cancer ${ }^{\text {b }}$ | 195 | 15 | 135 | 13 | 17 | 1 | 24 | 2 |
| COPD ${ }^{\text {c }}$ | 269 | 54 | 433 | 52 | 86 | 24 | 178 | 19 |
| Other respiratory diseases | 274 | 22 | 93 | 11 | 34 | 5 | 32 | 4 |
| Cardiovascular diseases | 848 | 24 | 476 | 12 | 143 | 6 | 223 | 4 |
| Other attributable medical causes | 0 | 0 | 0 | 0.0 | 0 | 0 | 0 | 0 |
| Other medical causes | 145 | 17 | 57 | 8 | 36 | 5 | 35 | 4 |
| Total medical | 2280 | 22 | 1556 | 18 | 410 | 6 | 587 | 5 |
| Non-medical (accidents and injuries) | 0 | 0 | 0 | 0.0 | 0 | 0 | 0 | 0 |
| Total mortality | 2280 | 19 | 1556 | 18 | 410 | 5 | 587 | 5 |
| ${ }^{\text {a }}$ See Table II.I for details on causes of death. |  |  |  |  |  |  |  |  |
| c The estimates also include ICD 495, which was not included in the CPS-II relative risk estimates, due to GBD grouping; normally this cause w causes. |  |  |  |  |  |  |  |  |

Figure II. 5 Distribution of mortality due to smoking by cause and development group, 2000
(a) Developing countries

Other medical causes

(b) Industrialized countries

Other medical causes


Figure II.5 Distribution of mortality due to smoking by cause and development group, 2000 (continued)


UADC Upper aerodigestive cancer.
Note: See Table II.I for details on causes of death.
was much higher in the former. As discussed below, this is because in general smoking-caused mortality in developing countries occurs at earlier ages than in industrialized nations accounting for a larger loss of life from premature mortality. Further, a comparison of Figures 11.5 and 11.6 indicates that the relative share of cancers in terms of the total burden of disease is lower than their comparative role in mortality because of the shorter morbidity associated with cancers compared to the other categories.

Lung cancer was the disease with the highest fraction attributable to smoking. Seventy-one per cent of all lung cancers or 0.85 million deaths ( $79 \%$ or 0.69 million deaths among men and $48 \%$ or 0.16 million deaths among women) were attributable to smoking. However, cardiovascular diseases were the largest cause of death due to smoking in terms of number of deaths. One million six hundred and ninety thousand cardiovascular disease deaths ( 1.37 million among men and 0.32 million

Figure II.6 Distribution of DALYs due to smoking by cause and development group, 2000
(a) Developing countries

(b) Industrialized countries


Figure II.6 Distribution of DALYs due to smoking by cause and development group, 2000 (continued)


UADC Upper aerodigestive cancer.
Note: See Table II.I for details on causes of death.
among women) were due to smoking, accounting for $35 \%$ of all smoking-attributable deaths ( $35 \%$ among men and $37 \%$ among women). Overall, $11 \%$ of all cardiovascular deaths in the world were attributable to smoking ( $17 \%$ among men and $4 \%$ among women). Only when all cancers are considered together did they approach cardiovascular diseases as the largest cause of death due to smoking. One million four hundred and seventy thousand neoplasm deaths ( $22 \%$ of all cancer deaths; 1.24 million or $33 \%$ of all adult male cancer deaths and 0.23 million or $8 \%$ of all adult female cancer deaths) were due to smoking, accounting for $30 \%$ of all smoking-attributable deaths ( $32 \%$ among men and $23 \%$ among women).

### 3.3 Mortality in industrialized countries

Table 11.13 provides the details of mortality due to smoking in industrialized countries by age, sex and causes of death. In the year 2000,
smoking caused an estimated 2.43 ( $95 \%$ CI 2.13-2.78) million deaths in industrialized countries for people aged $>30$ years, accounting for $19 \%$ of adult mortality. One million eight hundred and ten thousand ( $95 \%$ CI 1.62-2.02) deaths were among men ( $28 \%$ of total mortality of adult males) and 0.61 ( $95 \%$ CI $0.52-0.75$ ) million among women ( $10 \%$ of total mortality of adult females). The magnitude of the years of life lost due to premature mortality becomes more obvious when we note that 1.19 million, or approximately one half, of these deaths were among those aged 30-69 years.

In industrialized countries, smoking-caused deaths accounted for $33 \%$ of total mortality among males between the ages of 30 and 69 years ( 1.00 million deaths), $24 \%$ of total mortality among males aged $>70$ years ( 0.81 million deaths), $12 \%$ of total mortality among females between the ages of 30 and 69 years ( 0.19 million deaths), and $9 \%$ of total mortality among females aged $>70$ years ( 0.42 million deaths).

The fraction of smoking-attributable mortality among men was highest in the EUR-C and AMR-A subregions, causing 0.55 million and 0.35 million smoking-attributable male deaths, respectively. These were $32 \%$ and $28 \%$ of all adult male deaths ( $36 \%$ and $24 \%$ of all deaths for $30-69$ and $\geq 70$ age groups in EUR-C; $31 \%$ and $26 \%$ of all deaths for $30-69$ and $\geq 70$ age groups in AMR-A reflecting the fact that the two subregions are at different stages of the tobacco epidemic). Among women the highest fraction of smoking-attributable mortality was in AMR-A where 0.29 million deaths ( $22 \%$ of all mortality) $(27 \%$ and $20 \%$ of all deaths for the $30-69$ - and $\geq 70$-year age groups, respectively), were caused by smoking. The lowest fraction of smoking-attributable mortality among men in the industrialized world was in WPR-A ( $22 \%$ of all deaths; $18 \%$ and $24 \%$ of all deaths for $30-69-$ and $\geq 70$ year age groups, respectively) and among women in EUR-C (4\% of all deaths; $6 \%$ and $4 \%$ of all deaths for $30-69$ - and $\geq 70$-year age groups, respectively) and EUR-B ( $6 \%$ of all deaths; $10 \%$ and $5 \%$ of all deaths for the 30-69- and $\geq 70$-age groups, respectively).

For both males and females in all subregions and age groups, lung cancer was the cause of death with the largest fraction attributable to smoking, ranging from a low of $45 \%$ among females aged $\geq 30$ years in EUR-C to a high of $91-94 \%$ among males aged $\geq 30$ years in AMR-A, EUR-A, EUR-B and EUR-C. Overall $92 \%$ of all lung cancer deaths among adult ( $\geq 30$ years) males ( 0.40 million lung cancer deaths) and $71 \%$ of all lung cancer deaths among adult ( $\geq 30$ years) females ( 0.12 million lung cancer deaths) in industrialized countries were caused by smoking.

Despite the predominance of smoking as a cause, lung cancer accounted for only $22 \%$ ( 0.40 million) of smoking-attributable deaths among men and $19 \%$ ( 0.12 million) among women in industrialized countries. In fact, in terms of the fraction of all causes of death due
Table II.I3 Mortality due to smoking in industrialized countries by cause, sex and age, 2000

| Cause ${ }^{\text {a }}$ | Male |  |  |  | Female |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | No. of deaths (000s) | Fraction of total mortality (\%) | No. of deaths (000s) | Fraction of total mortality (\%) | No. of deaths (000s) | Fraction of total mortality (\%) | No. of deaths (000s) | Fraction of total mortality (\%) |
|  | 30-69 years |  | $\geq 70$ years |  | 30-69 years |  | $\geq 70$ years |  |
| Lung cancer | 216 | 91 | 187 | 92 | 50 | 70 | 67 | 72 |
| Upper aerodigestive cancer ${ }^{\text {b }}$ | 52 | 72 | 24 | 66 | 5 | 39 | 9 | 41 |
| Other cancer ${ }^{\text {b }}$ | 97 | 21 | 88 | 16 | 11 | 2 | 18 | 3 |
| COPD ${ }^{\text {c }}$ | 63 | 84 | 142 | 77 | 20 | 62 | 86 | 61 |
| Other respiratory diseases | 67 | 44 | 38 | 16 | 9 | 15 | 23 | 8 |
| Cardiovascular diseases | 455 | 40 | 298 | 17 | 77 | 13 | 192 | 7 |
| Other attributable medical causes | 55 | 32 | 33 | 13 | 17 | 14 | 28 | 7 |
| Other medical causes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total medical | 1005 | 39 | 810 | 24 | 189 | 13 | 423 | 9 |
| Non-medical (accidents and injuries) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| Total mortality | 1005 | 33 | 810 | 24 | 189 | 12 | 423 | 9 |
| a See Table II.I for details on causes of death. |  |  |  |  |  |  |  |  |
| c The estimates also include ICD 495, which was not included in the CPS-II relative risk estimates, due to GBD grouping; normally this cause w causes. |  |  |  |  |  |  |  |  |

to smoking, cardiovascular diseases led the grouping used in Table 11.1 and accounted for $42 \%$ and $44 \%$ of deaths caused by smoking among men and women, respectively ( 0.75 million male deaths and 0.27 million female deaths from cardiovascular diseases due to smoking). When all cancers are considered together, however, the numbers approached cardiovascular diseases for men. Neoplasm deaths accounted for $37 \%$ and $26 \%$ of all smoking caused deaths among men and women, respectively, in industrialized countries ( 0.66 million male deaths and 0.16 million female deaths from all cancers due to smoking accounting for $43 \%$ and $13 \%$ of all cancers among men and women, respectively).

### 3.4 Mortality in developing countries

Table 11.14 provides the details of mortality due to smoking in developing countries by age, sex and cause of death in 2000. The number of deaths attributable to smoking among people aged $>29$ years in developing countries in 2000 was 2.41 ( $95 \%$ CI 1.80-3.15) million accounting for $9 \%$ of total adult mortality in these countries. Of the smoking-attributable deaths, 2.02 ( $95 \%$ CI 1.56-2.50) million were among men ( $14 \%$ of total adult male mortality) and 0.38 ( $95 \%$ CI $0.25-0.65$ ) million among women ( $3 \%$ of total adult female mortality). About twice as many- 1.5 million deaths-deaths were among those between 30 and 69 years compared with 0.91 million among those aged $>69$ years.

In developing countries also, lung cancer was the disease with the highest fraction due to smoking. Fifty-five per cent of all lung cancers or 0.33 million deaths $(67 \%$ or 0.29 million deaths among men and $25 \%$ or 39000 deaths among women) were attributable to smoking. But lung cancer accounted for only $14 \%$ of all smoking-attributable mortality ( $14 \%$ among men and $10 \%$ among women) vs $21 \%$ in industrialized countries. As in the industrialized countries, cardiovascular diseases were the largest cause of death due to smoking, followed very closely by COPD. Six hundred and seventy thousand cardiovascular deaths (0.57 million among men and 97000 among women) were due to smoking, accounting for $28 \%$ of all smoking-attributable deaths ( $28 \%$ among men and $25 \%$ among women). Six hundred and fifty thousand COPD deaths ( 0.50 million among men and 0.16 million among women) were due to smoking accounting for $27 \%$ of all smoking-attributable deaths $(25 \%$ among men and $41 \%$ among women). This high contribution from COPD is consistent with direct observations of Liu et al. (1998) in China, where the high background (non-smoker) rates of COPD mortality due to other risk factors result in an even larger mortality due to smoking from this cause. Further, the lower contribution of cardiovascular diseases to smoking-caused mortality compared to the $42 \%$ in industrialized countries is likely to be due to lower overall cardiovascular disease mortality in these populations. Future changes in dietary risk factors and
Table II.I4 Mortality due to smoking in developing countries by cause, sex and age, 2000

| Cause ${ }^{\text {a }}$ | Male |  |  |  | Female |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | No. of deaths (000s) | Fraction of total mortality (\%) | No. of deaths (000s) | Fraction of total mortality (\%) | No. of deaths (000s) | Fraction of total mortality (\%) | No. of deaths (000s) | Fraction of total mortality (\%) |
|  | 30-69 years |  | $\geq 70$ years |  | 30-69 years |  | $\geq 70$ years |  |
| Lung cancer | 181 | 65 | 108 | 69 | 27 | 26 | 12 | 22 |
| Upper aerodigestive cancer ${ }^{\text {b }}$ | 100 | 38 | 42 | 35 | 12 | 10 | 6 | 7 |
| Other cancer ${ }^{\text {b }}$ | 98 | 11 | 48 | 9 | 7 | 1 | 6 | 1 |
| COPD ${ }^{\text {c }}$ | 206 | 49 | 290 | 45 | 65 | 20 | 93 | 12 |
| Other respiratory diseases | 207 | 18 | 55 | 9 | 26 | 4 | 9 | 1 |
| Cardiovascular diseases | 393 | 17 | 178 | 8 | 66 | 4 | 31 | 1 |
| Other attributable medical causes | 90 | 14 | 25 | 5 | 19 | 3 | 7 | 0 |
| Other medical causes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| Total medical | 1275 | 16 | 746 | 14 | 221 | 4 | 164 | 3 |
| Non-medical (accidents and injuries) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total mortality | 1275 | 14 | 746 | 14 | 221 | 3 | 164 | 3 |
| a See Table II.I for details on causes of death. <br> b The estimates of mortality for upper aerodigestive cancer include only ICD codes 140-I50. ICD code 161 is included with other cancers. <br> c The estimates also include ICD 495, which was not included in the CPS-II relative risk estimates, due to GBD grouping; normally this cause w causes. |  |  |  |  |  |  |  |  |

increased exposure to other cardiovascular disease risks in developing countries, even with similar attributable fractions, would result in a rise in tobacco-caused mortality in these subregions. As for the global total and in industrialized countries, when all cancers are considered together, they approach cardiovascular diseases and COPD as the largest cause of death due to smoking. Six hundred and fifty thousand neoplasm deaths ( $16 \%$ of all cancer deaths; 0.58 million or $26 \%$ of all adult male cancer deaths and 69000 or $4 \%$ of all adult female cancer deaths) were due to smoking, accounting for $27 \%$ of all smoking-attributable deaths $(29 \%$ among men and $18 \%$ among women).

In general, there was larger variation in mortality due to smoking among different subregions of developing countries than industrialized countries, because of the variability in the stages of the smoking epidemic. In the sections below, we discuss the estimates for different areas of developing countries, including comparisons with direct estimates where available.

## China

Liu et al. (1998) estimated that smoking caused 0.6 million deaths in China in 1990 and expected this number to rise to 0.8 million in 2000 if the fraction of mortality due to smoking remained unchanged. Since smoking had been rising rapidly in China since the 1970s, the fraction of deaths due to smoking was expected to rise, resulting in more deaths than the 0.8 million estimate. The 2000 estimates were based on a projected total adult ( $\geq 35$ years) mortality of 9 million in China in 2000. ${ }^{10}$

In fact, total adult mortality in China in 2000 from the GBD mortality database was 7.5 million deaths, reflecting recent health gains in China. Applying the 1990 cause-specific mortality and total mortality attributable fractions from Liu et al. (1998) to the 2000 mortality estimates in China would result in approximately 0.6 million smokingattributable deaths, with 0.51 million male deaths and 0.09 million female deaths. Actual mortality is expected to be higher, in particular among men, because of the rising trend of smoking (Corrao et al. 2000; WHO 1997). Further, since the proportional mortality analysis of Liu et al. (1998) cannot estimate mortality from those causes in the reference group, smoking-attributable mortality is underestimated (to zero) for these causes, to the extent that smoking caused some deaths from these diseases.

As described above in the section on methods, in estimating smoking-attributable mortality in China, we converted the 1990 relative risks for smoking estimated in Liu et al. (1998) to relative risks per unit of SIR. We then estimated the SIR for in China in 2000 to capture the impacts of the more recent increase in smoking. Mortality was then estimated by applying the relative risk estimates from 1990 (with some correction for potential confounding) to 2000 SIR estimates. The
exception is for diseases in the category "other medical causes" in Table 11.1, for which proportional mortality analysis does not estimate relative risks. For these causes, we used the relative risks of CPS-II from Table 11.1.

Using this method for China we estimated that smoking caused 0.59 million deaths, 0.49 million of which are among men ( $13 \%$ of total adult male mortality) and the remaining 0.10 million among women ( $3 \%$ of total adult female mortality). In China, deaths caused by smoking accounted for $11 \%$ of total mortality among males between the ages of 30 and 69 years ( 0.22 million deaths), $15 \%$ of total mortality among males aged $>69$ years ( 0.27 million deaths), $2 \%$ of total mortality among females between the ages of 35 and 69 years ( 32000 deaths), and 3\% of total mortality among females aged $>69$ years ( 66000 deaths).

The relatively small fraction of mortality among women due to smoking remained constant between 1990 and 2000, reflecting the stable, low prevalence of smoking among Chinese women. The fraction of mortality among men dropped for the 35-69-year age group from $13 \%$ to $11 \%$, and increased from $12 \%$ to $15 \%$ for the $\geq 70$-year age group. The increase among the older men is due to demographic shifts in smoking patterns. Tobacco consumption in China increased between 1970 and 1990 and then stabilized (Corrao et al. 2000; WHO 1997). Therefore, while many of the younger smokers in 1990 and in 2000 had started smoking around the same age, those in older age groups in 2000 are likely to have smoked for a longer time than those of similar age in 1990. This slight increase in the fraction of mortality attributable to smoking among the older age cohorts with the maturity of the smoking epidemic is consistent with historical trends in age-specific attributable fractions in Canada, the United Kingdom and the United States. The decline among the younger age groups was partially due to the $5 \%$ correction factor applied to the hazards for this country.

Lung cancer accounted for $20 \%$ and $18 \%$ of smoking-attributable mortality among men and women, respectively, in China, resulting in 98000 male deaths and 17000 female deaths. The fractions of smokingattributable mortality from all cancers were $44 \%$ ( 215000 deaths) for males and $29 \%$ (28000 deaths) for females. This suggests that the contribution of smoking to mortality from cancers other than lung cancer is larger in China than in industrialized countries (Lopez 1998). In China, COPD also accounted for a large fraction of smoking-attributable mortality with $33 \%$ ( 163000 deaths) and $61 \%$ ( 60000 deaths) of smokingcaused mortality among men and women respectively.

INDIA AND SEAR-D ${ }^{11}$
Seven hundred and fifty thousand deaths among adult men and 110000 among adult women were attributable to smoking in SEAR-D accounting for $18 \%$ and $3 \%$ of total mortality, respectively. Six hundred and thirty thousand of these deaths occured before the age of $70(13 \%$
of total mortality) and the remaining 230000 among those aged $\geq 70$ years ( $8 \%$ of total mortality).

Adult male lung cancer mortality, $82 \%$ of which ( 84000 deaths) in this subregion is caused by smoking, accounted for $11 \%$ of smokingattributable male mortality, the lowest fraction among males in any subregion. The 6000 lung cancer deaths among females attributed to smoking ( $26 \%$ of all female lung cancer deaths) was likewise the smallest fraction ( $6 \%$ ) of smoking-attributable deaths compared with other subregions. When all cancers are considered together, smoking caused 0.18 million neoplasm deaths in SEAR-D in 2000 ( 160000 among men and 14000 among women). Cardiovascular diseases, with 0.28 million deaths ( 240000 or $16 \%$ of all cardiovascular deaths among men and 34000 or $2 \%$ of all cardiovascular deaths among women), were the cause of death with the highest number due to smoking and accounted for $33 \%$ of all smoking-attributable deaths ( $33 \%$ among men and $31 \%$ among women), reflecting the large contribution of this cause to adult male mortality in this subregion.

No direct nationally representative study of mortality due to smoking was available from India or other countries in SEAR-D at the time of writing. Estimates from specific regions within India as well as indirect national estimates, however, are available and can be used for comparison with our results. Gupta and Mehta (2000) estimated that the relative risk for mortality from all causes in a mixed cohort of male smokers and non-smokers aged >34 years in India relative to non-users of tobacco is approximately 1.63. Assuming, as in the case of CPS-II relative risks, that $30 \%$ of the excess risk is due to confounding (because of covariates such as chewing tobacco, diet, etc.) this relative risk and a smoking prevalence of $40-50 \%$ imply that $15-18 \%$ of all male mortality is due to smoking, a result consistent with our estimate of $18 \%$ of male mortality attributable to tobacco. Applying the relative risk of 2.1 obtained by Gajalakshmi et al. (2003) with the same correction factor will result in an even higher attributable fraction (23-28\%) than ours. The estimates of total mortality in this chapter are lower than those by Gupta (1989) who attributes at least $19 \%$ of adult male mortality and $4 \%$ of adult female mortality to tobacco use, as well as those in a recent case-control study that finds an unadjusted attributable fraction of approximately $20 \%$ for all adult deaths among Indian men (Gajalakshmi et al. 2003).

## Additional oral cancer mortality due to oral tobacco use

Using the methods described above, we estimated that there were 60000 additional cases of oral cancer due to oral tobacco use (tobacco chewing) in SEAR-D, accounting for an additional $50 \%$ of oral cancers in the subregion. Of these, 39000 were among men ( $48 \%$ of male oral cancer deaths) and 22000 among women ( $54 \%$ of female oral cancer
deaths). We emphasize that these estimates are those cases of oral cancer that are caused by tobacco chewing in addition to smoking. Since many cases of oral cancer are likely to be affected by both habits, the overall effects of tobacco chewing are larger. Important sources of uncertainty in the estimates are the prevalence of oral tobacco use and relative risk estimates as well as the extent of overlap between tobacco smoking and chewing, especially for men.

## Other developing countries

The fraction of total adult mortality due to smoking ranged from a low of $2-4 \%$ in AFR-D, AMR-D and AFR-E to a high of $11 \%$ in SEAR-B and $18 \%$ in WPR-B (excluding China). For males the lowest fraction of total mortality due to smoking was in AMR-D (3\%), AFR-D (5\%) and AFR-E ( $6 \%$ ), reflecting the more recent smoking epidemic in these subregions. Given that the current prevalence of smoking among adult men is approximately $25-30 \%$ in AFR-D and $35-45 \%$ in AFR-E, this finding emphasizes the fact that current prevalence is a poor marker of accumulated smoking risks (see also Figure 11.4).

The highest fractions of adult male mortality due to smoking were in WPR-B (excluding china) ( $26 \%$ ), SEAR-B ( $19 \%$ ), EMR-B ( $15 \%$ ) and AMR-B $(15 \%)$. For females, the fraction of total mortality due to smoking in 2000 was equal to or below $2 \%$ in AFR-D, AFR-E, AMRD, EMR-D and SEAR-B. The highest fractions of female mortality were in AMR-B ( $6 \%$ ) and WPR-B (excluding China) ( $8 \%$ ), reflecting more recent increases in female smoking in these subregions, especially with increasing urbanization and economic development.

## 4. Discussion

We applied the indirect method of Peto et al. (1992), which uses absolute lung cancer mortality in a population as a marker for accumulated hazards of smoking, to estimate the mortality and disease burden due to smoking in different subregions of the world. We chose the parameters of the model, such as relative risks and non-smoker lung cancer mortality, based on direct estimates or by extrapolation from other subregions based on best available evidence, explicitly stating the assumptions and reasons for each choice.

Using this method, we estimated that in 2000, approximately 4.83 ( $95 \%$ CI $3.94-5.93$ ) million deaths worldwide were due to smoking, accounting for $12 \%$ of global adult mortality. Of these deaths, 2.41 ( $95 \%$ CI $1.80-3.15$ ) million were in developing countries, marking a transition to an era in which smoking killed as many people in developing countries as in industrialized nations. In fact, even in the earlier stages of the tobacco epidemic, more men died from smoking in developing countries than in the industrialized nations ( 2.02 million vs 1.81
million). In addition to those cases shared with smokers, there were an estimated 60000 deaths from oral tobacco use in SEAR-D. Premature mortality and morbidity caused by smoking accounted for an estimated $4.1 \%$ of the global burden of disease.

As we discussed under sources of uncertainty, using lung cancerwhich has a longer lag than cardiovascular diseases-as the marker for accumulated smoking hazard, would result in an overestimation of hazard where there has been sharp drops in smoking and underestimation of hazard where there has been large increases in smoking. The former is most likely to apply to North America and among males in some countries in western Europe where smoking has declined (partially or fully offset by a continued choice of conservative relative risk estimates). On the other hand, the underestimation scenario would be applicable to most developing countries where smoking has been on the rise in the past few decades.

Total male mortality in terms of numbers of deaths was considerably higher than female mortality- 3.0 fold in industrialized nations and 5.3 fold in developing countries. The decline of the male-to-female mortality ratio from 3.6 to 3.0 in industrialized countries between 1990 and 2000, however, reflects the recent relative increases in female smoking in these countries. 2.69 million deaths, more than one-half of the all global deaths due to smoking, were in the 30-69-year age group.

Due to differences in methodology and presentation, the estimates reported here are not fully comparable with those for previous years. The existing estimates of consumption, prevalence and mortality, however, generally indicate that mortality due to smoking (in terms of the fraction of cause-specific or all-cause mortality) has been relatively stable in industrialized countries over the past ten years. Some countries in the established market economies category have seen a small decline in male mortality while in most of these countries female mortality has increased, reflecting differential time trends in male and female smoking. Industrialized countries also have seen a small decline in the fraction of mortality in the $30-69$-year age group and an increase in the $\geq 70$-year group, confirming that in these countries as a whole, the smoking epidemic may be shifting with the effects increasingly being felt among the older age groups. There were, nonetheless, subregional differences, and the share of mortality at ages 30-69-years is in fact rising among females in some subregions including AMR-A and EUR-A.

Mortality and disease burden attributable to smoking, including its share of total mortality and sex or age patterns, varied importantly among different geographical regions of developing countries. This interregional variation, which is larger than that observed in industrialized countries, occurs because the nature and maturity of the smoking epidemic is highly affected by the varying economic and cultural determinants of smoking in these populations. A few general statements can nonetheless be made about the health effects of smoking in developing
countries. First, current hazards of smoking in these populations are highly concentrated among men. Given that the prevalence of smoking among women is still low in developing countries (with the exception of Latin America and the Caribbean and some countries in Asia), the current level of male mortality should provide an indicator of the large health losses that may well occur if female smoking increases over the next few decades. Second, relative to industrialized countries, developing countries have a higher proportion of smoking-attributable mortality in the $30-69$ age group than in the $\geq 70$ group ( $62 \%$ in developing countries vs $49 \%$ in industrialized countries). Coupled with the 1990-2000 trends in mortality for the two age groups in China, this suggests that as people (mostly men) who began smoking over the past three decades in developing countries become older, mortality due to smoking will continue to rise as a share of cause-specific mortality; and almost inevitably as a share of total mortality.

Smoking prevalence in some developing countries appears to have stabilized, albeit at very high levels. In others, it is still rising. Given the gradually shifting disease patterns and because most of the growth in global population is expected to take place in the developing world, the health effects of smoking, already one of the most important global health hazards, will continue to rise unless effective interventions and policies that curb and reduce smoking among males and prevent increases among females in these countries are implemented.

## 5. Projected future exposure

Many diseases caused by smoking, in particular various malignant neoplasms and COPD, occur after long delays. This motivated using SIR as the exposure variable for estimating the accumulated hazards of smoking. Disease burden due to smoking in the next few decades will depend on both past and future smoking patterns. There is therefore a need to link estimates of accumulated current exposure, which are in the form of SIR estimates, with future exposure, which is often in the form of projections of prevalence of smoking or tobacco consumption (whether under the business-as-usual scenario or some counterfactual). Further, the combination of past and projected future exposure must be presented in the form of a single exposure variable which accounts for hazard accumulation. We used the following steps to estimate future smoking prevalence and tobacco consumption and convert these to estimates of lung cancer mortality and SIR.

1. We estimated past and current age-sex-specific smoking prevalence under the business-as-usual scenario based on a descriptive model of the smoking epidemic (Lopez et al. 1994), calibrated to regional characteristics of the epidemic for different subregions. The tobacco epidemic was divided into five stages (early, rising, peak or maturity, declining and late)
as well as five transitional stages as seen in Figure 11.7. Historical evidence from multiple industrialized countries consistently shows that the youngest and oldest age groups at any stage of the epidemic have lower smoking prevalence (Gajalakshmi et al. 2000; Nicolaides-Bouman et al. 1993), the former probably because of economic and social constraints and the latter because of a higher mortality rate among smokers. The prevalence distribution is also based on the observation that in the rising stages of the epidemic, younger adults begin to smoke more than the older adults but as the epidemic matures the age-pattern becomes more stable (Gajalakshmi et al. 2000). Finally, as observed in historical data we assumed that the prevalence of female smoking is lower than that of males at every stage of the epidemic.

In calibrating the level of the prevalence curves for different regions, we used the available comparative data on current and historical smoking prevalence (Corrao et al. 2000; Gajalakshmi et al. 2000; WHO 1997). Recent data on smoking prevalence show that the male epidemic has peaked at lower levels in Latin America and the Caribbean as well as in sub-Saharan Africa compared to North America, Europe and Asia, possibly because of the economic crises in these regions in the last two decades of the 20th century. Although for most regions female smoking is still in the early stages of the epidemic, we assumed that the peak of the epidemic would be lower in developing countries than what was observed in industrialized countries in the past. This assumption was because in those developing countries where female smoking has been on the rise, prevalence is lower than levels previously observed at similar stages in industrialized countries.

In each country, males and females were assigned to one stage of the epidemic in 2000 (Table 11.15) based on the best available data on smoking patterns in recent years (Corrao et al. 2000; WHO 1997).
2. We divided country-level per capita consumption data into age-sexspecific per capita consumption of smoking based on the above estimates of prevalence. The Tobacco Free Initiative (TFI) of WHO provides timeseries estimates of tobacco consumption based on production, export, and import data from the Food and Agriculture Organization of the United Nations (FAO). The consumption numbers are reported as equivalent cigarettes per adult (aged $\geq 15$ years) for each country. We used the above estimates of prevalence to divide the past and current countrylevel per capita consumption numbers (corrected for smuggling and other sources of error whenever possible) into age-sex-specific per capita consumption.

In addition to differences in prevalence, male smokers smoke more cigarettes per day than female smokers (Gajalakshmi et al. 2000; Nicolaides-Bouman et al. 1993). Differences between age groups also exist (Gajalakshmi et al. 2000; Nicolaides-Bouman et al. 1993). The

Figure II. 7 Descriptive model for the main stages of the tobacco epidemic based on the parameter values in industrialized countries


Female


Note: For some developing regions, lower peak values were chosen based on observed prevalence data. The vertical axis scales are different for males and females to increase resolution. The number next to each curve indicates the epidemic stage. An additional transitional stage is also assumed between each pair. With appropriate policies, one can assume declining prevalence even beyond stage 5 . We assumed that the business-as-usual scenario would not include this achievement, which is considered as a part of our counterfactuals.

Table II.I5 Status of the tobacco epidemic in 2000 among (a) males and (b) females
(a)

| Epidemic stage ${ }^{\text {a }}$ | Country ${ }^{\text {b }}$ |
| :---: | :---: |
| 0.5-I | NA |
| 1.5-2 | AFR-D - all except Algeria, Mauritius, Seychelles; AFR-E - all except South Africa; EMR-B - Iran (Islamic Republic of); EMR-D - Afghanistan, Djibouti, Somalia, Sudan; EUR-B - Azerbaijan, Tajikistan, Turkmenistan, Uzbekistan |
| 2.5-3 | AFR-D - Algeria, Mauritius, Seychelles; AFR-E - South Africa; AMR-A - Cuba; AMR-B - all except Argentina, Brazil, Chile; AMR-D - all; EMR-B - all except Iran (Islamic Republic of); EMR-D - Egypt, Iraq, Morocco, Pakistan, Yemen; EUR-A - Croatia, Czech Republic, Greece, Portugal, Slovenia; EUR-B - all except Azerbaijan, Tajikistan, Turkmenistan, Uzbekistan; EUR-C - all; SEAR-B all; SEAR-D - all; WPR-A - Brunei Darussalam, Japan; WPR-B - all |
| 3.5-4 | AMR-B - Argentina, Brazil, Chile; EUR-A - Andorra, Austria, Denmark, France, Germany, Ireland, Israel, Italy, Luxembourg, Malta, Monaco, San Marino, Spain, Switzerland |
| 4.5-5 | AMR-A - Canada, USA; EUR-A - Belgium, Finland, Iceland, Netherlands, Norway, Sweden, United Kingdom; WPR-A - Australia, New Zealand, Singapore |

(b)

| Epidemic stage ${ }^{\text {a }}$ | Country ${ }^{\text {b }}$ |
| :--- | :--- |
| 0.5-I | AFR-D - all except Seychelles; AFR-E - all except South Africa; AMR-B - |
|  | Antigua and Barbuda, Bahamas, Barbados, Belize, Dominica, Grenada, Guyana, |
|  | Paraguay, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and Grenadines, |
|  | Suriname; EMR-B - all except Cyprus, Jordan, Lebanon, Syrian Arab Republic; |
|  | EMR-D - all except Morocco; Albania, Azerbaijan, Tajikistan, Turkmenistan, |
|  | Uzbekistan; SEAR-B - Indonesia, Sri Lanka; SEAR-D - all except Myanmar, |
|  | Nepal; WPR-A - Singapore; WPR-B - Cambodia, China, Malaysia, Mongolia, |
|  | Republic of Korea, Viet Nam |

NA Not applicable.
a The stages of the epidemic refer to those in Figure II.7.
b By subregion.

Table II.I6 Ratios of number of cigarettes smoked per day for various demographic groups from historical data in industrialized countries

|  | Ratios of number of cigarettes smoked |  |
| :--- | :--- | ---: |
| Age group (years) | Male | Female |
| $15-19$ | 1.00 | 1.00 |
| $20-29$ | 1.22 | 1.15 |
| $30-39$ | 1.42 | 1.24 |
| $40-49$ | 1.37 | 1.34 |
| $50-59$ | 1.26 | 1.21 |
| $60-69$ | 1.15 | 1.04 |
| $70-79$ | 1.15 | 1.04 |
| $\geq 80$ | 1.15 | 1.04 |
| Male | 1.0 |  |
| Female | $1 / 1.25$ (stages 4 and 5); ${ }^{\text {a }} \mathrm{I}$ I/I.5 (stage 3); |  |

a The stages of the epidemic refer to those in Figure II.7.
ratios for the number of cigarettes smoked per day between different age groups and men and women have been estimated for industrialized countries (Gajalakshmi et al. 2000) and are presented in Table 11.16. We assumed a lower female-to-male ratio in the earlier stages of the female epidemic.

We emphasize that because the prevalence estimates from the descriptive model are used to divide existing total per capita consumption into age-sex-specific estimates, it is only the various male-female and age ratios that affect the estimates, rather than the absolute values. Therefore, the age-sex-specific consumption estimates are not sensitive to the level of the prevalence curve or assumptions about the stage of the epidemic, provided that male-to-female and age ratios are close to actual values (i.e. under- or over-estimating male and female prevalence by the same factor would not affect the consumption estimates).
3. We projected age-sex-specific lung cancer mortality based on consumption projections and a statistical model of the relationship between lung cancer mortality and lagged consumption (Girosi and King 2002). Finally, the projections of lung cancer mortality were converted to the projections of SIR using the definition of SIR above.

## Acknowledgements

We thank R. Peto and J. Boreham for discussions on methodology and for providing data from the Chinese retrospective proportional mortal-
ity study; M. Thun and J. Henley for discussions on methodology and for additional re-analysis of data from the American Cancer Society CPSII cohort; S. Darby for data on lung cancer risk reversibility in the United Kingdom; E. Guindon and P. Jha for data on the prevalence of smoking and tobacco consumption; and P. Gupta for discussions on methods and data for oral tobacco use risk. The generous assistance of these individuals and their organizations greatly improved the data sources used in this chapter. The participants in the comparative risk assessment (CRA) project review meetings and anonymous reviewers provided valuable comments on methods and assumptions.

## Notes

1 See preface for an explanation of this term.
2 If the fraction of smokers in the mixture is $x$, then the lung cancer mortality for the mixture is $x S_{\dot{L} C}^{*}+(1-x) N_{\tilde{L} C}^{*}$. Since the SIR of the mixture has to equal that of the population, excess lung cancer mortality in the mixture has to be equal to that of the study population. Therefore $x S_{\tilde{L C}}^{*}+(1-x) N_{\tilde{L} C}^{*}-$ $N_{\tilde{L} C}^{*}=C_{L C}-N_{L C}$. Solving this equation gives $x=\frac{C_{L C}-N_{L C}}{S_{L C}^{*}-N_{L C}^{*}}$ which is equal to SIR when the study and reference populations have the same non-smoker lung cancer rates. When non-smoker lung cancer rates are not the same in the study and reference populations, the same result can be obtained with algebraic manipulation of Equation 2, accounting for differences in neversmoker rates.

3 The discussion in this section is based on age-specific risk estimates. In other words, a decline in risk as a result of cessation does not necessarily imply that risk stops rising in absolute terms. Rather, it implies that at all ages after cessation, relative risk is less than it would be if smoking continued.
4 These estimates also assume no new smokers in the cohort.
5 No disease with this latter characteristic is known.
6 Note that this applies only to current exposure and not to future exposure (avoidable risk) since the difference in decline time is accounted for in estimating risk reversibility, as described earlier.
7 Note that the estimates of oral cancer for smokers are based on SIR as the exposure variable to capture the accumulated hazards of smoking and not on prevalence. Therefore the fraction of smokers who also chew ( $p_{s c}=1-p_{s}$ ) was applied to SIR estimates rather than direct prevalence. Estimating chewing-caused oral cancer by increasing the relative risk for this condition in the SIR framework implicitly assumes that the accumulated hazards of the two habits are similar. If oral tobacco use has a longer history in the region, this would result in an underestimation of accumulated hazards.
8 Given that SIR values are estimated relative to reference populations of the same sex and age, female and male estimates are not directly comparable. At the same time, the large differences are illustrative of the relative magnitudes.

9 All fractions are based on mortality in the respective age groups $30-69, \geq 70$, and $\geq 30$ years.
10 Mortality estimates for China were combined with the remaining countries in WPR-B to obtain subregional estimates.

11 Eighty-three per cent of the population of SEAR-D lives in India. Therefore the estimates for the subregion are dominated by, and comparable in terms of fractions with, those from India.

## References

Alderson MR, Lee PN, Wang R (1985) Risks of lung cancer, chronic bronchitis, ischaemic heart disease, and stroke in relation to type of cigarettes smoked. Journal of Epidemiology and Community Health, 39:286-293.
Australia National Health and Medical Research Council (1997) The health effects of passive smoking. Commonwealth Department of Health and Family Services, Canberra.
Balaram P, Sridhar H, Rajkumar T et al. (2002) Oral cancer in southern India: the influence of smoking, drinking, paan-chewing and oral hygiene. International Journal of Cancer, 98:440-445.
Bennett NG, Horiuchi S (1984) Mortality estimation from registered deaths in less developed countries. Demography, 21:217-234.
Best EWR, Josie GH, Walker CB (1961) A Canadian study of mortality in relation to smoking habits: a preliminary report. Canadian Journal of Public Health, 52:99-106.
Bhonsle RB, Murti PR, Gupta PC (1992) Tobacco habits in India. In: Control of tobacco related cancers and other diseases. Gupta PC, Hamner JE, Murti PR, eds. Oxford University Press, Bombay.
Britton A, McKee M (2000) The relationship between alcohol and cardiovascular disease in eastern Europe: explaining the paradox. Journal of Epidemiology and Community Health, 54:328-332.
Bronnum-Hansen H, Juel K (2000) Estimating mortality due to cigarette smoking: two methods, same result. Epidemiology, 11:422-426.
Bruce N, Perez-Padilla R, Albalak R (2000) Indoor air pollution in developing countries: a major environmental and public health challenge. Bulletin of the World Health Organization, 78:1078-1092.
Cederlof R, Friberg L, Hrubec Z, Lorich U (1975) The relationship of smoking and some social covariables to mortality and cancer morbidity. Department of Environmental Hygiene, Karolinska Institute, Stockholm.
Cook DG, Pocock SJ, Shaper AG, Kussick SJ (1986) Giving up smoking and the risk of heart attacks: a report from the British Regional Heart Study. The Lancet, ii(8520):1376-1380.
Corrao MA, Guindon GE, Sharma N, Shokoohi DF, eds. (2000) Tobacco control: country profiles. American Cancer Society, Atlanta, GA.
Dhillon I, Gupta PC, Asma S (2000) Evidence for a causal link between smoking and tuberculosis: summary of the proceedings of the international scientific
expert meeting on the possible causality between smoking and tuberculosis. TATA Institute of Fundamental Research (with World Health Organization and Centers for Disease Control and Prevention). Thiruvanathapuram, Kerala.

Dikshit RP, Kanhere S (2000) Tobacco habits and risk of lung, oropharyngeal and oral cavity cancer: a population-based case-control study in Bhopal, India. International Journal of Epidemiology, 29:609-614.
Dobson AJ, Alexander HM, Heller RF, Lloyd DM (1991) How soon after quitting smoking does risk of heart disease decline? Journal of Clinical Epidemiology, 44:1247-1253.

Doll R (1978) Atmospheric pollution and lung cancer. Environmental Health Perspectives, 22:23-31.
Doll R (1986) Tobacco: an overview of health effects. In: Tobacco: a major international health hazard. (IARC Scientific Publication No. 74.) Zaridze DG, Peto R, eds. International Agency for Research on Cancer, Lyon.
Doll R (1998a) Diseases with lower risks in smokers. In: Report of the Scientific Committee on Tobacco and Health. UK Department of Health. HMSO, London.
Doll R (1998b) Uncovering the effects of smoking: historical perspective. Statistical Methods in Medical Research, 7:87-117.
Doll R, Hill AB (1950) Smoking and carcinoma of the lung. British Medical Journal, ii:739.
Doll R, Hill AB (1956) Lung cancer and other causes of death in relation to smoking. British Medical Journal, ii:1071-1081.

Doll R, Peto R (1976) Mortality in relation to smoking: 20 years' observation on male British doctors. British Medical Journal, ii:1525-1536.

Doll R, Peto R, Wheatly K, Gray R, Sutherland I (1994) Mortality in relations to smoking: 40 years' observation on male British doctors. British Medical Journal, 309:901-911.

Du YX, Cha Q, Chen XW et al. (1996) An epidemiological study of risk factors for lung cancer in Guangzhou, China. Lung Cancer, 14:S9-S37.
Dunn JE, Linden G, Brelow L (1960) Lung cancer mortality experience of men of certain occupations in California. American Journal of Public Health, 50:1475-1487.

English DR, Holman CDJ, Milne E et al. (1995) The quantification of drug caused morbidity and mortality in Australia, Part 2. Commonwealth Department of Human Services and Health, Canberra.

Environmental Protection Agency (1992) Respiratory health effects of passive smoking: lung cancer and other disorders. (EPA/600/6-90/006F.) U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

Ezzati M, Lopez AD (2003) Measuring the accumulated hazards of smoking: global and regional estimates for 2000. Tobacco Control, 12:79-85.

Finkel A (1990) Confronting uncertainty in risk management: a guide for decision makers. Resources for the Future, Washington, DC.

Flanders WD, Rothman KJ (1982) Interaction of alcohol and tobacco in laryngeal cancer. American Journal of Epidemiology, 115:371-379.
Fletcher CM, Peto R (1977) The natural history of chronic airflow obstruction. British Medical Journal, i:1645-1648.
Gajalakshmi CK, Jha P, Ranson K, Nguyen SN (2000) Global patterns of smoking and smoking-attributable mortality patterns. In: Tobacco control in developing countries. Jha P, Chaloupka FJ, eds. Oxford University Press, Oxford.
Gajalakshmi V, Peto R, Kanaka TS, Jha P (2003) Smoking and mortality from tuberculosis and other diseases in India: retrospective study of 43000 adult deaths and 35000 controls. Lancet, 363:507-515.

Garfinkel L (1985) Selection, follow-up, and analysis in the American Cancer Society prospective studies. In Selection, follow-up, and analysis in prospective studies: a workshop. (NCI Monograph 67.) Garfinkel L, Ochs O, Mushinkski M, eds. National Cancer Institute, Bethesda, MD.
Girosi F, King G (2002) Time series cross-sectional analysis with different explanatory variables in each cross-section. Center for Basic Research in the Social Sciences, Harvard University, Cambridge, MA.
Gupta PC (1989) An assessment of excess mortality caused by tobacco usage in India. (Proceedings of the UICC workshop "Tobacco or health".) In: Tobacco and health: the Indian scene. Sanghavi LD, Notani P, eds. Tata Memorial Centre, Bombay.
Gupta PC (1996) Survey of sociodemographic characteristics of tobacco use among 99598 individuals in Bombay, India using handheld computers. Tobacco Control, 5:114-120.

Gupta PC, Mehta HC (2000) Cohort study of all-cause mortality among tobacco users in Mumbai, India. Bulletin of the World Health Organization, 78:877-883.
Gupta PC, Pindborg JJ, Mehta FS (1982) Comparison of carcinogenecity of betel quid with and without tobacco: an epidemiological review. Ecology of Disease, 1:213-219.

Hackshaw AK, Law MR, Wald NJ (1997) The accumulated evidence on lung cancer and environmental tobacco smoke. British Medical Journal, 315:980-988.

Hammond EC (1966) Smoking in relation to the death rates of one million men and women. In: Epidemiological approaches to the study of cancer and other chronic disease. (National Cancer Institute Monograph No. 19.) Haenszel W, ed. National Cancer Institute, Bethesda, MD.
Hammond EC, Horn D (1958) Smoking and death rates: report of forty-four months of follow-up of 187783 men. Journal of American Medical Association, 166:1159-1172 and 1294-1308.

He XZ, Chen W, Liu ZY, Chapman RS (1991) An epidemiological study of lung cancer in Xuan Wei county, China: current progress. Case-control study on lung cancer and cooking fuel. Environmental Health Perspectives, 94:9-13.

Hecht SS (2003) Tobacco carcinogens, their biomarkers and tobacco-induced cancer. Nature Reviews Cancer, 3:733-744.

Higgins ITT, Mahan CM, Wynder EL (1988) Lung cancer among cigar and pipe smokers. Preventive Medicine, 17:116-128.

Hill K (1981) Estimating census and death registration completeness. Asian and Pacific Population Forum, 1:8-13.
Hirayama T (1977) Smoking and cancer: a prospective study on cancer epidemiology based on census population in Japan. Health Consequences, Education, Cessation Activities, and Governmental Action. Smoking and Health II. (Proceedings of the Third World Conference on Smoking and Health.) DHEW Publication No. [NIH] 77-1413. U.S. Government Printing Office, Washington, DC.
Hirayama T (1990) Life-style and mortality: a large-scale census-based cohort study in Japan, Karger, Tokyo.
International Agency for Research on Cancer (1984) IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Tobacco habits other than smoking; betel-quid and areca-nut chewing; and some related nitrosomines. International Agency for Research on Cancer, Lyon, France.
Jedrychowski W, Becher H, Wahrendorf J, Basa-Cierpialek Z (1990) A case-control study of lung cancer with special reference to the effect of air pollution in Poland. Journal of Epidemiology and Community Health, 44:114-120.

Jee SH, Suh I, Kim IS, Appel LJ (1999) Smoking and atherosclerotic cardiovascular disease in men with low levels of serum cholesterol: the Korea Medical Insurance Corporation study. Journal of the American Medical Association, 282:2149-2155.

Jha P, Ranson MK, Nguyen SN, Yach D (2002) Estimates of global and regional smoking prevalence in 1995 by age and gender. American Journal of Public Health, 92:1002-1006.
Kahn HA (1966) The Dorn study of smoking and mortality among U.S. veterans: report on eight and one half years of observation. In: Epidemiological approaches to the study of cancer and other chronic disease. (National Cancer Institute Monograph No. 19.) Haenszel W, ed. National Cancer Institute, Bethesda, MD.
Kawachi I, Colditz GA, Stampfer MJ et al. (1997) Smoking cessation and decreased risks of total mortality, stroke, and coronary heart disease incidence among women: a prospective cohort study. In: Changes in cigarette-related disease risks and their implications for prevention and control. (Smoking and Tobacco Control Monograph No. 8.) National Cancer Institute, ed. National Cancer Institute, Bethesda, MD.

Kinjo Y, Cui Y, Akiba S et al. (1998) Mortality risks of oesophageal cancer associated with hot tea, alcohol, tobacco and diet in Japan. Journal of Epidemiology, 8:235-243.

Ko YC, Lee CH, Chen MJ et al. (1997) Risk factors for primary lung cancer among non-smoking women in Taiwan. International Journal of Epidemiology, 26:24-31.
LaCroix AZ, Lang J, Scherr P et al. (1991) Smoking and mortality among older men and women in three communities. New England Journal of Medicine, 324:1619-1625.

Lam TH, Ho SY, Hedley AJ, Mak KH, Peto R (2001) Mortality and smoking in Hong Kong: Case-control study of all adult deaths in 1998. British Medical Journal, 323:361-363.
Law MR, Morris JK, Wald NJ (1997) Environmental tobacco smoke exposure and ischaemic heart disease: an evaluation of the evidence. British Medical Journal, 315:973-980.

Lee PN (1996) Mortality from tobacco in developed countries: are indirect estimates reliable? Regulatory Toxicology and Pharmacology, 24:60-68.

Levin ML, Goldstein H, Gerhardt PR (1950) Cancer and tobacco smoking. Journal of American Medical Association, 143:336-338.

Liu BQ, Peto R, Chen ZM et al. (1998) Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. British Medical Journal, 317:1411-1422.
Lopez AD (1998) Counting the dead in China. British Medical Journal, 317:1399.
Lopez AD (1999) Alcohol and smoking as risk factors. In: Health and mortality: issues of global concern. (Proceedings of the United Nations Symposium on Health and Mortality, Brussels, 19-22 November, 1997.) Chamie J, Cliquet RL, eds. United Nations Population Division, New York.
Lopez AD, Ahmad OB, Guillot M et al. (2002) World mortality in 2002: life tables for 191 countries. World Health Organization, Geneva.
Lopez AD, Collishaw NE, Piha T (1994) A descriptive model of the cigarette epidemic in developed countries. Tobacco Control, 3:242-247.

Lubin JH, Blot WJ, Berrino F et al. (1984) Modifying risk of developing lung cancer by changing habits of cigarette smoking. British Medical Journal, 288:1953-1956.

Malarcher AM, Schuman J, Epstein LA et al. (2000) Methodological issues in estimating smoking-attributable mortality in the United States. American Journal of Epidemiology, 152:573-584.
Mathers CD, Stein C, Ma Fat D et al. (2002) The global burden of disease 2000: version 2, methods and results. (GPE Discussion Paper No. 50.) World Health Organization, Global Programme on Evidence for Health Policy, Geneva. (Available at http://www.who.int/evidence).
Mills CA, Porter MM (1950) Tobacco smoking habits and cancer of the mouth and respiratory system. Cancer Research, 10:539-542.

Moolgavkar SH, Venzon DJ (1987) General relative risk regression models for epidemiologic studies. American Journal of Epidemiology, 126:949-961.
Muller FH (1939) Tobacco abuse and carcinoma of the lung. Zeitschrift fur Krebsforschung, 49:57-85.

Nandakumar A, Thimmasetty KT, Sreeramareddy NM et al. (1990) A popula-tion-based case-control investigation on cancers of the oral cavity in Bangalore, India. British Journal of Cancer, 62:847-851.
National Research Council (1994) Science and judgment in risk assessment, National Academy Press, Washington, DC.
Nicolaides-Bouman A, Wald N, Forey B, Lee P, eds. (1993) International smoking statistics: a collection of historical data from 22 economically developed countries. The Wolfson Institute of Preventive Medicine and Oxford University Press, London and Oxford.
Niu SR, Yang GH, Chen ZM et al. (1998) Emerging tobacco hazards in China: 2. Early mortality results from a prospective study. British Medical Journal, 317:1423-1424.

Nyberg F, Gustavsson P, Jarup L et al. (2000) Urban air pollution and lung cancer in Stockholm. Epidemiology, 11:487-495.
Parkin DM, Muir CS, Whelan SL, Gao YT, Ferlay J, Powell J, eds. (1992) Cancer incidence in five continents: Volume VI. (IARC Scientific Publications No. 120.) International Agency for Research on Cancer (IARC), Lyon.

Parkin DM, Whelan SL, Ferlay J, Black RJ, eds. (1997) Cancer incidence in five continents: Volume VII. (IARC Scientific Publications No. 143.) International Agency for Research on Cancer, Lyon.
Pathak DR, Samet JM, Humble CG, Skipper BJ (1986) Determinants of lung cancer risk in cigarette smokers in New Mexico. Journal of the National Cancer Institute, 76:597-604.

Percy C, Miller BA, Gloeckler Reis LA (1990) Effect of changes in cancer classification and the accuracy of cancer death certificates on trends in cancer mortality. Annals of New York Academy of Sciences, 609:87-97.

Percy C, Muir C (1989) The international comparability of cancer mortality data: results of an international death certificate study. American Journal of Epidemiology, 129:934-946.
Peto R (1978) Carcinogenic effects of chronic exposure to very low levels of toxic substances. Environmental Health Perspectives, 22:155-159.
Peto R (1986) Influence of dose and duration of smoking on lung cancer rates. In: Tobacco: a major international health hazard. Zaridze D, Peto R, eds. (IARC Scientific Publication No. 7) International Agency for Research on Cancer, Lyon.
Peto R, Darby S, Silcocks P, Whitley E, Doll R (2000) Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. British Medical Journal, 321:323-329.
Peto R, Doll R (1984) Keynote address: the control of lung cancer. In: Lung cancer: causes and prevention. Mizell M, Correa P, eds. Verlag Chemie International, Deerfield Beach, FL.
Peto R, Lopez AD, Boreham J, Thun M, Heath C Jr (1992) Mortality from tobacco in developed countries. The Lancet, 339:1268-1278.

Pope CA, III, Burnett RT, Thun MJ et al. (2002) Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Journal of American Medical Association, 287:1132-1141.

Preston SH, Heuveline P, Guillot M (2000) Demography: measuring and modelling population processes. Blackwell, Oxford.

Puddey IB, Rakic V, Dimmitt SB, Beilin LJ (1999) Influence of pattern of drinking on cardiovascular disease and cardiovascular risk factors: a review. Addiction, 94:649-663.
Rao DN, Ganesh B, Rao RS, Desa PB (1994) Risk assessment of tobacco, alcohol, and diet in oral cancer: a case-control study. International Journal of Cancer, 58:469-473.

Rogot E, Murray JL (1980) Smoking and causes of death among U.S. Veterans: 16 years of observation. Public Health Reports, 95:213-222.

Rosenberg L, Kaufman DW, Helmrich SP, Shapiro S (1985) The risk of myocardial infarction after quitting smoking in men under 55 years of age. New England Journal of Medicine, 313:1511-1514.
Rosenberg L, Palmer JR, Shapiro S (1990) The risk of myocardial infarction among women who stop smoking. New England Journal of Medicine, 322:213-217.

Sankaranarayanan R, Duffy SW, Day NE, Nair MK, Padmakumary G (1989a) A case-control investigation of cancer of the oral tongue and the floor of the mouth in southern India. International Journal of Cancer, 44:617-621.
Sankaranarayanan R, Duffy SW, Padmakumary G, Day NE, Nair MK (1990) Risk factors for cancer of the buccal and labial mucosa in Kerala, southern India. Journal of Epidemiology and Community Health, 44:286-292.
Sankaranarayanan R, Duffy SW, Padmakumary G, Day NE, Padmanabhan TK (1989b) Tobacco chewing, alcohol, and nasal snuff in cancer of the gingiva in Kerala, India. British Journal of Cancer, 60:638-643.
Schairer E, Schoniger E (1943) Lung cancer and tobacco use. Zeitschrift fur Krebsforschung, 54:261-269.
Schrek R, Baker LA, Ballard GP, Dolgoff S (1950) Tobacco smoking as an etiologic factor in disease. Cancer Research, 10:49-58.
Smith KR, Liu Y (1993) Indoor air pollution in developing countries. In: Epidemiology of lung cancer: lung biology in health and disease. Samet J, ed. Marcel Dekker, New York.
Smith KR, Shuhua G, Kun H, Daxiong Q (1993) One hundred million improved cookstoves in China: How was it done? World Development, 21:941-961.

Sobue T (1990) Association of indoor air pollution and lifestyle with lung cancer in Osaka, Japan. International Journal of Epidemiology, 19:S62-66.

Sobue T, Suzuki T, Fujimoto I et al. (1991) Lung cancer risk among ex-smokers. Japanese Journal of Cancer Research, 82:273-279.
Sterling TD, Rosenbaum WL, Weinkam JJ (1993) Risk attribution and tobaccorelated deaths. American Journal of Epidemiology, 138:128-139.

Thomas P, Zelikoff J (1999) Air pollutants: moderators of pulmonary host resistance against infection. In: Air pollution and health. Holgate ST, Samet JM, Maynard RL, Koren HS, eds. Academic Press, San Diego, CA.
Thun MJ, Apicella LF, Henley SJ (2000) Smoking vs other risk factors as the cause of smoking-attributable mortality. Journal of American Medical Association, 284:706-712.
Thun MJ, Day-Lally CA, Calle EE, Flanders WD, Heath CW Jr (1995) Excess mortality among cigarette smokers: changes in a 20-year interval. American Journal of Public Health, 85:1223-1230.
Thun MJ, Day-Lally CA, Myers DG et al. (1997a) Trends in tobacco smoking and mortality from cigarette use in cancer prevention studies I (1959 through 1965) and II (1982 through 1988). In: Changes in cigarette-related disease risks and their implications for prevention and control. (Smoking and Tobacco Control Monograph No. 8.) National Cancer Institute, ed. National Cancer Institute, Bethesda, MD.
Thun MJ, Myers DG, Day-Lally C et al. (1997b) Age and exposure-response relationships between cigarette smoking and premature death in cancer prevention study II. In: Age and exposure-response relationships between cigarette smoking and premature death in cancer prevention study II. National Cancer Institute, Bethesda.

Thun MJ, Peto R, Lopez AD et al. (1997c) Alcohol consumption and mortality among middle-aged and elderly U.S. adults. New England Journal of Medicine, 337:1705-1714.
U.S. Department of Health and Human Services (1986) The health consequences of using smokeless tobacco. A report of the advisory committee to the Surgeon General. (DHHS Publication No. [CDC] 89-8411.) USDHHS, Public Health Service, Centers for Disease Control and Prevention, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Bethesda, MD.
U.S. Department of Health and Human Services (1989) Reducing the consequences of smoking: 25 years of progress. A report of the Surgeon General, 1989. (DHHS Publication No. [CDC] 89-8411.) USDHHS, Public Health Service, Centers for Disease Control and Prevention, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Bethesda, MD.
U.S. Department of Health and Human Services (1990a) Chapter 4: smoking cessation and respiratory cancers. In: The bealth benefits of smoking cessation. A report of the Surgeon General, 1990. U.S. Department of Health and Human Services, ed. USDHHS, Public Health Service, Centers for Disease Control and Prevention, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Bethesda, MD.
U.S. Department of Health and Human Services (1990b) Chapter 6: Smoking cessation and cardiovascular disease. In: The health benefits of smoking cessation. A report of the Surgeon General, 1990. U.S. Department of Health and Human Services, ed. USDHHS, Public Health Service, Centers for Disease Control and Prevention, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Bethesda, MD.
U.S. Department of Health and Human Services (1990c) The health benefits of smoking cessation. A report of the Surgeon General, 1990. (DHHS Publication No. [CDC] 89-8411.) USDHHS, Public Health Service, Centers for Disease Control and Prevention, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Bethesda, MD.
U.S. Department of Health and Human Services (2001) Women and smoking. A report of the Surgeon General. (DHHS Publication No. [CDC] 89-8411.) USDHHS, Public Health Service, Centers for Disease Control and Prevention. Office on Smoking and Health, Bethesda, MD.

UK Department of Health, Scientific Committee on Tobacco and Health (1998) Report of the scientific committee on tobacco and health. UK Department of Health, London.

Valkonen T, van Poppel F (1997) Smoking. The contribution of smoking to sex differences in life expectancy. European Journal of Public Health, 7:302-310.
Vena JE (1982) Air pollution as a risk factor in lung cancer. American Journal of Epidemiology, 116:42-56.
Wang TJ, Zhou BS, Shi JP (1996) Lung cancer in nonsmoking Chinese women: a case control study. Lung Cancer, 14:S93-98.
Wassink WF (1948) Onstaansvoorwarden voor Longkanker [Etiology of lung cancer]. Tijdschrift voor Geneeskunde, 92:3732-3747.
WHO (1997) Tobacco or health: a global status report. World Health Organization, Geneva.

Wynder EL, Graham EA (1950) Tobacco smoking as a possible etiologic factor in bronchogenic carcinoma. Journal of American Medical Association, 143:329-336.

Yamaguchi N, Mochizuki-Kobayashi Y, Utsunomiya O (2000) Quantitative relationship between cumulative cigarette smoking and lung cancer mortality in Japan. International Journal of Epidemiology, 29:963-968.

Zaridze D, Peto R, eds. (1986) Tobacco: a major international health hazard. (IARC Scientific Publication No. 7.) International Agency for Research on Cancer, Lyon.

## Chapter I 2

Alcohol use<br>Jürgen Rehm, Robin Room, Maristela Monteiro, Gerhard Gmel, Kathryn Graham, Nina Rehn, Christopher T. Sempos, Ulrich Frick and<br>David Jernigan

## Summary

Alcohol has long been known as a risk factor for disease. The 1990 Global Burden of Disease (GBD) study (Murray and Lopez 1996a, 1996b) identified alcohol as one of the major global risk factors, accounting for $1.5 \%$ of global deaths, $2.1 \%$ of years of healthy life lost owing to premature mortality, $6.0 \%$ of years of life lost owing to disability and $3.5 \%$ of disability-adjusted life years (DALYs).

Based on the epidemiological literature, and modelling the relationship between alcohol exposure and disease, two dimensions of alcohol consumption were defined as exposure variables:

- average volume of alcohol consumption; and
- pattern of drinking.

Average volume of consumption was estimated using both existing estimates of country-specific adult per capita consumption (based on production and sales data) and self-reported alcohol consumption from general population surveys. Country-specific patterns of drinking were defined using indicators of high-volume drinking occasions and types of drinking situation (e.g. drinking with meals). These indicators were drawn from key informant surveys and general population surveys, where available. Optimal scaling procedures were used to determine whether drinking patterns formed a single dimension and the relative impact of the underlying indicators on the pattern value.

The relationship between exposure and various disease categories was estimated by the following methods. Estimates of the relationship between categories of average volume of alcohol consumed and chronic disease were based on disease-specific meta-analyses. Estimates of the relationship between average volume of alcohol consumption and acute disease were based on alcohol-attributable fractions (AAFs) published in the literature. To estimate the impact of patterns of drinking on the risk
of ischaemic heart disease (IHD) and injuries, multilevel modelling with random intercept and random slope was used.

Effects of alcohol on someone other than the drinker were either included in the AAFs of the literature (e.g. alcohol-related injury) or were modelled indirectly (e.g. in the case of the effect of alcohol on the newborn).

Both average volume of alcohol consumption and patterns of drinking varied markedly across subregions. ${ }^{1}$ Average volume of drinking was highest in EUR-A, EUR-C and AMR-A, and lowest in EMR-B, EMR-D and SEAR-D. Patterns were most detrimental in EUR-C, EUR-B, AMRD and AFR-E. Patterns were least detrimental in EUR-A and WPR-A.

Existing research indicates causal relationships between average volume of consumption and more than 60 International Statistical Classification of Diseases and Related Health Problems (ICD) codes, including both chronic diseases (malignant neoplasms, neuro-psychiatric conditions, cardiovascular diseases, gastrointestinal conditions) and injuries (intentional and unintentional). Although most of these relationships involve a detrimental impact of alcohol, there are beneficial relationships between alcohol and IHD, cerebrovascular disease and type II diabetes for certain combinations of average volume of consumption and patterns of drinking. Patterns of drinking were also associated with the level of injury burden from alcohol, although no pattern of drinking had beneficial effects on injury.

The present analysis found that alcohol-related burden of disease is considerable: $3.2 \%$ of global mortality and $4.0 \%$ of the global burden of disease measured in DALYs. In terms of alcohol-related mortality, almost half of the global burden is related to acute causes, i.e. unintentional and intentional injuries, particularly unintentional injuries. The next most important category comprises malignant neoplasms with $20 \%$ of the overall alcohol-related mortality burden, followed by cardiovascular diseases ( $15 \%$ of all alcohol-attributable deaths) and other noncommunicable diseases, primarily liver cirrhosis (13\%). However, although the overall proportion of cardiovascular deaths attributable to alcohol reflects a net result of $15 \%$, this figure does not give a clear picture of the underlying structure of the relationship between alcohol consumption and cardiovascular disease. In particular, although alcohol was estimated to cause a total of almost 600000 cardiovascular deaths in the year 2000, exceeding even the alcohol-related deaths of unintentional injuries, this figure was partly "offset" by the beneficial effects of alcohol on IHD and stroke. Across all diseases, more males than females die from the effects of alcohol, with a ratio of about $10: 1$.

In terms of DALYs, $4.0 \%$ of the overall global disease burden was attributable to alcohol. The biggest differential effect of alcohol on mortality vs morbidity was for neuro-psychiatric diseases. Neuropsychiatric diseases are often disabling, but not fatal, and this is reflected in the markedly higher proportion of overall disease burden caused by
this category compared to alcohol-attributable mortality $(38 \%$ of alcohol-attributable DALYs vs $6 \%$ of alcohol-attributable deaths). As with alcohol-related mortality, males have more than five times the alcohol-related disease burden in terms of DALYs than females.

Alcohol-attributable disease burden is expected to further increase in the future. This is due partly to increases in consumption in developing and emerging economies in south-east Asia and partly to shifting patterns of morbidity and mortality, in particular the increased significance of chronic diseases and injuries related to alcohol. This trend, however, could be reversed quickly, as much of the disease burden of alcohol is almost immediately preventable ( $40 \%$ of the overall alcohol-attributable burden is from acute conditions). While a total ban on alcohol is not realistic, there are other alcohol policy measures that could be implemented to reduce the resulting disease burden.

## 1. Introduction

### 1.1 Definition of alcohol as a risk factor

The relationship between alcohol consumption and health and social outcomes $^{2}$ is complex and multidimensional. Figure 12.1 gives an overview.

Alcohol consumption is linked to long-term biological and social consequences through three intermediate outcomes: intoxication,

Figure I2.I Model of alcohol consumption, intermediate outcomes and long-term consequences


[^44]dependence and direct biochemical effects. Examples of such biochemical effects are the promotion of blood clot dissolution and direct toxic effects on acinar cells triggering pancreatic damage. ${ }^{3}$ Figure 12.1 shows only the main causal pathways. Intoxication may, for example, lead to chronic social consequences (e.g. when a drunken driver kills somebody and thereafter loses his or her job and social standing). Most of the consequences of intoxication are nonetheless covered by acute health and social consequences.

- Direct biochemical effects of alcohol consumption may influence chronic disease, either beneficially or in a harmful way. Beneficial effects include the influence of moderate drinking on IHD by reducing plaque deposits in arteries, protecting against blood clot formation and promoting blood clot dissolution (Zakhari 1997). Examples of harmful effects include increasing the risk of high blood pressure, direct toxic effects on acinar cells triggering pancreatic damage (Apte et al. 1997) and hormonal disturbances (Emanuele and Emanuele 1997). The term "direct toxic and beneficial effects" is used to summarize all the biochemical effects of alcohol on body functions other than intoxication and dependence.
- Intoxication is a powerful mediator, mainly for acute outcomes such as accidents, intentional injuries or deaths, domestic conflict and violence, although episodes of intoxication can also be implicated in chronic health and social problems. The effects of alcohol on the central nervous system mainly determine the subjective feeling of intoxication. These effects are felt and can be measured even at consumption levels that are light to moderate (Eckardt et al. 1998).
- Alcohol dependence is a disorder in itself, but is also a powerful mechanism sustaining alcohol consumption and mediating its impact on both chronic and acute physiological and social consequences (Drummond 1990).

Biological mechanisms have historically been the most important criteria in establishing the causal link between alcohol consumption and health outcomes (English et al. 1995; Hill 1965; Rothman and Greenland 1998a).

Total consumption or average volume of consumption has been the usual measure of exposure linking alcohol to disease (Bruun et al. 1975). Average volume was linked to more than 60 disease conditions in a series of recent meta-analyses (English et al. 1995; Gutjahr et al. 2001; Ridolfo and Stevenson 2001; Single et al. 1999a).

As shown in Figure 12.1, average volume of consumption as a risk factor works mainly through biochemical effects or through dependence to produce long-term consequences. Although average volume is somewhat correlated with intoxication, this correlation is not of sufficient strength to adequately predict acute effects of alcohol related to injury
and death. Such effects are much better predicted by patterns of drinking (Rehm et al. 1996). For example, the same overall average volume of alcohol can be consumed in small quantities regularly with meals (e.g. two drinks a day with meals) or in large quantities on few occasions (e.g. two bottles of wine on a single occasion every Friday). Data on the influence of patterns of drinking are less available than data on overall consumption, but evidence is accumulating that patterns of drinking affect the link between alcohol and disease (Bondy 1996; Puddey et al. 1999; Rehm et al. 1996, 2003) and between alcohol and mortality (Rehm et al. 2001d). In other words, the impact of an average volume of consumption on mortality or morbidity is partly moderated by the way alcohol is consumed by the individual, which in turn is influenced by the social context (Room and Mäkelä 2000). It should be noted that patterns of drinking have been linked not only to acute health outcomes such as injuries (Greenfield 2001; Rossow et al. 2001) but also to chronic diseases such as IHD and especially sudden cardiac death (Britton and McKee 2000; Chadwick and Goode 1998; Puddey et al. 1999; Trevisan et al. 2001a, 2001b).

### 1.2 Choice of exposure variable

To determine the impact of alcohol on burden of disease, both average volume of consumption and pattern of drinking have to be considered and included in the analysis. Unfortunately, average volume of consumption and pattern of drinking are not independent at the level of the individual drinker, because average volume is often determined by heavy drinking occasions (Rehm and Gmel 2000a). Consider someone who drinks eight drinks per day. This person has by definition both a high average volume of consumption and many heavy drinking occasions. On the aggregate level, however, the two dimensions can be statistically independent, as described below in section 2.3.

Average volume of drinking is a relatively simple concept, at least on a theoretical level. It is more difficult to conceptualize patterns of drinking on a worldwide scale. For the comparative risk assessment (CRA) project (see also WHO 2002), aspects of drinking patterns likely to contribute to consequences were identified as a first step in this process (see also Rehm et al. 2001a, 2001b). Table 12.1 provides an overview of the results of this exercise.

### 1.3 Choice of theoretical minimum

Because the relationship between alcohol and disease or injury stems from two potentially interrelated dimensions-average volume of alcohol consumption and drinking pattern-the theoretical minimum or other counterfactual scenarios that provide a reference for hypothetical risk reduction should take both dimensions into account. One obvious and important counterfactual would be total abstinence from alcohol. If all effects of alcohol on health were negative, this would obviously be

Table I2.I Patterns of drinking relevant to CRA

| Pattern indicators | Link to disease |
| :---: | :---: |
| Proportion of the adult population who abstain from alcohol | The same adult consumption per capita will have more detrimental effects in countries where drinking is concentrated among fewer people. This variable was later dropped from the pattern analyses, as it has been incorporated into the average drinking categories |
| Heavy drinking occasions ${ }^{\text {a }}$ <br> High usual quantity of alcohol <br> per occasion <br> Proportion of drinkers who drink <br> daily or nearly daily <br> Proportion of drinking occasions <br> when drinkers get drunk <br> Festive drinking common-at <br> fiestas or community celebrations | The fewer occasions on which a given amount of alcohol is consumed, the more detrimental the consequences (Puddey et al. 1999; Room et al. 2002; Walsh and Rehm 1996). Heavy drinking occasions lead to an increase in injuries. Also, heavy drinking occasions have been shown to lead to detrimental cardiovascular outcomes |
| Drinking with meals-how common to drink with meals | Drinking with meals has been shown in epidemiological and biological research to be less detrimental than drinking at other times (Gentry 2000; Ramchandani et al. 2001; Trevisan et al. 2001a) |
| Drinking in public places-how common to drink in public places | Drinking in public often requires transportation, and thus has been linked to traffic accidents and injuries (Fahrenkrug and Rehm 1994) |
| Drinking linked to violence | Alcohol-related violence is an important cause of injuries. This variable confounds exposure and a potential consequence and was thus dropped in subsequent analyses |

a This is often termed "binge drinking". However, the definition of binge drinking varies widely (e.g. from heavy festive drinking with intoxication lasting more than one day to having five or more drinks on one occasion). It was therefore decided not to use the term in this work.
the ideal theoretical minimum. However, it has been shown that alcohol, if consumed in a regular pattern of light to moderate doses, has protective effects against IHD and potentially other ischaemic diseases (Ashley et al. 2000; Puddey et al. 1999).

The cardioprotective effect has the most relevance to countries with established market economies, which tend to have the longest life expectancy and the highest proportion of deaths from ischaemic disease (Murray and Lopez 1996a). In addition, the pattern of light regular drinking associated with this protective effect, when it occurs, is found mainly in these countries. Drinking to intoxication, heavy drinking occasions and other more detrimental drinking patterns, on the other hand, are often the prevalent drinking style outside established market economies ${ }^{4}$ (see Table 12.3 for an overview of country-by-country drinking patterns). Economic development and patterns of drinking are correlated to a higher degree than volume of alcohol consumption and economic development (Pearson correlation of 0.6 between patterns and per capita gross national product [GNP] on the 2001 data set of the CRA
with 89 countries, where the correlation between average volume of consumption and GNP is 0.2 ; see Rehm et al. 2001a, 2001b). Unfortunately, insufficient research has been carried out to identify the causal determinants of this relationship.

It is proposed to establish the following counterfactual scenarios for measuring the effects of alcohol consumption:

- total abstinence, which would mean an increase in the disease burden of some cardiovascular categories for established market economies and countries with a general pattern of light to moderate drinking;
- as a sensitivity analysis, the current status in patterns of drinking with different scenarios for volume of drinking;
- as a sensitivity analysis, the current status in average consumption with different values for patterns of drinking; and
- changing both average volume of consumption and drinking patterns towards light, regular drinking.

The overall attributable burden of disease should be calculated using abstainers as the comparison group, as this provides the global theoretical minimum. Owing to the large contribution of neuropsychological disease and injuries, neither of which benefits from alcohol consumption, it is also likely that at the population level in every subregion except AMR-A, EUR-A and WPR-A abstinence results in the lowest population risk. The cardioprotective effect will be included in established market economies with light to moderate drinking patterns by using relative risks smaller than 1, thus subtracting "prevented burden" from the burden of disease for these countries.

## 2. Estimating Risk Factor levels

### 2.1 Measuring average volume of alcohol consumption and patterns of drinking

To quantify the effects of alcohol on population health, it is necessary to measure the two key exposure variables, average volume of alcohol consumption and patterns of drinking, at the population level around the world.

## Average volume of alcohol consumption

There have been a number of attempts to gather country-level data on average volume of alcohol consumption, most recently by the World Health Organization (WHO) in the Global status report on alcohol (1999). Most attempts have tried to arrive at an aggregate figure per country, i.e. per capita consumption or adult per capita consumption (i.e. per capita consumption for all inhabitants aged $>15$ years). For several reasons, however, this aggregate figure is insufficient for health impact
assessment. First, as a global figure, this approach does not allow disaggregation into different groups (e.g. as defined by sex and age), which is needed for valid estimates of burden of disease. Second, per capita consumption estimates are usually based on production figures or sales data, and thus do not include consumption of home-made or illegally imported alcohol (see discussion of unrecorded consumption in Giesbrecht et al. 2000; Leifman 2001; Summer 2000).

On the positive side, the production and trade or sales data required to calculate per capita consumption have traditionally been collected by various sources (including the alcoholic beverage industry and international agencies such as the Food and Agriculture Organization of the United Nations [FAO]) and are available for most countries of the world. FAO collects production and trade data for different alcoholic beverages, mainly from ministries of agriculture and customs departments. For some countries, estimates include data on alcoholic beverages outside the usual beer, wine and spirits categories (e.g. palm wine and sorghum beer) but they do not systematically include home production or illicit production.

To use per capita consumption data for detailed risk analysis, we combined this source with survey data from various countries. Survey data are especially important in determining the proportion of abstainers in a country, as well as in dividing the overall volume into drinking categories by sex and age groups. In addition, some surveys can be used to estimate unrecorded consumption (e.g. Kühlhorn et al. 1999; and the current WHO-supported efforts in Brazil, China, India and Nigeria). ${ }^{5}$

Although survey data are essential for refining per capita estimates, they also have problems that preclude them from being used as the single source of data. In particular, survey data from many established market economies considerably underestimate total volume (Midanik 1988; Midanik and Harford 1994; Rehm 1998a; de Vries et al. 1999). ${ }^{6}$ In addition, survey data are less available than per capita consumption data on an international level (Rehm and Gmel 2000b; WHO 1999). Nevertheless, by combining both kinds of information one may arrive at disaggregated estimates. In this combination, the per capita estimates based on production and sales (combined with data on unrecorded consumption where available) are taken as the overall value. Surveys are then used to estimate the distribution of this overall volume among various groups, as defined by abstinence and different levels of drinking, and by sex and age.

## Patterns of drinking

Existing general population surveys cover some of the patterns of drinking listed in Table 12.1, but rarely would one find all features in one representative survey in a particular country. Moreover, not all surveys on drinking patterns are recent. Therefore, to develop drinking pattern estimates for the 14 subregions used in the CRA, key informant question-

Table I2.2 Countries with returned key informant questionnaires

| Subregion | Countries that returned questionnaires in at least one phase | Total no. of countries in subregion |
| :---: | :---: | :---: |
| AFR-D | Burkina Faso, Nigeria, Seychelles | 26 |
| AFR-E | Congo, Namibia, South Africa, Zambia | 20 |
| AMR-A | Canada, USA | 3 |
| AMR-B | Argentina, Brazil, Costa Rica, Mexico, Trinidad and Tobago | 26 |
| AMR-D | Peru | 6 |
| EMR-B | None | 13 |
| EMR-D | None | 9 |
| EUR-A | Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Malta, Netherlands, Norway, Portugal, Slovenia, Spain, Sweden, Switzerland, United Kingdom | 26 |
| EUR-B | Armenia, Bulgaria, Poland, Slovakia, The former Yugoslav Republic of Macedonia | 16 |
| EUR-C | Belarus, Estonia, Hungary, Latvia, Lithuania, Russian Federation | 9 |
| SEAR-B | Sri Lanka, Thailand | 3 |
| SEAR-D | India | 7 |
| WPR-A | Australia, Japan, New Zealand, Singapore | 5 |
| WPR-B | China, Fiji, Malaysia, Micronesia, Palau, Papua New Guinea, Philippines, Republic of Korea, Solomon Islands | 22 |
| Total |  | 191 |

naire studies were undertaken in early 2000 (Rehm et al. 2001b; for the key informant questionnaire see Appendix 1 in European Addiction Research 2001) and repeated in 2001, using a slightly modified questionnaire. ${ }^{7}$ In 2000, 61 questionnaires were sent out, 52 of which were returned, and in 2001 another 155 were sent out, 40 of which were returned. Table 12.2 lists, by subregion, the countries for which key informant surveys were received. Together with survey data, the responses from the survey on patterns provided sufficient data for a first estimate of patterns of drinking for all subregions.

## Summary of data sources used

The following country level measures of alcohol consumption were used in the analysis:

- adult per capita consumption based on sales data or production and trade data;
- unrecorded consumption based on various estimates;
- survey data on abstinence, average volume consumed in different sex and age groups, and patterns of drinking; and
- key informant information on various aspects of patterns of drinking.

Data for all years after 1998 were considered. Data were checked for consistency across time and for internal consistency (e.g. survey vs per capita estimates of average volume of consumption).

### 2.2 DATA SOURCES FOR AVERAGE VOLUME OF ALCOHOL CONSUMPTION

Data on adult per capita and unrecorded consumption were taken from the Global status report on alcohol (WHO 1999) and from the WHO Global Alcohol Database, created by the Marin Institute for the Prevention of Alcohol and Other Drug Problems and currently maintained by the Swiss Institute for the Prevention of Alcohol Problems. Surveys were also collected from this database, but additional surveys were accessed based on individual contacts (including several experts from each subregion) and by announcing this project on a specific WHO listserve and at the Annual Alcohol Epidemiology Symposia of the Kettil Bruun Society for Social and Epidemiological Research on Alcohol. Data on drinking patterns were collected from researchers and health officials known to WHO who had the knowledge to serve as key informants for their countries or regions. Key informant information was collected by surveys in early 2000 and in mid 2001.

Categorical levels for average volume of alcohol per day in relating consumption to chronic disease were selected to be consistent with previous meta-analyses (English et al. 1995; Gutjahr et al. 2001; Ridolfo and Stevenson 2001; Single et al. 1999a; for a discussion of the background to these categories see English et al. 1995; Holman et al. 1996) and were defined as follows:

- abstainer: a person not having had a drink containing alcohol within the last year;
- average volume drinking category I: for females $0-19.99 \mathrm{~g}$ pure alcohol daily; for males $0-39.99 \mathrm{~g}$ pure alcohol daily;
- average volume drinking category II: for females 20-39.99g pure alcohol daily; for males $40-59.99 \mathrm{~g}$ pure alcohol daily; and
- average volume drinking category III: for females 40 g or more pure alcohol daily; for males 60 g or more pure alcohol daily. ${ }^{8}$

This categorization of average drinking, rather than using a continuous measure of consumption, makes it possible to derive different shapes of risk curve (linear, J-shape, threshold, etc.) while, at the same time, allowing inclusion of data from studies in which only categorical information on levels of alcohol consumption were collected. Using per capita consumption data derived from production and trade or sales data plus unrecorded consumption as the first estimate of overall alcohol con-
sumption, the following strategy was adopted to generate age- and sexspecific prevalence rates.

- For each subregion, the average adult per capita consumption, including unrecorded consumption for the population aged $\geq 15$ years, was estimated as a population-weighted average of country-specific per capita consumption data. All entries per country after 1998 were taken and averaged to obtain a stable estimate for 2000 . The weights were derived from the average population aged $\geq 15$ years in each country for the years after 1998, on the basis of United Nations population data. Country-specific adult per capita data were estimated for 132 countries (see Table 12.3). Per capita consumption was known and adult per capita consumption could be calculated for more than $90 \%$ of the world's population. Survey information on abstinence was available for 69 countries, in particular from almost all countries with population larger than 100 million resulting in more than $80 \%$ of the world population with available survey data. This means that more than $50 \%$ of the countries for which data were available on per capita consumption also had survey data available.
- Country-specific survey data of the ratio of male to female consumption were used to allocate proportionally the overall adult per capita consumption to adult male and adult female per capita consumption.
- Based on surveys, the age-specific prevalence of drinking was calculated on the assumption that the average per capita consumption and the proportions of male and female abstainers were correct.


### 2.3 DATA SOURCES FOR PATTERNS OF DRINKING

Initial estimates of drinking patterns across a range of countries were based on two surveys of key informants selected by WHO staff, conducted in early 2000 and mid 2001. The surveys covered relevant drinking characteristics within different countries or regions. In most cases, respondents had access to national or regional survey data, although these data had not always been published in the international literature. In addition, the informants provided a confidence rating for their responses (i.e. whether based on surveys or just best guesses). This information was used for decisions about inclusion of data when conflicting information existed. As listed in Table 12.1, the survey considered five main areas of drinking patterns that might be expected to affect the impact of volume of drinking: proportion of abstainers, heavy drinking occasions, drinking with meals, drinking in public places and drinking linked to violence (later dropped from analyses because of its interference with outcome measures).

The key informant ratings were analysed using optimal scaling analysis (Bijleveld et al. 1998, chapter 2). This analysis is similar to factor analysis, but permits the simultaneous inclusion of ordinal and categor-
ical data. As with factor analysis, this statistical technique allows the analyst to determine the number of underlying dimensions and the relation of items to each dimension. In the analysis of patterns of drinking, one global dimension was identified and labelled as detrimental impact (for details see Rehm et al. 2001b).

The results of the optimal scaling analysis were very similar to a score derived simply by summing the ratings of the key informant survey (Pearson correlation: 0.93). To further simplify the pattern values into robust general categories based on these scale values, the countries were placed in four categories and assigned values from 1 to 4 . By the time the final pattern values were constructed, additional survey data were available as well as the second wave of key informant data, allowing refinement and corrections of estimates. Also, the proportion of abstainers was no longer included as one of the parameters of pattern weights, because rates of abstinence were taken into account separately as one of the average volume of consumption categories. The underlying variables and the scoring pattern can be found in Appendix A.

To apply pattern values to estimate the burden of disease attributable to alcohol, countries with missing data on drinking pattern values were assigned the same category as that of neighbouring countries, taking into consideration geographical and cultural proximity. The pattern values for more than 130 countries worldwide can be seen in Table 12.3. ${ }^{9}$

Patterns of drinking thus defined were found to be unrelated to volume: the overall Pearson correlation between pattern values and per capita consumption for the countries included is -0.126 and the more appropriate Spearman correlation is -0.072 . Both correlations do not achieve statistical significance; that is, they are not significantly different from zero (both correlations based on 132 countries with data on both variables). This suggests that drinking pattern may provide important unique information about the risks of drinking alcohol beyond that captured by per capita consumption.

Although this procedure allowed us to derive drinking pattern values from a combination of empirical data and expert judgement, these patterns still needed to be validated empirically to demonstrate that they were, in fact, related to outcomes. In other words, pattern values serve as a description of one aspect of exposure that is theoretically postulated to be related to harm, but such a relation still has to be empirically established. In addition, the degree of influence of patterns on harm (i.e. how much weight to assign to drinking pattern in calculating the burden of disease attributable to alcohol) had to be estimated. Moreover, the weight to be assigned to drinking pattern may vary by type of outcome, sex and age. A later section of this chapter describes how these weights were developed.

### 2.4 Methods for obtaining estimates where more than one data source exists

For estimating overall consumption, clear hierarchies were used for integrating per capita data into the WHO Global Alcohol Database and subsequently into this work.

- Scientifically derived and well documented local estimates (e.g. from the National Drug Research Institute in Australia, see Catalano et al. 2001) were given first priority.
- Production/sales data were used, such as the annual data on per capita consumption published by the alcohol industry (e.g. Productschap voor Gedistilleerde Dranken 1999, 2000).
- FAO production and trade data were used where other data were not available.

If more than one data point existed for the time after 1998 (e.g. per capita estimates for 1998 and 1999), data points were averaged. With respect to survey data, if more than one representative survey with more than 2000 persons existed, the most recent survey estimates were used. Survey data were always given priority over key informant estimates in estimating pattern values.

### 2.5 Methods for obtaining estimates where no data SOURCE EXISTS

Most regions had sufficient data for estimating prevalence of average volume of drinking, based on per capita consumption (including unrecorded consumption) plus the sex-specific ratio of abstinence derived from surveys. Survey data on abstinence were available for 69 countries ( $52.3 \%$ of all 132 countries included), and were additionally estimated for 39 countries ( $29.5 \%$ of all the countries). Estimation was based on abstinence rates in adjoining countries. Survey data on abstinence were available for many of the countries with large populations: Brazil, China, India, Mexico, Nigeria, the Russian Federation, South Africa, the United States of America and the major (western) European countries. These surveys were taken as indicators for the respective subregions. Regional numbers are to a large degree influenced by a limited number of highly populated countries. Thus, it is important to get the numbers for these countries correct by investing limited resources in estimating alcohol consumption for these countries, rather than attempting to improve the data for all countries in the world. ${ }^{10}$

Data were scarcer with respect to patterns of drinking. Only 44 countries ( $33.3 \%$ of 132 countries) had sufficient information on patterns to compute a pattern score. For another 88 countries ( $66.6 \%$ ), data on patterns had to be estimated. Clearly, it is a priority for future estimates that more data on drinking patterns be collected for use in these analyses (see
also Rehm and Gmel 2000b). As with per capita consumption, ratings on drinking patterns for countries for which no data were available were based on social and cultural factors (Muslim vs non-Muslim country, type of drinking culture, etc.) and on drinking patterns in surrounding countries.

### 2.6 DESCRIPTION OF DATABASES, INCLUDING METHODOLOGICAL QUALITIES

A description of the WHO Global Alcohol Database can be found in the Global status report on alcohol (WHO 1999). The Swiss Institute for the Prevention of Alcohol Problems constantly updates and expands the Database, which includes both adult per capita and survey data. In addition, many surveys were directly sent to the first author of this work, especially from countries that also supplied key informant information. Additional surveys from developing countries can be found in a WHO collection (WHO 2001). The data in the WHO Database are selected and scrutinized according to set criteria. The data from established market economies are usually based on more reliable sales data and better-quality surveys than those from developing countries.

### 2.7 Characteristics of excluded studies/Databases

Studies were excluded only if better survey information was available for the same country. Better survey information was defined in terms of how recently the survey had been conducted, the availability of a probability sample of at least 2000 respondents, or other quality criteria (e.g. a representative survey based on probabilistic sampling of the whole country vs a non-representative survey or regional survey; better alcohol measures, such as quantity-frequency measure vs frequency-only measure; or larger survey at the same time).

### 2.8 Estimates of COUNTRY DATA AND EXPOSURE DIMENSIONS BY subregion, age and sex

Table 12.3 provides country-specific information on alcohol consumption, drinking patterns, percentage abstainers and validity ratings for estimates. Also shown are the relevant subregion and the population aged $\geq 15$ years in 2000 .

Table 12.4 gives estimates of the proportion of population in each of the four categories of average alcohol consumption described above by subregion, sex and age. The CRA required two age groups for people aged $>70$ years: $70-79$ years and $\geq 80$ years. With a lack of evidence to differentiate between these two age categories, both were assumed to have the same prevalence and risk relations.

These figures assume no alcohol consumption leading to harm for young people under the age of 15 years. This is certainly not true for young people in established market economies, where drinking starts early and where alcohol-related harm can be found (though not with a
Table I2.3 Per capita alcohol consumption, patterns of drinking and abstinence in selected countries after 1998

| Country | Subregion | Adult per capita alcohol consumption ${ }^{2}$ | Unrecorded consumption ${ }^{\text {b }}$ | Pattern value ${ }^{\text {c }}$ | Validity pattern value ${ }^{\text {d }}$ | Percentage male abstainers ${ }^{\text {e }}$ | Percentage female abstainers ${ }^{f}$ | Validity of abstinence ${ }^{g}$ | Population aged <br> $\geq 15$ years (000s) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Albania | EUR-B | 4.77 | 3.00 | 3 | 0 | 12.0 | 36.0 | 0 | 2195.8 |
| Algeria | AFR-D | 0.47 | 0.16 | 3 | 0 | 80.0 | 98.0 | 0 | 19939.8 |
| Argentina | AMR-B | 16.30 | 1.00 | 2 | I | 7.0 | 21.0 | 1 | 26766.9 |
| Armenia | EUR-B | 2.88 | 1.44 | 2 | 1 | 9.9 | 60.0 | 1 | 2656.9 |
| Australia | WPR-A | 9.19 | 0.00 | 2 | 1 | 15.8 | 24.0 | 1 | 14988.4 |
| Austria | EUR-A | 13.90 | 1.00 | 1 | 0 | 13.0 | 33.0 | 1 | 6814.9 |
| Azerbaijan | EUR-B | 2.86 | 1.43 | 3 | 0 | 12.0 | 36.0 | 0 | 5520.9 |
| Barbados | AMR-B | 7.43 | -0.50 | 2 | 0 | 29.0 | 70.4 | 0 | 213.6 |
| Belarus | EUR-C | 12.22 | 4.90 | 4 | 0 | 2.0 | 4.0 | 1 | 8322.3 |
| Belgium | EUR-A | 11.45 | 0.50 | 1 | 0 | 9.6 | 20.5 | 1 | 8420.4 |
| Belize | AMR-B | 6.35 | 2.00 | 4 | 0 | 24.0 | 44.0 | 0.5 | 145.0 |
| Bolivia | AMR-D | 5.74 | 3.00 | 3 | 0 | 23.8 | 44.6 | 1 | 5028.6 |
| Bosnia and Herzegovina | EUR-B | 7.65 | 3.00 | 3 | 0 | 12.0 | 36.0 | 0 | 3223.8 |
| Botswana | AFR-E | 5.33 | 3.00 | 3 | 0 | 37.0 | 70.0 | 0 | 938.4 |
| Brazil | AMR-B | 8.59 | 3.00 | 3 | 1 | 36.4 | 57.0 | 1 | 121038.8 |
| Bulgaria | EUR-B | 13.08 | 5.00 | 2 | 1 | 8.0 | 16.0 | 1 | 6890.0 |
| Burkina Faso | AFR-D | 3.81 | 3.32 | 3 | 0 | - | - | - | 6287.6 |
| Burundi | AFR-E | 7.42 | 4.75 | 3 | 0 | - | - | - | 3591.1 |
| Cambodia | WPR-B | 0.36 | 0.12 | 3 | 0 | 74.4 | 96.0 | 0 | 6603.7 |
| Cameroon | AFR-D | 4.35 | 2.62 | 3 | 0 | - | - | - | 8524.9 |

Table I2.3 Per capita alcohol consumption, patterns of drinking and abstinence in selected countries after 1998 (continued)

| Country | Subregion | Adult per capita alcohol consumption ${ }^{\text {a }}$ | Unrecorded consumption ${ }^{\text {b }}$ | Pattern value ${ }^{\text {c }}$ | Validity pattern value ${ }^{\text {d }}$ | Percentage male abstainers ${ }^{\text {e }}$ | Percentage female abstainers ${ }^{f}$ | Validity of abstinence ${ }^{g}$ | Population aged $\geq 15$ years (000s) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Canada | AMR-A | 9.43 | 1.00 | 2 | 1 | 17.3 | 28.0 | 1 | 25248.1 |
| Central African Republic | AFR-E | 3.01 | 1.00 | 3 | 0 | - | - | - | 2078.4 |
| Chile | AMR-B | 8.34 | 1.00 | 3 | 0 | 31.4 | 46.5 | 1 | 10883.1 |
| China | WPR-B | 4.83 | 1.00 | 2 | 1 | 15.9 | 70.7 | 1 | 960300.9 |
| Colombia | AMR-B | 8.30 | 2.00 | 3 | 0 | 31.4 | 46.5 | 0.5 | 28470.8 |
| Costa Rica | AMR-B | 6.70 | 2.00 | 3 | 0 | 45.0 | 75.0 | 1 | 2721.4 |
| Croatia | EUR-A | 18.39 | 4.50 | 3 | 0 | 12.0 | 36.0 | 0 | 3709.2 |
| Cuba | AMR-A | 5.66 | 2.00 | 2 | 0 | 29.0 | 70.4 | 0 | 8823.4 |
| Cyprus | EMR-B | 9.29 | 1.00 | 1 | 0 | 1.2 | 15.4 | 0 | 603.0 |
| Czech Republic | EUR-A | 15.02 | 1.00 | 2 | 1 | 3.1 | 8.1 | 1 | 8547.7 |
| Democratic People's Republic of Korea | SEAR-D | 5.14 | 1.00 | 3 | 0 | - | - | - | 17399.1 |
| Denmark | EUR-A | 14.32 | 2.00 | 1 | 1 | 2.0 | 4.0 | 1 | 4341.6 |
| Djibouti | EMR-D | . 66 | 0.22 | 3 | 0 | - | - | - | 373.4 |
| Dominican Republic | AMR-B | 5.71 | 1.00 | 2 | 0 | 29.0 | 70.4 | 0 | 5687.8 |
| Ecuador | AMR-D | 5.49 | 3.66 | 3 | 0 | 20.0 | 40.0 | 0.5 | 8368.2 |
| Egypt | EMR-D | 0.92 | 0.46 | 2 | 0 | 70.0 | 98.0 | 0.5 | 44274.2 |
| El Salvador | AMR-B | 4.64 | 2.00 | 4 | 0 | 8.7 | 37.7 | 0 | 4041.9 |
| Eritrea | AFR-E | 2.55 | 0.85 | 3 | 0 | - | - | - | 2151.5 |
| Estonia | EUR-C | 11.70 | 5.00 | 3 | 0 | 5.0 | 10.0 | 1 | I 152.4 |

$$
\begin{array}{rr}
- & 33690.1 \\
\text { I } & 561.3 \\
1 & 4239.6 \\
1 & 48032.9 \\
- & 733.5 \\
0 & 3868.9 \\
1 & 69469.4 \\
- & 11489.3 \\
1 & 9051.8 \\
0.5 & 6420.1 \\
0 & 604.3 \\
0.5 & 4874.7 \\
0 & 3784.3 \\
1 & 8329.4 \\
0 & 215.4 \\
1 & 676054.8 \\
0 & 147181.7 \\
0 & 13558.9 \\
1 & 2938.4 \\
1 & 4492.5 \\
1 & 49133.0 \\
0 & 1780.9 \\
\hline & 107949.3 \\
\hline & \text { continued }
\end{array}
$$

Table I2.3 Per capita alcohol consumption, patterns of drinking and abstinence in selected countries after 1998 (continued)

| Country | Subregion | Adult per capita alcohol consumption ${ }^{\text {a }}$ | Unrecorded consumption ${ }^{\text {b }}$ | Pattern value ${ }^{\text {c }}$ | Validity pattern value ${ }^{\text {d }}$ | Percentage male abstainers ${ }^{\text {e }}$ | Percentage female abstainers ${ }^{f}$ | Validity of abstinence ${ }^{8}$ | Population aged $\geq 15$ years (000s) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Jordan | EMR-B | 0.28 | 0.14 | 2 | 0 | 74.0 | 98.0 | 0 | 3871.8 |
| Kazakhstan | EUR-C | 10.11 | 6.90 | 4 | 0 | 10.0 | 27.0 | 0 | 11751.4 |
| Kenya | AFR-E | 6.83 | 5.00 | 3 | 0 | 45.0 | 65.0 | 0 | 17137.1 |
| Kyrgyzstan | EUR-B | 5.89 | 4.00 | 3 | 0 | 60.0 | 80.0 | 0 | 3054.5 |
| Lao People's Democratic Republic | WPR-B | 5.82 | 2.00 | 3 | 0 | - | - | - | 3044.4 |
| Latvia | EUR-C | 16.48 | 7.00 | 3 | 0 | 15.0 | 46.2 | 1 | 1940.2 |
| Lebanon | EMR-B | 5.60 | 2.00 | 3 | 0 | - | - | - | 2208.6 |
| Lesotho | AFR-E | 3.16 | 1.58 | 3 | 0 | - | - | - | I 294.2 |
| Liberia | AFR-D | 4.54 | 2.00 | 3 | 0 | - | - | - | 1824.6 |
| Lithuania | EUR-C | 11.41 | 4.90 | 3 | 0 | 15.0 | 46.2 | 0 | 2964.8 |
| Luxembourg | EUR-A | 17.32 | -2.00 | 1 | 0 | 1.0 | 4.0 | 0.5 | 352.9 |
| Malaysia | WPR-B | 4.26 | 3.40 | 3 | 0 | 35.1 | 63.5 | I | 14678.4 |
| Mauritius | AFR-D | 15.62 | 11.00 | 3 | 0 | 22.0 | 53.0 | 1 | 865.2 |
| Mexico | AMR-B | 8.15 | 4.00 | 4 | 1 | 15.8 | 46.5 | 1 | 66105.2 |
| Mongolia | WPR-B | 4.45 | 2.00 | 3 | 0 | 20.2 | 62.7 | 0 | 1740.8 |
| Morocco | EMR-D | 1.16 | 0.58 | 3 | 0 | - | - | - | 19121.7 |
| Myanmar | SEAR-D | 0.62 | 0.42 | 2 | 0 | 45.0 | 93.5 | 0 | 32887.4 |
| Namibia | AFR-E | 5.40 | 1.80 | 3 | 1 | 60.9 | 47.1 | 1 | 1008.5 |
| Netherlands | EUR-A | 10.39 | 0.50 | I | I | 14.0 | 26.5 | 1 | 12926.5 |


| 1 | 2986.5 |
| :---: | :---: |
| 0 | 2905.3 |
| 1 | 63465.7 |
| 1 | 3588.2 |
| 0 | 91048.8 |
| 1 | 2948.4 |
| 0.5 | 3323.7 |
| 1 | 17094.4 |
| 1 | 48096.9 |
| 1 | 31240.1 |
| 1 | 8261.3 |
| 1 | 36775.6 |
| 0 | 3360.7 |
| 0 | 18364.0 |
| 1 | 120255.0 |
| - | 4224.2 |
| 0 | 12848.8 |
| - | 5244.9 |
| 1 | 69.0 |
| - | 2716.7 |
| 0 | 2778.7 |
| 1 | 4326.7 |



WPR-A
AMR-D
AFR-D
EUR-A
EMR-D
WPR-B
AMR-B
AMR-D
WPR-B
EUR-B
EUR-A
WPR-B
EUR-C
EUR-B
EUR-C
AFR-E
EMR-B
AFR-D
AFR-D
AFR-D
WPR-A
EUR-B
 Papua New Guinea Paraguay Peru Philippines Poland Portugal
Republic of Korea Republic of Moldova Romania Russian Federation Rwanda Saudi Arabia Senegal Seychelles Sierra Leone Singapore

Table I2.3 Per capita alcohol consumption, patterns of drinking and abstinence in selected countries after 1998 (continued)

| Country | Subregion | Adult per capita alcohol consumption ${ }^{\text {a }}$ | Unrecorded consumption ${ }^{\text {b }}$ | Pattern value ${ }^{\text {c }}$ | Validity pattern value ${ }^{\text {d }}$ | Percentage male abstainers ${ }^{\text {e }}$ | Percentage female abstainers ${ }^{f}$ | Validity of abstinence ${ }^{\text {g }}$ | Population aged $\geq 15 \text { years }(000 \mathrm{~s})$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Slovenia | EUR-A | 13.42 | 5.20 | 3 | 1 | 31.2 | 55.4 | 1 | 1669.0 |
| South Africa | AFR-E | 12.41 | 2.20 | 3 | 1 | 55.3 | 83.1 | 1 | 26236.4 |
| Spain | EUR-A | 13.28 | 1.00 | 1 | 1 | 7.0 | 24.0 | 1 | 33863.3 |
| Sri Lanka | SEAR-B | 0.57 | 0.38 | 3 | 0 | 74.4 | 96.0 | 1 | 13912.4 |
| Sudan | EMR-D | 0.69 | 0.46 | 3 | 0 | - | - | - | 17863.6 |
| Suriname | AMR-B | 5.96 | 0.00 | 3 | 0 | 30.0 | 55.0 | 0.5 | 290.0 |
| Swaziland | AFR-E | 7.89 | 4.05 | 3 | 0 | - | - | - | 574.6 |
| Sweden | EUR-A | 9.07 | 2.00 | 3 | 1 | 7.0 | 12.0 | 1 | 7288.2 |
| Switzerland | EUR-A | 12.49 | 0.50 | 1 | 1 | 11.0 | 27.0 | 1 | 6097.3 |
| Syrian Arab Republic | EMR-B | 0.70 | 0.36 | 2 | 0 | - | - | - | 9546.7 |
| Tajikistan | EUR-B | 5.23 | 3.95 | 3 | 0 | 60.0 | 80.0 | 0 | 3692.5 |
| Thailand | SEAR-B | 11.70 | 2.00 | 3 | 1 | 30.5 | 71.7 | 1 | 45909.9 |
| The former Yugoslav Republic of Macedonia | EUR-B | 8.56 | 3.20 | 3 | I | 20.0 | 40.0 | 1 | 1559.4 |
| Trinidad and Tobago | AMR-B | 2.36 | 0.00 | 2 | 1 | 29.0 | 70.4 | 1 | 970.9 |
| Tunisia | EMR-B | 1.80 | 0.50 | 2 | 0 | 70.0 | 95.0 | 0 | 6677.8 |
| Turkey | EUR-B | 4.30 | 2.70 | 3 | 0 | 35.0 | 55.0 | 0 | 47742.9 |
| Turkmenistan | EUR-B | 2.85 | 1.00 | 3 | 0 | 35.0 | 55.0 | 0 | 2780.1 |


| Uganda | AFR-E | 13.30 | 10.71 | 3 | 0 | 45.0 | 67.0 | 0.5 | 10877.2 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Ukraine | EUR-C | 8.00 | 4.00 | 3 | 1 | 12.0 | 36.0 | 0 | 41487.6 |
| United Arab Emirates | EMR-B | 3.68 | 1.00 | 2 | 0 | - | - | - | 1757.7 |
| United Kingdom | EUR-A | 11.88 | 2.00 | 2 | 1 | 8.0 | 14.0 | 1 | 47761.2 |
| United Republic of Tanzania | AFR-E | 6.47 | 2.00 | 3 | 0 | - | - | - | 18291.2 |
| United States | AMR-A | 9.47 | 1.00 | 2 | 1 | 28.1 | 42.6 | 1 | 218586.3 |
| Uruguay | AMR-B | 9.54 | 2.00 | 3 | 0 | 7.0 | 21.0 | 0 | 2509.7 |
| Uzbekistan | EUR-B | 2.92 | 1.90 | 3 | 0 | 60.0 | 80.0 | 0 | 15211.3 |
| Venezuela | EUR-C | 9.59 | 2.00 | 3 | 0 | 30.0 | 55.0 | 0.5 | 15942.8 |
| Viet Nam | WPR-B | 2.26 | 1.00 | 3 | 0 | - | - | - | 53320.3 |
| Zambia | AFR-E | 3.96 | 1.00 | 4 | 1 | 35.0 | 70.0 | 1 | 4837.4 |
| Zimbabwe | AFR-E | 12.65 | 9.00 | 4 | 0 | 7.0 | 36.0 | 0.5 | 6847.3 |
| No data. |  |  |  |  |  |  |  |  |  |
| Estimated annual per capita unrecorded consumption by adults ( $\geq 15$ years) after 1998 in litres of pure alcohol (based in part on the Global status reflect where estimated cross-border shopping by non-residents and drinking by tourists exceeds any estimated unrecorded consumption by resid |  |  |  |  |  |  |  |  |  |
| I denotes the least detrimental pattern value and 4 denotes the most detrimental. Pattern values are assumed to be relatively stable and were derived for the 1990s. |  |  |  |  |  |  |  |  |  |
| $0=$ imputed based on regional and cultural similarities, or based on data from the Global status report on alcohol; I = questionnaire available with sufter optimal scaling. |  |  |  |  |  |  |  |  |  |
| Proportion of adult males who were abstainers for the year before the survey. |  |  |  |  |  |  |  |  |  |
| Proportion of adult females who were abstainers for the year before the survey. |  |  |  |  |  |  |  |  |  |
| $0=$ imputed based on regional and cultural similarities; $0.5=$ based on survey but only for subsections of the country or not differentiated for sex sex. |  |  |  |  |  |  |  |  |  |

high prevalence) before the age of 15 years (Hibell et al. 2000). However, as data for this age group are scarce and as the prevalence rates for drinking and harm are low, the conservative approach of estimating no alcohol-related harm for this age group was adopted. The decision to model alcohol-related harm at zero for people's own drinking does not mean that there will be no alcohol-related harm estimated in this age group altogether. Rather, owing to the nature of effects, we must deviate from the standard epidemiological model and include effects of drinking on other externalities. To give just one example of this type of harm, a drunken driver may kill innocent bystanders or passengers driving in his car who are younger than 15 years, and this would be a fatality caused by alcohol.

The following distribution data were based on survey-based information on abstainers for 69 countries. For distributional information on drinkers, the following sources were considered:

- AFR-D, Nigeria (Mustonen et al. 2001 and newer survey data reported in the key informant questionnaire from I. Obot/O. Gureje);
- AFR-E, South Africa (Department of Health 1998, South African Demographic and Health Survey);
- AMR-A, Canada and the United States (surveys provided by E. Adlaf for Canada and T. Greenfield for the United States);
- AMR-B, Brazil (São Paulo) (Galduróz et al. 2000) and Mexico (surveys provided by the Mexican Institute of Psychiatry); see also WHO (2001) for country reports on Costa Rica and Mexico, and Jutkowitz and Hongsook (1994);
- AMR-D (survey data for Peru and Jutkowitz and Hongsook 1994);
- EMR-B and EMR-D (only per capita information and general distribution information as approximately log-normal ${ }^{11}$ with some data on abstinence);
- EUR-A (based on the average of many country surveys);
- EUR-B, Poland (based on survey information provided by key informants);
- EUR-C, the Russian Federation (published national and regional survey data from Bobak et al. 1999; Malyutina et al. 2001);
- SEAR-B, Sri Lanka and Thailand (survey estimates provided by key informants);
- SEAR-D, India (regional survey data provided by key informants);
- WPR-A, Australia and New Zealand (survey data, see English et al. 1995; Ridolfo and Stevenson for Australia and http://www. aphru.ac.nz/projects/Larger for New Zealand); and
- WPR-B, China (Wei et al. 1999 and additional information by the first author).

The numbers in Table 12.4 are given as a proportion of the age-sexspecific population. Thus, for females aged 15-29 years in WPR-B, $67.4 \%$ are estimated to abstain, $32.6 \%$ are estimated to drink the equivalent of between 0 g and 20 g pure alcohol per day, $0.1 \%$ are estimated to drink between 20 g and 40 g , and less than $0.05 \%$ are estimated to drink more than 40 g . In the oldest age group, $89.8 \%$ of females are estimated to abstain, $10.2 \%$ to drink the equivalent of up to 20 g pure alcohol per day, and less than $0.05 \%$ to drink more.

Currently, pattern values assigned to each country and then computed as a population-weighted average for each subregion are not specific by sex and age (see Table 12.5). This may change in the future when survey data on patterns become more available.

### 2.9 QUANTITATIVE AND QUALITATIVE SOURCES OF UNCERTAINTY

## UNCERTAINTY ANALYSIS

As described, alcohol consumption as a risk factor has two dimensions: average volume and patterns. It is thus proposed to base the uncertainty analysis on the weighted average of the available information for both dimensions for each subregion. Uncertainty analysis is undertaken to give an indication on uncertainty, based on different characteristics such as the source or variability of the estimated data. It will also be used to calculate the confidence intervals (CIs) of the alcohol-related burden. Classically, CIs for prevalence are determined by sample size, assuming that the underlying individual data are representative of the subregion. For some subregions, however, we do not have probabilistic samples. As alcohol consumption is a social activity and can vary markedly from one country to another within a region, one cannot automatically assume that the countries without surveys would have the same alcohol distribution as the countries with surveys. Thus, the procedure based on sample size cannot be used here. Moreover, prevalence was derived from both aggregate- (per capita consumption) and individual-level data in a triangulation of information, for which there is no statistical theory for readily deriving CIs.

The algorithms specified below were developed after extensive discussions with experts in the field. They are intended to reflect the quantity and quality of the underlying data sources. Aggregate-level data exist for all countries. To estimate average volume of alcohol consumption, we propose to base uncertainty analysis on the amount of survey information available in a subregion. The procedure allowed the estimation of lower or upper limits, even if no survey information existed. In Pakistan, for example, per capita alcohol consumption set clear upper boundaries for the highest drinking categories, since some values would

Table I2.4 Estimated proportions of population in average volume of consumption categories by sex, age and subregion, in $2000^{\text {a }}$

| Subregion | Sex | Average volume of consumption category ${ }^{\text {b }}$ | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | Abstinence | 0.510 | 0.462 | 0.514 | 0.565 | 0.616 | 0.616 |
|  |  | D I | 0.420 | 0.447 | 0.396 | 0.354 | 0.323 | 0.323 |
|  |  | D II | 0.063 | 0.073 | 0.072 | 0.063 | 0.044 | 0.044 |
|  |  | D III | 0.007 | 0.018 | 0.018 | 0.018 | 0.016 | 0.016 |
|  | Female | Abstinence | 0.746 | 0.697 | 0.697 | 0.746 | 0.796 | 0.796 |
|  |  | D I | 0.234 | 0.267 | 0.265 | 0.237 | 0.191 | 0.191 |
|  |  | D II | 0.019 | 0.030 | 0.029 | 0.010 | 0.010 | 0.010 |
|  |  | D III | 0.001 | 0.007 | 0.010 | 0.007 | 0.003 | 0.003 |
| AFR-E | Male | Abstinence | 0.432 | 0.375 | 0.428 | 0.482 | 0.536 | 0.536 |
|  |  | D I | 0.431 | 0.430 | 0.403 | 0.368 | 0.354 | 0.354 |
|  |  | D II | 0.125 | 0.161 | 0.135 | 0.116 | 0.082 | 0.082 |
|  |  | D III | 0.012 | 0.034 | 0.034 | 0.033 | 0.029 | 0.029 |
|  | Female | Abstinence | 0.715 | 0.664 | 0.664 | 0.715 | 0.766 | 0.766 |
|  |  | D I | 0.243 | 0.277 | 0.271 | 0.237 | 0.203 | 0.203 |
|  |  | D II | 0.036 | 0.046 | 0.046 | 0.036 | 0.025 | 0.025 |
|  |  | D III | 0.006 | 0.012 | 0.018 | 0.012 | 0.006 | 0.006 |
| AMR-A | Male | Abstinence | 0.240 | 0.221 | 0.268 | 0.374 | 0.431 | 0.431 |
|  |  | D I | 0.522 | 0.583 | 0.577 | 0.525 | 0.498 | 0.498 |
|  |  | D II | 0.164 | 0.146 | 0.105 | 0.095 | 0.064 | 0.064 |
|  |  | D III | 0.074 | 0.050 | 0.050 | 0.006 | 0.006 | 0.006 |
|  | Female | Abstinence | 0.360 | 0.331 | 0.409 | 0.584 | 0.633 | 0.633 |
|  |  | D I | 0.563 | 0.620 | 0.548 | 0.382 | 0.337 | 0.337 |
|  |  | D II | 0.050 | 0.033 | 0.027 | 0.025 | 0.023 | 0.023 |
|  |  | D III | 0.027 | 0.016 | 0.016 | 0.008 | 0.007 | 0.007 |
| AMR-B | Male | Abstinence | 0.220 | 0.203 | 0.172 | 0.199 | 0.358 | 0.358 |
|  |  | D I | 0.671 | 0.674 | 0.711 | 0.714 | 0.605 | 0.605 |
|  |  | D II | 0.035 | 0.045 | 0.045 | 0.031 | 0.025 | 0.025 |
|  |  | D III | 0.074 | 0.078 | 0.072 | 0.056 | 0.012 | 0.012 |
|  | Female | Abstinence | 0.463 | 0.444 | 0.444 | 0.488 | 0.538 | 0.538 |
|  |  | D I | 0.464 | 0.475 | 0.486 | 0.450 | 0.429 | 0.429 |
|  |  | D II | 0.028 | 0.028 | 0.026 | 0.030 | 0.020 | 0.020 |
|  |  | D III | 0.046 | 0.053 | 0.044 | 0.032 | 0.013 | 0.013 |
| AMR-D | Male | Abstinence | 0.298 | 0.264 | 0.304 | 0.398 | 0.498 | 0.498 |
|  |  | D I | 0.677 | 0.712 | 0.672 | 0.585 | 0.494 | 0.494 |
|  |  | D II | 0.017 | 0.016 | 0.016 | 0.013 | 0.008 | 0.008 |
|  |  | D III | 0.008 | 0.008 | 0.008 | 0.004 | 0.001 | 0.001 |
|  | Female | Abstinence | 0.431 | 0.393 | 0.487 | 0.548 | 0.628 | 0.628 |
|  |  | D I | 0.536 | 0.572 | 0.483 | 0.428 | 0.356 | 0.356 |
|  |  | D II | 0.024 | 0.025 | 0.021 | 0.019 | 0.012 | 0.012 |
|  |  | D III | 0.009 | 0.009 | 0.008 | 0.005 | 0.004 | 0.004 |
| EMR-B | Male | Abstinence | 0.789 | 0.838 | 0.838 | 0.887 | 0.976 | 0.976 |
|  |  | D I | 0.191 | 0.149 | 0.149 | 0.107 | 0.024 | 0.024 |
|  |  | D II | 0.013 | 0.010 | 0.010 | 0.004 | 0.000 | 0.000 |
|  |  | D III | 0.006 | 0.003 | 0.003 | 0.001 | 0.000 | 0.000 |

Table 12.4 Estimated proportions of population in average volume of consumption categories by sex, age and subregion, in 2000 ${ }^{\text {a }}$ (continued)

| Subregion | Sex | Average volume of consumption category ${ }^{\text {b }}$ | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EMR-D | Female | Abstinence | 0.937 | 0.968 | 0.968 | 0.998 | 1.000 | 1.000 |
|  |  | D I | 0.052 | 0.027 | 0.027 | 0.002 | 0.000 | 0.000 |
|  |  | D II | 0.008 | 0.005 | 0.005 | 0.000 | 0.000 | 0.000 |
|  |  | D III | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
|  | Male | Abstinence | 0.898 | 0.898 | 0.898 | 0.947 | 0.987 | 0.987 |
|  |  | D I | 0.101 | 0.102 | 0.102 | 0.052 | 0.013 | 0.013 |
|  |  | D II | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
|  |  | D III | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
|  | Female | Abstinence | 0.975 | 0.990 | 0.990 | 1.000 | 1.000 | 1.000 |
|  |  | D I | 0.024 | 0.010 | 0.010 | 0.000 | 0.000 | 0.000 |
|  |  | D II | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
|  |  | D III | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| EUR-A | Male | Abstinence | 0.074 | 0.074 | 0.093 | 0.139 | 0.186 | 0.186 |
|  |  | D I | 0.748 | 0.721 | 0.701 | 0.723 | 0.717 | 0.717 |
|  |  | D II | 0.087 | 0.102 | 0.096 | 0.069 | 0.042 | 0.042 |
|  |  | D III | 0.091 | 0.102 | 0.110 | 0.069 | 0.055 | 0.055 |
|  | Female | Abstinence | 0.138 | 0.138 | 0.173 | 0.259 | 0.346 | 0.346 |
|  |  | D I | 0.698 | 0.726 | 0.662 | 0.635 | 0.571 | 0.571 |
|  |  | D II | 0.123 | 0.103 | 0.124 | 0.085 | 0.065 | 0.065 |
|  |  | D III | 0.041 | 0.033 | 0.041 | 0.021 | 0.017 | 0.017 |
| EUR-B | Male | Abstinence | 0.237 | 0.284 | 0.284 | 0.331 | 0.379 | 0.379 |
|  |  | D I | 0.653 | 0.606 | 0.625 | 0.598 | 0.582 | 0.582 |
|  |  | D II | 0.054 | 0.054 | 0.049 | 0.038 | 0.016 | 0.016 |
|  |  | D III | 0.056 | 0.056 | 0.042 | 0.033 | 0.024 | 0.024 |
|  | Female | Abstinence | 0.410 | 0.461 | 0.512 | 0.614 | 0.614 | 0.614 |
|  |  | D I | 0.507 | 0.449 | 0.414 | 0.328 | 0.339 | 0.339 |
|  |  | D II | 0.068 | 0.069 | 0.053 | 0.047 | 0.037 | 0.037 |
|  |  | D III | 0.016 | 0.021 | 0.021 | 0.010 | 0.010 | 0.010 |
| EUR-C | Male | Abstinence | 0.088 | 0.088 | 0.110 | 0.164 | 0.219 | 0.219 |
|  |  | D I | 0.602 | 0.662 | 0.584 | 0.629 | 0.669 | 0.669 |
|  |  | D II | 0.188 | 0.159 | 0.185 | 0.132 | 0.067 | 0.067 |
|  |  | D III | 0.123 | 0.091 | 0.121 | 0.075 | 0.045 | 0.045 |
|  | Female | Abstinence | 0.140 | 0.140 | 0.175 | 0.263 | 0.351 | 0.351 |
|  |  | D I | 0.719 | 0.743 | 0.683 | 0.645 | 0.588 | 0.588 |
|  |  | D II | 0.115 | 0.096 | 0.116 | 0.079 | 0.051 | 0.051 |
|  |  | D III | 0.026 | 0.020 | 0.026 | 0.013 | 0.011 | 0.011 |
| SEAR-B | Male | Abstinence | 0.631 | 0.561 | 0.701 | 0.841 | 0.981 | 0.981 |
|  |  | D I | 0.359 | 0.410 | 0.295 | 0.156 | 0.018 | 0.018 |
|  |  | D II | 0.007 | 0.024 | 0.003 | 0.002 | 0.000 | 0.000 |
|  |  | D III | 0.002 | 0.006 | 0.001 | 0.001 | 0.000 | 0.000 |
|  | Female | Abstinence | 0.885 | 0.914 | 0.914 | 0.943 | 0.953 | 0.953 |
|  |  | D I | 0.103 | 0.078 | 0.078 | 0.052 | 0.047 | 0.047 |
|  |  | D II | 0.009 | 0.008 | 0.008 | 0.005 | 0.000 | 0.000 |
|  |  | D III | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 12.4 Estimated proportions of population in average volume of consumption categories by sex, age and subregion, in $2000^{a}$ (continued)

| Subregion | Sex | Average volume of consumption category ${ }^{\text {b }}$ | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| SEAR-D | Male | Abstinence | 0.717 | 0.637 | 0.796 | 0.956 | 1.000 | 1.000 |
|  |  | D I | 0.276 | 0.338 | 0.201 | 0.044 | 0.000 | 0.000 |
|  |  | D II | 0.006 | 0.020 | 0.002 | 0.001 | 0.000 | 0.000 |
|  |  | D III | 0.002 | 0.005 | 0.001 | 0.000 | 0.000 | 0.000 |
|  | Female | Abstinence | 0.937 | 0.968 | 0.968 | 0.998 | 1.000 | 1.000 |
|  |  | D I | 0.051 | 0.028 | 0.028 | 0.002 | 0.000 | 0.000 |
|  |  | D II | 0.007 | 0.004 | 0.004 | 0.000 | 0.000 | 0.000 |
|  |  | D III | 0.005 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| WPR-A | Male | Abstinence | 0.095 | 0.095 | 0.119 | 0.179 | 0.238 | 0.238 |
|  |  | D I | 0.834 | 0.849 | 0.812 | 0.776 | 0.730 | 0.730 |
|  |  | D II | 0.033 | 0.028 | 0.032 | 0.023 | 0.013 | 0.013 |
|  |  | D III | 0.037 | 0.028 | 0.037 | 0.023 | 0.018 | 0.018 |
|  | Female | Abstinence | 0.163 | 0.163 | 0.204 | 0.306 | 0.408 | 0.408 |
|  |  | D I | 0.806 | 0.812 | 0.766 | 0.675 | 0.578 | 0.578 |
|  |  | D II | 0.023 | 0.019 | 0.023 | 0.015 | 0.011 | 0.011 |
|  |  | D III | 0.008 | 0.006 | 0.008 | 0.004 | 0.003 | 0.003 |
| WPR-B | Male | Abstinence | 0.153 | 0.153 | 0.153 | 0.204 | 0.256 | 0.256 |
|  |  | D I | 0.769 | 0.761 | 0.761 | 0.719 | 0.684 | 0.684 |
|  |  | D II | 0.055 | 0.053 | 0.053 | 0.051 | 0.040 | 0.040 |
|  |  | D III | 0.022 | 0.032 | 0.032 | 0.026 | 0.020 | 0.020 |
|  | Female | Abstinence | 0.674 | 0.674 | 0.674 | 0.786 | 0.898 | 0.898 |
|  |  | D I | 0.326 | 0.326 | 0.326 | 0.214 | 0.102 | 0.102 |
|  |  | D II | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
|  |  | D III | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

[^45]Risk factor: alcohol-first dimension: average volume of consumption.
Units: grams of pure alcohol per day.
Definitions of categories of risk factor levels

| Level I: | Abstinence | abstainer |
| :--- | :--- | :--- |
| Level 2: | D I | drinking category I: women $>0-<20 \mathrm{~g} ;$ men $>0-<40 \mathrm{~g}$ |
| Level 3: | D II | drinking category II: women $20-<40 \mathrm{~g} ;$ men $40-<60 \mathrm{~g}$ |
| Level 4: | D III | drinking category III: women $>40 \mathrm{~g}$; men $>60 \mathrm{~g}$ |

not be physically possible within the overall volume consumed in this country.

In operationalizing these principles, the validity of the abstainer category was selected (for values in individual countries see Table 12.3). This variable was aggregated using the population aged $\geq 15$ years as a

Table 12.5 Patterns of drinking by subregion

| Subregion | Pattern value ${ }^{\mathrm{a}}$ | Validity of pattern value |
| :--- | :---: | :---: |
| AFR-D | 2.48 | 0.52 |
| AFR-E | 3.09 | 0.24 |
| AMR-A | 2.00 | 0.97 |
| AMR-B | 3.14 | 0.77 |
| AMR-D | 3.10 | 0.38 |
| EMR-B | 2.01 | 0.00 |
| EMR-D | 2.35 | 0.00 |
| EUR-A | 1.34 | 0.92 |
| EUR-B | 2.93 | 0.31 |
| EUR-C | 3.62 | 0.75 |
| SEAR-B | 2.50 | 0.22 |
| SEAR-D | 2.95 | 0.93 |
| WPR-A | 1.16 | 0.98 |
| WPR-B | 2.15 | 0.93 |

[^46]weight. The resulting values were treated in the uncertainty analysis as follows:

For aggregated validity values of 0 Base prevalence estimate on full range from 0 to twice the point estimate (i.e. point estimate plus $100 \%$ of the point estimate).

For aggregated validity values $>0$ and $<0.5$

For aggregated validity values $\geq 0.5$ and $<0.75$

Point estimate $\pm 50 \%$ of point estimate of prevalence.

Point estimate $\pm 25 \%$ of point estimate of prevalence.

For aggregated validity values $\geq 0.75$ Point estimate $\pm 10 \%$ of point estimate of prevalence.

This applies to both sexes, as most surveys include both males and females. Further corrections were made based on per capita consumption for subregions where certain distributions were not plausible because the known per capita consumption could not be derived with less than a certain proportion of abstainers (for EMR-B: maximum CI for abstainers $\pm 15 \%$, minimum: $\pm 1 \%$; for EMR-D: maximum $\pm 10 \%$, minimum: $\pm 1 \%$; for SEAR-B: maximum $\pm 10 \%$, minimum: $\pm 1 \%$; for SEAR-D and WPR-B: minimum: $\pm 0.2 \%)$. The results can be seen in Table 12.6.

Similarly, for constructing CIs around patterns of drinking (see Table 12.7), the validity of the underlying pattern values was used. Again, using the weighted average on the validity ratings of the pattern value (see Table 12.3 for the underlying country data), the following uncertainty ranges were proposed:

For validity of pattern value of 0 Uncertainty analysis on full range of pattern values (i.e. using pattern values 1 and 4 as bounds).

For validity of pattern value ranging from $>0$ and $<0.5$

For validity of pattern value ranging from $\geq 0.5$ and $<0.75$

For validity of pattern value ranging from $\geq 0.75$

Point estimate $\pm 0.5$.

Point estimate $\pm 0.25$.

The validity of pattern value ranged from 0 (e.g. only expert judgements for pattern in EMR-B and EMR-D with no underlying data) to values above 0.95 for several subregions (theoretically 1 if all the pattern values of constituent countries were derived from survey estimates).

The suggested procedure clearly reflects our range of knowledge. For example, we have no knowledge about patterns in EMR-B and EMR-D, where the uncertainty estimates consequently varied between 1 (best possible pattern) and 4 (most detrimental pattern), whereas we have good data for many other subregions (e.g. all the A mortality strata subregions), where the uncertainty estimates vary by only $\pm 0.25$ (e.g. EURA, between 1.09 and 1.59 ).

## Overall evaluation of quality for Exposure data

Data on production and trade or sales required for estimating adult per capita consumption have been collected systematically for decades in most countries. In this sense, alcohol as a risk factor for global health is privileged compared to some other risk factors, where exposure data are

Table I2.6 Estimated uncertainty range ( $\pm \%$ ) around point estimate for prevalence of average volume of consumption categories

| Subregion | Sex | Average volume of consumption category | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | Abstinence | 5.10 | 4.62 | 5.14 | 5.65 | 6.16 | 6.16 |
|  |  | D I | 4.20 | 4.47 | 3.96 | 3.54 | 3.23 | 3.23 |
|  |  | D II | 0.63 | 0.73 | 0.72 | 0.63 | 0.44 | 0.44 |
|  |  | D III | 0.07 | 0.18 | 0.18 | 0.18 | 0.16 | 0.16 |
|  | Female | Abstinence | 7.46 | 6.97 | 6.97 | 7.46 | 7.96 | 7.96 |
|  |  | D I | 2.34 | 2.67 | 2.65 | 2.37 | 1.91 | 1.91 |
|  |  | D II | 0.19 | 0.30 | 0.29 | 0.10 | 0.10 | 0.10 |
|  |  | D III | 0.01 | 0.07 | 0.10 | 0.07 | 0.03 | 0.03 |
| AFR-E | Male | Abstinence | 10.80 | 9.37 | 10.71 | 12.05 | 13.39 | 13.39 |
|  |  | D I | 10.77 | 10.75 | 10.08 | 9.21 | 8.84 | 8.84 |
|  |  | D II | 3.12 | 4.03 | 3.36 | 2.91 | 2.05 | 2.05 |
|  |  | D III | 0.31 | 0.85 | 0.84 | 0.83 | 0.72 | 0.72 |
|  | Female | Abstinence | 17.88 | 16.61 | 16.61 | 17.88 | 19.16 | 19.16 |
|  |  | D I | 6.07 | 6.93 | 6.78 | 5.92 | 5.07 | 5.07 |
|  |  | D II | 0.90 | 1.16 | 1.16 | 0.90 | 0.63 | 0.63 |
|  |  | D III | 0.15 | 0.30 | 0.45 | 0.30 | 0.15 | 0.15 |
| AMR-A | Male | Abstinence | 2.40 | 2.21 | 2.68 | 3.74 | 4.31 | 4.31 |
|  |  | D I | 5.22 | 5.83 | 5.77 | 5.25 | 4.98 | 4.98 |
|  |  | D II | 1.64 | 1.46 | 1.05 | 0.95 | 0.64 | 0.64 |
|  |  | D III | 0.74 | 0.50 | 0.50 | 0.06 | 0.06 | 0.06 |
|  | Female | Abstinence | 3.60 | 3.31 | 4.09 | 5.84 | 6.33 | 6.33 |
|  |  | D I | 5.63 | 6.20 | 5.48 | 3.82 | 3.37 | 3.37 |
|  |  | D II | 0.50 | 0.33 | 0.27 | 0.25 | 0.23 | 0.23 |
|  |  | D III | 0.27 | 0.16 | 0.16 | 0.08 | 0.07 | 0.07 |
| AMR-B | Male | Abstinence | 2.20 | 2.03 | 1.72 | 1.99 | 3.58 | 3.58 |
|  |  | D I | 6.71 | 6.74 | 7.11 | 7.14 | 6.05 | 6.05 |
|  |  | D II | 0.35 | 0.45 | 0.45 | 0.31 | 0.25 | 0.25 |
|  |  | D III | 0.74 | 0.78 | 0.72 | 0.56 | 0.12 | 0.12 |
|  | Female | Abstinence | 4.63 | 4.44 | 4.44 | 4.88 | 5.38 | 5.38 |
|  |  | D I | 4.64 | 4.75 | 4.86 | 4.50 | 4.29 | 4.29 |
|  |  | D II | 0.28 | 0.28 | 0.26 | 0.30 | 0.20 | 0.20 |
|  |  | D III | 0.46 | 0.53 | 0.44 | 0.32 | 0.13 | 0.13 |
| AMR-D | Male | Abstinence | 7.46 | 6.60 | 7.60 | 9.96 | 12.45 | 12.45 |
|  |  | D I | 16.93 | 17.80 | 16.80 | 14.62 | 12.34 | 12.34 |
|  |  | D II | 0.41 | 0.39 | 0.39 | 0.31 | 0.19 | 0.19 |
|  |  | D III | 0.20 | 0.20 | 0.20 | 0.10 | 0.02 | 0.02 |
|  | Female | Abstinence | 10.79 | 9.83 | 12.17 | 13.70 | 15.70 | 15.70 |
|  |  | D I | 13.40 | 14.31 | 12.09 | 10.70 | 8.90 | 8.90 |
|  |  | D II | 0.60 | 0.63 | 0.53 | 0.48 | 0.31 | 0.31 |
|  |  | D III | 0.22 | 0.23 | 0.21 | 0.12 | 0.09 | 0.09 |
| EMR-B | Male | Abstinence | 15.00 | 15.00 | 15.00 | 15.00 | 15.00 | 15.00 |
|  |  | D I | 19.55 | 15.25 | 15.25 | 11.00 | 2.65 | 2.65 |
|  |  | D II | 1.21 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  |  | D III | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

Table I2.6 Estimated uncertainty range ( $\pm \%$ ) around point estimate for prevalence of average volume of consumption categories (continued)

| Subregion | Sex | Average volume of consumption category | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EMR-D | Female | Abstinence | 15.00 | 15.00 | 15.00 | 15.00 | 15.00 | 15.00 |
|  |  | D I | 5.40 | 2.87 | 2.87 | 1.00 | 1.00 | 1.00 |
|  |  | D II | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  |  | D III | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | Male | Abstinence | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
|  |  | D I | 5.07 | 5.09 | 5.09 | 2.61 | 1.00 | 1.00 |
|  |  | D II | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  |  | D III | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | Female | Abstinence | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
|  |  | D I | 1.18 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  |  | D II | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  |  | D III | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| EUR-A | Male | Abstinence | 0.74 | 0.74 | 0.93 | 1.39 | 1.86 | 1.86 |
|  |  | D I | 7.48 | 7.21 | 7.01 | 7.23 | 7.17 | 7.17 |
|  |  | D II | 0.87 | 1.02 | 0.96 | 0.69 | 0.42 | 0.42 |
|  |  | D III | 0.91 | 1.02 | 1.10 | 0.69 | 0.55 | 0.55 |
|  | Female | Abstinence | 1.38 | 1.38 | 1.73 | 2.59 | 3.46 | 3.46 |
|  |  | D I | 6.98 | 7.26 | 6.62 | 6.35 | 5.71 | 5.71 |
|  |  | D II | 1.23 | 1.03 | 1.24 | 0.85 | 0.65 | 0.65 |
|  |  | D III | 0.41 | 0.33 | 0.41 | 0.21 | 0.17 | 0.17 |
| EUR-B | Male | Abstinence | 5.92 | 7.10 | 7.10 | 8.28 | 9.47 | 9.47 |
|  |  | D I | 16.34 | 15.14 | 15.63 | 14.94 | 14.55 | 14.55 |
|  |  | D II | 1.35 | 1.36 | 1.23 | 0.96 | 0.39 | 0.39 |
|  |  | D III | 1.39 | 1.40 | 1.05 | 0.82 | 0.59 | 0.59 |
|  | Female | Abstinence | 10.24 | 11.52 | 12.80 | 15.36 | 15.36 | 15.36 |
|  |  | D I | 12.66 | 11.22 | 10.35 | 8.20 | 8.47 | 8.47 |
|  |  | D II | 1.70 | 1.73 | 1.32 | 1.18 | 0.91 | 0.91 |
|  |  | D III | 0.39 | 0.53 | 0.53 | 0.26 | 0.26 | 0.26 |
| EUR-C | Male | Abstinence | 2.19 | 2.19 | 2.74 | 4.11 | 5.48 | 5.48 |
|  |  | D I | 15.05 | 16.55 | 14.60 | 15.72 | 16.72 | 16.72 |
|  |  | D II | 4.69 | 3.98 | 4.63 | 3.29 | 1.68 | 1.68 |
|  |  | D III | 3.07 | 2.27 | 3.03 | 1.88 | 1.12 | 1.12 |
|  | Female | Abstinence | 3.51 | 3.51 | 4.39 | 6.58 | 8.77 | 8.77 |
|  |  | D I | 17.97 | 18.58 | 17.07 | 16.12 | 14.69 | 14.69 |
|  |  | D II | 2.88 | 2.40 | 2.90 | 1.97 | 1.27 | 1.27 |
|  |  | D III | 0.64 | 0.51 | 0.64 | 0.33 | 0.27 | 0.27 |
| SEAR-B | Male | Abstinence | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
|  |  | D I | 17.96 | 20.48 | 14.73 | 7.80 | 1.00 | 1.00 |
|  |  | D II | 0.37 | 1.18 | 1.00 | 1.00 | 1.00 | 1.00 |
|  |  | D III | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | Female | Abstinence | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 | 10.00 |
|  |  | D I | 5.16 | 3.88 | 3.88 | 2.62 | 2.36 | 2.36 |
|  |  | D II | 0.46 | 0.41 | 0.41 | 1.00 | 1.00 | 1.00 |
|  |  | D III | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| SEAR-D | Male | Abstinence | 7.17 | 6.37 | 7.96 | 9.56 | 10.00 | 10.00 |
|  |  | D I | 2.76 | 3.38 | 2.01 | 0.44 | 0.20 | 0.20 |
|  |  | D II | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 |
|  |  | D III | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 |

Table I2.6 Estimated uncertainty range ( $\pm \%$ ) around point estimate for prevalence of average volume of consumption categories (continued)

| Subregion | Sex | Average volume of consumption category | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| WPR-A | Female | Abstinence | 9.37 | 9.68 | 9.68 | 9.98 | 10.00 | 10.00 |
|  |  | D I | 0.51 | 0.28 | 0.28 | 0.20 | 0.20 | 0.20 |
|  |  | D II | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 |
|  |  | D III | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 |
|  | Male | Abstinence | 0.95 | 0.95 | 1.19 | 1.79 | 2.38 | 2.38 |
|  |  | D I | 8.34 | 8.49 | 8.12 | 7.76 | 7.30 | 7.30 |
|  |  | D II | 0.33 | 0.28 | 0.32 | 0.23 | 0.13 | 0.13 |
|  |  | D III | 0.37 | 0.28 | 0.37 | 0.23 | 0.18 | 0.18 |
| WPR-B | Female | Abstinence | 1.63 | 1.63 | 2.04 | 3.06 | 4.08 | 4.08 |
|  |  | D I | 8.06 | 8.12 | 7.66 | 6.75 | 5.78 | 5.78 |
|  |  | D II | 0.23 | 0.19 | 0.23 | 0.15 | 0.11 | 0.11 |
|  |  | D III | 0.08 | 0.06 | 0.08 | 0.04 | 0.03 | 0.03 |
|  | Male | Abstinence | 1.53 | 1.53 | 1.53 | 2.04 | 2.56 | 2.56 |
|  |  | D I | 7.69 | 7.61 | 7.61 | 7.19 | 6.84 | 6.84 |
|  |  | D II | 0.55 | 0.53 | 0.53 | 0.51 | 0.40 | 0.40 |
|  |  | D III | 0.22 | 0.32 | 0.32 | 0.26 | 0.20 | 0.20 |
|  | Female | Abstinence | 6.74 | 6.74 | 6.74 | 7.86 | 8.98 | 8.98 |
|  |  | D I | 3.26 | 3.26 | 3.26 | 2.14 | 1.02 | 1.02 |
|  |  | D II | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 |
|  |  | D III | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 |

Table 12.7 Uncertainty analysis for patterns of drinking, by subregion

| Subregion | Pattern value | Validity of <br> pattern value | Lower boundary <br> for uncertainty | Upper boundary <br> for uncertainty |
| :--- | :---: | :---: | :---: | :---: |
| AFR-D | 2.48 | 0.52 | 1.98 | 2.98 |
| AFR-E | 3.09 | 0.24 | 2.09 | 4.00 |
| AMR-A | 2.00 | 0.97 | 1.75 | 2.25 |
| AMR-B | 3.14 | 0.77 | 2.89 | 3.39 |
| AMR-D | 3.10 | 0.38 | 2.10 | 4.00 |
| EMR-B | 2.01 | 0.00 | 1.00 | 4.00 |
| EMR-D | 2.35 | 0.00 | 1.00 | 4.00 |
| EUR-A | 1.34 | 0.92 | 1.09 | 1.59 |
| EUR-B | 2.93 | 0.31 | 1.93 | 3.93 |
| EUR-C | 3.62 | 0.75 | 3.37 | 3.87 |
| SEAR-B | 2.50 | 0.22 | 1.50 | 3.50 |
| SEAR-D | 2.95 | 0.93 | 2.70 | 3.20 |
| WPR-A | 1.16 | 0.98 | 1.00 | 1.66 |
| WPR-B | 2.15 | 0.93 | 1.90 | 2.40 |

available only in established market economies, if at all. Unfortunately, adult per capita consumption figures per se, as described above, cannot be used as exposure data for global burden of disease because these figures do not give any specification about who consumed alcohol, when and in what quantities. This can only be specified by using representative surveys of the general population. Thus, for the current exercise, adult per capita figures had to be supplemented by survey data and, where such data were not available, by expert judgements. This procedure may have introduced errors at different points.

- Errors in adult per capita consumption estimates. Because per capita estimates are derived mainly from production and trade or sale, other sources (home brewing, cross-border smuggling, illegal production, etc.) are often ignored. There is a long tradition of trying to estimate this unrecorded consumption, but mainly in established market economies where the proportion of unrecorded to recorded is low relative to other parts of the world. For example, see the recent European Comparative Alcohol Study (ECAS) estimates for Europe (http://www.fhi.se/pdf/ECAS_2.pdf; Leifman 2001) compared to the estimates of unrecorded consumption for sub-Saharan Africa given below. In addition, estimation of unrecorded consumption is particularly difficult in regions where alcohol is prohibited for religious reasons. In summary, although these sources of error exist, the global aggregate estimates on adult per capita consumption are among the best and most reliable sources for global risk factors.

Adult per capita consumption figures then were used to derive prevalence rates for specific drinking categories for different age-sex groups. This derivation was based on survey data with the following potential sources of error.

- Survey measurement error. Alcohol surveys are subject to measurement errors that apply generally to surveys (for an overview see Groves 1989). However, the most important sources of error relating to estimating alcohol consumption include non-probabilistic sampling, sampling schemes that exclude groups with a high alcohol consumption, and measurement bias. The major problem with survey estimates of alcohol consumption is that surveys generally account for $66 \%$ or less of alcohol produced or sold (Midanik 1988; Midanik and Harford 1994; Rehm 1998a; de Vries et al. 1999). However, combining survey and aggregate (i.e. per capita) estimates may ameliorate the problem of underreporting that is characteristic of survey estimates.
- Errors in combining adult per capita and survey data. Although combining survey and aggregate data can rectify the underreporting of survey data, other sources of error may be created by this procedure. This is especially true if the difference in overall adult consumption between aggregate and survey estimates is large. There are theoreti-
cally several ways in which the survey estimates could be used to arrive at the per capita figures. For example, the prevalence of all drinking categories could be proportionally increased to cover the difference between survey and per capita estimates. This is not possible, however, as the rate of abstainers would automatically change as well. Thus, a procedure must be selected to keep the proportion of abstainers from the survey fixed and proportionally increase the higher consumption categories at the expense of the middle category. Even if this procedure is plausible it may introduce error, such as when survey underreporting differs for different age-sex groups. Nevertheless, the overall benefits of combining surveys and adult per capita data seem to outweigh this disadvantage.

There are other potential sources of error in the exposure estimates.

- Errors in estimating missing survey data. Survey data are available for only some countries, so that regional figures had to be estimated from selected countries where surveys were available. This introduces error, and overall the lack of data for some countries is probably the most important source of error for exposure estimates. The more survey data are lacking, the more severe is this problem. Thus it is least severe for those who abstain from alcohol, because even when survey data on the amount of alcohol consumed were not available, many countries had survey data in which respondents were asked whether or not they consumed alcohol. The problem of error in estimating missing survey data was most severe for estimates of drinking patterns, where expert judgements had to be used to supplement survey results, and where no production and trade or sale estimates were available for triangulation.

Overall, this listing of sources for potential errors and biases clearly indicates that the results of CRA, while based on the best available sources, are accompanied by large uncertainties, only some of which could be quantified.

## 3. ESTIMATING RISK FACTOR-DISEASE RELATIONSHIPS

### 3.1 OUTCOMES TO BE ASSESSED, EVIDENCE OF CAUSALITY and ExClusions

Average volume of alcohol consumption has been related to more than 60 categories of the ninth revision of the ICD (ICD-9). This review restricts itself to categories that have already been identified in systematic meta-analyses, together with depression, which we describe in more detail below. Specifically, the following meta-analytical reviews were used: Gutjahr et al. 2001; English et al. 1995; Ridolfo and Stevenson 2001; Single et al. 1996, 1999a.

The categories that have been selected in modelling the impact of patterns of drinking are based on our own literature reviews. The analyses involving drinking patterns include the two main ICD categories for which sufficient evidence of a causal link to drinking patterns has been established: IHD and injuries. In addition, we restricted modelling of the protective effects of moderate regular drinking on type II diabetes and stroke to established market economies with the best drinking patterns (AMR-A, EUR-A and WPR-A), as there is evidence that patterns influence these diseases as well (for detailed reasoning, see below).

## ASSESSMENT OF CAUSALITY

Following the procedure described by English et al. (1995), the evidence of causality between alcohol consumption and disease outcomes (including both harmful and protective effects for particular diseases) is assessed in accordance with the Australian National Health and Medical Research Council's (NHMRC) Guidelines for the development, implementation and evaluation of clinical practice guidelines, which are the most used in the alcohol field and close to the criteria of Hill (1965). Sufficient evidence of causality includes outcomes for which the evidence indicates that an association (positive or negative) exists between alcohol consumption and the disease or injury and that chance, confounding variables and other bias can with reasonable confidence be ruled out as factors in this association. This judgement was made using the usual criteria for establishing causality in epidemiology (Hill 1965; Rothman and Greenland 1998a), with the most weight placed on the following four criteria:

- consistency across several studies;
- established experimental biochemical evidence of mediating processes, or at least physiological plausibility;
- strength of the association (effect size); and
- temporality (i.e. cause before effect).

Two examples of judgements regarding somewhat controversial outcomes may illustrate this process. For lung cancer, meta-analysis showed a consistent effect with a relatively large effect size (English et al. 1995; but see the meta-analysis of Bagnardi et al. 2001, which came to a different conclusion), after adjusting for smoking in at least some studies. However, evidence for the possible biological mechanisms is not conclusive at present (Bandera et al. 2001) and residual confounding from smoking cannot be excluded as an alternative explanation. Thus, consistent with a recent review of the epidemiological evidence, the evidence for alcohol causing lung cancer was not judged sufficient to establish causality according to the criteria listed above (Bandera et al. 2001), and thus lung cancer was excluded as an alcohol-related disease outcome in this work.

On the other hand, although English et al. (1995) concluded that there was not sufficient evidence linking alcohol consumption and breast cancer, recent advances both in biological and epidemiological research have changed this evaluation (especially Smith-Warner et al. 1998; Singletary and Gapstur 2001; see below for detailed reasoning), so that breast cancer now is included in the list of alcohol-related outcomes.

It was concluded that there was limited evidence of causality when an association (positive or negative) was observed between alcohol consumption and the disease or injury for which a causal interpretation was considered to be credible, although chance, confounding variables or other bias cannot be ruled out with reasonable confidence. These diseases were not included in determining either alcohol-related mortality or burden of disease.

It was concluded that there was inadequate evidence of causality when the available studies were of insufficient quantity, consistency or statistical power to permit a conclusion regarding the presence or absence of a causal connection.

It was concluded that there was evidence suggesting lack of causality when several adequate studies, covering the full range of levels of alcohol consumption in the population, indicated a lack of a relationship (positive or negative) between alcohol consumption and the disease or injury. This conclusion is inevitably limited to diseases and injuries, levels of consumption and lengths of observations covered by the available studies, and the possibility of very small risks at the levels of exposure studied can never be excluded.

## EXCLUDED OUTCOMES AND REASONS FOR EXCLUSION

A list of outcomes considered and excluded can be found in English et al. (1995). Some of these outcomes had been reconsidered in subsequent reviews (Gutjahr et al. 2001; Ridolfo and Stevenson 2001; Single et al. 1999a) and we followed Gutjahr et al. (2001) for the final selection of outcomes. Otherwise, the only restriction was one of age. Outcomes for people under 15 years of age who had been drinking themselves and subsequent alcohol-related consequences were excluded for two reasons:

- there are not enough global data on drinking in these age groups; and
- the epidemiological basis linking drinking to health outcomes is scarce.

Based on these considerations, the conservative choice was made not to include these outcomes.

As stated above, this does not mean that no alcohol-related outcomes are calculated for the age group under 15 years. On the contrary, socalled second-hand effects of alcohol (somebody else's drinking causing alcohol-related harm to a person) are included in the estimation.

### 3.2 Overview of methods

Meta-analyses on average volume of Drinking and disease
Meta-analyses were the bases for estimating the risk relationships between average volume of alcohol consumption and chronic disease. In alcohol epidemiology, there is a tradition of conducting such metaanalyses as part of social cost studies (English et al. 1995; Gutjahr et al. 2001; Ridolfo and Stevenson 2001; Single et al. 1999a). We used the results of the most recent existing meta-analyses (Gutjahr et al. 2001) for most outcomes, and Ridolfo and Stevenson (2001) for breast cancer and the subtypes of stroke. Details such as inclusion and exclusion of studies are described below.

## Multilevel modelling to determine pattern weights

For estimating the contribution of patterns of drinking applied to IHD and injuries, a new methodology had to be developed. Details of this method are described and discussed in Rehm and Gmel (2000b) and Gmel et al. (2001). It consists of combining multilevel ${ }^{12}$ models with pooled cross-sectional time series models, and aims at maximizing methodological rigour and practical feasibility for the data available to undertake such an analysis. The principal variables for the present study are time series of adult per capita alcohol consumption and mortality for different countries, and one value measuring drinking patterns in a country. One could imagine conducting time series analyses separately for each country and then relating the patterns measured with the different estimates obtained for the relationship between alcohol consumption and mortality. However, time series data on both mortality and alcohol consumption for any single country are commonly too short (Rehm and Gmel 2001a) to perform reliable estimation and hypothesis testing, and there are no time-specific data on patterns of drinking within a country to determine pattern weights in this manner. Even for established market economies, there are at most 40 consecutive years of data from the same source (see e.g. Table 12.12 for the length of time series on IHD mortality and adult per capita consumption). In many developing and emerging economies, however, if data are available at all, they cover less than 10 consecutive years. Similarly, for many countries, owing to social and political changes (e.g. the countries of the former Soviet Union), longer time series inherently do not exist.

A common strategy to overcome such data shortcomings for parameter estimation is pooling of data across countries ${ }^{13}$ to increase overall sample size, and to set constraints for several parameters (Greene 2000). Examples of such constraints could be to assume similar (i.e. differences statistically not significant) error variances (homoscedasticity) across countries, or to estimate the same regression parameters across all countries, or to assume a distribution for varying parameters (e.g. fixed vs random).

The approach adopted for the present study consists of pooling time series of differing lengths across countries and conducting multilevel analysis. For this approach, statistical problems arise from three major sources.

1. Pooling of data may violate the independence assumption in regression analysis. This concerns at least two aspects. First, the variability of measurements made in different countries is usually much greater than the variability of measurements within one country. Thus, data points are not independent of each other. In addition, within-country variability itself may differ among countries (heteroscedastic variances). Second, as countries provide different numbers of disease and consumption data points, countries with more data points (mainly established market economies) would have a higher impact on an overall estimate of the association between alcohol consumption and mortality than countries for which fewer data points exist (mainly developing and emerging economies). Such a disproportionate contribution to an overall regression estimate would be aggravated by the fact that the population sizes of countries would be often inversely related to the number of data points contributed to the regression analysis by such countries. For IHD mortality, for example, Switzerland ( 31 data points) and Norway ( 33 data points) would far outweigh the more populous South Africa (3 data points) and the Russian Federation ( 8 data points), even though they do not contribute substantially to the global burden of disease.
2. The estimation of time series needs further attention as regards to stationarity (stochastic or deterministic trends in the data) and the correlation of residuals within a country over time (autocorrelation).
3. The use of pooled cross-sectional time series analysis complicates estimation compared to a single time series alone, as errors (or residuals, both are used synonymously in this chapter) may co-vary across sections, and the degree of stationarity or autocorrelation may vary across sections (see below).

It should be noted that it was not possible to estimate all potential effects simultaneously. Either estimation of parameters failed to converge (potential explanations will be discussed below), the available standard software did not include the estimation of more complex models, or estimation of models was not applicable in the present context. To give an example of the latter, the estimation of cross-sectional correlation of errors would need time series that are equal in length (so-called balanced panels), at least with the available software used for estimation (Shazam: Whistler et al. 2001; STATA: StataCorp 1999; HLM: Bryk et al. 1996; MLWin: Yang et al. 1999). It was not possible to test models that include
cross-sectional autocorrelation in the present context, since balanced panel data are not available.

As not all potential shortcomings and drawbacks of multilevel and pooled cross-sectional time series models could be considered in one single model, sensitivity analysis was performed in two phases.

In phase 1, to determine pattern weights, multilevel analyses were conducted using a pilot sample of 29 European countries ${ }^{14}$ for which data were available for at least three consecutive years in the 1990s on each of the following variables: adult per capita alcohol consumption, standardized mortality and per capita GNP (level-1 variables) and an estimate of patterns of drinking (level-2 variable assumed to be timeinvariant). Calendar year was used to control for omitted variable bias and the time structure (Gmel et al. 2001); ${ }^{15}$ per capita GNP was included to control for economic strength as a potential confounder. The analyses to elucidate pattern weights for IHD and injuries have, for the present work, used the same methodology (for details see Gmel et al. 2001). For the sensitivity analysis and for the final pattern weights to calculate burden of disease in this chapter, data were taken from the following sources.

- Mortality data (either all-cause mortality in the sensitivity analysis or IHD/injury) were obtained from the WHO Mortality Database and age-standardized using United Nations population estimates. Direct standardization of mortality rates was performed using the latest WHO world standard population (Ahmad et al. 2000). The reference population is quite "young" with regard to the age distributions of populations in established market economies, but better reflects developing and emerging economies. On the other hand, the new WHO standard takes into account the reduced mortality rates in the older age groups nowadays, which have made the distribution a little "older" than the formerly widely used SEGI standard (Segi 1960).
- Adult per capita alcohol consumption data were again taken from WHO Global Alcohol Database described above.
- Per capita GNP data were taken from World Bank statistics, which used the Atlas method to arrive at standardized, de-inflated values in US dollars for the year 2000.

For all countries in this phase, time series data were collected from 1962 onwards on the four level- 1 variables-standardized mortality (either all-cause mortality in the sensitivity analyses or IHD and injury in the final estimates), calendar year, adult per capita alcohol consumption and per capita GNP.

In phase 2 , different models were run to especially take into account the cross-sectional time series structure, i.e. to assess the impact of eventually violating statistical assumption in hierarchical multilevel models.

For this exercise, data on countries where a long time series exists were used to try to estimate effects such as heteroscedasticity and error structure (autocorrelation) with some precision. Per capita consumption and data on all-cause mortality for 15 countries were obtained from ECAS, and are extensively described in the February 2001 supplement to Addiction (e.g. Norström 2001). Briefly, per capita alcohol consumption, measured in litres of pure alcohol per inhabitant aged $\geq 15$ years, was obtained from the Brewers Association of Canada (1997). Age-specific data on all-cause mortality were obtained from the WHO Mortality Database and standardized to the WHO 1998 standard population (Ahmad et al. 2000). For all countries, time series were available for a time span of at least 40 years. Worldwide, few if any countries exist that would be able to provide longer time series (say 60 or more years) of annual data.

## Sensitivity analysis, phase 1: multilevel models

The main difference between multilevel models and simple regression models is that the association between the outcome (mortality) and the independent variables (adult per capita consumption; calendar year as a control variable) is not fixed but varies across cross-sections (i.e. countries). Therefore, multilevel modelling analytically takes into account the lack of independence stemming from a potentially lower variability of data within a country compared to the variability between countries.

Essentially, this problem is the same as in cluster sampling, where variation between clusters is usually also higher than variation within clusters. Two approaches to account for cluster sampling have been suggested (see Lehtonen and Pahkinen 1994). One, the design-based approach, treats cluster sampling and the effect of it on parameter estimates as a nuisance. The information on variation across different clusters and variation within clusters is not included in the model. The estimation of CIs for relevant parameters (e.g. slopes and intercepts in regression analysis) without considering cluster sampling, however, is usually biased. Commonly, standard errors of parameters are underestimated, i.e. CIs are estimated to be narrower than they really are. The design-based approach accounts for these effects by adjusting the standard errors for the impact of the design, e.g. the cluster sampling, but does not model the effects explicitly. The other approach, the modelbased approach, treats the different variability across clusters as one parameter of interest in the study and explicitly models this variability. For this approach, multilevel (random coefficient) models are used. Comparison of the two approaches has shown that they yield similar results, but that the model-based (random coefficient) approach yields additional information as it is partly able to model the variability of estimates, and therefore to explain it (Skinner et al. 1989).

Our modelling is one variant of the model-based random coefficient approach. The rationale of multilevel models in the present context is described in detail by Rehm and Gmel (2000b); see also Bryk and Raudenbush (1992); Hox (1995); Kreft and de Leeuw (1998). Briefly, intercept and slopes are assumed to vary randomly ${ }^{16}$ across sections. In a two-level analysis this variation can be predicted by variables at the level of the cross-sections (e.g. per subregion). In the present study the variation in the slopes of adult per capita consumption to predict mortality will be explained by drinking patterns per country. Thus, it is assumed that drinking patterns moderate the association between adult per capita consumption and mortality. The corresponding model (without control variables) can be described as follows:

$$
\begin{equation*}
\text { mortality_rate }_{\mathrm{tc}}=\beta_{0 \mathrm{c}}+\beta_{1 \mathrm{c}} \times \text { alcohol }_{\mathrm{tc}}+\varepsilon_{\mathrm{tc}} \tag{1}
\end{equation*}
$$

where $t=$ index of time
$c=$ index of countries
The coefficients $\beta_{0 c}$ and $\beta_{1 c}$ symbolize random variables, which vary across countries. Therefore, the simplest way to model this random variation is given by the following equation:

$$
\begin{equation*}
\beta_{00}=\gamma_{0 c}+\mu_{0 c} \tag{2}
\end{equation*}
$$

where $\quad \gamma_{00}=$ global intercept of mortality rate
$\mu_{0 c}=$ country-specific variation of intercepts of mortality rates
similarly $\beta_{1 \mathrm{c}}=\gamma_{10}+\mu_{1 c}$
where $\quad \gamma_{10}=$ global slope of impact of alcohol (level-2 intercept of alcohol)
$\mu_{1 \mathrm{c}}=$ country specific variation of alcohol impact
Influences of country-specific drinking patterns on the slopes of per capita consumption can then be modelled as:

$$
\begin{equation*}
\text { pattern_weight }=\beta_{1 \mathrm{c}}=\gamma_{10}+\gamma_{11} \times \text { pattern_value }{ }_{c}+\mu_{1 \mathrm{c}} \tag{3}
\end{equation*}
$$

The impact of one unit of per capita consumption on mortality is thus assumed constant in time but specific for each country, and is denoted by $\beta_{1 c}$. This impact itself (throughout this chapter called "pattern weight") is regarded both as a dependent variable and as a predictor variable, for which the value not only varies across countries but also systematically depends on drinking pattern (i.e. pattern value) observed in the respective country.

The coefficient $\gamma_{10}$ can be regarded as a baseline measure for the impact of per capita consumption on mortality. ${ }^{17}$ The country-specific deviations
are introduced by the patterns of drinking in the country and a country specific error term $\mu_{1 c}$.

Note that if we estimated, in a simple one-level analysis, a global coefficient $\beta_{1}$ (not varying across countries), this measure would not necessarily be the same as $\gamma_{10}$ in a random coefficients model including level-2 variables. The coefficient $\gamma_{11}$ can be regarded as the contribution of drinking patterns to modifying the detrimental effects of per capita consumption on mortality. This modifying effect per unit of drinking patterns is treated as being the same for all countries. But clearly drinking patterns vary between countries, and therefore the impact on mortality $\beta_{1 \mathrm{c}}$ is specific to each country. There may remain unexplained variation in these country-specific impact coefficients. This is expressed by the level- 2 error term $\mu_{1 \mathrm{c}}$.

Figure 12.2 shows the modifications in residuals for different models. Figure 12.2(a) reflects a simple regression model in which all countries were simply pooled into one data set and analysis was performed without any multilevel modelling (i.e. standard multiple regression; in terms of multilevel modelling this would be a model with constant intercepts and constant slopes). Vertical lines in Figure 12.2 separate residuals within a country from those in other countries. It can be seen that residuals $\left(\varepsilon_{\mathrm{c}}\right)$ are clearly not independent, as for some countries all residuals were greater than zero whereas for other countries all residuals were below zero. In addition, residuals within a country clearly exhibit trends over time, which also violates the independence assumption. In Figure 12.2(b), a random intercept model was used. This means that for each country a separate intercept was estimated but the association between alcohol consumption and mortality was constant (i.e. it was estimated that the global level of mortality differed across countries, but the association between alcohol consumption and mortality was assumed to be the same across all countries). In this model residuals were more clustered around zero (owing to the random intercept), although there were still trends in the residuals. In the final model also, slopes were allowed to vary across countries (i.e. the association between alcohol consumption and mortality could vary across countries). In addition, control for confounding (GNP as constant slope term) and for omitted variable bias and potential deterministic trend stationarity (calendar year as a random slope term) was included in the multilevel model. The residuals of such a model (Figure 12.2 [c]) look very much as they should, as they exhibit neither visible trends nor autocorrelation. Thus, such a model not only turned out to be feasible but it also seemed to account for most of the undesirable aspects of residuals in time series analysis, such as nonstationarity and autocorrelated residuals. Nevertheless, the variability of errors still seems to be different across countries (heteroscedasticity).

Figure I2.2 Level-I residuals of multilevel models: a) constant intercept and constant alcohol slope; b) random intercept and constant alcohol slope; c) random intercept and random slope (including control for confounding, i.e. constant slope for GNP and random slope for calendar year)

| (a) |  <br> Data points |
| :---: | :---: |
| (b) |  |
| (c) |  |

[^47]
## Sensitivity analysis, phase 2: pooled cross-sectional time series models

The models discussed above could also be seen as pooled cross-sectional time series models in so far as data used consisted of time series. The models discussed here, however, include special features to deal with the structure of residuals, namely correlations in time. For the models discussed above, the order of data points in time within a country is not taken into account and could be randomly rearranged within each country. In time series analysis, however, it is often assumed that (within each country) values that are closer in time may be more correlated than values that are more separated in time. Thus, the time span between two values must be taken into account. This could not be done by the multilevel modelling, only indirectly by including calendar year into the models.

The sensitivity analyses of different pooled cross-sectional time series models used data on 15 countries from the ECAS project. As a starting point we analysed the data with multilevel models as outlined for Figure 12.2(c), using three different estimation techniques with three different statistical software packages ( 2 -stage OLS with STATA; iterative GLS with MLWin; restricted ML with HLM). Parameter estimates across the three models were comparable and differed less than those from models discussed below (for details of findings see Gmel et al. 2001; for details of differences in estimators see Kreft and de Leeuw 1998).

The simplest model of a pooled time series design, called the "constant coefficient model" (Sayrs 1989) or "population-averaged model with independent errors" (StataCorp 1999), would stack all observations across time points and cross-sections into one data file and analyse the combined data by standard regression techniques (e.g. OLS regression for an interval-scaled dependent variable). Such a model would assume that observations across time and cross-sections are completely independent of each other. This model equals a "multilevel" model with a single fixed intercept and a single fixed slope. This means that neither the ordering in time nor the grouping within cross-sections (countries) must be obeyed and, hence, that there is no association between the time points within a cross-section or between time points over cross-sections, and that there is no relationship between the cross-sections within a time point or between time points. The constant coefficient model often serves as a reference model only. It could be written as follows.

1. Zero expectation of errors for all cross-sections:

$$
\mathrm{E}\left(\varepsilon_{\mathrm{ct}}\right)=0 \quad \text { for all } \mathrm{c}, \mathrm{t}
$$

2. Constant error variance for all cross-sections:

$$
\mathrm{V}\left(\varepsilon_{\mathrm{ct}}\right)=\sigma^{2} \quad \text { for all } \mathrm{c}, \mathrm{t}
$$

3. Uncorrelatedness of errors within and across cross-sections:

$$
\operatorname{COV}\left(\varepsilon_{i t}, \varepsilon_{\mathrm{it}}\right)=0 \quad \text { for any } \mathrm{i}, \mathrm{j}, \mathrm{t}
$$

where $t=$ index of time
$c, i, j=$ index of countries, where it $\neq j t$
Assumptions 1 and 2 would be violated when, for example, mortality in a country is always higher than the overall prediction across all countries and all time points. For such a country all errors would be positive. Similarly, the error variances may vary across countries (heteroscedasticity) owing to the fact that, for example, mortality is measured with different reliability in different countries. Typical for time series data, errors may be correlated within a country (autocorrelation within cross-section) but also at the same time point across countries, which would violate assumption 3. An example of such a pattern may be seen in the Nordic countries, where there are similar alcohol policies.

Related to the corresponding procedures in STATA (StataCorp 1999) two sets of models were run. The first set uses GEE estimation (Liang and Zeger 1986) and allows different descriptions of the correlation matrix within cross-sections, subject to the constraint that the same correlation matrix applies to all cross-sections. The following models were used, with $R_{t, s}$ being the $t, s$ element of the correlation matrix, where $t$ and $s$ describe time points (here, years).

1. The independence structure (i.e. $\mathrm{R}_{\mathrm{t}, \mathrm{s}}=1$ for $\mathrm{t}=\mathrm{s}$ and 0 otherwise) is equivalent to a model for which all observations are pooled into one file and analysed as if all data come from the same underlying population.
2. The autoregressive (AR) structure (i.e. for an autoregressive structure of order $1, R_{t, s}=1$ for $t=s$ and $\rho^{|t-s|}$ otherwise) models an exponentially decaying correlation in time within a cross-section, hence assuming that the less observations are correlated the more they are separated in time.
3. The stationary structure (i.e. for stationarity of order $1, R_{t, s}=1$ for $\mathrm{t}=\mathrm{s}, \rho$ for $|\mathrm{t}-\mathrm{s}|=1$, and 0 otherwise) permits a correlation only between two consecutive time points.
4. The nonstationary structure (i.e. for nonstationarity of order $\mathrm{g}, \mathrm{R}_{\mathrm{t}, \mathrm{s}}=1$ for $t=s, \rho_{t s}$ for $\left.g \geq|t-s|>0\right)$ permits correlations for all observations separated by up to $g$ time points. The correlation may differ with the number of time points between two observations, and with the location in time (i.e. a different correlation between 1950 and 1951 and between 1974 and 1975). A completely unconstrained correlation matrix would be a nonstationary structure with $\mathrm{g}=\mathrm{n}-1$ (number of time points).

The second set of models uses GLS estimation. Compared with the GEE models, this method relaxes the restrictions of sameness within cross-section correlation matrices. Hence, it permits the estimation of heteroscedastic variances across sections, and the estimation of cross-
sectional specific autoregression (i.e. each country is allowed to have its own error variance and its own magnitude of autocorrelation). The models are restricted to autoregressive models of order 1.

Thus, the GEE models in STATA permit a greater flexibility in estimating error structures, but assume that these structures are constant for all countries. On the other hand, the GLS models in STATA allow the coefficients of autocorrelated residuals and the error variance to vary across countries, but are restricted to an autocorrelation of order 1 only. None of these models allows for testing of random coefficient models. Thus, the association between alcohol consumption and mortality is assumed to be constant across all countries. The same is true of variables to adjust for confounding, omitted variable bias and time trends (GNP and calendar time). Models were used with and without inclusion of calendar time.

The findings can be summarized as follows.

- Measurements within countries are highly correlated in such a way that accounting for correlations only at lags 1 or 2 were not sufficient (e.g. models with stationary or nonstationary structure up to order 2). Thus, the correlations of higher order remained significant. Better fits were obtained, for example, under the assumption of an autoregressive structure, i.e. exponentially decaying correlations with increasing time span between measurements. In general, point estimates of coefficients for the alcohol-mortality association increased with better control of this autoregressive structure, indicating that insufficient control would bias estimates of this association downwards.
- The inclusion of calendar time as a constant coefficient control variable acted in the same direction as controlling for autoregression and clearly performed better than controlling for nonstationarity alone; controlling for nonstationarity did not further improve the estimation of the alcohol coefficient when used in addition to calendar time. As a conclusion, the use of calendar time as a continuous variable is at least as good as other methods (e.g. differencing) to account for nonstationarity in the present context. However, inclusion of autoregression further improved estimation even when calendar time was included.
- Whether the autoregressive structure was estimated to be constant across countries or to be country-specific resulted in smaller differences than those between an autoregressive structure and a structure restricted to low-order lag correlations (nonstationary or stationary). It should be noted, however, that the estimation of higher order (>2) nonstationary structures (which allow more flexibility of the correlation matrix) did not converge to a solution. This was probably because of the number of parameters that needed estimation. For a
model with an autoregressive structure (e.g. of order 1), although correlations for higher lags can be significant, only 1 parameter must be estimated as it is assumed that correlations of time points with higher lag orders follow an exponentially decaying function of the correlation at lag 1. In nonstationary models, correlations with a higher lag order can vary freely and must be estimated separately.
- The inclusion of heteroscedasticity (unequal variances) did not improve models, or did so only marginally.

As already mentioned, it appeared that the better the error structure of time series was captured, the higher were the point estimates for the relationship between alcohol consumption and mortality. In none of these models, however, could varying coefficients for this relationship across countries be modelled. The underlying assumption of these pooled cross-sectional models is that there is a constant relationship and that pooling is used to increase sample size for efficient estimation, while adjusting for a more complex error structure than independence. However, tests for constant coefficients showed that there is significant variation across countries for both the coefficient of alcohol consumption and for calendar year, pointing to geographically differing and not altogether changing associations. Although the autocorrelated error structure could not be accounted for, random coefficient models yielded the highest association for alcohol consumption, pointing to the possibility that the use of random coefficient models better captured the data and the error structure than constant coefficient models accounting for autocorrelated residuals, by estimating a constant effect of alcohol consumption and a constant effect of calendar time. Clearly, random coefficient models including the autoregressive error structure may further improve estimation.

There is software that can in principle handle autocorrelation and heteroscedasticity in multilevel models (Bryk et al. 1996; Yang et al. 1999). It should be noted, however, that in the present study the inclusion of autocorrelated disturbances failed to converge. The possible reason for the nonconvergence might be that the model was already well specified with the inclusion of random coefficient time trends, and the additional inclusion of further parameters might therefore have resulted in colinearity problems. This seems to be a general problem in analysis of aggregate data in the alcohol field, where relatively low variability within series is coupled with relatively short series. It is, however, not likely that the omission of autocorrelated disturbances in random coefficient models may have greatly distorted the findings. Although including autocorrelation of errors resulted in further improvement, this improvement was relatively minor in pooled cross-sectional time series models with constant alcohol coefficients that adjusted for a constant coefficient of calendar time, compared to the improvement from random coefficients models. In addition, in random coefficient models without including autocorrelation
the impact of calendar year was not estimated as being constant across all countries, and might therefore further account for autocorrelation, as it separately adjusted in each country and therefore better captured the country-specific confounding than a country-averaged model.

The advantage of random coefficient models was that not only could the association between consumption and mortality vary across countries, but so could the associations with the control variable time. Thus, the adjustment of time could be analysed in a country-specific manner. Finally, none of the pooled cross-sectional time series models could analyse variations in the relationship between alcohol consumption and mortality as being moderated by patterns.

In conclusion, all sensitivity analysis performed indicated that multilevel modelling outperformed pooled cross-sectional time series models, although the autocorrelated error structure could not be taken into account directly. As estimation of such a structure was not feasible even for the longer time series in the field, there was no hope that this could be done when more countries with relatively short series (fewer than 10 data points) were added. On the contrary, adding other countries would increase the cross-country variability and reduce the impact of autocorrelation within country, especially as this may have little impact given the few data points.

### 3.3 CRITERIA FOR IDENTIFYING RELEVANT STUDIES

The criteria listed in Table 12.8 apply to all meta-analyses used to estimate the relationship between per capita consumption and specific diseases (English et al. 1995).

Alcohol-related consequences were thus identified by reviewing and evaluating large-scale epidemiological studies on alcohol and health, including epidemiological input into major reviews (Collins and Lapsley 1991; Corrao et al. 1999; Devlin et al. 1997; English et al. 1995; Gurr 1996; Harwood et al. 1998; Klingeman and Gmel 2001; Rice et al. 1991; Ridolfo and Stevenson 2001; Single et al. 1996, 1999a, 1999b; Stinson et al. 1993; U.S. Department of Health and Human Services 2000). Papers were collected primarily from the peer-reviewed international literature. As indicated above under the discussion on causality, we followed the accepted guidelines established in the first major review (English et al. 1995). All conditions for which evidence of a causal relationship was conclusive were included in the final list. Discussion of disease conditions where causality was not judged to be sufficient can be found in section 3.5.

### 3.4 Description of studies, including METHODOLOGICAL QUALITIES

More than 6000 studies were included in the different analyses on which the estimates for this chapter were based. For further descriptions we refer to the original publications of the meta-analyses (English et al.

Table 12.8 Exclusion criteria for studies used in determining the relationship between average volume of consumption and mortality/morbidity in CRA
$\left.\begin{array}{ll}\text { Reason for exclusion } & \text { Explanation and/or examples } \\ \hline \text { Restricted study population } & \begin{array}{l}\text { The study was carried out in a sample that was difficult to } \\ \text { generalize for the entire population (for example, a cohort } \\ \text { of persons suffering from a particular disorder) }\end{array} \\ \text { Inappropriate comparison } & \begin{array}{l}\text { The control group was contaminated by a high exposure to } \\ \text { the risk factor under consideration or other factors related } \\ \text { to the condition }\end{array} \\ \text { group } & \begin{array}{l}\text { Exposure was alcohol dependence rather than } \\ \text { alcohol consumption, or alcohol consumption was undefined } \\ \text { or had poorly defined nominal categories (for example, } \\ \text { "regular" vs "non-regular" drinkers) or was limited to } \\ \text { frequency of consumption (without considering quantity) or }\end{array} \\ \text { measure } & \begin{array}{l}\text { was measured in a manner not representative of general } \\ \text { exposure (for example, with meals or in the last 24 hours) }\end{array} \\ \text { A study was reported in more than one paper, and was }\end{array}\right\}$

1995; Gutjahr et al. 2001; Ridolfo and Stevenson 2001; Single et al. 1999a). Nevertheless, we provide here a general methodological description of the studies used and their assumptions.

Most studies reviewed used either a cohort or case-control design (Rothman and Greenland 1998b). For epidemiological studies on alcohol consumption using these designs, the following limitations generally apply.

- Alcohol use is mostly measured by only a few questions, either separating frequency of drinking and quantity per occasion or combining them into one modified frequency question in the tradition of nutritional epidemiology (e.g. Rehm 1998a, 1998b for further descriptions and a critique). Research has shown that such questions may lead to underestimates of true consumption. Also, data on the relationship
between patterns of drinking and health outcomes are limited, as typical questions in epidemiological surveys cannot be used to measure drinking patterns (Rehm 2000). Studies with more complete alcohol assessment are mostly cross-sectional, or the baseline assessment was carried out very recently so that no longitudinal results can yet be reported.
- The relationships between alcohol consumption and chronic disease outcomes are often based on outcomes assessed at follow-up, regressed on several variables from the baseline assessment. These procedures assume that the baseline variables are stable over time, or that they are somehow good indicators of the postulated theoretical relationship (e.g. Rehm et al. 1996). For example, in assessing the relationship between volume of consumption and liver cirrhosis, it must be assumed that heavy consumption persists after baseline and is a good indicator for overall tissue exposure, which is the theoretical determinant (Lelbach 1975, 1976). Work on the regression dilution bias has shown that the size of the real effect is often underestimated by using only the baseline assessment (Clarke et al. 1999) if the exposure is somewhat constant or preserves rank order over time. Moreover, there is evidence from longitudinal studies that individual drinking patterns are not stable over time (Fillmore 1988; Vaillant and Hiller-Sturmhofel 1996), which may obscure the apparent relationship even further.
- The relationship between alcohol consumption and outcomes must be assumed to be fairly constant across settings, in that results of epidemiological studies from a few countries are applied to others. While this may be justified with biologically based relationships (e.g. alcohol and breast cancer), ${ }^{18}$ such an assumption is more problematic with casualties and injuries as outcomes, since these are much more context-dependent.

All of these points indicate that the results of the meta-analyses on which our estimates are based will have limits with regard to precision. These limits are not captured by the CI for combined relative risks.

### 3.5 Relative risks and attributable fractions

Some conditions, such as alcoholic psychosis or alcohol dependence syndrome, are by definition causally related and wholly attributable to alcohol (i.e. they would not exist in the absence of alcohol consumption) (Table 12.9). For most conditions, however, alcohol is a contributory rather than a sufficient cause (Rothman and Greenland 1998a). Pooling of risk estimates for these diseases from individual studies was performed by means of precision-based weighting (English et al. 1995). Methods and results of the pooling procedure (meta-analysis) have been described in more detail elsewhere (Gutjahr and Gmel 2001).

In contrast, the alcohol-attributable fractions (AAFs) for acute consequences such as injuries are usually directly determined from the blood alcohol concentration (BAC) at the time of the injury. For example, road accidents are attributed to alcohol according to whether the driver responsible for the accident tested positive for alcohol and to what degree (e.g. BAC $\geq 0.05 \%$ ). ${ }^{19}$ In the case of traffic accidents we have relative risk estimates, based on case-control studies, for different levels of BAC (Ridolfo and Stevenson 2001; see also McLeod et al. 1999). But relative risk estimates are usually rare for acute consequences other than traffic injuries. Thus, for the purpose of the present study, AAFs from the international literature were used (English et al. 1995; Gutjahr and Gmel 2001; Ridolfo and Stevenson 2001; Single et al. 1996; Stinson et al. 1993).

To structure the presentation and discussion of results, alcohol-related health consequences will be categorized as follows.

1. Chronic harmful effects of alcohol consumption, excluding depression and IHD:

- wholly alcohol-attributable outcomes;
- cancers (neoplasms);
- cardiovascular diseases;
- liver cirrhosis;
- effects of prenatal alcohol exposure;
- neuropsychological conditions; and
- other chronic diseases.

2. Chronic beneficial effects of alcohol consumption, excluding IHD:

- ischaemic stroke; and
- other conditions (type II diabetes, gallstones).

3. IHD as a chronic condition where alcohol has harmful and beneficial consequences:

- depression; and
- acute adverse effects:
- unintentional injuries (motor vehicle accidents, poisonings, falls, drownings, other unintentional injuries)
- intentional injuries (self-inflicted injuries, homicide, other intentional injuries).

Table I2.9 Disease conditions that are by definition fully alcoholattributable $(\mathrm{AAF}=\mathrm{I})$

| ICD-9 code | Disease |
| :--- | :--- |
| 291 | Alcoholic psychoses |
| 303 | Alcohol dependence |
| 305.0 | Alcohol abuse |
| 357.5 | Alcoholic polyneuropathy |
| 425.5 | Alcoholic cardiomyopathy |
| 535.3 | Alcoholic gastritis |
| $571.0-571.3$ | Alcoholic fatty liver, acute alcoholic hepatitis, alcoholic cirrhosis of |
| 790.3 | liver, unspecified alcoholic liver damage |
| $980.0,980.1$ | Elevated blood alcohol level |
|  | Toxic effect of ethyl alcohol, toxic <br> effect of methyl alcohol |

### 3.6 Chronic harmful effects of alcohol consumption, EXCLUDING DEPRESSION AND IHD

## Wholly alcohol-attributable diseases

A number of diseases are by definition fully attributable to alcohol (AAF $=1$ ). These are listed in Table 12.9.

## DISEASES WITH A CONTRIbUTORY ROLE

## Cancer

Oropharyngeal, oesophageal and liver cancers. Alcohol has consistently been related to the risk of cancer of the mouth (lip, tongue), pharynx, larynx, hypopharynx, oesophagus and liver (Corrao et al. 1999; English et al. 1995; Gurr 1996; Single et al. 1999; U.S. Department of Health and Human Services 2000; WHO 2000). The relationship between average volume of alcohol consumed and cancer incidence is usually characterized as increasing almost monotonically (Bagnardi et al. 2001), but this may partially be an artefact of the methods used (Single et al. 1999). Evidence for the role of alcohol in these cancers has accumulated from case-control and cohort studies. Recently, much emphasis has been placed on investigating biochemical mechanisms in laboratory studies to explain the carcinogenic behaviour of alcohol (U.S. Department of Health and Human Services 2000). ${ }^{20}$

Female breast cancer. Much research has been conducted over the last decade on breast cancer. Prior to 1995, it was usually concluded that evidence of a causal relationship with alcohol was insufficient (English et al. 1995; Rosenberg et al. 1993; Schatzkin and Longnecker 1994). Recent studies and reviews have shown, however, that not only haz-
ardous or harmful drinking but also even moderate alcohol consumption can cause female breast cancer (Single et al. 1999a). A meta-analysis by Smith-Warner et al. (1998) found a clear linear relationship over the whole continuum of consumption. Other original studies supported this finding (Bowlin et al. 1997; Corrao et al. 1999; Nasca et al. 1994; Royo-Bordonada et al. 1997; Swanson et al. 1997; Van den Brandt et al. 1995; Wingo et al. 1997). In contrast to the weight of evidence, Zhang et al. (1999) concluded from their investigation that moderate intake did not increase the risk of breast cancer, and that a low level of drinking was associated with a protective effect. This finding, however, appears to be a notable outlier (Longnecker 1999) and, so far, has not been corroborated. Recent studies have focused on plausible biological mechanisms, including alcohol's effect on hormones and tissue, its contribution to the initiation, progression and promotion of breast cancer, and its interaction with nutritional factors (for an overview see Singletary and Gapstur 2001; Soler et al. 1998; U.S. Department of Health and Human Services 2000).

Cancers of the stomach, pancreas, colon, rectum and prostate. Many recent research projects have investigated whether these cancers are alcohol-related. Overall, evidence for a causal relationship between alcohol and cancer of these sites, if any was found, was weak and inconclusive (Bode and Bode 1997; Boutron et al. 1995; De Stefani et al. 1998; Gapstur et al. 1994; Harnack et al. 1997; Ji et al. 1996; Longnecker and Enger 1996; Lundberg and Passik 1997; Piette et al. 1998; Sarles et al. 1996; Seitz et al. 1998a, 1998b; Singborg 1998; Soler et al. 1998). On prostate cancer, again most studies did not report observing an increased risk (Breslow and Weed 1998; Ellison et al. 1998; Hiatt et al. 1994; Tavani et al. 1994), whereas two cohort studies (Ajani et al. 1998; Putnam et al. 1998) and one case-control study (Hayes et al. 1996) reported a small increased risk in men who consume even moderate amounts of alcohol. In conclusion, evidence for a causal relationship between alcohol and cancer of the stomach, pancreas, colon, rectum and prostate has not so far produced consistent results, especially with regard to physiological pathways. Thus, we did not include these cancers, even though some of them showed significantly elevated risks in a recent metaanalysis (Bagnardi et al. 2001).

Cancer of salivary glands, ovary, endometrium, bladder. It has been hypothesized that alcohol might constitute a risk factor for cancer of the major salivary glands (Horn-Ross et al. 1997; Muscat and Wynder 1998), ovary, endometrium (Bradley et al. 1998; Longnecker and Enger 1996; Newcomb et al. 1997; Parazzini et al. 1995) and bladder (Bruemmer et al. 1997; Donato et al. 1997; Longnecker and Enger 1996; Yu et al. 1997). For each of these sites, results were scarce or conflicting, and
the effects, if any, were not statistically significant. Moreover, there is no knowledge of physiological pathways for these sites.

There is an almost linear dose-response relationship between volume of drinking and the relative risk of alcohol-related cancers. Although there have been speculations about the impact of patterns of drinking, especially for breast cancer (Kohlmeier and Mendez 1997), the current state of knowledge does not suggest that these play an important role in the etiology of cancer. ${ }^{21}$ Thus, the alcohol-attributable burden for cancer will be modelled exclusively on average volume.

## Cardiovascular disease

The role of alcohol, as both a risk and protective factor for cardiovascular disease, has been studied extensively in the past decade. IHD has been the focus of most research and is discussed separately below. Most studies suggest that low-level consumption also offers some protection against ischaemic stroke, and this condition is therefore also discussed in the section below on the beneficial effects of alcohol.

In contrast, hypertension and other cardiovascular disorders such as cardiac arrhythmias, heart failure and ill-defined descriptions and complications of heart disease are adversely affected by alcohol (see e.g. Friedman 1998; Klatsky 1995; Puddey et al. 1999; Rosenqvist 1998; U.S. Department of Health and Human Services 1997; Wood et al. 1998). The weight of evidence suggests that daily consumption of more than 30 g pure alcohol for men (and presumably lower levels for women) causes hypertension (Beilin et al. 1996; Curtis et al. 1997; English et al. 1995; Grobbee et al. 1999; Keil et al. 1997; Klatsky 1996). Low-level intake, however, was not associated with hypertension in men, and may even confer a small protective effect in women (English et al. 1995). There are some indications that hypertension may be related to the pattern of heavy drinking occasions (Murray et al. 2002; Puddey et al. 1999; Wannamethee and Shaper 1991).

For haemorrhagic stroke, the weight of evidence suggests an increase in risk for males even at low levels of consumption (Berger et al. 1999; Jackson 1994; Sacco et al. 1999; You et al. 1997). For females, the most recent meta-analyses of Ridolfo and Stevenson (2001) suggested a protective effect for drinking categories I and II but an 8 -fold increased risk for drinking the equivalent of more than 40 g pure alcohol daily (see Table 12.10). Patterns of drinking not only play a role in any protective effects of alcohol on IHD but are also relevant for risk of stroke (Hillbom et al. 1998) and sudden cardiovascular death or cardiovascular death in general (Kauhanen et al. 1997a, 1997b; Kosarevic et al. 1982; Poikolainen et al. 1983; Wannamethee and Shaper 1992), with heavy drinking occasions and intoxication resulting in increased risk. Patterns of drinking should therefore be included in future estimates of harmful cardiovascular outcomes.

## Liver cirrhosis

Alcohol consumption has been identified as the leading cause of liver cirrhosis in established market economies (Corrao et al. 1997, 1998; English et al. 1995). Whereas the association with alcoholic liver cirrhosis is clear, with all cases being attributable to alcohol, debate remains as to whether this equally applies to unspecified liver cirrhosis. Several authors contend that, empirically, it is extremely difficult to separate alcoholic from unspecified liver cirrhosis, and that the term "unspecified liver cirrhosis" is applied when no specific etiological factor is reported or identified (English et al. 1995). Research in the United States and in Central and South American countries has indicated that an appreciable proportion of deaths from cirrhosis without mention of alcohol was in fact attributable to alcohol (Haberman and Weinbaum 1990; Puffer and Griffith 1967; Room 1972). ${ }^{22}$

On the other hand, applying AAFs of liver cirrhosis to other countries can be extremely misleading. In many countries (e.g. China and India), liver cirrhosis is mainly caused by other factors such as viral infections. The corresponding AAFs have been shown to vary between less than 10\% (China) and 90\% (Finland) (WHO 2000).

The relationship between alcohol consumption and liver cirrhosis seems to depend mainly on volume of drinking and is independent of pattern of drinking (Lelbach 1975, 1976). However, some research also indicates a potential effect of occasions of heavy drinking (Rhodés et al. 1993). Moreover, there is some indication that spirits are especially harmful in causing liver cirrhosis (Gruenewald and Ponicki 1995; Kerr et al. 2000; Longnecker et al. 1981; Schmidt 1991). The problem with this research is that it is almost entirely based on ecological studies and thus describes only correlations, which may have other causes (Morgenstern 1998; Rehm and Gmel 2001a).

## Effects of prenatal alcohol exposure

Today, there is ample evidence that alcohol consumption during pregnancy is related to various risks to the fetus, which include gross congenital anomalies and fetal alcohol syndrome (FAS) (Alvear et al. 1998; Church et al. 1997; Faden et al. 1997; Habbick and Snyder 1997; Larkby and Day 1997; Larroque and Kaminski 1996; Mattson et al. 1997; Passaro and Little 1997; Passaro et al. 1996; Polygenis et al. 1998; Roebuck et al. 1998; Shu et al. 1995; Windham et al. 1995). FAS has been characterized as a continuum, with minor physical malformations at one end and serious neurobiological dysfunctions, including mental retardation, at the other (Connor and Streissguth 1996). The prenatal teratogenic effects of alcohol also include lethal outcomes comprising spontaneous abortion, low birth weight, fetal damage, prematurity and intrauterine growth retardation (Abel 1997; Bradley et al. 1998; Windham et al. 1997). These can occur even at low average
volumes of consumption, particularly during the first trimester of pregnancy.

## Mental conditions

The co-morbidity of alcohol dependence with other mental conditions is high, both in clinical and in general population samples (e.g. Grant and Harford 1995; Merikangas et al. 1998). The crucial question in this respect is about causation. We have included depression in this review only where we believe the evidence to be sufficient to conclude a causal role for alcohol. Since this relationship is controversial, it is discussed below in a separate section.

## Other chronic conditions

Other risks of alcohol consumption currently discussed in the literature include epilepsy (see e.g. Jallon et al. 1998; Leone et al. 1997; Martin et al. 1995), acute and chronic pancreatitis and psoriasis. Whereas for pancreatitis the causal role of alcohol seems to be clear, Amman et al. (1996) and Skinazi et al. (1995) contend that the discrimination between acute and chronic pancreatitis is not justifiable, since the overwhelming majority of patients presenting with acute pancreatitis at the same time have an underlying chronic pancreatitis (Robles-Diaz and Gorelick 1997; Thakker 1998). On psoriasis, our search did not yield any recent studies. English et al. (1995) found that the results of the pooled estimates were consistent with a moderately strong and statistically significant effect of average volume of consumption.

Table 12.10 summarizes the relative risks at different levels of consumption for alcohol-related chronic consequences.

### 3.7 BENEFICIAL HEALTH EFFECTS OF ALCOHOL CONSUMPTION on disease

## ISCHAEMIC STROKE

Cerebrovascular disease (stroke) consists of several subtypes, the most common being ischaemic stroke and haemorrhagic stroke, which are affected differently by alcohol. For ischaemic stroke (the predominant type) the weight of evidence, including biological mechanisms, suggests effects similar to those for IHD, namely that low to moderate consumption may offer some protection (Beilin et al. 1996; Hillbom 1998; Keil et al. 1997; Kitamura et al. 1998; Knuiman and Vu 1996; Sacco et al. 1999; Thun et al. 1997; Yuan et al. 1997; Wannamethee and Shaper 1996). It seems that this protective effect is more pronounced in females (Table 12.11).

## Other beneficial health effects

Alcohol consumption may offer some protection against type II diabetes and cholelithiasis (gallstones) (see also Ashley et al. 2000 for a recent

Table 12.10 Relative risks for chronic harmful alcohol-related disease for different drinking categories (relative to abstainers)

| Disease or condition | ICD-9 codes | Relative risk |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Drinking category I |  | Drinking category II |  | Drinking category III |  |
|  |  | Males | Females | Males | Females | Males | Females |
| Lip and oropharyngeal cancer | $\begin{aligned} & 140,141, \\ & 143-146, \\ & 148,149, \\ & 230.0 \end{aligned}$ | 1.45 | 1.45 | 1.85 | 1.85 | 5.39 | 5.39 |
| Oesophageal cancer | 150, 230.1 | 1.80 | 1.80 | 2.38 | 2.38 | 4.36 | 4.36 |
| Liver cancer | 155, 230.8 | 1.45 | 1.45 | 3.03 | 3.03 | 3.60 | 3.60 |
| Laryngeal cancer | 161, 231.0 | 1.83 | 1.83 | 3.90 | 3.90 | 4.93 | 4.93 |
| Female breast cancer, $<45$ years | 174, 233.0 | NA | 1.15 | NA | 1.41 | NA | 1.46 |
| Female breast cancer, $\geq 45$ years |  | NA | 1.14 | NA | 1.38 | NA | 1.62 |
| Epilepsy | 345 | 1.23 | 1.34 | 7.52 | 7.22 | 6.83 | 7.52 |
| Hypertension ${ }^{\text {a }}$ | 401-405 | 1.40 | 1.40 | 2.00 | 2.00 | 4.10 | 2.00 |
| Cardiac arrhythmias | $\begin{aligned} & \text { 427.0, 427.2, } \\ & 427.3 \end{aligned}$ | 1.51 | 1.51 | 2.23 | 2.23 | 2.23 | 2.23 |
| Heart failure and ill-defined complications of heart disease ${ }^{\text {b }}$ | 428, 429 |  |  |  |  |  |  |
| Haemorrhagic stroke | 430-432 | 1.27 | 0.59 | 2.19 | 0.65 | 2.38 | 7.98 |
| Oesophageal varices | 456.0-456.2 | 1.26 | 1.26 | 9.54 | 9.54 | 9.54 | 9.54 |
| Gastro-oesophageal haemorrhage ${ }^{\text {c }}$ | 530.7 | NA | NA | NA | NA | NA | NA |
| Unspecified liver cirrhosis | 571.5-571.9 | 1.26 | 1.26 | 9.54 | 9.54 | 13.00 | 13.00 |
| Acute and chronic pancreatitis ${ }^{\text {a }}$ | 577.0, 577.1 | 1.30 | 1.30 | 1.80 | 1.80 | 3.20 | 1.80 |
| Spontaneous abortion | 634 | NA | 1.20 | NA | 1.76 | NA | 1.76 |
| Low birth weight | 656.5 | 1.00 | 1.00 | 1.40 | 1.40 | 1.40 | 1.40 |
| Psoriasis | 696.1 | 1.58 | 1.58 | 1.60 | 1.60 | 2.20 | 2.20 |
| Prematurity | 764 | 0.93 | 0.93 | 1.36 | 1.36 | 1.36 | 1.36 |
| Intrauterine growth retardation | 765 | 0.99 | 0.99 | 1.68 | 1.68 | 1.68 | 1.68 |


| NA | Not applicable. |
| :--- | :--- |
| a | Relative risk estimates taken from Corrao et al. (1999); most major cost studies derived AAFs |
| for acute and chronic pancreatitis directly (e.g. Australia, for 577.0: 0.24; for 577.I: 0.84 [English |  |
| et al. 1995]). |  |
| b | Heart failure AAF determined indirectly from other circulatory diseases. |
| c | Relative risks not applicable because AAFs were usually obtained directly (e.g. Switzerland, for <br>  <br> 530.7: 0.47). |
| Sources: | unless otherwise indicated Gutjahr et al. (200I); Ridolfo and Stevenson (200I). |

overview on beneficial effects of alcohol). The Australian meta-analysis by English et al. (1995) concluded that there was some evidence that alcohol may protect against the onset of type II diabetes. Since then, the findings from a cohort of more than 40000 male health professionals showed that moderate alcohol consumption may reduce the risk of type II diabetes, perhaps through the effects of alcohol on insulin sensitivity (Rimm et al. 1995). In addition, findings from the British Regional Heart Study indicated a protective effect (Perry et al. 1995). Further, a followup of men enrolled in the United States Physicians Study revealed a marked negative association of incident type II diabetes with alcohol consumption (Ajani et al. 1999). In a recent prospective study, a U-shaped association was found between alcohol and type II diabetes (Wei et al. 2000).

On the other hand, Kao et al. (1998) found evidence (based on small numbers) of an inverse relationship between alcohol consumption and the risk of type II diabetes for women. This relationship was not found in men; indeed, men consuming more than 21 units of alcohol per week were at increased risk of type II diabetes. A protective effect of moderate alcohol consumption against type II diabetes may be mediated through the effects of alcohol on glucose tolerance and insulin resistance. Moderate alcohol drinking has been shown to increase insulin sensitivity (Facchini et al. 1994; Kiechl et al. 1996) and lower insulin resistance (Lazarus et al. 1997), even in young adult drinkers (Flanagan et al. 2000). Finally, there is some evidence that inflammatory processes may mediate alcohol-induced diabetes (Imhof et al. 2001; Pradhan et al. 2001). In summary, there is growing evidence from cohort studies that moderate alcohol consumption reduces the risk of diabetes and a plausible underlying biological mechanism has been identified. This was the reason for including this effect as a beneficial effect in subregions with beneficial drinking patterns (established market economies with best mortality pattern: AMR-A, EUR-A, WPR-A). For all other subregions, no effect was modelled for these drinking categories. However, evidence for the relationship between alcohol consumption and diabetes is far from conclusive at present.

There is evidence that alcohol consumption may offer some protection against gallstones (English et al. 1995; Holman et al. 1996). These findings have been substantiated by recent large-scale cohort and case-control studies, which reported an inverse relationship (Attili et al. 1998; Caroli-Bosc et al. 1998; Chen et al. 1999; Leitzmann et al. 1998). Table 12.11 gives an overview of diseases for which alcohol potentially has beneficial effects.

Table 12.1I Relative risks for chronic beneficial alcohol-related health effects for different drinking categories (compared to abstainers)

| Disease or condition | ICD-9 code | Relative risk |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Drinking category I |  | Drinking category II |  | Drinking category III |  |
|  |  | Males | Females | Males | Females | Males | Females |
| Type II diabetes | 250 | 0.99 | 0.92 | 0.57 | 0.87 | 0.73 | 1.13 |
| Ischaemic stroke | 433-435 | 0.94 | 0.52 | 1.33 | 0.64 | 1.65 | 1.06 |
| Cholelithiasis | 574 | 0.82 | 0.82 | 0.68 | 0.68 | 0.50 | 0.50 |

Source: Gutjahr et al. 2001; Ridolfo and Stevenson 2001.

### 3.8 IHD AS A CHRONIC CONDITION FOR WHICH ALCOHOL HAS HARMFUL AND BENEFICIAL CONSEQUENCES

## Epidemiology-average volume of consumption

$\mathrm{IHD}^{23}$ is one of the leading causes of death in the world (Murray and Lopez 1996a). The most important health benefits of alcohol in terms of IHD have been found at low to moderate levels of average volume of consumption (Beaglehole and Jackson 1992; Doll 1998; Edwards et al. 1994; Fuchs et al. 1995; Goldberg et al. 1995; Hillbom 1998; Holman et al. 1996; Jackson 1994; Rehm et al. 1997; Single et al. 1999a; Svärdsudd 1998). Only a few individual-level studies have failed to substantiate this association in men (Hart et al. 1999) or women (Fillmore et al. 1998; Maskarinec et al. 1998).

While some studies have suggested that alcohol may offer protection against IHD across the entire continuum of consumption (Camargo et al. 1997; Doll et al. 1994; Keil et al. 1997; Rehm et al. 1997 [males only]; Kitamura et al. 1998; Thun et al. 1997), they nevertheless show that most of the protective effect is gained at low levels of consumption, such as one drink every other day.

Overall, average volume of drinking and IHD show a J-shape relationship in the usual medical epidemiological cohort studies in established market economies (Corrao et al. 2000). Compared to abstinence, low to moderate average consumption of alcohol has been found to confer a lower risk of IHD incidence and mortality. For higher levels of average volume of consumption, the risk relationship is reversed (e.g. Corrao et al. 2000; Friedman and Kimball 1986; Rehm et al. 1997), with heavy average consumption being associated with a risk larger than that for abstainers. In the most recent meta-analysis Corrao et al. (2000) demonstrated the described J-shape, and also demonstrated several other characteristics from the literature on average volume of consumption and IHD.

- There was a pronounced sex effect, showing that women were less protected for a given level of consumption, with an earlier upturn of the curve.
- The beneficial effect of alcohol was less pronounced for fatal outcomes.
- The study results were inconsistent, especially with respect to relative risk for higher intake, indicating additional influencing factors not controlled for.
- The better quality studies placed the maximum beneficial effect at lower levels of average alcohol intake. The maximum protective effect was measured at 20 g pure alcohol per day; the relative risk $=1$ line, equivalent to abstainers' risk, was crossed at 72 g per day; and there was a significant detrimental effect over 89 g per day.
- More specifically, cohort studies, wholly adjusted studies, studies that compared drinkers with lifetime abstainers and those that excluded sick subjects at baseline showed less beneficial effects than case-control studies, unadjusted or partially adjusted studies, studies that compared drinkers with current abstainers and those that included sick subjects.
- Mediterranean countries showed more protective effects for the same levels of average consumption.

The epidemiological evidence that light to moderate average alcohol consumption protects against IHD is strengthened by substantial evidence concerning the biological mechanisms by which a protective effect could be mediated (Rankin 1994; Renaud et al. 1993; Single et al. 1999b; Svärdsudd 1998). First, moderate alcohol intake has been linked to favourable lipid profiles, especially an increase in high-density lipoproteins (HDL) (Baraona and Lieber 1998). It has been estimated that as much as $40-50 \%$ of the protective effect may be attributable to this mechanism (Criqui et al. 1987; Criqui and Ringel 1994; Suh et al. 1992). Second, moderate alcohol intake favourably affects coagulation profiles, particularly through its effects on platelet aggregation (McKenzie and Eisenberg 1996; Rubin 1999) and fibrinolysis (Reeder et al. 1996). Third, low to moderate consumption of alcohol has been shown to favourably affect insulin resistance (Kiechl et al. 1996; Lazarus et al. 1997; Rankin 1994). Fourth, it has been postulated that alcohol could protect against IHD through its effect on hormonal profiles, particularly its estrogen effects (Svärdsudd 1998). Fifth, the alcohol metabolite acetate has been postulated to protect against IHD by promoting vasodilation (U.S. Department of Health and Human Services 1997). Sixth, alcohol affects inflammation and, through this pathway, can influence IHD (Imhof et al. 2001; Jacques et al. 2001; Morrow and Ridker 2000; Ridker 2001). Finally, it is possible that some of the effect is mediated through the anti-
oxidative constituents of alcohol beverages, especially wine (Reinke and McCay 1996). Nevertheless, most of the protective effect appears to be linked to ethanol per se.

The protective effect of light to moderate consumption has been questioned on several grounds. The role of the comparison group has been questioned (Shaper 1990a, 1990b; Shaper et al. 1988), it being suggested that the abstainer group includes people who have stopped drinking because of health reasons and these are responsible for the elevated disease risk compared to light and moderate drinkers. Many subsequent studies controlled for this effect by taking lifetime abstainers as the comparison group (Rehm and Sempos 1995). Nevertheless, in most established market economies, where most of the research on alcohol and IHD has taken place, abstainers constitute only a minority of the general population and the possibility that they have other behavioural characteristics responsible for the elevated IHD risk cannot be excluded. No alternative explanation has ever been empirically demonstrated, however. For instance, social isolation has been theoretically claimed to confound the alcohol-mortality relationship (Skog 1996), but empirical research has not been able to substantiate this effect (Murray et al. 1999).

In conclusion, the relationship between average volume of drinking and IHD seems to be J-shaped. Light to moderate drinking is associated with a lower IHD risk than abstinence or heavy drinking. However, the results are inconsistent, indicating that factors other than those included in the study may also determine the relationship. One of the main factors may be pattern of drinking (i.e. the way in which the same average amount of alcohol is consumed). In this respect two patterns deserve mentioning: irregular heavy drinking and drinking with meals.

## EPIDEMIOLOGY—PATTERNS OF DRINKING

## Heavy drinking occasions

Heavy drinking occasions have been linked to adverse cardiovascular events for some time (Poikolainen 1983). However, many studies had used wider endpoints than IHD (Kauhanen et al. 1997b) or samples of problem drinkers or persons with alcohol-use disorders (Dyer et al. 1977; Rosengren et al. 1987; Rossow and Amundsen 1997), where heavy drinking patterns are confounded with volume.

Some of the more recent studies have controlled for (average) volume of drinking. A case-control study in Australia (McElduff and Dobson 1997) compared 11511 cases of acute myocardial infarction or coronary death with 6077 randomly selected controls. If people drank in binges (usually five or more drinks on an occasion for women, nine or more drinks on an occasion for men), there were no protective effects for coronary events and relative risks were mainly larger than 1 compared to abstainers (indicating higher risks for major coronary events). This
elevated risk was present even in groups with low overall volume of drinking. As expected, the authors also found a protective effect of daily drinking, which was most pronounced for regular light to moderate drinkers.

Similarly, Murray et al. (2002) evaluated the cardiovascular consequences of binge drinking (eight or more drinks at a sitting) and usual (non-binge) drinking of alcohol in a longitudinal, population-based study. Interview data from 1154 men and women aged 18 to 65 years in Winnipeg, Canada were linked to health care utilization and mortality records in an eight-year follow-up period. Cox proportional hazards regressions were estimated separately for men and women. The outcomes included first event for physician visits, and hospitalizations and deaths due to IHD, hypertension, or other cardiovascular disease. Binge drinking increased the risk of IHD in men (hazard ratio [HR] of 2.3, $95 \% \mathrm{CI}$ $1.2-4.2$ ) and women (HR of $1.1,95 \%$ CI 1.02-1.2) and increased the risk of hypertension in men (HR of 1.6, $95 \%$ CI 1.04-2.4) but not women. Binge drinking had no effect on the risk of other cardiovascular disease. All of these results were controlled for average volume of drinking. Again, the expected cardioprotective effects were confirmed in both men and women. The harmful effects of heavy drinking occasions on IHD morbidity and mortality could thus be disaggregated from the effect of average volume of drinking. Finally, Trevisan et al. (2001a) found in a case-control design that, after adjustment for average volume of consumption, weekend drinking by men was significantly related to risk of myocardial infarction compared to men drinking less than once a week (logistic regression: OR 1.9, 95\% CI 1.2-3.2).

In addition to the effect on IHD, there appears to be a relationship between irregular heavy drinking occasions and other forms of cardiovascular death, especially sudden cardiac death (Kauhanen et al. 1997b; Wannamethee and Shaper 1992; Wood et al. 1998). This is consistent with the physiological mechanisms of increased clotting and reducing the threshold for ventricular fibrillation after heavy drinking occasions, which have been reviewed by McKee and Britton (1998). Specifically, heavy drinking occasions have been shown to increase the blood level of low-density lipoproteins (LDL), which in turn have been linked to negative cardiovascular outcomes. Contrary to low or moderate steady drinking, heavy irregular drinking occasions are not associated with an increase in levels of HDL, which have been linked to favourable cardiovascular outcomes. In addition, irregular drinking is associated with increased risk of thrombosis after cessation of drinking (Renaud and Ruf 1996). Finally, irregular heavy drinking seems to predispose to histological changes in the myocardium and conducting systems, as well as to a reduction in the threshold for ventricular fibrillation. In conclusion, irregular heavy drinking occasions are mainly associated with physiological mechanisms that increase the risk of sudden cardiac death and other cardiovascular outcomes, in contrast to the physiological
mechanisms triggered by regular low to moderate consumption that are linked to favourable cardiac outcomes. Nevertheless, individual-level studies are still scarce and some studies show no effects (Murray et al. 1998).

## Drinking with meals

Trevisan et al. (2001b) reported on drinking with meals and IHD mortality based on the Risk Factor and Life Expectancy Study, a pooled series of epidemiological studies conducted in Italy with 8647 males and 6521 females aged 30-59 years at baseline and free of cardiovascular disease. Subjects were followed up for an average of seven years. Alcohol consumption showed a protective effect on IHD, and drinking wine with meals was linked to more positive outcomes than drinking wine outside meals. Compared to drinking with meals, drinking wine outside meals had a relative risk of $1.8,95 \%$ CI $0.97-3.5$ for IHD in males, adjusted for average volume of drinking and other potential confounders. There were not enough IHD cases to conduct a similar analysis for females, but the effects for all-cause mortality for females showed a five-fold risk for wine outside meals compared to wine with meals (relative risk of 5.0, 95\% CI 1.5-10.9).

Another study (Trevisan et al. 2001a), using a case-control design, examined 443 male myocardial infarction survivors and 922 healthy controls aged 35-69 years. Compared to non-drinkers the age, education and smoking-adjusted odds ratios for former drinkers and current drinkers were $0.67,95 \%$ CI $0.32-1.38$ and $0.47,95 \%$ CI $0.24-0.95$, respectively, confirming the overall cardioprotective effect of alcohol consumption (see above). Men who reported drinking without food at least $75 \%$ of the time had an odds ratio of $1.5,95 \%$ CI $0.96-2.3$ compared to those who drank mainly with meals and snacks, after adjustment for age, education and volume of alcohol consumed.

The potential mechanisms linking consumption of alcoholic beverages with meals to a lower IHD risk, compared to consumption between meals, remain to be fully clarified. However a few mechanisms have been hypothesized. A study by Trevisan et al. (1987) in a large sample of Italian men and women found a significant association between drinking between meals and higher prevalence of hypertension, compared to drinking with meals, even after adjustment for differences in alcohol consumption between these drinking pattern categories. These findings were recently confirmed in another study using a population-based sample in the United States (Wu and Trevisan 2001). Finally, Foppa et al. (1999) found in a controlled randomized trial that moderate consumption of wine with a meal reduced postprandial blood pressure. Drinking with meals has also been shown to positively affect fibrinolysis (Hendriks et al. 1994) and lipid levels (Veenstra et al. 1990).

Other potential physiological links between drinking with meals and these IHD risk factors include a reduced absorption of alcohol owing to
the presence of food in the gastrointestinal tract (Gentry 2000). Another physiological link may be that food increases the alcohol elimination rate (Ramchandani et al. 2001).

## AgGregate-LEVEL STUDIES ON PATTERNS AND AVERAGE CONSUMPTION

Given the described relative scarcity of individual-level studies, it is not surprising that much of the argumentation is based on aggregate-level studies, especially on Russian experience with the natural experiment of the Gorbachev anti-alcohol campaign. The Russian Federation is generally considered to be one of the countries with the highest rates of irregular heavy drinking (Bobak et al. 1999; Malyutina et al. 2001). Thus, if a heavy drinking style has an adverse impact on cardiovascular disease in general and on IHD in particular, such effects should have become evident at the population level in the experience of the anti-alcohol campaign during the last years of the Soviet Union. In the period 1984-1987, when estimated total alcohol consumption in the Russian Federation fell by about $25 \%$ (Shkolnikov and Nemtsov 1997), age-adjusted male deaths from circulatory disease fell by $9 \%$ (Leon et al. 1997). After the end of the campaign, the death rate rose again quite dramatically. The role of alcohol in the recent drastic increases in mortality in the Russian Federation, however, remains controversial, as many other changes occurred in the late 1980s and early 1990s (Bobak and Marmot 1999; Britton and McKee 2000; Leon et al. 1997; McKee et al. 2001; Notzon et al. 1998; Shkolnikow et al. 2001). There seems to be general agreement, however, that alcohol has played an important role in increasing mortality rates, although the level of impact is unclear.

There is another indirect line of research on the effect of heavy drinking on IHD. Countries with a tradition of heavier or binge drinking on weekends show proportionately high IHD or cardiovascular disease mortality on or immediately after the weekend (Germany, IHD: Willich et al. 1994; Moscow, cardiovascular disease events: Chenet et al. 1998; Lithuania, IHD events: Chenet et al. 2001; Scotland, IHD events: cf. Evans et al. 2000). Other aggregate-level research on per capita alcohol consumption and IHD have failed to find effects, even for countries with the best drinking pattern (i.e. drinking pattern $=1$ ) in a time series analysis with differenced data, controlling for tobacco (Hemström 2001). Finally, Skog (1983) also found no significant effects in a time series analysis on differenced data for Norway.

## Summary of The epidemiological evidence

In conclusion, the relationship between drinking and IHD is complex, and the epidemiological literature on it is evolving. There seems to be a clear beneficial effect, supported by biochemical evidence, of regular light to moderate drinking, but it is unclear how many people actually drink in a manner that will provide them with these benefits. Also, there are
indications that irregular bouts of heavy drinking are linked to physiological mechanisms, which are in turn linked to negative IHD outcomes, as well as to other negative cardiovascular disease outcomes. Better surveys are needed, including the measurement of biomarkers indicative of the relationships specified above. In these surveys, the relevant variables such as irregular heavy drinking or drinking with meals should also be included.

Many of the individual-level cohort studies on the relationship between average volume of drinking and IHD have been carried out on special populations such as nurses, doctors or other health professionals, mostly in established market economies, who have relatively regular and potentially beneficial drinking patterns. As a result, the effect of average volume of alcohol consumption may be overstated in the usual meta-analysis. Thus, the results of these analyses cannot be applied worldwide, since more detrimental patterns of drinking prevail in the majority of countries. Moreover, the impact of pattern of drinking has to be included, in addition to the effects of average volume of drinking.

Unfortunately, as described above, patterns of drinking cannot be modelled by meta-analysis owing to the scarcity of data on the relationship between exposure and outcome. Thus, we used the multilevel modelling approach described earlier to incorporate patterns of drinking into the estimates.

MEASURING THE EFFECT OF AVERAGE VOLUME OF CONSUMPTION AND DRINKING PATTERN ON IHD MORTALITY
The details of determining the impact of average volume and pattern of drinking on IHD are described earlier. The characteristics of the data set are given in Table 12.12.

Table 12.13 gives an overview of the most important results with respect to the differential impact of alcohol consumption on IHD mortality. As predicted, in countries with consumption pattern 1, alcohol had beneficial effects on the incidence of IHD. For countries with pattern 2, the impact on IHD varied around zero (i.e. no marked impact of alcohol). In countries with pattern 3, alcohol showed a detrimental impact on IHD for males only. For countries with pattern 4, the detrimental impact of alcohol was pronounced for both males and females. Assuming interval scales between the categories of the pattern variable, the models were estimated as in Table 12.14.

Again, in countries with pattern 1, beneficial effects appeared for both males ( -0.0214 ) and females $(-0.0376)$. These consumption models indicate that the overall effect of IHD at an average pattern level ( 2.51 in the sample used) is detrimental for males and not significantly different from zero for females. In other words, the results from individual-level studies could be replicated in this aggregate-level study. This means that for countries such as France and Italy we expect a beneficial effect of alcohol on IHD mortality; for countries such as Slovenia and the United
Table I2.I2 Characteristics of the data set for calculating the relationship between per capita consumption, patterns of drinking and IHD mortality (per 1000 population per year)

| Country | IHD mortality males | IHD mortality females | Average per capita alcohol consumption | Pattern of drinking | Per capita GNP | Year of first data | Year of latest data | Number of years |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Albania | 0.99 | 0.40 | 2.20 | 3 | 585.71 | 1992 | 1998 | 7 |
| Argentina | 1.37 | 0.70 | 27.48 | 2 | 3472.00 | 1966 | 1996 | 25 |
| Armenia | 3.34 | 2.30 | 2.53 | 2 | 507.14 | 1992 | 1998 | 7 |
| Australia | 2.75 | 1.36 | 11.19 | 2 | 10341.76 | 1964 | 1997 | 34 |
| Austria | 1.92 | 1.00 | 14.16 | 1 | 11136.05 | 1962 | 1999 | 38 |
| Azerbaijan | 4.10 | 2.40 | 2.10 | 3 | 746.67 | 1990 | 1999 | 9 |
| Bahamas | 0.95 | 0.46 | 14.33 | 2 | 7067.69 | 1969 | 1995 | 13 |
| Bahrain | 2.04 | 1.72 | 5.33 | 2 | 7423.33 | 1985 | 1988 | 3 |
| Barbados | 0.93 | 0.56 | 7.88 | 2 | 3357.42 | 1964 | 1995 | 31 |
| Belarus | 4.03 | 2.20 | 9.68 | 4 | 2651.11 | 1989 | 1998 | 9 |
| Belgium | 1.44 | 0.63 | 12.35 | I | 9487.10 | 1964 | 1994 | 31 |
| Belize | 0.53 | 0.38 | 6.01 | 4 | 1127.86 | 1964 | 1995 | 28 |
| Bulgaria | 2.14 | 1.28 | 11.97 | 2 | I 898.24 | 1982 | 1998 | 17 |
| Canada | 2.45 | 1.22 | 9.74 | 2 | II 367.65 | 1964 | 1997 | 34 |
| Chile | 1.16 | 0.74 | 11.58 | 3 | 1562.26 | 1964 | 1994 | 31 |
| Colombia | 1.09 | 0.70 | 5.49 | 3 | 993.68 | 1967 | 1994 | 19 |
| Costa Rica | 1.30 | 0.84 | 4.48 | 4 | 1268.13 | 1964 | 1995 | 32 |

Table I2.12 Characteristics of the data set for calculating the relationship between per capita consumption, patterns of drinking and IHD mortality (per 1000 population per year) (continued)

| Country | IHD mortality males | IHD mortality females | Average per capita alcohol consumption | Pattern of drinking | Per capita GNP | Year of first data | Year of latest data | Number of years |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Croatia | 1.78 | 1.03 | 13.24 | 3 | 3638.33 | 1993 | 1998 | 6 |
| Czech Republic | 2.76 | 1.43 | 15.42 | 2 | 3915.71 | 1986 | 1999 | 14 |
| Denmark | 2.59 | 1.32 | 10.91 | 2 | 13139.70 | 1964 | 1996 | 33 |
| Dominican Republic | 0.46 | 0.31 | 2.91 | 2 | 729.52 | 1965 | 1985 | 21 |
| Ecuador | 0.37 | 0.24 | 2.30 | 3 | 882.90 | 1964 | 1995 | 31 |
| El Salvador | 0.38 | 0.25 | 1.95 | 4 | 564.74 | 1964 | 1993 | 19 |
| Estonia | 4.16 | 2.18 | 13.43 | 3 | 3107.50 | 1992 | 1999 | 8 |
| Finland | 3.25 | 1.37 | 7.42 | 3 | 10642.42 | 1964 | 1996 | 33 |
| France | 0.82 | 0.35 | 23.15 | 1 | 11120.29 | 1964 | 1997 | 34 |
| Germany | 1.59 | 0.80 | 13.63 | 1 | 26661.67 | 1993 | 1998 | 6 |
| Greece | 0.94 | 0.44 | 9.69 | 2 | 5027.14 | 1964 | 1998 | 35 |
| Guatemala | 0.26 | 0.21 | 2.42 | 4 | 644.71 | 1964 | 1984 | 17 |
| Guyana | 1.17 | 0.69 | 8.42 | 3 | 540.00 | 1977 | 1994 | 5 |
| Honduras | 0.20 | 0.13 | 2.27 | 4 | 418.67 | 1966 | 1981 | 15 |
| Hungary | 2.54 | 1.30 | 15.85 | 3 | 2837.83 | 1977 | 1999 | 23 |
| Iceland | 2.30 | 1.06 | 5.42 | 3 | 12691.82 | 1964 | 1996 | 33 |
| Ireland | 2.80 | 1.39 | 10.86 | 3 | 5865.76 | 1964 | 1996 | 33 |
| Israel | 1.89 | 1.19 | 3.04 | 2 | 8236.36 | 1975 | 1996 | 22 |

$$
\begin{array}{r}
33 \\
7 \\
34 \\
10 \\
5 \\
10 \\
5 \\
31 \\
33 \\
35 \\
34 \\
34 \\
35 \\
17 \\
33 \\
18 \\
21 \\
35 \\
13 \\
10 \\
8 \\
\hline \text { continued }
\end{array}
$$


8418.18
927.14
13654.41
1691.00
488.00
3149.00
2380.00
19366.77
4004.85
1473.71
1877.94
10900.29
7426.29
448.24
14216.06
828.89
454.76
3844.86
6664.62
1432.00
2683.75

$n-\quad+m m m--m+-N+m m m-m m+$
$\underset{\sim}{\infty} \underset{\sim}{\infty} \underset{\sim}{\infty} \underset{\sim}{\infty} \underset{\sim}{\infty} \underset{\sim}{\infty} \underset{\sim}{\infty} \underset{\sim}{N} \underset{\sim}{N}$ N


Table I2.I2 Characteristics of the data set for calculating the relationship between per capita consumption, patterns of drinking

| Country | $\begin{aligned} & \text { IHD mortality } \\ & \text { males } \end{aligned}$ | IHD mortality females | Average per capita alcohol consumption | Pattern of drinking | Per capita GNP | Year of first data | Year of latest data | Number of years |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Singapore | 1.44 | 0.79 | 2.30 | 2 | 9106.86 | 1964 | 1998 | 35 |
| Slovakia | 2.85 | 1.57 | 12.54 | 3 | 2437.50 | 1992 | 1995 | 4 |
| Slovenia | 1.27 | 0.66 | 14.80 | 2 | 8912.00 | 1994 | 1998 | 5 |
| South Africa | 1.01 | 0.46 | 10.30 | 3 | 3606.67 | 1993 | 1995 | 3 |
| Spain | 0.83 | 0.40 | 17.20 | 1 | 5554.17 | 1962 | 1997 | 36 |
| Sri Lanka | 0.45 | 0.22 | . 27 | 3 | 268.46 | 1964 | 1986 | 13 |
| Suriname | 0.92 | 0.54 | 6.43 | 3 | 1749.05 | 1964 | 1992 | 21 |
| Sweden | 2.54 | 1.27 | 7.57 | 3 | 13042.42 | 1964 | 1996 | 33 |
| Switzerland | 1.43 | 0.68 | 14.99 | 1 | 15815.16 | 1964 | 1994 | 31 |
| Thailand | 0.02 | 0.01 | 3.58 | 3 | 541.20 | 1964 | 1994 | 25 |
| The former Yugoslav Republic of Macedonia | 1.22 | 0.58 | 5.76 | 3 | 1350.00 | 1994 | 1997 | 4 |
| Trinidad and Tobago | 2.05 | 1.28 | 4.76 | 2 | 3254.33 | 1962 | 1994 | 30 |
| Ukraine | 4.19 | 2.48 | 4.64 | 3 | 1276.00 | 1990 | 1999 | 10 |
| United Kingdom | 2.67 | 1.23 | 9.49 | 2 | 9210.00 | 1964 | 1998 | 35 |
| United States | 2.68 | 1.39 | 9.46 | 2 | 14028.53 | 1964 | 1997 | 34 |
| Uruguay | 1.77 | 0.96 | 8.08 | 3 | 1680.00 | 1966 | 1990 | 24 |
| Uzbekistan | 3.72 | 2.67 | 1.07 | 3 | 733.33 | 1996 | 1998 | 3 |
| Venezuela | 1.38 | 0.94 | 8.31 | 3 | 2992.50 | 1969 | 1994 | 24 |

Table 12.13 Average effect of consumption of I litre of pure alcohol per capita ${ }^{\text {a }}$ for different drinking patterns

| Pattern of drinking | Males | Females |
| :--- | :---: | ---: |
| I | -0.016227 | -0.038174 |
| 2 | 0.004050 | -0.014323 |
| 3 | 0.053951 | 0.001908 |
| 4 | 0.084529 | 0.035584 |

[^48]Table 12.14 Effect of per capita consumption and drinking patterns on standardized IHD rates, assuming equal intervals between pattern values

|  | Coefficient | SE | $t$-value | $d f$ | $P$ |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Males <br> Effect of per capita consumption <br> at average pattern |  |  |  |  |  |
| Deviations from average patterns | 0.033503 | 0.036353 | 0.011622 | 3.128 | 72 | 0.002

[^49]States there will be no overall effect; for countries such as the Republic of Korea and Ukraine we expect an overall detrimental effect of alcohol for males; and for countries such as the Russian Federation we expect a detrimental effect for both males and females. The results from this model are consistent with the finding that the anti-alcohol campaign in the Gorbachev era had such an impact on IHD mortality in the Russian Federation.

The model has limitations, however. Most notably, each country has been assigned one pattern value that is assumed to be stable over time. Whereas patterns have been found to be stable in the past (Room 1992; Simpura 2001), clearly this assumption is an oversimplification. Moreover, it is clear that in most countries there are people with different drinking patterns. Thus, future research should be based on distributions of patterns by sex and age, rather than on one pattern value per country.

The results of this model are not consistent with a recent ecological analysis by Hemström (2001), using a time series approach with differenced data. Hemström (2001) found a random distribution of insignificant negative (beneficial) and positive (detrimental) alcohol effect estimates. He used only per capita data without any control for drinking pattern. However, based on individual-level studies and the results of the present analysis, beneficial effects would have been expected for countries with a consumption pattern of 1 , such as France, Italy, Portugal and Spain. We can only speculate on the difference in results between the two analyses. Hemström's time series were relatively short (45 years per country), not allowing some of the tests for correct model specification (Rehm and Gmel 2001a). There also may have been some overdifferencing problems with co-integration or the joint use of differencing and taking the logarithm, which may have obscured the effect (Greene 2000; Hatanaka 1996; Yaffee 2000).

On the other hand, the present analysis draws on relatively limited time series in a wide range of drinking cultures, which increases the ability to generalize but which also has limitations. Our findings for the pattern- 1 subgroup have the strong advantage of convergent data from individual-level studies, supported by biological plausibility. We have therefore used our finding of a protective effect at the population level for the countries with beneficial patterns (pattern 1) as a current best estimate of effects (for comparison with results using coefficients from meta-analysis).

Given the limits of ecological analysis, individual-level data should always be given priority in judgements of causality, although not necessarily in terms of estimates of the overall effect of changes in a risk factor at the population level (Skog 1996). This applies to our analysis as well.

## Determining AAFs for IHD, including patterns of DRinking

To arrive at the AAFs, the results of the dummy regression were taken and only effects larger than $\pm 0.015$ were modelled. However, since the
multilevel analysis could only control for per capita GNP as a confounder, and since there are general limits in controlling for confounding in such analyses (Morgenstern 1998), the effects of alcohol indicators were halved in order to be conservative and to adjust for any other confounding. This approach, though arbitrary in the choice of halving, is consistent with that taken in risk analysis for other risk factors (e.g. tobacco) to avoid residual confounding.

Taking into consideration the distribution of patterns in all countries within the 14 subregions, and standardized mortality rates after 1994, the AAFs can be derived (Table 12.15).

Table 12.15 shows the effects of alcohol-attributable IHD mortality based on aggregate-level analysis, halving the effects to adjust for potential confounding. This can easily be done by using the respective values in the formulas above.

Alternatively, one may use the relative risks from the individual-level studies as a basis and apply them to the subregions AMR-A, EUR-A and WPR-A, as almost all of the studies come from these subregions. Such an approach leads to higher cardioprotective effects (see Table 12.16) and may well contain overestimates based on the usual cohort composition, where people with more regular drinking styles are overrepresented. However, individual-level studies are usually preferred to ecological studies both for establishing causality and for making estimates of impact (e.g. Morgenstern 1998). Further, with respect to overall disease burden attributable to alcohol, this provides a more conservative approach to estimating.

The resulting AAFs, using individual-level estimates of relative risk for AMR-A, EUR-A and WPR-A, would be $-0.11,-0.15$ and -0.15 , respectively. The cardioprotective effect estimated in this way is larger than that given above (see Table 12.15). We used these estimates for the final calculations. The numbers for EUR-B are given only for sensitivity analysis. Overall, the average patterns for this subregion (2.9) are much closer to the average pattern for EUR-C (3.6) than that for EURA (1.3). Thus, we did not feel justified in using the estimates based on studies almost exclusively from countries with beneficial patterns (A subregions).

### 3.9 Depression

## BACKGROUND AND EPIDEMIOLOGY

Alcohol is implicated in a variety of mental disorders that are not alcohol-specific. However, no major overview on alcohol-attributable burden of disease has yet included these disorders (English et al. 1995; Gutjahr et al. 2001; Rehm et al. 2001a, 2001b; Ridolfo and Stevenson 2001; Single et al. 1999). While the causality of the relation is hard to define, sufficient evidence now exists for us to include an estimate of the causal role of alcohol in depression, a major mental disorder.

Table I2.15 AAFs predicted by the multilevel analysis

| Subregion | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | All | 15-29 | 30-44 | 45-59 | 60-69 | $\geq 70$ |
| Males |  |  |  |  |  |  |
| AFR-D | 0.02 | 0.03 | 0.03 | 0.03 | 0.03 | 0.02 |
| AFR-E | 0.07 | 0.07 | 0.08 | 0.08 | 0.07 | 0.07 |
| AMR-A ${ }^{\text {a }}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| AMR-B | 0.16 | 0.15 | 0.15 | 0.16 | 0.17 | 0.15 |
| AMR-D | 0.08 | 0.09 | 0.10 | 0.11 | 0.10 | 0.08 |
| EMR-B | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| EMR-D | 0.01 | 0.01 | 0.01 | 0.02 | 0.01 | 0.00 |
| EUR-A ${ }^{\text {a }}$ | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 |
| EUR-B | 0.11 | 0.12 | 0.12 | 0.12 | 0.12 | 0.11 |
| EUR-C | 0.15 | 0.14 | 0.15 | 0.14 | 0.14 | 0.15 |
| SEAR-B | 0.01 | 0.02 | 0.03 | 0.04 | 0.02 | 0.00 |
| SEAR-D | 0.04 | 0.06 | 0.12 | 0.14 | 0.04 | 0.00 |
| WPR-A ${ }^{\text {a }}$ | -0.07 | -0.07 | -0.07 | -0.08 | -0.07 | -0.07 |
| WPR-B | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| Females |  |  |  |  |  |  |
| AFR-D | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| AFR-E | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| AMR-A ${ }^{\text {a }}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| AMR-B | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| AMR-D | 0.03 | 0.03 | 0.03 | 0.04 | 0.03 | 0.03 |
| EMR-B | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| EMR-D | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| EUR-A ${ }^{\text {a }}$ | -0.10 | -0.10 | -0.10 | -0.10 | -0.10 | -0.10 |
| EUR-B | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| EUR-C | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| SEAR-B | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| SEAR-D | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| WPR-A ${ }^{\text {a }}$ | -0.10 | $-0.10$ | -0.11 | -0.11 | -0.10 | -0.10 |
| WPR-B | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

a The estimates for AMR-A, EUR-A and WPR-A serve only as sensitivity analysis as the AAFs for these subregions will be based on individual-level studies (see Table 12.16 below).

In the general population, alcohol dependence and major depression co-occur over-proportionally, on both a 12 -month and a lifetime basis (Kessler et al. 1996, 1997; Lynskey 1998). Among alcohol consumers in the general population, higher volume of consumption is associated with more symptoms of depression (Graham and Schmid 1999; Mehrabian 2001; Rodgers et al. 2000). Compared to moderate drinkers, both higher
Table 12.16 Ischaemic heart disease AAFs for selected subregions, applying relative risk estimates ${ }^{\text {a }}$

| Subregion | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | 70-79 |  | $\geq 80$ |  | Total |  |  |
|  | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | All |
| AMR-A | -0.14 | -0.12 | -0.15 | -0.13 | -0.14 | -0.11 | -0.12 | -0.08 | -0.11 | -0.07 | -0.11 | -0.07 | -0.13 | -0.08 | -0.11 |
| EUR-A | -0.18 | -0.16 | -0.17 | -0.17 | -0.17 | -0.16 | -0.17 | -0.14 | -0.16 | -0.13 | -0.16 | -0.13 | -0.16 | -0.13 | -0.15 |
| EUR-B ${ }^{\text {b }}$ | -0.15 | -0.11 | -0.13 | -0.10 | -0.14 | -0.09 | -0.13 | -0.07 | -0.12 | -0.07 | -0.12 | -0.07 | -0.13 | -0.08 | -0.11 |
| WPR-A | -0.18 | -0.17 | -0.19 | -0.17 | -0.18 | -0.16 | -0.17 | -0.14 | -0.15 | -0.12 | -0.15 | -0.12 | -0.17 | -0.13 | -0.15 |
| a Relative risks are $0.82,0.83$ and I .12 for drinking categories I , II and III for men and $0.82,0.83$ and I .00 for women. <br> b EUR-B estimates are only given for reasons of sensitivity analysis (see section 3.6). |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

(Bjork et al. 1999) and lower (Rodgers et al. 2000) levels of depressive symptoms have been found among abstainers. Among patients under treatment for alcohol abuse and dependence, the prevalence of major depression is higher than in the general population (Lynskey 1998; Schuckit et al. 1997a). Similarly the prevalence of alcohol abuse and dependence is higher for patients under treatment for depression (Alpert et al. 1999; Blixen et al. 1997).

This suggests that alcohol abuse is linked to depressive symptoms, and that alcohol dependence and depressive disorders co-occur to a larger degree than expected by chance. However, it is not clear in the individual case whether depression caused alcohol problems, whether alcohol consumption or alcohol problems caused depression, or whether both could be attributed to a third cause (Vaillant 1993). The pathway from depression to problematic alcohol use and alcohol dependence has long been discussed under the heading of self-medication (i.e. the use of alcohol to alleviate depressive symptoms). In addition, a shared third cause could be certain neurobiological mechanisms (see Markou et al. 1998) or genetic predisposition. Moreover, all three pathways of causation may co-exist at the same time, with different proportions of the cooccurring morbidities being attributable to each, and of course some co-occurrence being simply due to chance.

## Establishing the causal Link between alcohol AND DEPRESSION

Causal relations in epidemiology are usually based on several criteria (Hill 1965; Rothman and Greenland 1998a). Using these criteria, we will review the evidence that part of the burden of depression is caused by alcohol. As indicated above, this does not preclude the possibility that part of the burden of alcohol dependence is also caused by depression, or that part of the co-occurrence is causally related to some third factor. Following the majority of the literature, we base our argumentation mainly on alcohol dependence rather than alcohol consumption per se. This is done mostly for reasons of data availability: depression and alcohol dependence are usually part of the same mental health surveys, and almost all estimates of co-occurrence stem from this kind of survey. It is not usual to include questions on alcohol consumption per se in this type of survey, nor is it usual to include a diagnosis of depression in alcohol surveys.

Prevalences of alcohol dependence or alcohol use disorders (i.e. alcohol dependence and harmful use of alcohol) were thus used as indicator of alcohol consumption drinking categories, assuming a constant relationship between the two variables. This indicator relationship was also made possible by the fact that alcohol use disorders by definition have an AAF of 1.0 -that is, under the counterfactual scenario (Murray and Lopez 1999) of no alcohol available at all, there would be no burden due to alcohol dependence. Moreover, in different regions of the world,
average volume of drinking and prevalence of alcohol dependence are correlated to a high degree ( $r=0.86$ ) (Rehm and Eschmann 2002).

## Temporal order

Causal factors must precede consequences. Logically, therefore, only that fraction of depression in which the onset of alcohol problems preceded the onset of depression can be caused by alcohol problems. The fraction in which alcohol problems came first is an upper bound for the proportion of depressive disorders caused by alcohol dependence. This upper bound is summarized in Table 12.17 for different forms of depression in several countries, based on the International Consortium in Psychiatric Epidemiology (ICPE) (Merikangas et al. 1998).

Clearly, in all areas (i.e. countries, provinces and cities) mentioned in Table 12.17, the proportion of depressive disorders preceded by alcohol dependence is higher for males than for females. This corresponds to the higher prevalence of alcohol dependence in these areas. In fact, the proportion of rates of depressive disorders and alcohol problems correlate to 0.80 (major depression) and 0.82 (other depressions) for these areas (Pearson correlations; alcohol dependence rates from World bealth report 2001 and Rehm and Eschmann 2002). This means that at least $64 \%$ of the variation in the proportion of depressive disorders in the various subregions can be statistically "explained" by the variation in alcohol dependence. Of course, the remarks made above on possible forms of causal pathway still apply. This relationship provides a basis for predicting such rates for other subregions where data are lacking (see below).

Another indicator of the role of alcohol dependence in causing some depressive disorders is the comparison between onsets for different mental disorders co-occurring with alcohol disorders. In the US National Comorbidity Survey (Kessler et al. 1996) the proportion of disorders preceded by alcohol dependence was higher for depression than for any other disorder.

## Consistency

Epidemiological studies are very consistent, both in the general population and in clinical samples, in showing that alcohol dependence and depressive disorders co-occur to a higher degree than might be expected by chance. This has been observed in several countries and regions (e.g. Merikangas et al. 1998; Swendson et al. 1998). In fact, we do not know of any study where alcohol dependence and depressive disorders did not co-vary to a larger degree than that expected by chance.

The co-variation between the proportion of people with depressive disorders with a preceding alcohol dependence and the prevalence of alcohol dependence is also consistent across countries (see above).

Table 12.17 Percentages of the population in which the onset of alcohol problems ${ }^{\text {a }}$ occurred prior to diagnosis of depression, by country, age and sex

| Country/Area | Age group (years) | Major depression ${ }^{\text {b }}$ |  |  |  |  | Other depressions ${ }^{\text {c }}$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Males |  | Females |  | $\frac{\text { Total }}{\%}$ | Males |  | Females |  | $\frac{\text { Total }}{\%}$ |
|  |  | $n$ | \% | $n$ | \% |  | $n$ | \% | $n$ | \% |  |
| USA (NCS) | 15-29 | 119 | 31.1 | 241 | 15.6 | 20.7 | 39 | 28.6 | 67 | 11.6 | 17.8 |
|  | 30-44 | 168 | 47.0 | 286 | 21.1 | 30.6 | 68 | 51.7 | 121 | 22.4 | 33.0 |
|  | 45-59 | 59 | 30.8 | 117 | 10.5 | 17.3 | 35 | 32.8 | 65 | 13.9 | 20.6 |
|  | 60-69 | - | - | - | - | - | - | - | - | - | - |
|  | $\geq 70$ | - | - | - | - | - | - | - | - | - | - |
|  | Total | 345 | 38.7 | 644 | 17.1 | 24.7 | 142 | 40.7 | 253 | 17.4 | 25.8 |
| USA | 15-29 | 46 | 20.8 | 78 | 13.7 | 16.3 | 25 | 34.8 | 32 | 17.0 | 24.8 |
| CA | 30-44 | 37 | 20.9 | 63 | 7.5 | 12.5 | 40 | 32.8 | 23 | 12.5 | 25.3 |
| Fresno | 45-59 | 11 | 1.8 | 24 | 21.4 | 15.1 | 9 | 14.3 | 10 | 0.0 | 6.7 |
|  | 60-69 | - | - | - | - | - | - | - | - | - | - |
|  | $\geq 70$ | - | - | - | - | - | - | - | - | - | - |
|  | Total | 94 | 18.5 | 165 | 12.5 | 14.7 | 74 | 31.3 | 65 | 12.9 | 22.7 |
| Mexico | 15-29 | 23 | 8.2 | 36 | 2.9 | 5.0 | 14 | 18.6 | 7 | 0.0 | 12.1 |
| Mexico | 30-44 | 12 | 30.1 | 50 | 2.0 | 7.3 | 4 | 40.4 | 11 | 0.0 | 11.7 |
| City | 45-59 | 6 | 31.8 | 14 | 0.0 | 9.8 | 4 | 20.4 | 8 | 0.0 | 6.3 |
|  | 60-69 | - | - | - | - | - | - | - | - | - | - |
|  | $\geq 70$ | - | - | - | - | - | - | - | - | - | - |
|  | Total | 41 | 17.9 | 99 | 2.0 | 6.7 | 22 | 23.4 | 26 | 0.0 | 10.6 |
| Canada | 15-29 | 56 | 20.6 | 123 | 20.3 | 20.4 | 43 | 26.8 | 54 | 19.7 | 22.8 |
| Ontario | 30-44 | 98 | 47.0 | 177 | 6.1 | 20.7 | 47 | 46.7 | 62 | 11.9 | 26.9 |
|  | 45-59 | 29 | 33.6 | 85 | 3.0 | 10.8 | 9 | 22.4 | 36 | 2.8 | 6.7 |
|  | 60-69 | - | - | - | - | - | - | - | - | - | - |
|  | $\geq 70$ | - | - | - | - | - | - | - | - | - | - |
|  | Total | 183 | 36.8 | 386 | 10.0 | 18.6 | 98 | 35.8 | 152 | 12.5 | 21.7 |
| Netherlands | 15-29 | 70 | 26.9 | 177 | 10.1 | 14.9 | 46 | 17.4 | 76 | 6.4 | 10.6 |
|  | 30-44 | 180 | 35.9 | 317 | 13.6 | 21.7 | 72 | 15.8 | 146 | 7.8 | 10.5 |
|  | 45-59 | 129 | 31.9 | 180 | 7.2 | 17.5 | 58 | 32.2 | 118 | 3.7 | 13.2 |
|  | 60-69 | 18 | 33.7 | 33 | 4.0 | 14.4 | 11 | 0.0 | 26 | 0.0 | 0.0 |
|  | $\geq 70$ | - | - | - | - | - | - | - | - | - | - |
|  | Total | 397 | 32.9 | 707 | 10.7 | 18.7 | 187 | 20.4 | 366 | 5.7 | 10.6 |
| Brazil | 15-29 | 10 | 61.0 | 23 | 2.5 | 20.2 | 4 | 0.0 | 6 | 37.8 | 23.4 |
|  | 30-44 | 21 | 10.0 | 53 | 17.0 | 15.1 | 7 | 28.6 | 22 | 22.2 | 23.8 |
|  | 45-59 | 14 | 17.6 | 36 | 6.5 | 9.5 | 15 | 21.1 | 15 | 4.0 | 12.8 |
|  | 60-69 | 2 | 66.7 | 10 | 0.0 | 12.8 | 2 | 100.0 | 6 | 22.2 | 38.5 |
|  | $\geq 70$ | 2 | 50.0 | 5 | 0.0 | 12.7 | 1 | 0.0 | 5 | 0.0 | 0.0 |
|  | Total | 48 | 26.7 | 126 | 9.4 | 14.2 | 29 | 23.9 | 54 | 16.9 | 19.4 |
| Germany | 15-29 | 126 | 14.6 | 213 | 4.8 | 8.4 | 39 | 17.7 | 84 | 2.4 | 7.3 |
|  | 30-44 | - | - | - | - | - | - | - | - | - | - |
|  | 45-59 | - | - | - | - | - | - | - | - | - | - |
|  | 60-69 | - | - | - | - | - | - | - | - | - | - |
|  | $\geq 70$ | - | - | - | - | - | - | - | - | - | - |
|  | Total | 126 | 14.6 | 213 | 4.8 | 8.4 | 39 | 17.7 | 84 | 2.4 | 7.3 |

Table 12.17 Percentages of the population in which the onset of alcohol problems ${ }^{\mathrm{a}}$ occurred prior to diagnosis of depression, by country, age and sex (continued)

| Country/Area |  | Age group (years) | Major depression ${ }^{\text {b }}$ |  |  |  |  | Other depressions ${ }^{\text {c }}$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Males | Females |  | $\frac{\text { Total }}{\%}$ | Males |  | Females |  | $\frac{\text { Total }}{\%}$ |
|  |  | $n$ | \% | $n$ |  | \% | $n$ | \% | $n$ |  | \% |
| Japan |  |  | 15-29 | 2 | 100.0 | 7 | 0.0 | 20.1 | 0 | NA | 5 | 0.0 | 0.0 |
|  |  |  | 30-44 | 5 | 23.7 | 3 | 0.0 | 15.1 | 4 | 32.2 | 2 | 0.0 | 21.0 |
|  |  | 45-59 | 4 | 0.0 | 5 | 0.0 | 0.0 | 2 | 0.0 | 6 | 0.0 | 0.0 |
|  |  | 60-69 | 3 | 0.0 | 2 | 0.0 | 0.0 | I | 0.0 | 3 | 0.0 | 0.0 |
|  |  | $\geq 70$ | I | 100.0 | I | 0.0 | 49.8 | 1 | 100.0 | I | 0.0 | 49.8 |
|  |  | Total | 14 | 24.9 | 17 | 0.0 | 11.4 | 7 | 27.4 | 16 | 0.0 | 8.3 |
| Chile |  | 15-29 | 37 | 28.6 | 67 | 11.3 | 17.4 | 11 | 0.0 | 47 | 11.4 | 9.3 |
|  |  | 30-44 | 30 | 27.2 | 53 | 2.0 | 11.1 | 11 | 2.2 | 70 | 5.7 | 5.2 |
|  |  | 45-59 | 13 | 44.0 | 42 | 6.7 | 15.4 | 18 | 39.3 | 50 | 1.7 | 11.6 |
|  |  | 60-69 | 8 | 21.6 | 6 | 0.0 | 12.0 | 8 | 69.1 | 11 | 1.0 | 29.3 |
|  |  | $\geq 70$ | 2 | 7.2 | 6 | 0.0 | 1.9 | I | 20.4 | 5 | 0.0 | 2.8 |
|  |  | Total | 89 | 29.2 | 173 | 6.6 | 14.3 | 48 | 26.5 | 182 | 5.6 | 10.0 |
| NCS National Comorbidity Survey. | National Comorbidity Survey. |  |  |  |  |  |  |  |  |  |  |  |
| - No data. |  |  |  |  |  |  |  |  |  |  |  |  |
| a | For NCS, Brazil, Chile and the Netherlands, defined as having an alcohol problem, alcohol dependence or alcohol abuse. In all other countries/areas, defined as either having alcohol dependence or alcohol abuse. |  |  |  |  |  |  |  |  |  |  |  |
| ${ }^{\text {b }}$ | In all countries, major depression is defined with exclusion and without hierarchy. |  |  |  |  |  |  |  |  |  |  |  |
|  | In all countries, other depression is defined as having dysthymia or bipolar disorder with exclusions and without hierarchy. |  |  |  |  |  |  |  |  |  |  |  |
| Source: Table numbers were calculated using the International Consortium for Psychiatric Epidemio data (see also Merikangas et al. 1998). |  |  |  |  |  |  |  |  |  |  |  |  |

## Strength of association

It has been consistently found that alcohol-dependent individuals demonstrate a two- to three-fold increase in risk of depressive disorders (e.g. Hilarski and Wodarki 2001; Schuckit 1996; Swendson et al. 1998). This is an effect comparable in size with many other causal effects (Gutjahr et al. 2001) (Table 12.19).

## Reversibility (remission during abstinence)

Key evidence for a causal effect of alcohol dependence on depressive disorders comes from studies that analyse what happens to rates of depressive disorders when patients with clinical symptoms abstain from drinking. Most of these studies come to the conclusion that many depressive syndromes markedly improve within days or weeks of abstinence (Brown and Schuckit 1988; Dackis et al. 1986; Davidson 1995; Gibson and Becker 1973, Penick et al. 1988; Pettinati et al. 1982; Willenbring 1986). Of course, other things change within therapy, so not all of the
effect is necessarily due to the pharmacological effect of drinking. In addition, experimental studies have found that more symptoms of depression are reported during heavy drinking episodes (Isbell et al. 1955; Schuckit et al. 1997b; Tamarin et al. 1970; Weiner et al. 1971). In conclusion, there is sufficient evidence that abstinence substantially removes symptoms of depression in alcohol-dependent persons within a short time.

## Family patterns

Several studies have tried to separate alcohol-dependent persons with primary (sometimes called "independent") depressive disorders from those with secondary (sometimes called "induced") depressive disorders by examining different family patterns of both alcohol dependence and depressive disorders. For instance, Hesselbrock et al. (1983) and Schuckit et al. (1997b) found higher rates of depressive disorders, rather than alcoholism, in close relatives of alcohol-dependent patients with primary depression than in patients with secondary depression. Other studies, both in clinical samples and in the general population, did not find different rates of alcohol dependence or affective disorders based on the primary-secondary distinction (Grant and Pickering 1997; Hasegawa et al. 1991). Thus, the studies on genetic vulnerability suggest differences between primary and secondary transmission, but are not conclusive.

## Biological mechanisms

There are several plausible mechanisms by which alcohol dependence may cause depressive disorders (Markou et al. 1998), but research is not yet conclusive. It should be noted that there is a biological link via intoxication or heavy use rather than via dependence alone.

## Dose-response relationship

Merikangas et al. (1998) found that there is a continuum in the magnitude of co-morbidity as a function of position on the spectrum of substance use (use, problems, dependence). While there are relationships at the level of symptoms in the general population, the relationships are strongest between alcohol dependence and depressive disorders (see above).

## Potential alternative explanations

There may be other explanations for the co-occurrence of alcohol dependence and depressive disorders, either from genetic disposition or the environment. For example, Grant and Pickering (1997) suggested that alcoholism and major depression might be alternative manifestations of the same underlying disorder. Nevertheless, there is no explanation for the finding that depressive symptoms increase markedly during bouts of heavy drinking and disappear during periods of abstinence, even if no antidepressant medication is given. This is the strongest indication of a causal effect.

## Summary of evidence on causality

Overall, we find sufficient evidence of causality for the influence of alcohol dependence on depressive disorders. The evidence indicates that a clear and consistent association exists between alcohol dependence and depressive disorders and that chance, confounding variables and other bias can be ruled out with reasonable confidence as factors in this association. Consistent with the assessments for other disease and injury categories, most weight was placed on consistency across several studies, strength of the association, reversibility, temporal order and the fact that the effect was at least physiologically plausible.

## Estimating AAFs of depressive disorders

Quantitative estimates of the proportion of depressive disorders attributable to alcohol can be derived from the high correlation between alcohol dependence and the proportion of depressive disorders with preceding alcohol-use disorders (see above), using alcohol dependence rates in different subregions of the world (Table 12.18, columns 3 and 4).

The empirical data on proportion of depressive disorders with preceding alcohol-use disorders were regressed on survey results of alcohol dependence for the same subregions without a constant. The omission of the constant was due to the fact that in a situation without any alcohol dependence, the proportion of alcohol-attributable depressive disorders should also be zero. Thus, the regression line must pass through the origin. Since the relationship is quite close, the respective regression coefficients became highly significant even with few data points.

Clearly, the proportions in columns 3 and 4 of Table 12.18 are the upper limit of depressive disorders attributable to alcohol. In order to derive a realistic proportion of alcohol-attributable depressive disorders, we need to subtract the proportion of co-occurrences due to chance. In a situation of chance, the occurrence of alcohol-use disorders in depressed persons should be exactly equal to the occurrence of alcohol-use disorders in non-depressed persons (i.e. the general population). Thus, the prevalence of alcohol-use disorders was first subtracted from the upper limit to derive AAFs. The prevalence of alcohol dependence was taken from The world health report 2001, which itself was based on a pooled analysis of survey results. The relationship between alcohol dependence and alcohol abuse (i.e. harmful use in ICD-10) was derived from the US National Comorbidity Survey (Kessler 1998) and was assumed to be constant across subregions. This assumption was necessary as there were fewer data on alcohol abuse than on alcohol dependence, and as the diagnostic systems used vary considerably for this diagnosis.

To control for possible confounding, the effects were halved as has been done elsewhere (e.g. IHD analysis above) as well as for other risk factors (Peto et al. 1992 for smoking; see also chapter 11). The resulting estimates for AAFs of depressive disorders can be found in Table 12.18, columns 5 and 6.

Table I2.18 Prevalence of alcohol dependence by subregion and sex, and AAFs (percentage of depressive disorders) in people aged $\geq 15$ years

|  | Prevalence | Upper limit | Upper limit |  |  |
| :---: | :---: | :---: | :---: | :---: | :--- |
| Subregion | of alcohol | dependence (\%) | for AAF, major | depression | for AAF, other |
| depression | AAF, major | AAF, other |  |  |  |
| depression | depression |  |  |  |  |

Males

| AFR-D | 1.37 | 5.17 | 5.31 | 1.68 | 1.76 |
| :--- | ---: | ---: | ---: | ---: | ---: |
| AFR-E | 2.89 | 10.93 | 11.24 | 3.56 | 3.71 |
| AMR-A | 8.04 | 30.37 | 31.22 | 9.89 | 10.31 |
| AMR-B | 5.72 | 21.60 | 22.20 | 7.04 | 7.34 |
| AMR-D | 5.13 | 19.39 | 19.93 | 6.32 | 6.59 |
| EMR-B | 0.07 | 0.25 | 0.26 | 0.09 | 0.09 |
| EMR-D | 0.07 | 0.26 | 0.26 | 0.09 | 0.09 |
| EUR-A | 5.61 | 21.21 | 21.80 | 6.91 | 7.20 |
| EUR-B | 1.16 | 4.39 | 4.52 | 1.43 | 1.49 |
| EUR-C | 8.22 | 31.05 | 31.91 | 10.11 | 10.54 |
| SEAR-B | 0.77 | 2.91 | 2.99 | 0.95 | 0.99 |
| SEAR-D | 1.58 | 5.96 | 6.13 | 1.94 | 2.03 |
| WPR-A | 3.12 | 11.77 | 12.10 | 3.84 | 4.00 |
| WPR-B | 1.78 | 6.73 | 6.91 | 2.19 | 2.29 |


| Females |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- |
| AFR-D | 0.10 | 0.37 | 0.38 | 0.12 | 0.12 |
| AFR-E | 0.31 | 1.17 | 1.20 | 0.37 | 0.38 |
| AMR-A | 2.14 | 8.10 | 8.33 | 2.52 | 2.63 |
| AMR-B | 1.18 | 4.46 | 4.59 | 1.39 | 1.45 |
| AMR-D | 1.19 | 4.50 | 4.63 | 1.40 | 1.46 |
| EMR-B | 0.00 | 0.02 | 0.02 | 0.01 | 0.01 |
| EMR-D | 0.01 | 4.46 | 0.02 | 0.01 | 0.01 |
| EUR-A | 1.18 | 0.88 | 4.58 | 1.39 | 1.45 |
| EUR-B | 0.23 | 5.24 | 5.91 | 0.28 | 0.29 |
| EUR-C | 1.39 | 0.26 | 0.26 | 1.63 | 1.70 |
| SEAR-B | 0.07 | 0.20 | 0.21 | 0.08 | 0.09 |
| SEAR-D | 0.05 | 4.22 | 4.34 | 1.31 | 1.37 |
| WPR-A | 1.12 | 0.19 | 0.19 | 0.06 | 0.06 |
| WPR-B | 0.05 |  |  | 0.07 |  |

a From World health report 2001, based on survey results (Rehm and Eschmann 2002) and then estimated consistently with DisMod, ${ }^{24}$ taking into account case fatality, duration and/or incidence.

These results show that AAFs of depressive disorders vary substantially in different subregions of the world. They reflect differences in rates of heavy alcohol consumption and alcohol dependence. Thus, AAFs are considerably larger for males than for females. They are highest in the

Russian Federation and its surrounding countries (EUR-C) and in North America (AMR-A), and almost nonexistent in Muslim-dominated areas (EMR-B and EMR-D).

## CONCLUSIONS ON ALCOHOL AND DEPRESSION

Based on standard criteria of causality, we conclude that there is sufficient evidence for a causal relation between alcohol-use disorders and depressive disorders. We suspect that careful examination would also reveal a relationship between heavy drinking and depressive disorders, although heavy drinking unfortunately is not usually measured as an endpoint in epidemiological cohort studies.

The status of alcohol abuse as a causal agent in depression is not as clear. There are co-occurrences between alcohol abuse and depressive disorders that are larger than chance (Kessler et al. 1996, 1997), but the relationship is weaker compared to the relationship with alcohol dependence. This may have to do with the less clear conceptual status of alcohol abuse, which is defined in the current version of the Diagnostic and statistical manual of mental disorders (DSM) (American Psychiatric Association 1994) nosology to include social and legal responses to the patient's drinking, and which is not a category of disorder in ICD. On the other hand, many studies found only one factor when analysing criteria of alcohol dependence and alcohol abuse, or a division of factors that did not correspond to the division between the diagnoses in DSM. Thus, the relationship between alcohol abuse and depressive disorders should be clarified in future research. It may have relevance, beyond burden of disease research, in helping to clarify the status of alcohol abuse in general.

### 3.10 Summary of relative risk for chronic diseases, using CRA disease categories

Relative risk estimates are summarized in Table 12.19.

### 3.11 Acute adverse health consequences

Alcohol use has been associated with increased risk of injury in a wide variety of settings, including road traffic accidents (involving vehicles, bicycles and pedestrians), falls, fires, injuries related to sports and recreational activities, self-inflicted injuries and injuries resulting from interpersonal violence (Cherpitel 1992; Freedland et al. 1993; Hingson and Howland 1987, 1993; Hurst et al. 1994; Martin 1992; Martin and Bachman 1997; U.S. Department of Health and Human Services 1997,2000 ). There is also some evidence that the presence of alcohol in the body at the time of injury may be associated with a greater severity of injury and a less positive outcome (Fuller 1995; Li et al. 1997).
Table I2.19 Relative risk for major chronic disease categories by sex and average drinking category

| Disease | ICD-9 (4-digit) | ICD-I0 (4-digit) | Males |  |  | Females |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Drinking category I | Drinking category II | Drinking category III | Drinking category $I$ | Drinking category II | Drinking category III |
| Conditions arising during the perinatal period | $\begin{aligned} & 760-779 \\ & \text { except } 771.3 \end{aligned}$ | P00-P96 |  |  |  |  |  |  |
| Low birth weight | 764-765 | P05-P07 | 1.00 | 1.40 | 1.40 | 1.00 | 1.40 | 1.40 |
| Malignant neoplasms | 140-208 | C00-C97 |  |  |  |  |  |  |
| Mouth and oropharynx cancers | 140-149 | $\mathrm{C} 00-\mathrm{Cl} 4$ | 1.45 | 1.85 | 5.39 | 1.45 | 1.85 | 5.39 |
| Oesophagus cancer | 150 | CI5 | 1.80 | 2.38 | 4.36 | 1.80 | 2.38 | 4.36 |
| Liver cancer | 155 | C22 | 1.45 | 3.03 | 3.60 | 1.45 | 3.03 | 3.60 |
| Female breast cancer* | 174 | C50 | NA | NA | NA | 1.14 | 1.41 | 1.59 |
| $>45$ years of age* |  |  | NA | NA | NA | 1.15 | 1.41 | 1.46 |
| $\geq 45$ years of age* |  |  | NA | NA | NA | 1.14 | 1.38 | 1.62 |
| Other neoplasms | 210-239 | D00-D48 | 1.10 | 1.30 | 1.70 | 1.10 | 1.30 | 1.70 |
| Type II diabetes | 250 | EIO-EI4 | 1.00 | 0.57 | 0.73 | 0.92 | 0.87 | 1.13 |

Neuro-psychiatric conditions
Unipolar major depression Epilepsy
Alcohol-use disorders Cardiovascular diseases
Hypertensive disease
Ischaemic heart disease Cerebrovascular disease Ischaemic stroke*
Haemorrhagic stroke* Digestive diseases
Cirrhosis of the liver
NA Not applicable.
Sources: Gutjahr et al. 2001; Ridolfo and Stevenson 2001; if indicated by * the category III estimates for IHD were based on Corrao et al. 2000.

## Unintentional injuries

Alcohol consumption produces effects that are often perceived as positive, as evidenced by the widespread popularity of drinking. But it also leads to actions that result in unintentional injury and death. This section highlights research findings on causality of alcohol involvement and findings relevant to establishing dose-response relationships and drinking patterns. It focuses on traffic injuries, as most of the research has been conducted in this area, and traffic accidents are the most important component of unintentional injuries (Rehm et al. 2003b).

Studies relating average volume of drinking to risk of injury have found the risk of injury to be positively related to increasing average intake levels of alcohol, with the risk increasing at relatively low volumes of intake (Cherpitel et al. 1995). Two studies of injury among older adults reported a U-shaped relationship between alcohol use and occupational injury (Zwerling et al. 1996) and traumatic deaths (Ross et al. 1990). However, abstinence could be related to existing health problems or cognitive deficits that are, in turn, related to accident risk (Zwerling et al. 1996). Hence the higher risk among abstainers is likely to be purely spurious.

Several patterns of drinking have been related to risk of injury. Frequent heavy drinking and frequent subjective drunkenness are both associated with injury, particularly injury resulting from violence (Cherpitel 1996a). Frequency of heavy drinking has also been associated with a greater likelihood of death due to injury, relative to other causes (Li et al. 1994). One important line of research in this area has empirically defined a parameter of usual drinking pattern that is most closely associated with the risk of injury and drunk driving behaviour, after adjusting for other drinking pattern variables and characteristics of the drinker (Gruenewald and Nephew 1994; Gruenewald et al. 1996a, 1996b; Treno and Holder 1997; Treno et al. 1997). The greatest risk was found in individuals who consume relatively large amounts on some occasions, and whose highest amounts are markedly greater than their average amount per occasion.

Several retrospective studies have compared BAC in individuals who have experienced a collision or trauma, compared with selected individuals not involved in trauma, using a case-control design (Cherpitel 1992; Freedland et al. 1993; Fuller 1995; Hurst et al. 1994; Stoduto et al. 1993; U.S. Department of Health and Human Services 1997). One of the most influential case-control series was the Grand Rapids Study of 5985 collisions (Borkenstein et al. 1964; Hurst et al. 1994). Statistically adequate re-analysis of the Grand Rapids Study indicates that all levels of BAC are associated with an increased risk of crashes, relative to a BAC of zero, with an accelerating slope in which the risk of injury increases markedly with high BACs (Hurst et al. 1994).

There are clear reasons why alcohol is related to injury. Moderate doses of alcohol have been demonstrated in controlled experimental studies to have cognitive and psychomotor effects that are relevant to the risk of injury, such as reaction time, cognitive processing, coordination and vigilance (Eckardt et al. 1998; Kruger et al. 1993; Moskowitz and Robinson 1988; U.S. Department of Health and Human Services 1997). The comprehensive recent review by Eckardt et al. (1998) concluded that the threshold dose for negative effects on psychomotor tasks is generally found at around $40-50 \mathrm{mg} \%$ (equivalent to $0.04-0.05 \%$ ). The authors also stated, "injury can occur as a result of alcohol's disruption of psychomotor function in individuals at BACs of approximately 10 mM ", which is equal to a BAC of little less than $50 \mathrm{mg} \%$.

Dose-response curves observed in experimental data are not always monotonic. For example, a recent experimental study (Lloyd and Rogers 1997) assessed the effects of low doses of alcohol given with a meal, and found that 8 g of absolute alcohol (about 0.25 litre of beer) resulted in improved performance of complex cognitive tasks relative to no alcohol, but that 24 g of absolute alcohol produced impaired performance. Such J -shaped or U-shaped effects of low ethanol doses on task-specific performance are explicable pharmacologically (Eckardt et al. 1998). The Grand Rapids Study also found that dose-response curves varied somewhat between novices and frequent, experienced drinkers.

In summary, the evidence indicates that the amount consumed per occasion, and more specifically the blood alcohol content, is the critical feature in determining risk of injury. Blood alcohol concentrations as low as $40-50 \mathrm{mg} \%$ may cause psychomotor impairment, leading to increased risk of injury in circumstances such as driving or operating machinery.

Thus, despite methodological problems, there is evidence of causality for the most researched injury category (traffic accidents). Table 12.20 gives the AAFs for different kinds of injuries in four recent reviews. The reviews based their estimates on meta-analyses or other summaries of the relations found in published studies. It should be recognized that, while there are many such studies, they are mostly from a relatively small range of countries. Most of the AAFs were directly derived, for example from police statistics, although there are case-control studies as well (McLeod et al. 1999).

Causality, at least for traffic accidents, can be established since:

- alcohol is clearly associated with the outcome;
- there is a dose-response relationship: the higher the BAC, the higher the chance of injury;
- there is a biochemical explanation for the relationship; and
- with suitable interventions to reduce alcohol consumption, the outcome is reduced as well. Thus, in a meta-analysis, Shults et al.
Table 12.20 AAFs of acute alcohol-related health effects in the adult general population

| Injury | ICD-9 code | Review |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Stinson et al.1993; USA |  | English et al. 1995; Australia |  | Single et al. 1996; Canada |  | Ridolfo and Stevenson 2001; Australia |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females |
| Motor vehicle traffic accidents | E810-E819 | 0.42 | 0.42 | 0.37 | 0.18 | 0.43 | 0.43 | $\begin{aligned} & 0.33(\mathrm{~d}) ;{ }^{\mathrm{a}} \\ & 024(\mathrm{~h}) ;{ }^{\mathrm{a}} \\ & \text { pedestrians } \\ & 0.40 \text { (d); } \\ & 0.37 \text { (h) } \end{aligned}$ | $\begin{aligned} & 0.11 \text { (d) } \\ & \text { and (h); } \\ & \text { pedestrians } \\ & 0.17 \text { (d); } \\ & 0.06 \text { (h) } \end{aligned}$ |
| Motor vehicle nontraffic accidents | E820-E825 | 0.42 | 0.42 | 0.37 | 0.18 | 0.43 | 0.43 |  |  |
| Bicycle accident injuries | E826 | 0.20 | 0.20 | 0.37 | 0.18 | 0.20 | 0.20 |  |  |
| Other road vehicle accident injuries | E829 | 0.20 | 0.20 | 0.37 | 0.18 | 0.20 | 0.2 |  |  |
| Water transport accident injuries | E830-E839 | 0.20 | 0.20 | - | - | 0.20 | 0.20 | - | - |
| Air-space transport accident injuries | E840-E845 | 0.16 | 0.16 | - | - | 0.16 | 0.16 | - | - |
| Accidental ethanol and methanol poisoning | E860.0-E860.2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Accidental fall injuries | E880-E888 | 0.35 | 0.35 | 0.34 | 0.34 | 0.20-0.34 ${ }^{\text {b }}$ | $0.13-0.34^{\text {b }}$ | $\begin{aligned} & 0.22 \text { for } \\ & \text { age }<65 \text {; } \\ & 0.12 \text { for age } \\ & \geq 65 \end{aligned}$ | 0.14 for age <65; 0.04 for age $\geq 65$ |

Arson injuries
Accidental excessive cold Accidental drowning Accidental aspiration
Striking against/struck by objects Caught in/between objects Occupational and machine injuries Accidental firearm missile injuries Suicide, self-inflicted injuries Victim fight, brawl, rape Victim assault, firearms E890-E899
E9010
E91I
E917
E918
E919-E920
E922
E950-E959
E960
E965
E966
E967
E968 E890-E899
E910
E91I
E917
E918
E919-E920
E922
E950-E959
E960
E965
E966
E967
E968 E890-E899
E901
E910
E917
E918
E919-E920
E922
E950-E959
E960
E965
E966
E967
E968 o E890-E899 E990-E899
E910
E91I
E917
E918
E919-E920
E922
E950-E959
E960
E965
E966
E967
E968 E990-E899
E910
E91I
E917
E918
E919-E920
E922
E950-E959
E960
E965
E966
E967
E968 E890-E899
E9010
E91I
E917
E918
E919-E920
E922
E950-E959
E960
E965
E966
E967
E968 E890-E899
E9010
E91I
E917
E918
E919-E920
E922
E950-E959
E960
E965
E966
E967
E968 E890-E899
E9010
E91I
E917
E918
E919-E920
E922
E950-E959
E960
E965
E966
E967
E968 E890-E899
E9010
E91I
E917
E918
E919-E920
E922
E950-E959
E960
E965
E966
E967
E968 E890-E899
E9010
E91I
E917
E918
E919-E920
E922
E950-E959
E960
E965
E966
E967
E968 E890-E899
E9010
E91I
E917
E918
E919-E920
E922
E950-E959
E960
E965
E966
E967
E968












 0.44
-
0.34
1.00
-
-
0.07
-
0.29
0.47
0.47
0.47
0.16
0.47
0.47






[^50]Late effects of injuries by another
(2001) found random breath-testing programmes or selective breathtesting checkpoints to be effective in reducing mortality from traffic accidents by $18 \%$ and $20 \%$, respectively.

## INTENTIONAL INJURIES

Alcohol is strongly associated with violent crime (Graham and West 2001), although this association varies considerably across settings (Murdoch et al 1990; Room and Rossow 2001). Studies on violence have repeatedly shown that alcohol consumption precedes violent events, and that the amount of drinking is related to the severity of the subsequent violence. In addition, the experimental literature suggests that alcohol plays a causally contributing role ${ }^{25}$ in aggression. Meta-analyses of experimental studies suggest a small to moderate effect size of about 0.22 (Bushman 1997) in the overall relationship between alcohol consumption and aggression. Some effort has been made to separate pharmacological effects from expectations, ${ }^{26}$ but the general conclusion is that expectations form part of the "psycho-pharmacological" effects of alcohol (Bushman 1997; Graham et al. 1998), and neither can nor should be separated in attempting to understand the effects of alcohol. Alcohol bathes the brain in chemicals and it is likely that a number of different effects of alcohol contribute to the increased likelihood of aggressive behaviour. First, alcohol seems to have an effect on the serotonin (5HT) and GABA brain receptors, similar to that produced by some benzodiazepines (Pihl et al. 1993). The subjective experience of this effect may be a reduced level of fear and anxiety about social, physical or legal consequences of one's actions. This reduced fear or anxiety may result in increased risk-taking by some drinkers. This particular causal pathway has received support from animal research linking alcohol, GABA receptors and aggression (Miczek et al. 1993) and from experimental and observational research showing higher risk taking associated with alcohol intoxication (Graham et al. 2000; Pihl and Peterson 1993). Alcohol also affects cognitive functioning (Peterson et al. 1990), leading to impaired problem solving in conflict situations (Sayette et al. 1993) and overly emotional responses or emotional lability (Pihl et al. 1993). Other behavioural and attitudinal effects of alcohol related to aggression have been identified, although at this point not necessarily linked to particular pharmacological effects on the brain. These include a narrow and tenacious focus on the present (Graham et al. 2000; Washburne 1956), also described as "alcohol myopia" (Steele and Josephs 1990) and increased concerns with demonstrating personal power, at least for men (Graham et al. 2000; McClelland et al. 1972; Tomsen 1997).

Alcohol-related violence involves more complex issues of social interaction than would be relevant to drink-driving and other alcohol-related accidental injuries. In particular, the effects of alcohol are moderated by both the environment and the characteristics of the drinker (Chermack and Giancola 1997; Lipsey et al. 1997; Rossow et al. 2001; U.S. Depart-
ment of Health and Human Services 2000). For example, meta-analysis of experimental research on alcohol and aggression found that the effects of environmental manipulation to increase aggression were stronger for intoxicated than for sober participants. In another meta-analysis, Ito et al. (1996) found that the effects of alcohol were greater in situations characterized by greater anxiety, inhibition conflict and frustration, while differences between sober and intoxicated persons were smaller in situations involving high provocation or self-focused attention. Further, given sufficient disincentives for aggression the effects of alcohol on aggression can be reduced or even eliminated altogether (Hoaken et al. 1998; Jeavons and Taylor 1985).

As with alcohol-related accidents, some proportion of violence that occurs after people have been drinking might have occurred anyway, without the involvement of alcohol. Alcohol-related violence involves an interaction of the effects of alcohol on one or more people, the environment and the personality of the drinker (Graham et al. 1998). However, the environment for alcohol-related aggression is not independent of drinking. ${ }^{27}$ For example, in environments devoted to drinking (e.g. bars, pubs), it does not make sense to try to determine the proportion of violence that would have occurred even if the person had not been drinking, because this particular environment does not exist without drinking. Although a few incidents that occur in bars involve interpersonal conflict between friends or couples that might have occurred in another setting, almost all incidents of aggression that occur in bars are unplanned, emerge from the social interaction in the bar (Graham and Wells 2001) and often involve strangers. Therefore, it seems reasonable to assume that close to $100 \%$ of incidents of violence occurring in bars and other environments where drinking is the main activity should be considered attributable to alcohol, either directly through the pharmacological effects of alcohol or indirectly through the social norms related to drinking.

Estimating the proportion of violence in other settings that should be attributed to alcohol is more problematic. There are pharmacological effects of alcohol, as described above, that make aggressive interactions more likely. This is more likely to be the case if all those involved have been drinking, owing to the interaction of the effects of alcohol on each person (Leonard 1984). In addition, alcohol is known to increase the likelihood of the escalation of conflict (Martin and Bachman 1997; Sharps et al. 2001). On the other hand, marital violence, for example, often occurs when neither party has been drinking. Therefore, the assessment of the exact proportion of alcohol-related violent injuries and death that should be attributable to alcohol is often difficult, and needs to be assessed from different sources, such as time series analyses, natural experiments, case-control studies, emergency-room studies, general population surveys and experimental designs (Pernanen 2001).

## DERIVING THE AAFs FOR DIFFERENT SUBREGIONS FOR INJURIES (BOTH INTENTIONAL AND UNINTENTIONAL)

Injuries are also influenced by average volume of alcohol consumption and by patterns of drinking, especially by acute levels of BAC or intoxication. To model this relationship, a multilevel analysis identical to the one on IHD was used (for statistical derivation and specification of the model see section 3.2; for formulas see notes below Table 12.13). Table 12.21 gives an overview of the underlying data. Table 12.22 gives the main results of the analysis.

The effect of alcohol on injury at pattern 1 is 0.013 for males and 0.010 for females. The effect for pattern 2 populations is the same, as the coefficients did not significantly differ. For pattern 3, the effect is 0.056 for males and 0.014 for females. For pattern 4 , the respective effects are 0.196 for males and 0.027 for females.

The results can be summarized as follows.

- Average volume of drinking has a significant detrimental effect on risk of injury even at consumption pattern 1 , independent of sex. The impact is larger in males.
- The impact of per capita consumption on injury is different between different countries, as shown by the significant variance component.
- No significant difference in the effect of drinking on injury risk was found between patterns 1 and 2 .
- Pattern 3 has a significantly higher injury risk for both sexes, but the impact is much stronger in males (about 10 times).
- Pattern 4 has the highest injury risk for both sexes, and again the impact is much stronger for males.

To estimate AAFs for injuries we used the Australian AAFs, since these reflected the most up-to-date information (Ridolfo and Stevenson 2001). These were converted into odds ratios, ${ }^{28}$ which were applied to estimated exposure prevalence for all subregions, as with other diseases, multiplicatively adjusted by pattern weights from Table 12.22 and average volume of alcohol consumption. This procedure must be regarded as a crude approximation, yet the best attainable at present. Given the potential variation in the role of alcohol in casualties across settings, there is an urgent need for empirical studies of the relationship in different world regions, using a variety of methods. The WHO Collaborative Study on Alcohol and Injuries (www.who.int/ substance_abuse/topic_alcohol_injuries.htm) constitutes a step forward on this.

Appendix B gives an overview of the derived AAFs for mortality for major categories of accidental and intentional injuries, as defined for the CRA project.

Table I2.2I Characteristics of data set to calculate the relationship between per capita consumption, patterns of drinking and injury mortality

| Country | Injury mortality, males per 1000 | Injury mortality, females per 1000 | Average per capita alcohol consumption |  | Year of first data | Year of last data | Number of years |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Albania | 0.74 | 0.21 | 2.20 | 3 | 1992 | 1998 | 7 |
| Argentina | 0.87 | 0.28 | 27.48 | 2 | 1966 | 1996 | 25 |
| Armenia | 0.37 | 0.10 | 2.53 | 2 | 1992 | 1998 | 7 |
| Australia | 0.78 | 0.31 | 11.19 | 2 | 1964 | 1997 | 34 |
| Austria | 1.07 | 0.39 | 14.16 | I | 1962 | 1999 | 38 |
| Azerbaijan | 0.68 | 0.14 | 2.10 | 3 | 1990 | 1999 | 9 |
| Bahamas | 0.97 | 0.30 | 14.30 | 2 | 1969 | 1995 | 13 |
| Bahrain | 0.33 | 0.09 | 5.33 | 2 | 1985 | 1988 | 3 |
| Barbados | 0.63 | 0.17 | 7.88 | 2 | 1964 | 1995 | 31 |
| Belarus | 1.55 | 0.35 | 9.68 | 4 | 1989 | 1998 | 9 |
| Belgium | 0.87 | 0.41 | 12.35 | 1 | 1964 | 1994 | 31 |
| Belize | 0.59 | 0.17 | 6.01 | 4 | 1964 | 1995 | 28 |
| Bulgaria | 0.89 | 0.26 | 11.97 | 2 | 1982 | 1998 | 17 |
| Canada | 0.81 | 0.31 | 9.74 | 2 | 1964 | 1997 | 34 |
| Chile | 1.44 | 0.33 | 11.58 | 3 | 1964 | 1994 | 31 |
| Colombia | 1.72 | 0.32 | 5.49 | 3 | 1967 | 1994 | 19 |
| Costa Rica | 0.94 | 0.25 | 4.48 | 4 | 1964 | 1995 | 32 |
| Croatia | 0.93 | 0.31 | 13.24 | 3 | 1993 | 1998 | 6 |
| Czech Republic | 0.92 | 0.40 | 15.42 | 2 | 1986 | 1999 | 14 |
| Denmark | 0.77 | 0.40 | 10.91 | 2 | 1964 | 1996 | 33 |
| Dominican Republic | 0.63 | 0.20 | 2.91 | 2 | 1965 | 1985 | 21 |
| Ecuador | 1.23 | 0.33 | 2.30 | 3 | 1964 | 1995 | 31 |
| El Salvador | 1.99 | 0.36 | 1.95 | 4 | 1964 | 1993 | 19 |
| Estonia | 2.50 | 0.58 | 13.43 | 3 | 1992 | 1999 | 8 |
| Fiji | 0.50 | 0.26 | 2.98 | 3 | 1978 | 1978 | 1 |
| Finland | 1.21 | 0.35 | 7.42 | 3 | 1964 | 1996 | 33 |
| France | 1.02 | 0.43 | 23.15 | 1 | 1964 | 1997 | 34 |
| Germany | 0.54 | 0.21 | 13.63 | 1 | 1993 | 1998 | 6 |
| Greece | 0.58 | 0.23 | 9.69 | 2 | 1964 | 1998 | 35 |
| Guatemala | 1.49 | 0.26 | 2.33 | 4 | 1964 | 1984 | 14 |
| Guyana | 0.90 | 0.24 | 9.50 | 3 | 1979 | 1994 | 4 |
| Honduras | 0.62 | 0.10 | 2.62 | 4 | 1976 | 1979 | 4 |
| Hungary | 1.39 | 0.54 | 15.85 | 3 | 1977 | 1999 | 23 |
| Iceland | 0.85 | 0.29 | 5.42 | 3 | 1964 | 1996 | 33 |
| Ireland | 0.60 | 0.24 | 10.86 | 3 | 1964 | 1996 | 33 |
| Israel | 0.56 | 0.29 | 3.04 | 2 | 1975 | 1996 | 22 |
| Italy | 0.64 | 0.24 | 18.48 | I | 1964 | 1996 | 33 |
| Jamaica | 0.41 | 0.10 | 3.12 | 2 | 1964 | 1985 | 7 |
| Japan | 0.68 | 0.27 | 5.78 | I | 1964 | 1997 | 34 |

Table 12.2I Characteristics of data set to calculate the relationship between per capita consumption, patterns of drinking and injury mortality (continued)

| Country | Injury mortality, males per 1000 | Injury mortality, females per 1000 | Average per capita alcohol consumption |  | Year of first data | Year of last data | Number of years |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Kazakhstan | 1.65 | 0.44 | 7.24 | 4 | 1989 | 1998 | 10 |
| Kyrgyzstan | 0.91 | 0.25 | 1.88 | 3 | 1995 | 1999 | 5 |
| Latvia | 2.09 | 0.52 | 8.53 | 3 | 1989 | 1998 | 10 |
| Lithuania | 2.26 | 0.51 | 5.76 | 3 | 1989 | 1998 | 5 |
| Luxembourg | 0.97 | 0.37 | 17.37 | 1 | 1967 | 1997 | 31 |
| Malta | 0.35 | 0.14 | 5.27 | I | 1965 | 1998 | 33 |
| Mauritius | 0.80 | 0.26 | 3.30 | 3 | 1964 | 1998 | 35 |
| Mexico | 1.60 | 0.33 | 3.94 | 4 | 1962 | 1995 | 34 |
| Netherlands | 0.51 | 0.27 | 9.39 | I | 1964 | 1997 | 34 |
| New Zealand | 0.81 | 0.35 | 11.36 | 2 | 1964 | 1998 | 35 |
| Nicaragua | 1.25 | 0.28 | 3.29 | 4 | 1964 | 1994 | 17 |
| Norway | 0.69 | 0.27 | 5.11 | 3 | 1964 | 1996 | 33 |
| Peru | 0.67 | 0.21 | 6.13 | 3 | 1966 | 1989 | 16 |
| Philippines | 0.67 | 0.16 | 2.59 | 3 | 1964 | 1993 | 18 |
| Portugal | 0.98 | 0.28 | 18.85 | 1 | 1964 | 1998 | 35 |
| Qatar | 0.47 | 0.17 | . 84 | 2 | 1995 | 1995 | I |
| Republic of Korea | 1.09 | 0.38 | 8.90 | 3 | 1985 | 1997 | 13 |
| Romania | 1.11 | 0.33 | 10.78 | 3 | 1989 | 1998 | 10 |
| Russian | 2.05 | 0.50 | 8.88 | 4 | 1991 | 1998 | 8 |
| Federation |  |  |  |  |  |  |  |
| Singapore | 0.64 | 0.23 | 2.30 | 2 | 1964 | 1998 | 35 |
| Slovakia | 1.04 | 0.31 | 12.54 | 3 | 1992 | 1995 | 4 |
| Slovenia | 1.15 | 0.38 | 14.80 | 2 | 1994 | 1998 | 5 |
| Spain | 0.60 | 0.19 | 17.20 | I | 1962 | 1997 | 36 |
| Sri Lanka | 1.01 | 0.40 | . 27 | 3 | 1964 | 1986 | 13 |
| Suriname | 1.05 | 0.37 | 6.43 | 3 | 1964 | 1992 | 21 |
| Sweden | 0.72 | 0.30 | 7.57 | 3 | 1964 | 1996 | 33 |
| Switzerland | 0.93 | 0.38 | 14.99 | I | 1964 | 1994 | 31 |
| Tajikistan | 0.58 | 0.16 | 2.52 | 3 | 1992 | 1992 | 1 |
| Thailand | 0.92 | 0.26 | 3.58 | 3 | 1964 | 1994 | 25 |
| The former Yugoslav Republic of Macedonia | 0.47 | 0.15 | 5.76 | 3 | 1994 | 1997 | 4 |
| Trinidad and Tobago | 0.95 | 0.27 | 4.79 | 2 | 1962 | 1994 | 29 |
| Turkmenistan | 0.43 | 0.18 | 2.13 | 3 | 1995 | 1995 | 1 |
| Ukraine | 1.41 | 0.32 | 4.64 | 3 | 1990 | 1999 | 10 |
| United Kingdom | 0.47 | 0.23 | 9.49 | 2 | 1964 | 1998 | 35 |
| United States | 0.92 | 0.32 | 9.46 | 2 | 1964 | 1997 | 34 |
| Uruguay | 0.84 | 0.28 | 8.08 | 3 | 1966 | 1990 | 24 |
| Uzbekistan | 0.46 | 0.15 | 1.07 | 3 | 1996 | 1998 | 3 |
| Venezuela | 1.29 | 0.33 | 8.31 | 3 | 1969 | 1994 | 24 |

Table I 2.22 Effects of per capita consumption and patterns of drinking on risk of injury mortality

|  | Coefficient | SE | $t$-value | df | P | Variance component | df | $\chi^{2}$ | P |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Males |  |  |  |  |  |  |  |  |  |
| Average volume $=$ per capita consumption |  |  |  |  |  |  |  |  |  |
| Not adjusted | 0.045 | 0.013 | 3.46 | 79 | 0.001 | 0.0106 | 72 | 584.1 | 0.000 |
| Adjusted by GNP, year | 0.047 | 0.012 | 3.90 | 79 | 0.000 | 0.0097 | 72 | 620.6 | 0.000 |
| After inclusion of pattern dummy variables on second level | 0.013 | 0.006 | 2.05 | 76 | 0.040 | 0.0080 | 69 | 532.1 | 0.000 |
| Patterns of drinking ${ }^{\text {a }}$ |  |  |  |  |  |  |  |  |  |
| Pattern 2 | -0.009 | 0.011 | $<1$ | 76 | 0.391 |  |  |  |  |
| Pattern 3 | 0.043 | 0.014 | 3.13 | 76 | 0.002 |  |  |  |  |
| Pattern 4 | 0.173 | 0.062 | 2.77 | 76 | 0.006 |  |  |  |  |
| Females |  |  |  |  |  |  |  |  |  |
| Average volume $=$ per capita consumption |  |  |  |  |  |  |  |  |  |
| Not adjusted | 0.0097 | 0.0023 | 4.19 | 79 | 0.000 | 0.00028 | 72 | 398.7 | 0.000 |
| Adjusted by GNP, year | 0.0121 | 0.0021 | 5.91 | 79 | 0.000 | 0.00021 | 72 | 496.2 | 0.000 |
| After inclusion of pattern dummy variables on second level | 0.0096 | 0.0021 | 4.46 | 76 | 0.000 | 0.00020 | 69 | 412.8 | 0.000 |
| Patterns of drinking ${ }^{\text {a }}$ |  |  |  |  |  |  |  |  |  |
| Pattern 2 | -0.0024 | 0.0031 | <1 | 76 | 0.440 |  |  |  |  |
| Pattern 3 | 0.0041 | 0.0030 | 1.40 | 76 | 0.163 |  |  |  |  |
| Pattern 4 | 0.0171 | 0.0077 | 2.24 | 76 | 0.025 |  |  |  |  |
| a Patterns 2, 3 and 4 are compared to pattern 1 . |  |  |  |  |  |  |  |  |  |

Another point concerns the relation of alcohol with type of outcomemorbidity vs mortality. In general, more severe outcomes are more related to alcohol than less severe outcomes (Rehm et al. 2003b; Single et al. 1999b). Consequently, the AAFs for mortality should be higher than the AAFs for morbidity. Unfortunately, most research to determine AAFs for injury did not explicitly separate mortality and morbidity (see, e.g. Table 12.20). Ridolfo and Stevenson (2001) explicitly separated the AAFs for motor vehicle accidents, and for males found 0.328 for deaths and 0.247 for hospitalizations; they lacked sufficient data for females. Based on their work and that of Cherpitel (1994, 1996b), we determined the ratio of AAF for morbidity as two thirds of the AAF for mortality. The ratio for other kinds of injury is lower (Cherpitel 1994, 1996b). To
be conservative, these ratios were set at 0.44 (or two thirds of the ratio for motor vehicle accidents) (Cherpitel 1994, 1996b).

### 3.12 Quantitative and qualitative sources of uncertainty

Since most of the chronic disease relationships with alcohol depend on biochemical processes linked to average volume of consumption over time, their hazards have been fairly stable across countries (Corrao et al. 2000). On the other hand, injuries are context-dependent to a much larger degree. A good example is the difference between liver cancer and traffic accidents. Based on biochemical evidence, there are reasons to believe that the relationship between average volume of alcohol consumption and liver cancer is relatively stable across different countries and societies, even though epidemiological work tends to be concentrated in established market economies. The most notable exception for chronic disease has been IHD, where patterns of drinking play a decisive role in determining the impact of average volume of drinking. On the other hand, the number of accidents (and alcohol-related traffic accidents in particular) depends on many background variables, as illustrated above. Thus, the risk relations between injuries and alcohol are much less stable and their transferability is more questionable. Where it has to be done, it should carry wider CIs.

Based on the above considerations, the following pertain.

- For chronic diseases, estimates of relative risk are usually based on meta-analyses of more than 20 studies with relatively small CIs. The uncertainty introduced by cross-population transfer of data is not that large, as the relationships depend on biochemical mechanisms. It is therefore suggested that $\pm 15 \%$ of the point estimate be used as the standard in an uncertainty analysis. This applies to all chronic disease categories where AAFs are directly derived from prevalence and relative risk.
- In all cases where AAFs are derived in other ways (e.g. injury), these fractions are more influenced by contextual differences from one region to another and should thus be modelled with more uncertainty. We suggest $\pm 30 \%$ of the point estimate to account for additional assumptions.
- IHD is the notable exception. In this case, estimates differ considerably (see the heterogeneity in the meta-analysis of Corrao et al. 2000), with respect not only to the magnitude but also to the direction of the relation. To account for this uncertainty and the difference of estimates in different models, we suggest the values set out in Table 12.23.
Table 12.23 is based on the following assumptions.
- For AMR-A, EUR-A and WPR-A, the results from the individual-level meta-analysis (Corrao et al. 2000) were taken as best estimates. For

Table I2.23 AAF estimates and uncertainty intervals for IHD by subregion

|  | Best estimates |  |
| :--- | :---: | ---: |
| Subregion | Males |  |
| AFR-D | $0.02(-0.05,0.05)$ | Females |
| AFR-E | $0.07(0.00,0.09)$ | $0.00(-0.03,0.03)$ |
| AMR-A | $-0.13(-0.17,0.00)$ | $0.00(-0.03,0.03)$ |
| AMR-B | $0.16(0.00,0.21)$ | $-0.08(-0.12,0.00)$ |
| AMR-D | $0.08(0.00,0.12)$ | $0.02(-0.05,0.05)$ |
| EMR-B | $0.00(-0.03,0.03)$ | $0.03(-0.05,0.05)$ |
| EMR-D | $0.01(-0.03,0.03)$ | $0.00(-0.03,0.03)$ |
| EUR-A | $-0.16(-0.21,0.10)$ | $0.00(0.03,0.03)$ |
| EUR-B | $0.11(-0.13,0.15)$ | $-0.13(-0.17,0.04)$ |
| EUR-C | $0.15(0.10,0.20)$ | $0.00(-0.08,0.05)$ |
| SEAR-B | $0.01(-0.03,0.03)$ | $0.03(0.02,0.04)$ |
| SEAR-D | $0.04(-0.03,0.03)$ | $0.00(-0.03,0.03)$ |
| WPR-A | $-0.17(-0.23,-0.10)$ | $0.00(-0.03,0.03)$ |
| WPR-B | $0.01(-0.05,0.05)$ | $-0.13(-0.17,-0.07)$ |

a Best estimates derived from relative risk and not from multilevel estimates.
these subregions the results from the aggregate multilevel analysis were taken as upper limits, and $30 \%$ lower than the best estimate as lower limits.

- For EUR-B, the aggregate multilevel results were taken as best estimates, and the results from the individual-level analysis was taken as lower limit, and $30 \%$ higher than the best estimate as upper limit.
- For all other subregions, the aggregate multilevel analysis results were taken as best estimates and, based on pattern and volume of the subregion, uncertainty intervals were chosen as follows:
- $\pm 0.03$ in the case of low-volume drinking (average $<3 \mathrm{~g} /$ day) and average pattern values lower than 3.5;
- $\pm 0.05$ in the case of females for volumes $>3 \mathrm{~g} /$ day and average pattern values lower than 3.5 (this restriction for females was intended to account for the higher uncertainty of pattern values for females); and
- for all other estimates zero was taken as the lower bound, and the best estimate plus $30 \%$ as the upper bound (except for EURC, the only subregion with an average pattern value greater than 3.5).

A number of points need to be emphasized when interpreting these results. The underlying research for chronic disease is quite heterogeneous with respect to quality. In particular, measurement of alcohol provides limited information on patterns of drinking and for characteristics of abstainers. Most studies have just one time measurement of exposure. Often cohorts were selected with respect to minimizing loss to followup, and thus samples with more regular, low-to-moderate drinking styles were used. This constitutes a problem for estimating the effects of patterns of drinking, as well as for estimating the effects of continuous heavy drinking.

There is also a problem of measuring exposure with respect to acute consequences, although slightly different because often the BAC is given as the only indicator. Such a measure does not allow one to differentiate between the effects of pattern of drinking and average volume of alcohol consumption, as a heavy drinking occasion may be the exception or the norm. But for the population level, we need both types of information, as numbers of injuries will depend on both (see above). In addition, the BAC alone does not allow one to determine if alcohol was a contributing causal factor or not, only in combination with other information on control conditions, i.e. series of BACs in accident and non-accident conditions (Borkenstein et al. 1964). Unfortunately, such control conditions are lacking in most research (Gmel and Rehm 2003). Thus, with the exception of traffic accidents, the overall quality of the underlying research for most alcohol-related acute outcomes is of poor quality and derived AAFs may be subject to considerable error. This is reflected in the wide uncertainty margins suggested above.

### 3.13 Estimates of risk reversibility

Part of the risk from alcohol is immediately reversible: all acute risks can be completely reversed if alcohol is removed. Chronic diseases often depend on lifetime exposure, and thus risk is often reduced but not completely eliminated by removal of alcohol.

On the other hand, there are indications that a reduction of alcohol consumption in populations is associated with a fairly rapid decrease in chronic diseases such as liver cirrhosis. For example, time series analyses showed that decreases in per capita consumption were associated with considerable concurrent reductions in liver cirrhosis (e.g. Ramstedt 2001; Skog 1980; and especially Cook and Tauchen 1982).

Another example of a chronic condition with rapid, sometimes almost immediate remission is depression. In fact, most studies come to the conclusion that many depressive syndromes markedly improve within days to weeks of abstinence (Brown and Schuckit 1988; Dackis et al. 1986; Davidson 1995; Gibson and Becker 1973, Penick et al. 1988; Pettinati et al. 1982; Willenbring 1986).

It is not clear what effect alcohol removal would have on alcohol use disorders. Clearly, some criteria of both alcohol dependence and harmful
use of alcohol would no longer apply (e.g. continued use despite harmful consequences).

## 4. DISCUSSION OF ESTIMATES OF ALCOHOLATTRIBUTABLE BURDEN

### 4.1 Mortality

Alcohol-related burden of disease is considerable: $3.2 \%$ of global mortality and $4.0 \%$ of global burden of disease as measured in DALYs. In terms of alcohol-related mortality, almost half of the global burden $(46 \%)$ is related in acute causes, i.e. unintentional and intentional injuries (see Table 12.24; details in Appendix C). Within this mortality burden, for acute causes, unintentional injuries are by far the most important. The next important category is malignant neoplasms with $20 \%$ of the overall alcohol-related mortality burden, followed by cardiovascular diseases ( $15 \%$ of all alcohol-attributable deaths) and other noncommunicable diseases, a category almost entirely made up of liver cirrhosis ( $13 \%$ ). Cardiovascular deaths are a special case in that different patterns of drinking lead to beneficial and detrimental outcomes. Thus, the net result of $15 \%$ does not give a clear picture of the underlying structure. Going beyond the net result, alcohol was estimated to cause a total of almost 600000 cardiovascular deaths in the year 2000, exceeding even the alcohol-related burden of unintentional injuries. This figure was partly "offset" by the beneficial effects of alcohol on IHD and stroke. More males than females die of the effects of alcohol, with a ratio of about $10: 1$.

Table I2.24 Global deaths (000s) attributable to alcohol by major disease and injury categories, 2000

| Disease or injury | Males | Females | Total | Percentage of all <br> alcohol-attributable deaths |
| :--- | ---: | :---: | :---: | :---: |
| Conditions arising during the <br> perinatal period | 2 | 1 | 3 | 0 |
| Malignant neoplasms | 269 | 86 | 355 | 20 |
| Neuro-psychiatric conditions | 91 | 19 | 111 | 6 |
| Cardiovascular diseases | 392 | -124 | 268 | 15 |
| Other noncommunicable diseases | 193 | 49 | 242 | 13 |
| (type II diabetes, liver cirrhosis) |  |  |  |  |
| Unintentional injuries | 484 | 92 | 577 | 32 |
| Intentional injuries | 206 | 42 | 248 | 14 |
| Alcohol-related mortality (all causes) | 1638 | 166 | 1804 | 100 |
| All deaths | 29232 | 26629 | 55861 | In comparison, estimated |
| Percentage of all deaths that can | 5.6 | 0.6 | 3.2 | total for I990: I.5 |
| be attributable to alcohol |  |  |  |  |

The overall relationship between average volume of alcohol consumption and all-cause mortality is thus J-shaped in established market economies for age groups under 45 years, where benefits of light to moderate consumption on IHD apply (Holman et al. 1996; Rehm et al. 2001c). In countries with a predominant pattern of irregular heavy drinking, no J-shape can be expected and the shape between alcohol and all-cause mortality is expected to increase monotonically.

The estimated percentage of alcohol-attributable mortality (3.2\%) is more than double that estimated in the 1990 GBD study $(3.2 \%$ vs $1.5 \%)$. There are several reasons for this increase. First, alcohol consumption has increased overall, especially in the very populous SEAR-B, SEAR-D and WPR-B subregions, including China and India. In addition, in these subregions we do not expect benefits of drinking, unless the current patterns of drinking change to the positive. Second, the relative impact of injuries and chronic disease on overall mortality, both of which are related to alcohol, has increased over the past 10 years. Third, the methodologies are not comparable between the two estimates. The 2000 estimate differs in the following three major respects.

- It is much more disaggregated, with respect both to burden categories and to regional data. Thus, the present work has included adult per capita data for almost all countries, and much more survey-related data than available for the 1990 estimates.
- The present exercise explicitly includes quantifiable patterns of drinking for both IHD and injuries, whereas the 1990 estimates were almost entirely based on volume of consumption. This difference is most striking with regard to IHD, where the 1990 study considered only beneficial effects. The current exercise estimates both beneficial and detrimental effects, depending on patterns of drinking.
- The meta-analyses on average volume of consumption and different disease outcomes have become much more refined in terms of methodology (compare, e.g. Corrao et al. 1999, 2000 with the methodology of English et al. 1995).

Finally, the estimates in this work are restricted to GBD disease categories. Several diseases related to alcohol could not be accounted for, most notably cardiac arrhythmias and heart failure. In the GBD disease categories, cardiac arrhythmias and heart failure would be part of "other cardiac conditions". We had neither epidemiological studies on the hazards for various diseases in this broad category nor any data on the relative proportion of cardiac arrhythmias and heart failure among other cardiac conditions by subregion and sex, in order to separate these diseases. In addition, oesophageal varices, acute and chronic pancreatitis and several conditions occurring during the perinatal period could not be included for similar reasons. However, their alcohol-attributable mor-
tality burden would be likely to be minor compared to the "other cardiac conditions" mentioned above.

### 4.2 DALYs

Alcohol-attributable DALYs are summarized in Table 12.25. See Appendix D for details by subregion.

The biggest shift in the relative impact of disease categories compared to the pattern for alcohol-caused mortality is seen for neuropsychiatric diseases. Neuropsychiatric diseases are often disabling, but rarely fatal, and this is reflected in the markedly higher proportion of overall disease burden due to alcohol ( $38 \%$ ) in this category compared to alcohol-attributable mortality ( $6 \%$ ). Males have far more ( $>5$-fold) alcohol-related disease burden than females. The mortality and burden of disease figures presented here are net figures, where the alcohol-related beneficial effects on disease have been subtracted from its harmful effects. Therefore, the detrimental effects of alcohol on mortality, and disease burden in general, far outweigh the beneficial effects.

What are the most striking differences between subregions? Clearly alcohol-related burden is most detrimental in the developed world. Here $9.2 \%$ of the entire disease burden is attributable to alcohol, only exceeded by the burden attributable to tobacco and blood pressure (see Table 12.26 and WHO 2002). Here also, the ratio of males to females is lowest. However, as Table 12.26 indicates, alcohol also places a toll on health in the developed world, with relatively low mortality

Table I 2.25 Global burden of disease in 2000 attributable to alcohol according to major disease categories (DALYs in 000s)

| Disease or injury | Males | Females | Total | Percentage of all alcohol-attributable DALYs |
| :---: | :---: | :---: | :---: | :---: |
| Conditions arising during the perinatal period | 68 | 55 | 123 | 0 |
| Malignant neoplasm | 3180 | 1021 | 4201 | 7 |
| Neuro-psychiatric conditions | 18090 | 3814 | 21904 | 38 |
| Cardiovascular diseases | 4411 | -428 | 3983 | 7 |
| Other non-communicable diseases (type II diabetes, liver cirrhosis) | 3695 | 860 | 4555 | 8 |
| Unintentional injuries | 14008 | 2487 | 16495 | 28 |
| Intentional injuries | 5945 | 1117 | 7062 | 12 |
| Alcohol-related disease burden all causes (DALYs) | 49397 | 8926 | 58323 | 100 |
| All DALYs | 761562 | 693911 | 1455473 | In comparison, estimated total |
| Percentage of all DALYs that can be attributable to alcohol | 6.5 | 1.3 | 4.0 | for 1990: 3.5 |

Table I2.26 Burden of disease in 2000 attributable to tobacco, alcohol and drugs, by development status and sex

|  | $\begin{gathered} \text { High-mortality developing } \\ \text { subregions } \\ \text { (AFR-D, AFR-E, AMR-D, EMR-D, SEAR-D) } \end{gathered}$ |  |  | Low-mortality developing subregions <br> MR-B, EMR-B, SEAR-B, WPR-B) |  |  | Developed subregions <br> (AMR-A, EUR-A, EUR-B, EUR-C, WPR-A) |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Males | Females | Total | Males | Females | Total | Males | Females | Total |
| Total DALYs (000s) | 420711 | 412052 | 832763 | 223181 | 185316 | 408497 | 117670 | 96543 | 214213 |
| Smoking and oral tobacco | 3.4\% | 0.6\% | 2.0\% | 6.2\% | 1.3\% | 4.0\% | 17.1\% | 6.2\% | 12.2\% |
| Alcohol | 2.6\% | 0.5\% | 1.6\% | 9.8\% | 2.0\% | 6.2\% | 14.0\% | 3.3\% | 9.2\% |
| Illicit drugs | 0.8\% | 0.2\% | 0.5\% | 1.2\% | 0.4\% | 0.8\% | 2.4\% | 1.2\% | 1.8\% |

patterns. Here the disease burden attributable to alcohol is the highest of all 26 risk factors examined in the CRA of the GBD study in 2000 (Ezzati et al. 2002). In high-mortality developing subregions, in Africa and parts of south-east Asia, alcohol is not yet one of the major risk factors. Here, the most important risk factors are underweight, unsafe sex, unsafe water, sanitation and hygiene, and other environmental factors. However, if past developments can help predict the future, we can expect that the alcohol-attributable burden will increase in these subregions along with economic development (see also section 5).

### 4.3 Conclusions

Alcohol causes a considerable burden of disease, in terms both of mortality and disability. While the total elimination of alcohol is not realistic, there are evidence-based policy measures that could substantially reduce the burden of alcohol. The recent review by Ludbrook et al. (2001) on measures to reduce alcohol misuse assessed the quality of evidence for four types of intervention aimed at reducing alcohol use and its consequences. Their findings coincide with a number of earlier reviews (e.g. Bruun et al. 1975; Edwards et al. 1994) and with the overview of Babor et al. (2003). In sum, the following measures were found quite effective:

- policy and legislative interventions, including taxation on alcohol sales, drink-driving laws, restricted licensing of outlets and advertising controls;
- law enforcement, for example random breath-testing of drivers;
- community interventions; and
- brief interventions.

On the other hand, mass media and awareness campaigns were not found to be very effective, although they seemed to be somewhat more popular with politicians and policy-makers.

Since these interventions exist and have been empirically shown to reduce the burden of both chronic and acute disease caused by alcohol, and also alcohol-attributable social harm, there is no justification for alcohol-related disease to remain at such a high level in many parts of the world.

## 5. Projections of the future

Quantitative projections regarding future exposure to alcohol are feasible only for average volume of drinking, since it is extremely difficult if not impossible to judge how drinking patterns will alter over time (see Figures 12.3-12.5).

Adult per capita consumption in EUR-A and EUR-B seems to be driven by long-term trends (Mäkelä et al. 1981; Simpura 1998). There

Figure I 2.3 Adult per capita consumption in litres of absolute alcohol for the AFR, EMR and EUR subregions


Figure 12.4 Adult per capita consumption in litres of absolute alcohol for the SEAR and WPR subregions

are indications that the current downward trend is levelling off. However, it is very hard to determine the period of long-term waves with short-term time series since 1960 or so. Thus, we predict a stable exposure for EUR-A and EUR-B at about the average level of consumption of the past 10 years. For EUR-C, the curve shows the most change. The dip at the end of the 1980s is due to the anti-alcohol campaign of the Gorbachev period in the former Soviet Union (White 1996). At the end

Figure I 2.5 Adult per capita consumption in litres of absolute alcohol for the AMR subregions

of the campaign, alcohol consumption rose again to former levels. Again, since there is no clear trend after the campaign, we predict about the same level of consumption as the average of the years after the end of the Gorbachev campaign. No trends are apparent for EMR-B, EMR-D and AFR-D. Therefore, the most reasonable projection of future alcohol consumption would be the level of current consumption. To obtain more stable estimates, the average of the 1990s has been used.

For the South-East Asia and Western Pacific subregions, the following predictions appear justified based on the trend data shown in Figure 12.4. For WPR-A, there was an upward trend that seems to have stopped at the end of the 1980s; thus, the average volume of the 1990s was used as the best projection. For WPR-B and SEAR-B, consumption clearly increased and we modelled future consumption by a linear upward trend. The upward trend in SEAR-D was less pronounced, but nonetheless present, and we therefore again used the linear upward trend.

In the Americas, there has been a long wave of increasing and then decreasing consumption for North America (AMR-A), almost parallel to the European consumption. For AMR-B and AMR-D there are slight downward trends. Since the long-term wave seems to turn upwards again, and since the downward trends did not reach significance, it seems prudent to model the future for all three subregions at about the same as the average of the 1990 s, as a stable estimate of the status quo.

Trends in drinking patterns have been studied in very special subsets of populations, such as youth in Europe (Hibell et al. 2000). But at the global level, for pattern of drinking, there is not even enough reliable data concerning the current situation, and thus quantitative predictions
are not possible. Therefore we assumed constant drinking patterns under a "business-as-usual" scenario. Projections of future adult per capita alcohol consumption are given in Table 12.27.

In three subregions (SEAR-B, SEAR-D and WPR-B) unrecorded consumption has to be added in order to predict burden, and these data need to be converted into drinking categories as above, based on surveys. Since we have neither survey data on the future nor any predictions for survey or unrecorded consumption, we suggest modelling future consumption as follows.

- For all subregions except SEAR-B, SEAR-D and WPR-B, predictions are based on 2000 data on proportions of drinking categories.
- Proportions of drinking categories II and III are increased by $2.1 \%$ over those for 2000 for SEAR-B and WPR-B, and by $0.7 \%$ for SEARD (Table 12.28). Of course, there are limits of linear increase, and we considered the current levels of consumption in the A subregions as upper limits.

In summary, the best estimates predict global increases in average consumption of alcohol, triggered by increases in developing and emerging economies in the South-East Asia and Western Pacific regions.

Table 12.27 Projections of adult per capita consumption by subregion, in litres of pure alcohol, excluding SEAR-B, SEAR-D and WPR-B

|  | Subregion |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
|  | AFR-D | AFR-E | AMR-A | AMR-B | AMR-D |  |
| Mean | 3.543 | 4.127 | 8.276 | 6.669 | 3.462 |  |
| 95\% Cl upper | 3.608 | 4.258 | 8.428 | 6.826 | 3.558 |  |
| 95\% Cl lower | 3.478 | 3.996 | 8.125 | 6.512 | 3.366 |  |
| SD | 0.085 | 0.171 | 0.197 | 0.204 | 0.125 |  |
| Trend | No trend | No trend | Long waves | No trend | No trend |  |
|  |  |  |  |  |  |  |
| EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | WPR-A |  |
| Mean | 1.248 | 0.238 | 12.632 | 5.372 | 8.316 | 7.012 |
| $95 \% ~ C l ~ u p p e r ~$ | $I .355$ | 0.245 | 13.008 | 5.552 | 8.845 | 7.127 |
| $95 \%$ Cl lower | $I .142$ | 0.231 | 12.255 | 5.192 | 7.787 | 6.897 |
| SD | 0.139 | 0.009 | 0.490 | 0.234 | 0.688 | 0.149 |
| Trend | No trend | No trend | Long waves | No trend | No trend | Long waves |

Table 12.28 Projections of adult per capita consumption by subregion, in litres of pure alcohol, for SEAR-B, SEAR-D and WPR-B

|  | Subregion |  |  |
| :--- | :---: | :---: | :---: |
|  | SEAR-B | SEAR-D | WPR-B |
| Increase per year in litres <br> of pure alcohol after 2000 | 0.063 | 0.014 | 0.108 |
| $95 \% \mathrm{Cl}$ upper | 0.068 | 0.016 | 0.115 |
| $95 \% \mathrm{Cl}$ lower | 0.057 | 0.012 | 0.102 |
| Increase in adult per capita <br> consumption 2000 (\%) | 2.1 | 0.7 | 2.1 |
| Trend | Linear upward | Linear upward | Linear upward |
| Shared variation: year and 1960-1999a <br> consumption | $88.0 \%$ | $93.2 \%$ | $96.8 \%$ |

The shared variation or "explained variance" denotes a measure of strength of the relationship, i.e. how much of the variation of adult per capita consumption is explained by the linear trend.

## Acknowledgements

Parts of an earlier version of the contributions were made available on the WHO listserve for discussion, and we should like to thank all who responded. In addition, the authors would like to thank the following participants in the WHO expert meetings on alcohol as a risk factor for the burden of disease, held in Geneva on 29-30 January and 7-8 May 2001, for their valuable comments when revising prior versions: Mary Jane Ashley, Mary Assunta, Vivek Benegal, Susan Bondy, Ruth Bonita, Maximilian de Courten, Majid Ezzati, Oye Gureje, Wei Hao, Jennifer Hillebrand, Claudio Jeronimo da Silva, Håkan Leifman, Alan Lopez, Maria-Elena Medina-Mora, Zofia Mielecka-Kubien, Alexander Nemtsov, Ian B. Puddey, Margret Rihs-Middel, Leanne Riley, Kari J. Rummukainen, Shekhar Saxena, Raquel Shaw Moxam, Jussi Simpura, Bedirhan Üstün and Justin Willis.

Many people, too numerous to be listed here, contributed with data, and we wish to thank all of them for their help. Special thanks go to Ed Adlaf, Tom Greenfield, Wei Hao, Håkan Leifman, Maria-Elena MedinaMora and Tim Stockwell for conducting data analyses or data searches specifically for this exercise. Special thanks also go to Ron Kessler and his team, who provided specific estimates for the point on alcohol and depression, without which it would have been impossible to estimate alcohol-attributable burden on depression.

WHO, the Swiss Federal Office of Public Health (contract no. 00.001588), the Swiss National Science Foundation, the Addiction Research Institute in Zurich, Switzerland, the Swiss Institute for the Prevention of Alcohol and Other Drug Problems in Lausanne, and the

Centre for Addiction and Mental Health in Toronto, provided financial and/or technical support for this study.

Parts of the assessment of relations between alcohol and disease were undertaken for the report The social costs of alcohol consumption in Switzerland and received financial support from the Swiss Federal Office of Public Health (grant no. 98-000794/8120).

Several anonymous reviewers, as well as the members of the editorial committee (Majid Ezzati, Alan Lopez, Anthony Rodgers), deserve special thanks for their help in finalizing the text, as does Steve Vander Hoorn for the repeated burden calculations.

Finally, the authors would like to thank the staff at the Swiss Institute for the Prevention of Alcohol and Other Drug Problems in Lausanne and the Addiction Research Institute in Zurich, especially Irene Bosshard and Elisabeth Grisel, for technical assistance.

## Notes

1 See preface for an explanation of this term.
2 Social outcomes of alcohol consumption are defined as changes that affect the social behaviour of individuals, or their interaction with partners and other family members, or their circumstances (Rehm 2001). Social outcomes would include family problems, public disorder, or workplace problems (for overviews see Gmel and Rehm 2003; Klingeman and Gmel 2001). Social outcomes or consequences are not addressed in this chapter unless they are included in ICD-10. The majority of these problems are not covered by ICD-10, even though health is broadly defined by WHO to include well-being.

3 Intoxication and dependence are of course also influenced by biochemistry. However, since these two intermediate outcomes are central in shaping the effect of alcohol on many health and social outcomes, they are discussed separately. The other effects (e.g. on promotion of blood clot dissolution) are often specific for one disease or a limited group of diseases. Both intoxication and dependence are defined as health outcomes in ICD-10.
4 This is not to imply that there is no drinking to intoxication or occasions of heavy drinking in countries with established market economies; it is simply to say that this pattern of drinking is more common in countries with developing or emerging economies.
5 There are other ways of estimating unrecorded consumption, such as those based on available raw materials (see the estimates for the Russian Federation by Nemtsov 1998, 2000, 2002).
6 Surveys do not necessarily underestimate the recorded per capita consumption, even though the literature sometimes appears to imply it. For some countries, e.g. Mexico, adding up the figures from the survey may lead to higher estimates than the recorded per capita consumption.
7 This questionnaire can be obtained from the first author on request. It was finalized at a WHO expert meeting in Geneva, May 2001.

8 For comparison, a $75-\mathrm{cl}$ bottle of wine contains about 70 g of pure alcohol.
9 As part of the process of developing these ratings, an earlier list of derived and assigned pattern values as shown in Table 12.3 was made available on a WHO listserve to a large number of key informants for critical assessment. This process resulted in the identification of local surveys, which helped improve the estimates.
10 This reasoning was also behind the list of those invited to a WHO workshop on unrecorded consumption in May 2001, where experts from Brazil, China, India, Nigeria and the Russian Federation met with other experts on the methodology of estimating unrecorded consumption to discuss current estimates and develop a methodology to improve data gathering.

11 Ledermann had been the first to claim that the distribution of alcohol consumption among drinkers is log-normal. Subsequent research found the exact shape to be different but still approximately log-normally distributed (e.g. Duffy 1986 and rejoinders).
12 These analyses are sometimes also called hierarchical linear analyses (Bryk and Raudenbush 1992). Since the term "hierarchical" is ambivalent (in sociology it has also been used to describe stepwise regression), we exclusively use the term "multilevel" in this chapter.

13 In the statistical literature units are called sections, hence the method used is called cross-sectional time series analysis. In our case, countries are sections.

14 Europe was taken as a pilot as data are most available there. The current analysis included data from 81 countries for injuries and 74 countries for IHD, most of them outside Europe.
15 Year is only controlling for the linear part of the time structure. However, sensitivity analyses were carried out to estimate the performance of the method used.

16 "Random" does not mean that the underlying relationships are completely random. Effects may be partly deterministic owing, for example, to different policies. The term "random" here means that effects across sections or countries cannot be estimated without error, and the errors are assumed to have a random distribution.

17 If the patterns were estimated as deviations from mean patterns, then the value of $\gamma_{10}$ would reflect the average impact of adult per capita consumption.
18 Even for biologically based relationships, the relationship could be moderated by other factors such as diet (e.g. alcohol may be related to breast cancer through hormonal effects, but diet also affects hormonal levels and this may have an influence on the alcohol-breast cancer relationship). However, except for IHD, meta-analyses on alcohol and chronic disease have yielded fairly similar effects for different populations, so the assumption of applying the same effect for average volume of drinking is probably justified.
19 Such "categorical" attribution is quite different from the statistical estimation used in other epidemiological studies. For the usual derivation of alcohol-attributable fractions see Rothman and Greenland (1998b).

20 Type of beverage has been excluded so far from our consideration of patterns of drinking. While the evidence is not conclusive on the effect of beverage on all-cause mortality or on cardiovascular disease (e.g. Gruenewald et al. 2000; Kerr et al. 2000; Rimm et al. 1996), there are some indications that cancers of the gastrointestinal tract are differentially influenced by alcohol in higher concentration.
21 Part of this lack of an influence on patterns of cancer risk may be due to methodological reasons. Most epidemiological studies measure only volume of consumption and model only monotonically increasing trends, and thus could not detect any influence of patterns of drinking even if they were present.
22 In the Inter-American Investigation of Mortality, which studied 4000 deaths in each of 12 cities in 1962-1964, the final assignment of all deaths from cirrhosis "with mention of alcoholism" was $80.4 \%$ of all cirrhosis deaths. About half of these had been "without mention" on the death certificate. Only in Santiago and Mexico City was the final assignment for "with mention" less than twice the initial number. The study used searches of medical records and interviews with decedents' families and attending physicians to reassign deaths from the initial classification (Room 1972, based on Puffer and Griffith 1967).
23 IHD is used here for denoting all diseases with ICD-9 rubrics 410-414 (ICD10: I20-I25). The same categories have also been labelled coronary heart disease (CHD).
24 DisMod is a software tool that may be used to check the internal consistency of epidemiological estimates of incidence, prevalence, duration and case fatality for diseases. The latest version (DisMod II) is distributed by WHO: http://www3.who.int/whosis/burden/burden_dismod/burden_dismod_dismo d2.cfm?path=whosis,burden, burden_dismod,burden_dismod_dismod2\& language=english.
25 The notion of a causal contributing role is at the heart of the epidemiological concept of causality (Rothman and Greenland 1998). According to such standards, as explained in the text, the causal role of alcohol in intentional injuries is established. According to criteria used in criminology, this may not be the case.

26 Part of the effect of alcohol on aggression or on other more social outcomes is due to psychological variables. The term "expectations" denotes the individual predictions and expectancies of what will happen after consumption of alcohol. Experimental research has demonstrated that such psychological variables play a role in determining outcome.
27 Often traffic accidents are described as if the environment's effects are totally independent of the person's behaviour. However, tired drunk-drivers on the road at 03:00 might well not have been there if they had not been drinking. Intentionality (and alcohol's effects on it) forms part of what are called "accidents".

28 Odds $=\mathrm{p} /(1-\mathrm{p})$.

## References

Abel EL (1997) Maternal alcohol consumption and spontaneous abortion. Alcohol and Alcoholism, 32:211-219.
Ahmad OB, Boschi-Pinto C, Lopez AD, Murray CJL, Lozano R, Inoue M (2000) Age standardization of rates: a new WHO world standard. (GPE Discussion Paper No. 31. EIP/GPE/EBD). World Health Organization, Geneva.
Ajani UA, Cook N, Hebert P et al. (1998) Alcohol consumption and incidence of malignant neoplasms: the physicians' health study. American Journal of Epidemiology, 147:S45.

Ajani UA, Hennekens C, Spelsberg A, Manson J (1999) Alcohol consumption and risk of type two diabetes mellitus. American Journal of Epidemiology, 149:S3.

Alpert JE, Fava M, Uebelacker LA et al. (1999) Patterns of axis I comorbidity in early-onset versus late-onset major depressive disorder. Biological Psychiatry, 46:202-211.
Alvear J, Andreani S, Cortés F (1998) Fetal alcohol syndrome and fetal alcohol effects: importance of early diagnosis and nutritional treatment. Revista Medica de Chile, 126:407-412.

American Psychiatric Association (1994) DSM-IV Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association, Washington, DC.

Ammann RW, Heitz PU, Klöppel G (1996) Course of alcoholic chronic pancreatitis: a prospective clinicomorphological long-term study. Gastroenterology, 111:224-231.

Apte MV, Wilson JS, Korsten MA (1997) Alcohol-related pancreatic damage: mechanisms and treatment. Alcohol Health and Research World, 21:1320.

Ashley MJ, Rehm J, Bondy S, Single E, Rankin J (2000) Beyond ischemic heart disease: are there other health benefits form drinking alcohol? Contemporary Drug Problems, 27:735-777.

Attili AF, Scafato E, Marchioli R, Marfisi RM, Festi D (1998) Diet and gallstones in Italy: the cross-sectional MICOL results. Hepatology, 27:1492-1598.

Babor T, Caetano R, Casswell S et al. (2003) Alcohol: no ordinary commodity. Research and public policy. Oxford University Press, Oxford.

Bagnardi V, Blangiardo M, La Vecchia C, Corrao G (2001) Alcohol consumption and the risk of cancer. Alcohol Research and Health, 25:236-270.

Bandera EV, Freudenheim JL, Vena JE (2001) Alcohol and lung cancer: a review of the epidemiologic evidence. Cancer Epidemiology, Biomarkers and Prevention, 10:813-821.

Baraona E, Lieber CS (1998) Alcohol and lipids. In: The Consequences of Alcoholism. Galanter M, ed. Plenum Press, New York.
Beaglehole R, Jackson R (1992) Alcohol, cardiovascular diseases and all causes of death: a review of the epidemiological evidence. Drug and Alcohol Review, 11:175-290.

Beilin LJ, Puddey IB, Burke V (1996) Alcohol and hypertension—kill or cure? Journal of Human Hypertension, 10:S1-5.

Berger K, Ajani UA, Kase CS et al. (1999) Light-to-moderate alcohol consumption and the risk of stroke among U.S. male physicians. New England Journal of Medicine, 341:1557-1564.
Bijleveld CCJH, van der Kamp LJT (1998) Longitudinal data analysis: designs, models and methods. Sage Publications Ltd., London.
Biork JM, Dougherty DM, Moeller FG (1999) Symptomatology of depression and anxiety in female "social drinkers". American Journal of Drug and Alcohol Abuse, 25:173-182.
Bland JM, Altman DG (1994) Statistics notes: correlation, regression, and repeated data. British Medical Journal, 308:896.
Blixen CE, McDougall GJ, Suen LJ (1997) Dual diagnosis in elders discharged from a psychiatric hospital. International Journal of Geriatric Psychiatry, 12:307-313.

Bobak M, McKee M, Rose R et al. (1999) Alcohol consumption in a national sample of the Russian population. Addiction, 94:857-866.

Bobak M, Marmot M (1999) Alcohol and mortality in Russia: is it different than elsewhere? Annals of Epidemiology, 9:335-338.
Bobak M, Skodova Z, Marmot M (2000) Effect of beer drinking on risk of myocardial infarction: population based case-control study. British Medical Journal, 320:1378-1379.
Bode C, Bode JC (1997) Alcohol's role in gastrointestinal tract disorders. Alcohol Health and Research World, 21:76-83.
Bondy SJ (1996) Drinking patterns and their consequences: report from an international meeting-overview of studies on drinking patterns and consequences. Addiction, 91:1663-1674.

Bondy S, Rehm J (1998) The interplay of drinking patterns and other determinants of health. Drug and Alcohol Review, 17:399-411.

Borkenstein RF, Crowther RF, Shumate RP (1964) The role of the drinking driver in traffic accidents. Department of Police Administration, Bloomington, IN.
Boutron MC, Faivre J, Dop MC, Quipourt V, Senesse P (1995) Tobacco, alcohol, and colorectal tumors, a multistep process. American Journal of Epidemiology, 141:1038-1046.
Bowlin SJ, Leske MC, Varma A, Nasca P, Weinstein A, Caplan L (1997) Breast cancer risk and alcohol consumption: results from a large case-control study. International Journal of Epidemiology, 26:915-923.
Bradley KA, Badrinath S, Bush K, Boyd-Wickizier J, Anawalt B (1998) Medical risks for women who drink alcohol. Journal of General Internal Medicine, 13:627-639.

Breslow RA, Weed DL (1998) Review of epidemiologic studies of alcohol and prostate cancer: 1971-1996. Nutrition and Cancer, 30:1-13.
Brewers Association of Canada (1997) Alcoholic beverage taxation and control policies. 9th edn. Brewers Association of Canada, Ottawa.

Britton A, McKee M (2000) The relation between alcohol and cardiovascular disease in Eastern Europe: explaining the paradox. Journal of Epidemiology and Community Health, 54:328-332.
Brown SA, Schuckit MA (1988) Changes in depression among abstinent alcoholics. Journal of Studies on Alcohol, 49:412-417.
Bruemmer B, White E, Vaughan TL, Cheney CL (1997) Fluid intake and the incidence of bladder cancer among middle-aged men and women in a three-county area of western Washington. Nutrition and Cancer, 29:163168.

Bruun K, Edwards G, Lumio M et al. (1975) Alcohol control policies in public health perspective. Finnish Foundation for Alcohol Studies, Helsinki.
Bryk AS, Raudenbush SW (1992) Hierarchical linear models: applications and data analysis methods. Advanced Quantitative Techniques in the Social Sciences. Sage Publications, Newbury Park, CA.

Bryk AS, Raudenbush SW, Congdon R (1996) HLM: Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs. Scientific Software International, Inc., Chicago, IL.
Bushman BJ (1997) Effects of alcohol on human aggression: validity of proposed mechanisms. In: Recent developments in alcoholism. Alcohol and violence. Galanter M, ed. Plenum, New York.
Camargo CA, Hennekens C, Gaziano JM, Glynn RJ, Manson J, Stampfer MJ (1997) Prospective study of moderate alcohol consumption and mortality in US male physicians. Archives of Internal Medicine, 157:79-85.
Caroli-Bosc FX, Deveau C, Peten EP et al. (1998) Cholelithiasis and dietary risk factors: an epidemiologic investigation in Vidauban, southeast France. General practitioner's group of Vidauban. Digestive Diseases and Sciences, 43:2131-2137.

Catalano P, Chikritzhs T, Stockwell T, Webb M, Rohlin C-J, Dietze P (2001) Trends in per capita alcohol consumption in Australia 1990/91-1998/99. No. 4. National Drug Research Institute, Perth.

Chadwick DJ, Goode JA (1998) Alcohol and cardiovascular diseases. John Wiley \& Sons, Chichester.
Chen CY, Lu CL, Lee PC, Wang SS, Chang FY, Lee SD (1999) The risk factors for gallstone disease among senior citizens: an Oriental study. Hepatogastroenterology, 46:1607-1612.
Chenet L, McKee M, Leon D, Shkolnikov V, Vassin S (1998) Alcohol and cardiovascular mortality in Moscow; new evidence of a causal association. Journal of Epidemiology and Community Health, 52:772-774.

Chermack ST, Giancola HM (1997) The relation between alcohol and aggression: An integrated biopsychosocial conceptualization. Clinical Psychology Review, 17:621-649.
Cherpitel CJ (1992) The epidemiology of alcohol-related trauma. Alcohol Health and Research World, 16:191-196.
Cherpitel CJ (1994) Alcohol and casualties: a comparison of emergency room and coroner data. Alcohol and Alcoholism, 29:211-218.

Cherpitel CJ (1996a) Drinking patterns and problems and drinking in the event: an analysis of injury by cause among casualty patients. Alcoholism, Clinical and Experimental Research, 20:1130-1137.

Cherpitel CJ (1996b) Alcohol in fatal and nonfatal injuries: a comparison of coroner and emergency room data from the same county. Alcoholism, Clinical and Experimental Research, 20:338-342.
Cherpitel CJ, Tam T, Midanik L, Caetano R, Greenfield T (1995) Alcohol and non-fatal injury in the U.S. general population: A risk function analysis. Accident Analysis Prevention, 27:651-661.
Church MW, Eldis F, Blakley BW, Bawle EV (1997) Hearing, language, speech, vestibular, and dentofacial disorders in fetal alcohol syndrome. Alcoholism, Clinical and Experimental Research, 21:227-237.

Clarke R, Shipley M, Lewington S et al. (1999) Underestimation of risk associations due to regression dilution in long term follow-up of prospective studies. American Journal of Epidemiology, 150:341-353.
Collins D, Lapsley H (1991) Estimating the economic costs of drug abuse in Australia. Vol. No. 15. Commonwealth of Australia, Canberra.
Connor PD, Streissguth AP (1996) Effects of prenatal exosure to alcohol across the life span. Alcohol Health and Research World, 20:170-174.
Cook PJ, Tauchen G (1982) The effect of liquor taxes on heavy drinking. Addiction, 13:379-390.
Corrao G, Arico S, Zambon A, Torchio P, Di Orio F (1997) Female sex and the risk of liver cirrhosis. Collaborative groups for the study of liver diseases in Italy. Scandinavian Journal of Gastroenterology, 32:1174-1180.

Corrao G, Bagnardi V, Zambon A, Arico S (1999) Exploring the dose-response relationship between alcohol consumption and the risk of several alcoholrelated conditions: a meta-analysis. Addiction, 94:1551-1573.

Corrao G, Bagnardi V, Zambon A, Torchio P (1998) Meta-analysis of alcohol intake in relation to risk of liver cirrhosis. Alcohol and Alcoholism, 33:381-392.

Corrao G, Rubbiati L, Bagnardi V, Zambon A, Poikolainen K (2000) Alcohol and coronary heart disease: a meta-analysis. Addiction, 95:1505-1523.
Corrao G, Zambon A, Torchio P et al. (1998) Attributable risk for symptomatic liver cirrhosis in Italy. Journal of Hepatology, 28:608-614.
Criqui MH, Cowan LD, Tyroler JA et al. (1987) Lipoprotein mediators for the effects of alcohol consumption and cigarette smoking on cardiovascular mortality. Results from the lipid research clinics follow-up study. American Journal of Epidemiology, 126:629-637.

Criqui MH, Ringek BL (1994) Does diet or alcohol explain the French paradox? The Lancet, 344:1719-1723.

Curtis AB, Sherman JA, Strogatz DS, Raghunathan TE, Siobàn H (1997) Alcohol consumption and changes in blood pressure among African Americans-The Pitt county study. American Journal of Epidemiology, 146:727-733.

Dackis CA, Gold MS, Pottash ALC, Sweeney DR (1986) Evaluating depression in alcoholics. Psychiatry Research, 17:105-109.

Davidson KM (1995) Diagnosis of depession in alcohol dependence: changes in prevalence with drinking status. British Journal of Psychiatry, 166:199-204.

De Stefani E, Boffetta P, Carzoglio J, Mendilaharsu S, Deneo-Pellegrini H (1998) Tobacco smoking and alcohol drinking as risk factors for stomach cancer: a case-control study in Uruguay. Cancer Causes Control, 9:321-329.
de Vries JHM, Lemmens P, Pietinen P, Kok FP (1999) Assessment of alcohol consumption. In: Health issues related to alcohol consumption. 2nd edn. Macdonald I, ed. Blackwell Science Ltd., London.
Devlin NJ, Scuffham PA, Bung LJ (1997) The social costs of alcohol abuse in New Zealand. Addiction, 92:1491-1505.
Doll R (1998) The benefit of alcohol in moderation. Drug and Alcohol Review, 17:353-363.

Doll R, Peto R, Hall E, Whealey K, Gray R (1994) Mortality in relation to consumption of alcohol: 13 years' observations on male British doctors. British Medical Journal, 309:911-918.

Donato F, Boffetta P, Fazioli R, Aulenti V, Gelatti U, Porru S (1997) Bladder cancer, tobacco smoking, coffee and alcohol drinking in Brescia, northern Italy. European Journal of Epidemiology, 13:795-800.
Drummond DC (1990) The relationship between alcohol dependence and alcohol-related problems in a clinical population. British Journal of Addiction, 85:357-366.
Duffy J-C (1986) Distribution of alcohol consumption: 30 years on. British Journal of Addiction, 81:735-741.
Dyer AR, Stamler J, Paul O et al. (1977) Alcohol consumption, cardiovascular risk factors, and mortality in two Chicago epidemiologic studies. Circulation, 56:1067-1074.

Eckardt MJ, File SE, Gessa GL et al. (1998) Effects of moderate alcohol consumption in the central nervous system. Alcoholism, Clinical and Experimental Research, 22:998-1040.
Edwards G, Anderson P, Babor TF et al. (1994) Alcohol policy and the public good. Oxford University Press, Oxford.
Ellison LF, Stokes J, Gibbons L et al. (1998) Monographs series on aging-related diseases: X. Prostate cancer. Chronic Diseases in Canada, 19:1-18.
Emanuele N, Emanuele MA (1997) The endocrine system: alcohol alters critical hormonal balance. Alcohol Health and Research World, 21:53-64.
English DR, Holman CDJ, Milne E et al. (1995) The quantification of drug caused morbidity and mortality in Australia 1995. Commonwealth Department of Human Services and Health, Canberra.

European Addiction Research (2001) Special issue on alcohol addiction treatment. European Addiction Research, 7: 95-157.

Evans C, Chalmers J, Capewell S et al. (2000) "I don't like Mondays" day of the week of coronary heart disease deaths in Scotland: study of routinely collected data. British Medical Journal, 320:218-219.

Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJL, the Comparative Risk Assessment Collaborating Group (2002) Selected major risk factors and global and regional burden of disease. The Lancet, 360:1347-1360.
Facchini F, Chen YD, Reaven GM (1994) Light-to-moderate alcohol intake is associated with enhanced insulin sensitivity. Diabetes Care, 17:115-119.
Faden VB, Graubard BI, Dufour M (1997) The relationship of drinking and birth outcome in a US national sample of expectant mothers. Paediatric and Perinatal Epidemiology, 11:167-180.

Fahrenkrug H, Rehm J (1994) Trinkkontexte und Freizeitaktivitäten in der Vorphase alkoholbezogener Strassenverkehrsunfälle junger Fahrerinnen und Fahrer in der Schweiz. Sucht, 41:169-180.

Fillmore KM (1988) Alcohol use across the life course: a critical review of 70 years of international longitudinal research. Addiction Research Foundation, Toronto.

Fillmore KM, Golding JM, Graves KL et al. (1998) Alcohol consumption and mortality. III. Studies of female populations. Addiction, 93:219-229.
Flanagan DEH, Moore VM, Godsland IF, Cockington RA, Robins JS, Phillips DI (2000) Alcohol consumption and insulin resitance in young adults. European Journal of Clinical Investigation, 30:297-301.
Foppa M, Fuchs FD, Preissler L, Andrighetto A, Rosito GA, Duncan BB (1999) Wine with the noon meal lowers post meal blood pressure: randomized trial in centrally obese hypertensive patients. Hypertension, 33:119.

Freedland E, McMicken D, D'Onofrio G (1993) Alcohol and trauma. Emergency Medicine Clinics of North America, 11:225-239.
Friedman HS (1998) Cardiovascular effects of alcohol. In: The Consequences of Alcoholism. Galanter M, ed. Plenum Press, New York.

Friedman L-A, Kimball A-W (1986) Coronary heart disease mortality and alcohol consumption in Framingham. American Journal of Epidemiology, 124:481-489.
Fuchs CS, Stampfer MJ, Colditz GA et al. (1995) Alcohol consumption and mortality among women. New England Journal of Medicine, 332:12451250.

Fuller M (1995) Alcohol use and injury severity in trauma patients. Journal of Affective Disorders, 14:47-53.
Galduróz JCF, Noto AR, Nappo SA, Carlini EA (2000) I levantomento domicilar nacional sobre o uso de drogas psicotrópicas. Parte A: Estudo envolvendo as 24 majores cidades do estado de São Paulo 1999. CEBRID, São Paulo.

Gapstur SM, Potter JD, Folsom AR (1994) Alcohol consumption and colon and rectal cancer in postmenopausal women. International Journal of Epidemiology, 23:50-57.

Gentry RT (2000) Effect of food on the pharmocokinetics of alcohol consumption. Alcoholism, Clinical and Experimental Research, 24:403-404.

Gibson S, Becker J (1973) Changes in alcoholics' self-reported depression. Quarterly Journal of Studies on Alcohol, 34:829-836.

Giesbrecht N, Greenfield TK, Lemmens P, Österberg E (2000) Estimating alcohol consumption: measurement and policy issues related to legal and illegal sources of alcohol. Contemporary Drug Problems, 27:221-233.

Gmel G, Gutjahr E (2001) Alcohol consumption and social harm: quantitative research methodology. In: Mapping the social consequences of alcohol consumption. Klingeman H, Gmel G, eds. Kluwer Academic Publishers, Dordrecht.
Gmel G, Rehm J (2003) Harmful alcohol use. Alcohol Research and Health, 27:52-62.

Gmel G, Rehm J, Frick U (2001) Methodological approaches to conducting pooled cross-sectional time series analysis: the example of the EU states. European Addiction Research, 7:128-137.

Gmel G, Rehm J, Room R, Greenfield TK (2000) Dimensions of alcohol-related social harm in survey research. Journal of Substance Abuse, 12:113-138.
Goldberg DM, Hahn SE, Parkes JG (1995) Beyond alcohol: beverage consumption and cardiovascular mortality. Clinica Chimica Acta, 237:155-187.
Graham K, Leonard KE, Room R et al. (1998) Current directions in research on understanding and preventing intoxicated aggression. Addiction, 93:659676.

Graham K, Schmidt G (1999) Alcohol use and psychosocial well-being among older adults. Journal of Studies on Alcohol, 60:345-351.
Graham K, Wells S (2001) Aggression among young adults in the social context of the bar. Addiction Research, 9:193-219.

Graham K, West P (2001) Alcohol and crime. In: International handbook of alcohol dependence and problems. Heather N, Peters TJ, Stockwell T, eds. John Wiley \& Sons, London.

Graham K, West P, Wells S (2000) Evaluating theories of alcohol-related aggression using observations of young adults in bars. Addiction, 95:847-863.
Grant BF, Harford TC (1995) Comorbidity between DSM-IV alcohol use disorders and major depression: results of a national survey. Drug and Alcohol Dependence, 39:197-206.
Grant BF, Pickering RP (1997) Familial aggregation of DSM-IV alcohol disorders: examination of the primary-secondary distinction in a general population sample. Journal of Nervous and Mental Disease, 185:335-343.
Greene WH (2000) Econometric analysis. Prentice Hall, Upper Saddle River, NJ.

Greenfield TK (2001) Individual risk of alcohol-related disease and problems. In: International handbook of alcohol dependence and problems. Heather N, Peters TJ, Stockwell T, eds. Wiley \& Sons, Chichester.

Grobbee DE, Rimm EB, Keil U, Renaud SC (1999) Alcohol and the cardiovascular system. In: Health issues related to alcohol consumption. Macdonald I, ed. Blackwell \& ILSI, Oxford.

Groves RM (1989) Survey errors and survey costs. John Wiley \& Sons, New York.

Gruenewald PJ, Hermos JA, Rehm J et al. (2000) Commentaries on Kerr et al.'s "Beverage-specific alcohol consumption and cirrhosis mortality in a group of English-speaking beer-drinking countries". Addiction, 95:347-358.
Gruenewald PJ, Mitchell PR, Treno AJ (1996) Drinking and driving: drinking patterns and drinking problems. Addiction, 91:1649.
Gruenewald PJ, Nephew T (1994) Drinking in California: theoretical and empirical analysis of alcohol consumption patterns. Addiction, 89:707-723.
Gruenewald PJ, Ponicki WR (1995) Relationship of alcohol sales to cirrhosis mortality. Journal of Studies on Alcohol, 56:635-641.

Gruenewald PJ, Treno AJ, Mitchell PR (1996) Drinking patterns and drinking behaviours: theoretical models of risky acts. Contemporary Drug Problems, 23:407-440.

Gurr M (1996) Alcohol: health issues related to alcohol consumption. 2nd edn. ILSI Europe, Brussels.
Gutjahr E, Gmel G (2001) Die sozialen Kosten des Alkoholkonsums in der Schweiz: Epidemiologische Grundlagen 1995-1995. Forschungsbericht Nr. 36. Schweizerische Fachstelle für Alkohol- und andere Drogenprobleme (SFA), Lausanne.
Gutjahr E, Gmel G, Rehm J (2001) Relation between average alcohol consumption and disease: an overview. European Addiction Research, 7:117127.

Habbick BF, Snyder RE (1997) Mortality in foetal alcohol syndrome. Canadian Journal of Public Health, 88:181-183.
Haberman PW, Weinbaum DF (1990) Liver cirrhosis with and without mention of alcohol as cause of death. British Journal of Addiction, 85:217-222.

Harnack LJ, Anderson KE, Zheng W, Folsom AR, Sellers TA, Kushi LH (1997) Smoking, alcohol, coffee, and tea intake and incidence of cancer of the exocrine pancreas: the lowa women's health study. Cancer Epidemiology, Biomarkers and Prevention, 6:1081-1086.
Hart CL, Smith GD, Hole DJ, Hawthorne VM (1999) Alcohol consumption and mortality from all causes, coronary heart disease, and stroke: results from a prospective cohort study of Scottish men with 21 years of follow up. British Medical Journal, 318:1725-1729.

Harwood H, Fountain D, Livermore G, The Lewin Group (1998) The economic costs of alcohol and drug abuse in the United States, 1992. Vol. No. 98-4327. U.S. Department of Health and Human Services, National Institue on Drug Abuse (NIDA), Rockville, MD.
Hasegawa K, Mukasa H, Nakazawa Y, Kodama H, Nakamura K (1991) Primary and secondary depression in alcoholism: Clinical features and family history. Drug and Alcohol Dependence, 27:275-281.

Hatanaka M (1996) Time-series-based econometrics. Unit root and co-intergration. Oxford University Press, New York.

Hayes RB, Brown LM, Schoenberg JB et al. (1996) Alcohol use and prostate cancer risk in US blacks and whites. American Journal of Epidemiology, 143:692-697.
Hemström Ö (2001) Per capita alcohol consumption and ischaemic heart disease mortality. Addiction, 96:S93-112.
Hendriks HF, Veenstra J, Velthuis-te Wierik EJ, Schaafsma G, Kluft C (1994) Effect of moderate dose of alcohol with evening meal on fibrinolytic factors. British Medical Journal, 308:1003-1006.
Hesselbrock V, Tennen H, Stabenau J, Hesselbrock M (1983) Affective disorder in alcoholism. International Journal of the Addictions, 18:435-444.
Hiatt RA, Armstrong MA, Klatsky AL, Sidney S (1994) Alcohol consumption, smoking, and other risk factors and prostate cancer in a large health plan cohort in California (United States). Cancer Causes Control, 5:66-72.

Hibell B, Andersson B, Ahlström S et al. (2000) The 1999 ESPAD Report. Alcohol and other drug use among students in 30 European countries. Pompidou Group; Council of Europe, Stockholm.
Hilarski C; Wodarki JS (2001) Comorbid substance abuse and mental illness: Diagnosis and treatment. Journal of Social Work Practice in the Addictions, 1:105-119.
Hill AB (1965) The environment and disease: association or causation? Proceedings of the Royal Society of Medicine, 58:295-300.
Hillbom M (1998) Alcohol consumption and stroke: benefits and risks. Alcoholism, Clinical and Experimental Research, 22:S352-358.
Hillbom M, Juvela S, Karttunen V (1998) Mechanisms of alcohol-related strokes. In: Alcohol and Cardiovascular Diseases. Chadwick DJ, Goode JA, eds. John Wiley \& Sons, Inc., New York.
Hingson R, Howland J (1987) Alcohol as a risk factor for injury or death resulting from accidental falls: a review of the literature. Journal of Studies on Alcohol, 48:212-219.
Hingson R, Howland J (1993) Alcohol and non-traffic unintended injuries. Addiction, 88:877-883.
Hoaken PNS, Assaad JM, Phil RO (1998) Cognitive functioning and the inhibition of alcohol-induced aggression. Journal of Studies on Alcohol, 59:599-607.
Holman CD, English DR, Milne E, Winter MJ (1996) Meta-analysis of alcohol and all-cause mortality: a validation of NHMRC recommendations. Medical Journal of Australia, 164:141-145.
Horn-Ross PL, Ljung BM, Morrow M (1997) Environmental factors and the risk of salivary gland cancer. Epidemiology, 8:414-419.
Hox JJ (1995) Applied multilevel analysis. 2nd edn. TT-Publikaties, Amsterdam.
Hurst PM, Harte WJ, Firth WJ (1994) The Grand Rapids dip revisited. Accident, Analysis and Prevention, 26:647-654.

Imhof A, Froehlich M, Brenner H, Boeing H, Pepys MB, Koenig W (2001) Effect of alcohol consumption on systemic markers of inflammation. The Lancet, 357:763-767.

Isbell H, Fraser HF, Wikler A, Belleville RE, Eisenman AJ (1955) Experimental study of the etiology of "rum fits" and delirium tremens. Quarterly Journal of Studies on Alcohol, 16:1-33.
Ito TA, Miller N, Pollock VE (1996) Alcohol and aggression: a meta-analysis on the moderating effects of inhibitory cues, triggering events, and self-focused attention. Psychological Bulletin, 120:60-82.
Jackson R (1994) Cardiovascular disease and alcohol consumption: evidence of benefit of epidemiological studies. Contemporary Drug Problems, 21:5-24.

Jacques PF, Bostom AG, Wilson PW, Rich S, Rosenberg IH, Selhub J (2001) Determinants of plasma total homocysteine concentration in the Framingham Offspring cohort. American Journal of Clinical Nutrition, 73:613-621.

Jallon P, Smadja D, Cabre P et al. (1998) Crises épileptiques, épilepsies et facteurs de risque: expérience d'une enquête en Martinique. Revue Neurologique, 154:408-411.
Jeavons CM, Taylor SP (1985) The control of alcohol-related aggression: redirecting the inebriate's attention to socially appropriate conduct. Aggressive Behavior, 11:93-101.
Ji BT, Chow WH, Yang G et al. (1996) Influence of cigarette smoking, alcohol, and green tea consumption on the risk of carcinoma of the cardia and distal stomach in Shanghai, China. Cancer, 77:2449-2457.

Jutkovitz JM, Hongsook E (1994) Drug prevalence in Latin America and Caribbean countries: a cross-national analysis. Drugs: Education, Prevention and Policy, 1:199-251.

Kao WHL, Brancati FL, Boland LL, Watson RL, Puddey IB (1998) Gender differences in the association of alcohol consumption and the risk of type 2 diabetes mellitus: the Atherosclerosis Risk in Communities (ARIC) study. American Journal of Epidemiology, 147:S60.
Kauhanen J, Kaplan GA, Goldberg DE et al. (1997b) Frequent hangovers and cardiovascular mortality in middle-aged men. Epidemiology, 8:310-314.
Kauhanen J, Kaplan GA, Goldberg DE, Salonen JT (1997a) Beer drinking and mortality: results from the Kuopio ischemic heart disease risk factors study, a prospective population based study. British Medical Journal, 315:846-851.

Keil U, Chambless LE, Döring A, Filipiak B, Stieber J (1997) The relation of alcohol intake to coronary heart disease and all-cause mortality in a beerdrinking population. Epidemiology, 8:150-156.
Kerr WC, Fillmore KM, Marvy P (2000) Beverage-specific alcohol consumption and cirrhosis mortality in a group of English-speaking beer-drinking countries. Addiction, 95:339-346.
Kessler RC (1998) Sex differences in DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. Journal of the American Medical Women's Association, 53:148-158.

Kessler RC, Crum RM, Warner LA, Nelson CB, Schulenberg J, Anthony JC (1997) Lifetime co-occurrence of DSM-III-R alcohol abuse and dependence with other psychiatric disorders in the National Comorbidity Survey. Archives of General Psychiatry, 54:313-321.
Kessler RC, Nelson CB, McGonagle KA, Edlund MJ, Frank RG, Leaf PJ (1996) Epidemiology of co-occuring addictive and mental disorders: Implications for prevention and service utilization. American Journal of Orthopsychiatry, 66:17-31.

Kiechl S, Willeit J, Poewe W et al. (1996) Insulin sensitivity and regular alcohol consumption: large, prospective, cross-sectional population study (Bruneck study). British Medical Journal, 313:1040-1044.

Kitamura A, Iso H, Sankai T et al. (1998) Alcohol intake and premature coronary heart disease in urban Japanese men. American Journal of Epidemiology, 147:59-65.
Klatsky AL (1995) Cardiovascular effects of alcohol. Scientific American. Science and Medicine, 2:28-37.
Klatsky AL (1996) Alcohol, coronary heart disease, and hypertension. Annual Review of Medicine, 47:149-160.
Klingeman H, Gmel G (2001) Mapping the social consequences of alcohol consumption. Kluwer Academic Publishers, Dordrecht.
Knuiman MW, Vu HT (1996) Risk factors for stroke mortality in men and women: the Busselton Study. Journal of Cardiovascular Risk, 3:447-452.
Kohlmeier L, Mendez M (1997) Controversies surrounding diet and breast cancer. Proceedings of the Nutrition Society, 56:369-382.

Koskinen P, Kupari M, Leinonen H, Luomanmaki K (1987) Alcohol and new onset atrial fibrillation: a case-control study of a current series. British Heart Journal, 57:468-473.
Kozarevic D, Demirovic JTG et al. (1982) Drinking habits and coronary heart disease. The Yugoslavia cardiovascular disease study. American Journal of Epidemiology, 116:748-758.
Kreft IGG, de Leeuw J (1998) Introducing multilevel modelling. ISM Introducing Statistical Methods. 2nd edn. Sage Publications, London.
Krüger HP, Utzelmann H-D, Berghaus G, Kroj G (1993) Effects of low alcohol dosage: a review of the literature. In: Alcohol, drugs and trafic safety-T'92. Krüger HP, Utzelmann H-D, Berghaus G, Kroj G, eds. Verlag TÜV Rheinland, Cologne.

Kühlhorn E, Hibell B, Larsson S, Ramstedt M, Zetterberg HL (1999) Alkoholkonsumtionen i Sverige under 1990-talet. [Alchol consumption in Sweden in the 1990s.] OAS, Socialdepartmentet [Ministry of Health and Social Affairs], Stockholm.
Larkby C, Day N (1997) Effects of prenatal alcohol exposure. Alcohol Health and Research World, 21:192-198.
Larroque B, Kaminski M (1996) Alcohol consumption during pregnancy. Annals of Epidemiology, 6:169.

Lazarus R, Sparrow D, Weiss ST (1997) Alcohol intake and insulin levels. The Normative Aging Study. American Journal of Epidemiology, 145:909-916.

Lehtonen R, Pahkinen EJ (1994) Practical methods for design and analysis of complex surveys. John Wiley \& Sons Ltd., Chichester.
Leifman H (2001) Estimations of unrecorded alcohol consumption levels and trends in 14 European countries. Nordic Studies on Alcohol and Drugs, 18: S54-70.

Leitzmann MF, Willett WC, Giovannucci EL et al. (1998) Alcohol intake and the risk of symptomatic gallstones in men. American Journal of Epidemiology, 147:S54.
Lelbach WK (1975) Cirrhosis in the alcoholic and its relation to volume of alcohol abuse. Annals of the New York Academy of Sciences, 252:85-105.
Lelbach WK (1976) Epidemiology of liver disease. In: Progress in liver diseases. Popper H, Schaffner F, eds. Grune \& Stratton, New York.
Leon DA, Chenet L, Shkolnikov V et al. (1997) Huge variation in Russian mortality rates 1984-94: artefact, alcohol, or what? The Lancet, 350:383-388.

Leonard KE (1984) Alcohol consumption and escalatory aggression in intoxicated and sober dyads. Journal of Studies on Alcohol, 45:75-80.

Leone M, Bottacchi E, Beghi E et al. (1997) Alcohol use is a risk factor for a first generalized tonic clonic seizure. Neurology, 48:614-620.

Li G, Keyl PM, Smith GS, Baker SP (1997) Alcohol and injury severity: reappraisal of the continuing controversy. Journal of Trauma, 42:562-569.
Li G, Smith GS, Baker SP (1994) Drinking behaviour in relation to cause of death among U.S. adults. American Journal of Public Health, 84:1402-1406.
Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika, 73:13-22.
Lipsey MW, Wilson DB, Cohen MA, Derzon JH (1997) Is there a causal relationship between alcohol use and violence? In: Recent developments in alcoholism, Vol. 13, Alcohol and violence. Galanter M, ed. Plenum, New York.
Lloyd HM, Rogers PJ (1997) Mood and cognitive performance improved by a small amount of alcohol given with a lunchtime meal. Behavioural Pharmacology, 8:88-195.

Longnecker MP (1999) Invited commentary: the Framingham results on alcohol and breast cancer. American Journal of Epidemiology, 149:102.

Longnecker MP, Enger SM (1996) Epidemiological data on alcohol beverage consumption and risk of cancer. Clinica Chimica Acta, 246:121-141.
Longnecker MP, Wolz M, Parker DA (1981) Ethnicity, distilled spirits consumption and mortality in Pennsylvania. Journal of Studies on Alcohol, 42:791-796.
Ludbrook A, Godfrey C, Wyness L et al. (2002) Effective and cost-effective measures to reduce alcohol misuse in Scotland: a literature review. Health Economics Research Unit, Aberdeen.
Lundberg JC, Passik SD (1997) Alcohol and cancer: a review for psycho-oncologists. Psychooncology, 6:253-266.

Lynskey MT (1998) The comorbidity of alcohol dependence and affective disorders: treatment implications. Drug and Alcohol Dependence, 52:201209.

MacAndrew C, Edgerton RB (1969) Drunken comportment. A social explanation. Aldine, Chicago, IL.
Malyutina S, Bobak M, Kurilovitch S, Ryizova E, Nikitin Y, Marmot M (2001) Alcohol consumption and binge drinking in Novosibirsk, Russia, 1985-1995. Addiction, 96:987-995.
Markou A, Kosten TR, Koob GF (1998) Neurobiological similarities in depression and drug dependence: a self-medication hypothesis. Neuropsychopharmacology, 18:135-174.
Martin SE (1992) Epidemiology of alcohol-related interpersonal violence. Alcohol Health and Research World, 16:230-237.

Martin SE, Bachman R (1997) The relationship or alcohol to injury in assault cases. In: Recent developments in alcoholism, Vol. 13, Alcohol and violence. Galanter M, ed. Plenum, New York.

Martín R, Díaz C, Nuévalos C, Matías-Guiu J (1995) Etiological and prognostic factors in the late onset epilepsy. Revista de Neurologica, 23:285-289.
Maskarinec G, Meng L, Kolonel LN (1998) Alcohol intake, body weight, and mortality in a multiethnic prospective cohort. Epidemiology, 9:654-661.
Mattson SN, Riley EP, Gramling L, Delis DC, Jones KL (1997) Heavy prenatal alcohol exposure with or without physical features of fetal alcohol syndrome leads to IQ deficits. Journal of Pediatrics, 131:718-721.
Mäkelä K, Room R, Single E, Sulkunen P, Walsh B, with 13 others (1981) Alcohol, society, and the state: 1. A comparative history of alcohol control. Addiction Research Foundation, Toronto.
McClelland DC, Davis WN, Kalin R, Wanner E (1972) The drinking man. Alcohol and human motivation. Collier-Macmillan, Toronto.

McElduff P, Dobson AJ (1997) How much alcohol and how often? Population based case-control study of alcohol consumption and risk of a major coronary event. British Medical Journal, 314:1159
McKee M, Britton A (1998) The positive relationship between alcohol and heart disease in eastern Europe: potential physiological mechanisms. Journal of the Royal Society of Medicine, 91:402-407.
McKee M, Shkolnikov V, Leon DA (2001) Alcohol is implicated in the fluctuations in cardiovascular disease in Russia since the 1980s. Annals of Epidemiology, 11:1-6.
McKenzie CR, Eisenberg PR (1996) Alcohol, coagulation, and arterial thrombosis. In: Alcohol and the cardiovasular system. Zakhari S, Wassef M, eds. National Institues of Health, National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD.

McLeod R, Stockwell T, Stevens M, Philips M (1999) Relationship between alcohol consumption patterns and injury. Addiction, 94:1719-1734.

Mehrabian A (2001) General relations among drug use, alcohol use, and major indexes of psychopathology. Journal of Psychology, 135:71-86.

Merikangas KR, Mehta RL, Molnar BE et al. (1998) Comorbidity of substance use disorders with mood and anxiety disorders: results of the International Consortium in Psychiatric Epidemiology. Addictive Behaviors: An International Journal, 23:893-907.
Miczek KA, Weerts EM, DeBold JF (1993) Alcohol, benzodiazepine-GABA receptor complex and aggression: ethiological analysis of individual differences in rodents and primates. Journal of Studies on Alcohol, 11:S170-179.
Midanik LT (1988) Validity of self-reported alcohol use: a literature review and assessment. British Journal of Addiction, 83:1019-1029.

Midanik LT, Harford TC (1994) Overview of alcohol consumption measurement papers. Addiction, 89:393-394.

Morgenstern H (1998) Ecologic studies. In: Modern epidemiology. Rothman KJ, Greenland S, eds. Lippincott-Raven Publishers, Philadelphia, PA.

Morrow DA, Ridker PM (2000) C-reactive protein, inflammation, and coronary risk. Medical Clinics of North America, 84:149-161.

Moskowitz H, Robinson C (1988) Effects of low doses of alcohol on drivingrelated skills: a review of the evidence. National Highway Traffic Safety Administration, Washington, DC.
Mukamal KJ, Rimm EB (2001) Alcohol's effects on the risk for coronary heart disease. Alcohol Research and Health, 25:255-261.
Murdoch DD, Pihl RO, Ross D (1990) Alcohol and crimes of violence: present issues. International Journal of the Addictions, 25:1065-1081.
Murray CJL, Lopez A, eds (1996a) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, Vol. I. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Murray CJL, Lopez A (1996b) Quantifying the burden of disease and injury attributable to ten major risk factors. In: The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, Vol. I. Murray CJL, Lopez A, eds. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Murray CJL, Lopez A (1999) On the comparable quantification of health risk: lessons from the Global Burden of Disease Study. Epidemiology, 10:594-605.

Murray RP, Connett JE, Tyas SL et al. (2002) Alcohol volume, drinking pattern, and cardiovascular disease morbidity and mortality: is there a U-shaped function? American Journal of Epidemiology, 155:242-248.
Murray RP, Istvan JA, Daniels K, Beaudoin CM (1998) Alcohol and morbidity in the Lung Health Study. Journal of Studies on Alcohol, 59:250-257.
Murray RP, Rehm J, Shaten J, Connett JE (1999) Does social integration confound the relation between alcohol consumption and mortality in the

Multiple Risk Factor Intervention Trial (MRFIT)? Journal of Studies on Alcohol, 60:740-745.

Muscat JE, Wynder EL (1998) A case-control study of risk factors for major salivary gland cancer. Otolaryngology and Head and Neck Surgery, 118: 195-198.

Mustonen H, Beukes L, Du Preez V (2001) Alcohol drinking in Namibia. In: Surveys of drinking patterns and problems in seven developing countries. World Health Organization, Geneva.
Nasca PC, Liu S, Baptiste MS, Kwon CS, Jacobson H, Metzger BS (1994) Alcohol consumption and breast cancer: estrogen receptor status and histology. American Journal of Epidemiology, 140:980-987.

Nemtsov AV (1998) Alcohol-related harm and alcohol consumption in Moscow before, during and after a major anti-alcohol campaign. Addiction, 93:15011510.

Nemtsov AV (2000) Estimates of total alcohol consumption in Russia, 1980-1994. Drug and Alcohol Dependence, 58:133-142.

Nemtsov AV (2002) Alcohol-related human losses in Russia in the 1980s and 1990s. Addiction, 97:1413-1425.

Newcomb PA, Trentham-Dietz A, Storer BE (1997) Alcohol consumption in relation to endometrial cancer risk. Cancer Epidemiology, Biomarkers and Prevention, 6:775-778.

Norström T (2001) Per capita alcohol consumption and all-cause mortality in 14 European countries. Addiction, 96:S113-128.
Notzon FC, Komarov YM, Ermakov SP, Sempos CT, Marks JS, Sempos EV (1998) Causes of declining life expectancy in Russia. Journal of the American Medical Association, 279:793-800.

Parazzini F, La Vecchia C, D’Avanzo B, Moroni S, Chatenoud L, Ricci E (1995) Alcohol and endometrial cancer risk: findings from an Italian case-control study. Nutrition and Cancer, 23:55-62.

Passaro K-AT, Little RE (1997) Childbearing and alcohol use. In: Gender and alcohol: individual and social perspectives. Wilsnack RW, Wilsnack SC, eds. Rutgers Center of Alcohol Studies, New Brunswick, NJ.
Passaro KT, Little RE, Savitz DA, Noss J (1996) The effect of maternal drinking before conception and in early pregnancy on infant birthweight. The ALSPAC Study Team. Avon Longitudinal Study of Pregnancy and Childhood. Epidemiology, 7:377-383.

Penick EC, Powell BJ, Liskow BI, Jackson JO, Nickel EJ (1988) The stability of coexisting psychiatric syndromes in alcoholic men after one year. Journal of Studies on Alcohol, 49:395-405.

Pernanen K (2001) What is meant by "alcohol-related" consequences? In: Mapping the social consequences of alcohol consumption. Klingeman H, Gmel G, eds. Kluwer Academic Publishers, Dordrecht.
Perry IJ, Wannamethee SG, Walker MK, Thomson AG, Whincup PH, Shaper AG (1995) Prospective study of risk factors for development of non-insulin
dependent diabetes in middle aged British men. British Medical Journal, 310:560-564.

Peterson JB, Rothfleisch J, Zelazo P, Pihl RO (1990) Acute alcohol intoxication and neuropsychological functioning. Journal of Studies on Alcohol, 51:114-122.

Peto R, Lopez AD, Boreham J, Thun M, Heath CW Jr. (1992) Mortality from tobacco in developed countries: indirect estimation from national vital statistics. The Lancet, 339:1268-1278.

Pettinati H, Sugerman A, Maurer HS (1982) Four-year MMPI changes in abstinent and drinking alcoholics. Alcoholism, Clinical and Experimental Research, 6:487-494.

Piette JD, Barnett PG, Moos RH (1998) First-time admissions with alcoholrelated medical problems: a 10-year follow-up of a national sample of alcoholic patients. Journal of Studies on Alcohol, 89-96.

Pihl RO, Peterson J (1993) Alcohol/drugs use and aggressive behavior. In: Mental disorder and crime. Hodgkins S, ed. Sage, London.

Pihl RO, Peterson J, Lau M (1993) A biosocial model of the alcohol-aggression relationship. Journal of Studies on Alcohol, 11:S128-139.
Poikolainen K (1983) Inebriation and mortality. International Journal of Epidemiology, 12:151-155.
Polygenis D, Wharton S, Malmberg C et al. (1998) Moderate alcohol consumption during pregnancy and the incidence of fetal malformations: a metaanalysis. Neurotoxicology and Teratology, 20:61-67.
Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM (2001) C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. Journal of the American Medical Association, 286:327-234.

Productschap voor Gedistilleerde Dranken (1999) World drink trends, 1999. International beverage consumption and production trends. NTC, Schiedam, the Netherlands.

Productschap voor Gedistilleerde Dranken (2000) World drink trends, 2000. International beverage consumption and production trends. NTC, Schiedam, the Netherlands.

Puddey IB, Rakic V, Dimmitt SB, Beilin LJ (1999) Influence of drinking on cardiovascular disease and cardiovascular risk factors: a review. Addiction, 94:649-663.

Puffer R, Griffith GW (1967) Patterns of urban mortality. No. 151. Pan American Health Organization, Washington, DC.

Putnam SC, Cerhan JR, Parker AS, Wallace RB, Cantor KP, Lynch CF (1998) Alcohol consumption and prostate cancer in a cohort of Iowa males. American Journal of Epidemiology, 147:42.

Ramchandani VA, Kwo PY, Li TK (2001) Effect of food and food composition on alcohol elimination rates in healthy men and women. Journal of Clinical Pharmacology, 41:1345-1350.

Ramstedt M (1996a) Per capita alcohol consumption and liver cirrhosis mortality in 14 European countries. Addiction, 96:19-34.

Ramstedt M (1996b) Alcohol and suicide in 14 European countries. Addiction, 96:59-75.

Rankin J (1994) Biological mechanisms at moderate levels of alcohol consumption that may affect coronary heart disease. Contemporary Drug Problems, 21:45-57.

Reeder VC, Aikens ML, Li X-N, Booyse FM (1996) Alcohol and the fibrinolytic system. In: Alcohol and the cardiovascular system. Zakhari S, Wassef M, eds. National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD.
Rehm J (1998a) Measuring quantity, frequency and volume of drinking. Alcoholism, Clinical and Experimental Research, 22:S4-14.

Rehm J (1998b) Measuring alcohol consumption: how about adopting usual epidemiological standards. Addiction, 93:970-972.

Rehm J (2000) Re: "Alcohol intake assessment: the sober facts". American Journal of Epidemiology, 151:436-438.

Rehm J (2001) Concepts, dimensions and measures of alcohol-related social consequences. A basic framework for alcohol-related benefits and harm. In: Mapping social consequences of alcohol consumption. Klingeman H, Gmel G, eds. Kluwer, Dordrecht.
Rehm J, Ashley MJ, Room R et al. (1996) Drinking patterns and their consequences: report from an international meeting. Addiction, 91:1615-1621.
Rehm J, Bondy S, Sempos CT, Vuong CV (1997) Alcohol consumption and coronary heart disease morbidity and mortality. American Journal of Epidemiology, 146:495-501.
Rehm J, Eschmann S (2002) Global monitoring of average volume of alcohol consumption. Zeitschrift für Sozial- und Präventivmedizin, 47:48-58.
Rehm J, Gmel G (2000a) Aggregating dimensions of alcohol consumption to predict medical and social consequences. Journal of Substance Abuse, 12: 155-168.

Rehm J, Gmel G (2000b) Gaps and needs in international alcohol epidemiology. Journal of Substance Use, 5:6-13.
Rehm J, Gmel G (2001a) Aggregate time-series regression in the field of alcohol. Addiction, 96:945-954.
Rehm J, Gmel G (2001b) Applying principles of comparative risk analysis to substance-abuse related burden. European Addiction Research, 7:95-97.
Rehm J, Gmel G, Room R et al. (2001a) Alcohol as a risk factor for burden of disease. World Health Organization, Geneva.
Rehm J, Gmel G, Sempos CT, Trevisan M (2003b) Alcohol-related mortality and morbidity. Alcohol Research and Health, 27:39-51.
Rehm J, Greenfield TK, Rogers JD (2001d) Average volume of alcohol consumption, patterns of drinking and all-cause mortality. Results from the U.S. National Alcohol Survey. American Journal of Epidemiology, 153:64-71.

Rehm J, Gutjahr E, Gmel G (2001c) Alcohol and all-cause mortality: a pooled analysis. Contemporary Drug Problems, 28:337-361.

Rehm J, Monteiro M, Room R et al. (2001b) Steps towards constructing a global comparative risk analysis for alcohol consumption: determining indicators and empirical weights for patterns of drinking, deciding about theoretical minimum, and dealing with different consequences. European Addiction Research, 7:138-147.
Rehm J, Sempos CT (1995) Alcohol consumption and mortality-questions about causality, confounding and methodology. Addiction, 90:493-498.
Rehm J, Sempos CT, Trevisan M (2003a) Alcohol and cardiovascular diseasemore than one paradox to consider. Average volume of alcohol consumption, patterns of drinking and risk of coronary heart disease: a review. Journal of Cardiovascular Risk, 10:15-20.

Reinke LA, McCay PB (1996) Interaction between alcohol and antioxidants. In: Alcohol and the Cardiovascular System. Zakhari S, Wassef M, eds. National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD.
Renaud SC, Criqui MH, Farchi G, Veenstra J (1993) Alcohol drinking and coronary heart disease. In: Health issues related to alcohol consumption. Verschuren PM, ed. International Life Sciences Institute, Washington, DC.
Renaud SC, Ruf JC (1996) Effects of alcohol on platelet function. Clinica Chimica Acta, 246:77-89.

Rhodés J, Salaspuro M, Sorenson TIA (1993) Alcohol and liver diseases. In: Health issues related to alcohol consumption. Verschuren PM, ed. International Life Sciences Institute, Washington, DC.

Rice DP, Kelman S, Miller LS (1991) The economic cost of alcohol abuse. Alcohol Health and Research World, 15:307-316.
Ridker PM (2001) Role of inflammatory biomarkers in prediction of coronary heart disease. The Lancet, 358:946-948.
Ridolfo B, Stevenson C (2001) The quantification of drug-caused mortality and morbidity in Australia 1998. Australian Institute of Health and Welfare, Canberra.

Rimm EB, Chan J, Stampfer MJ, Colditz GA, Willett WC (1995) Prospective study for cigarette smoking, alcohol use, and the risk of diabetes in men. British Medical Journal, 310:555-559.
Rimm EB, Klatsky A, Grobbe D, Stampfer M (1996) Review of moderate alcohol consumption and reduced risk of coronary heart disease: is the effect due to beer, wine or spirits. British Medical Journal, 312:731-735.
Rimm EB, Willett WC, Hu FB et al. (1998) Folate and vitamin B6 from diet and supplements in relation to risk of coronary heart disease among women. Journal of the American Medical Association, 279:359-364.
Rimm EB, Williams P, Fosher K, Criqui MH, Stampfer MJ (1999) Moderate alcohol intake and lower risk of coronary heart disease: meta-analysis of effects on lipids and haemostatic factors. British Medical Journal, 319: 1523-1528.

Robles-Diaz G, Gorelick FS (1997) Alcohol and acute pancreatitis. Yale Journal of Biology and Medicine, 70:77-87.

Rodgers B, Korten AE, Jorm AF, Jacomb PA, Christensen H, Henderson AS (2000) Non-linear relationships in associations of depression and anxiety with alcohol use. Psychological Medicine, 30:421-432.
Roebuck TM, Mattson SN, Riley EP (1998) A review of the neuroanatomical findings in children with fetal alcohol syndrome or prenatal exposure to alcohol. Alcoholism, Clinical and Experimental Research, 22:339-344.
Room R (1972) Alcohol-related urban mortality: a cross-national study. Drinking and Drug Practices Surveyor, 5:11-13.
Room R (1992) The impossible dream? Routes to reducing alcohol problems in a temperance culture. Journal of Substance Abuse, 4:91-106.
Room R, Jernigan D, Carlini-Marlatt B et al. (2002) Alcohol in developing societies: a public health approach. Finnish Foundation for Alcohol Studies in collaboration with the World Health Organization, Helsinki.

Room R, Mäkelä K (2000) Typologies of the cultural position of drinking. Journal of Studies on Alcohol, 61:475-483.
Room R, Rossow I (2001) The share of violence attributable to drinking. Journal of Substance Use, 6:218-228.
Rosenberg L, Metzger LS, Palmer JR (1993) Alcohol consumption and risk of breast cancer: a review of the epidemiologic evidence. Epidemiologic Reviews, 15:133-144.

Rosengren A, Wilhelmsen L, Pennert K, Berglund G, Elmfeldt D (1987) Alcoholic intemperance, coronary heart disease and mortality in middle-aged Swedish men. Acta Medica Scandinavica, 222:201-213.
Rosenqvist M (1998) Alcohol and cardiac arrhythmias. Alcoholism, Clinical and Experimental Research, 22:S318-322.

Ross R, Bernstein L, Trent L, Henderson B, Paganini-Hill A (1990) A prospective study of risk factors for traumatic deaths in a retirement community. Alcoholism, Clinical and Experimental Research, 20:1130-1137.

Rossow I, Amundsen A (1997) Alcohol abuse and mortality: a 40-year prospective study of Nowegian conscripts. Social Science and Medicine, 44:261-267.

Rossow I, Pernanen K, Rehm J (2001) Alcohol, suicide and violence. In: Mapping the Social Consequences of Alcohol Consumption. Klingeman H, Gmel G, eds. Kluwer Academic Publishers, Dordrecht.
Rothman KJ, Greenland S (1998b) Modern epidemiology. 2nd edn. LippincottRaven Publishers, Philadelphia, PA.
Rothman KJ, Greenland S (1998a) Causation and causal inference. In: Modern epidemiology. 2nd edn. Rothman KJ, Greenland S, eds. Lippincott-Raven Publishers, Philadelphia, PA.

Royo-Bordonada MA, Martin-Moreno JM, Guallar E et al. (1997) Alcohol intake and risk of breast cancer: the Euramic Study. Neoplasma, 44:150-156.

Rubin R (1999) Effect of ethanol on platelet function. Alcoholism, Clinical and Experimental Research, 23:1114-1118.

Sacco RL, Elkind M, Boden-Albala B et al. (1999) The protective effect of moderate alcohol consumption in ischemic stroke. Journal of the American Medical Association, 281:53-60.

Sarles H, Bernard JP, Johnson CD (1996) Alcohol and the pancreas. In: Alcohol misuse: a european perspective. Peters TJ, ed. Harwood Academic Publishers, Amsterdam.
Sayette MA, Wilson T, Elias MJ (1993) Alcohol and aggression: a social information processing analysis. Journal of Studies on Alcohol, 54:399-407.
Sayrs LW (1989) Pooled time series analysis. Sage Publications, Newbury Park, CA.

Schatzkin A, Longnecker MP (1994) Alcohol and breast cancer: where are we now and where do we go from here? Cancer, 74:1101-1110.

Schmidt DN (1991) Apparent risk factors for chronic and acute pancreatitis in Stockholm county: spirits but not wine and beer. International Journal of Pancreatology, 8:45-50.

Schuckit MA (1996) Alcohol, anxiety, and depressive disorders. Alcohol Health and Research World, 20:81-85.

Schuckit MA, Tipp JE, Bergman M, Reich W, Hesselbrock VM, Smith TL (1997b) Comparison of induced and independent major depressive disorders in 2,945 alcoholics. American Journal of Psychiatry, 154:948-957.
Schuckit MA, Tipp JE, Bucholz KK et al. (1997a) The life-time rates of three major mood disorders and four major anxiety disorders in alcoholics and controls. Addiction, 92:1289-1304.
Segi M (1960) Cancer mortality for selected sites in 24 countries (1950-57). Department of Health, Tohoku University of Medicine, Sendai, Japan.

Seitz HK, Poschl G, Simanowski UA (1998a) Alcohol and cancer. Recent Developments in Alcoholism, 14:67-95.

Seitz HK, Simanowski UA, Homann N, Waldherr R (1998b) Cell proliferation and its evaluation in the colorectal mucosa: effect of ethanol. Zeitschrift für Gastroenterologie, 36:645-655.

Shaper AG (1990a) Alcohol and mortality: a review of prospective studies. British Journal of Addiction, 85:837-847.

Shaper AG (1990b) A response to commentaries: the effects of self-selection. British Journal of Addiction, 85:859-861.

Shaper AG, Wannamethee SG, Walker M (1988) Alcohol and mortality in British men: explaining the U-shaped curve. The Lancet, 2:1267-1273.
Sharps PW, Campbell D, Gary F, Webster D (2001) The role of alcohol use in intimate partner femicide. American Journal on Addictions, 10:122-135.
Shkolnikov V, McKee M, Leon DA (2001) Changes in live expectancy in Russia in the mid-1990s. The Lancet, 357:917-921.
Shkolnikov VM, McKee M, Chervyakov VV, Kyrianov NA (2002) Is the link between alcohol and cardiovascular death among young Russian men due to misclassification of acute alcohol intoxication? Evidence from the city of Izhevsk. Journal of Epidemiology and Community Health, 56:171-174.

Shkolnikov VM, Nemtsov A (1997) The anti-alcohol campaign and variations in Russian mortality. In: Premature Death in the New Independent States. Bobadilla JL, Costello CA, Mitchell F, eds. National Academy Press, Washington DC.
Shu I, Shaten BJ, Cutler JA, Kuller LH (1992) Alcohol use and mortality from coronary heart disease: the role of high-density lipoprotein cholesterol. The Multiple Risk Factor Intervention Trial Research Group. Annals of Internal Medicine, 116:881-887.

Shu XO, Hatch MC, Mills J, Clemens J, Susser M (1995) Maternal smoking, alcohol drinking, caffeine consumption, and fetal growth from a prospective study. Epidemiology, 6:115-120.

Shults RA, Elder RW, Sleet DA, et al. (2001) Reviews of evidence regarding interventions to reduce alcohol-impaired driving. American Journal of Preventive Medicine, 21:S66-88.

Simpura J (1998) Mediterranean mysteries: mechanisms of declining alcohol consumption. Addiction, 93:1301-1304.
Simpura J (2001) Trends in alcohol consumption and drinking patterns: sociological and economic explanations and alcohol policies. Nordic Studies on Alcohol and Drugs, 18:S3-13.
Single E, Ashley MJ, Bondy S, Rankin J, Rehm J (1999) Evidence regarding the level of alcohol consumption considered to be low-risk for men and women. Final report submitted to the Australian Commonwealth Department of Health and Aged Care.

Single E, Robson L, Rehm J, Xie X (1999) Morbidity and mortality attributable to alcohol, tobacco, and illicit drug use in Canada. American Journal of Public Health, 89:385-390.

Single E, Robson L, Xie X, Rehm J (1996) The cost of substance abuse in Canada. Canadian Centre on Substance Abuse, Ottawa, Ontario.
Singletary KW, Gapstur SM (2001) Alcohol and breast cancer: Review of epidemiologic and experimental evidence and potential mechanisms. JAMA: Journal of the American Medical Association, 86:2143-2151.
Skinaze F, Lévy P, Bernades P (1995) Les pancréatities aguiës alcooliques rélèventelles toujours une pancréatite chronique? Gastroentérologie Clinique et Biologique, 19:266-269.

Skinner CJ, Holt D, Smith TMF (1989) Analysis of complex surveys. John Wiley $\&$ Sons Ltd., Chichester.

Skog O-J (1980) Liver cirrhosis epidemiology: some methodological problems. British Journal of Addiction, 75:227-243.

Skog O-J (1983) Methodological problems in the analysis of temporal covariation between alcohol consumption and ischaemic heart disease. British Journal of Addiction, 78:157-172.
Skog O-J (1996) Public health consequences of the J-curve hypothesis of alcohol problems. Addiction, 91:325-337.

Smith-Warner SA, Spiegelman D, Yaun S-S et al. (1998) Alcohol and breast cancer in women: a pooled analysis of cohort studies. Journal of the American Medical Association, 279:535-540.

Soler M, Chatenoud L, La Vecchia C et al. (1998) Diet, alcohol, coffee and pancreatic cancer: final results from an Italian study. European Review of Cancer Prevention, 7:455-460.

StataCorp (1999) Stata reference manual, release 6-volume 4 Su-Z. Stata Press, College Station, Texas.
Steele CM, Josephs RA (1990) Alcohol myopia: its prized and dangerous effects. American Psychologist, 45:921-933.
Stinson FS, Dufour MC, Steffens RA et al. (1993) Alcohol-related mortality in the United States, 1979-1989. Alcohol Health and Research World, 17:251-260.

Stoduto G, Vingilis E, Kapur B et al. (1993) Alcohol and drug use among motor vehicle collision victims admitted to a regional trauma unit: demographic, in-jury and crash characteristics. Accident, Analysis and Prevention, 25:411420.

Svärdsudd K (1998) Moderate alcohol consumption and cardiovascular disease: is there evidence for a preventive effect? Alcoholism, Clinical and Experimental Research, 22:S307-314.
Swanson CA, Coates FJ, Malone KE et al. (1997) Alcohol consumption and breast cancer risk among women under age 45 years. Epidemiology, 8:231237.

Swendsen JD, Merikangas KR, Canino GJ et al. (1998) Comorbidity of alcoholism with anxiety and depressive disorders in four geographic communities. Comprehensive Psychiatry, 39:176-184.

Tamerin JS, Weiner S, Mendelson JH (1970) Alcoholics’ expectancies and recall of experiences during intoxication. American Journal of Psychiatry, 126:1697-1704.

Tavani A, Negri E, Franceschi S et al. (1994) Alcohol consumption and risk of prostate cancer. Nutrition and Cancer, 21:25-31.
Thakker KD (1998) An overview of health risks and benefits of alcohol consumption. Alcoholism, Clinical and Experimental Research, 22:S285-298.
Thun JM, Peto R, Lopez AD et al. (1997) Alcohol consumption and mortality among middle-aged and elderly U.S. adults. New England Journal of Medicine, 337:1705-1714.

Tomsen SA (1997) A top night out-social protest, masculinity and the culture of drinking violence. British Journal of Criminology, 37:990-1002.

Treno AJ, Gruenewald PJ, Ponicki WR (1997) The contribution of drinking patterns to the relative risk of injury in six communities: A self-report based probability approach. Journal of Studies on Alcohol, 58:372-381.
Treno AJ, Holder HD (1997) Measurement of alcohol-involved injury in community prevention: the search for a surrogate III. Alcoholism, Clinical and Experimental Research, 21:1695-1703.

Trevisan M, Krogh V, Farinaro E et al. (1987) Alcohol consumption, drinking pattern and blood pressure: analysis of data from the Italian National Research Council Study. International Journal of Epidemiology, 16:520-527.

Trevisan M, Ram M, Hovey K et al. (2001b) Alcohol drinking patterns and myocardial infaction. American Journal of Epidemiology, 153:S97.
Trevisan MT, Schisterman E, Mennotti A et al. (2001a) Drinking pattern and mortality: The Italian Risk Factor and Life Expectancy Pooling Project. Annals of Epidemiology, 11:312-319.
Trotter RT, Rehm J, Chatterji S et al. (2001) Cross-Cultural Applicability Research (CAR) Methods. In: Disability and culture: universalism and diversity. Üstün TB, Chatterji S, Bickenbaach JE, et al., eds. Hogrefe \& Huber Publishers, Göttingen.
U.S. Department of Health and Human Services (1997) Ninth special report to the US Congress on alcohol and health (Vol. No. 97-4017). U.S. Department of Health and Human Services, National Institute on Alcohol Abuse and Alcoholism (NIAAA), Rockville, MD.
U.S. Department of Health and Human Services (2000) 10th special report to the US Congress on alcohol and health: highlights from current research, Vol. No. 97-4017. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health (NIH), National Institute on Alcohol Abuse and Alcoholism (NIAAA), Rockville, MD.
Vaillant GE (1993) Is alcoholism more often the cause or the result of depression? Harvard Review of Psychiatry, 1:94-99.
Vaillant GE, Hiller-Sturmhofel S (1996) Natural history of alcoholism. Alcohol Health and Research World, 20:152-161.
Van den Brandt PA, Goldbohm RA, Van't Veer P (1995) Alcohol and breast cancer: results from the Netherland Cohort Study. American Journal of Epidemiology, 141:907-915.
Veenstra J, Ockhuizen T, van de Pol et al. (1990) Effects of moderate dose of alcohol on blood lipids and lipoproteins postprandially and in the fasting state. Alcohol, 25:371-377.
Vikhert A, Tsiplenkova VG, Cherpachenko N (1986) Alcoholic cardiomyopathy and sudden cardiac death. Journal of the American College of Cardiology, 8:A3-11.

Walsh G, Rehm J (1996) Daily drinking and harm. Contemporary Drug Problems, 23:465-478.

Wannamethee SG, Shaper AG (1991) Alcohol intake and variations in blood pressure by day of examination. Journal of Human Hypertension, 5:59-67.

Wannamethee SG, Shaper AG (1992) Alcohol and sudden cardiac death. British Heart Journal, 68:443-448.

Wannamethee SG, Shaper AG (1996) Patterns of alcohol intake and risk of stroke in middle-aged British men. Stroke, 27:1033-1039.
Washburne C (1956) Alcohol, self and the group. Quarterly Journal of Studies on Alcohol, 12:108-123.

Washburne C (1961) Primitive drinking. A study of the use and functions of alcohol in preliterate societies. College and University Press, New York, NY.

Wei H, Derson Y, Shuiyuan X et al. (1999) Alcohol consumption and alcoholrelated problems: Chinese experience from six area samples. Addiction, 94:1467-1476.

Wei M, Gibbons LW, Mitchell TL et al. (2000) Alcohol intake and incidence of type 2 diabetes in men. Diabetes Care, 23:18-22.
Weiner S, Tamerin JS, Steinglass P (1971) Familial patterns in chronic alcoholism: a study of a father and son during experimental intoxication. American Journal of Psychiatry, 127:82-87.
Whistler D, White KJ, Wong SD et al. (2001) Shazam: the econometric computer program, version 9. User's reference manual. Northwest Econometrics, Ltd., Vancouver, BC.

White S (1996) Russia goes dry: alcohol, state and society. Cambridge University Press, Cambridge, MA.

WHO (1999) Global status report on alcohol. World Health Organization, Substance Abuse Department, Geneva.

WHO (2000) International guide for monitoring alcohol consumption and related harm. World Health Organization, Department of Mental Health and Substance Dependence. Copenhagen.
WHO (2001) Surveys of drinking patterns and problems in seven developing countries. World Health Organization, Department of Mental Health and Substance Dependence. Geneva.
WHO (2002) World Health Report 2002: reducing risks, promoting healthy life. World Health Organization, Geneva.
Willenbring ML (1986) Measurement of depression in alcoholics. Journal of Studies on Alcohol, 49:412-417.

Willich SN, Lowel H, Lewis M et al. (1994) Weekly variation of acute myocardial infarction. Increased Monday risk in the working population. Circulation, 90:87-93.

Windham GC, Fenster L, Hopkins B et al. (1995) The association of moderate maternal and paternal alcohol consumption with birthweight and gestational age. Epidemiology, 6:591-597.
Windham GC, Von Behren J, Fenster L et al. (1997) Moderate maternal alcohol consumption and risk of spontaneous abortion. Epidemiology, 8:509-514.
Wingo PA, Calle EE, Thun JM et al. (1997) Alcohol consumption and the risk of fatal breast cancer in a prospective cohort of United States women. American Journal of Epidemiology, 145:S28.
Wood D, de Backer G, Faegeman O et al. (1998) Prevention of coronary heart disease in clinical practice. Recommendations of the Second Joint Task Force of European and Other Societies in Coronary Prevention. Atherosclerosis, 140:199-270.

Wu T, Trevisan M (2001) Drinking pattern and hypertension in a populationbased sample. American Journal of Epidemiology, 153:S59.

Yaffee R (2000) Introduction to time-series analysis and forecasting. Academic Press Inc., San Diego, CA.

Yang M, Rasbash J, Goldstein H et al. (1999) MLwiN macros for advanced multilevel modelling. Version 2.0. Multilevel Models Project, Institute of Education, London.

You R, McNeil J, O’Malley H et al. (1997) Risk factors for stroke due to cerebral infarction in young adults. Stroke, 28:1913-1918.
Yu Y, Hu J, Wang PP et al. (1997) Risk factors for bladder cancer: a case-control study in northeast China. European Review of Cancer Prevention, 6:363369.

Yuan J-M, Ross RK, Gao Y-T et al. (1997) Follow-up study of moderate alcohol intake and mortality among middle-aged men in Shanghai, China. British Medical Journal, 314:18-23.

Zakhari S (1997) Alcohol and the cardiovascular system: molecular mechanisms for beneficial and harmful action. Alcohol Health and Research World, 21:21-29.

Zhang Y, Kreger BE, Dorgan JF et al. (1999) Alcohol consumption and risk of breast cancer: the Framingham study revisited. American Journal of Epidemiology, 149:93-101.
Zwerling C, Sprince N, Wallace R et al. (1996) Alcohol and occupational injuries among older workers. Accident, Analysis and Prevention, 28:371-376.

Appendix A: "Pattern of drinking" variables and their
Relative weights
Heavy drinking occasions
(Maximum of 11 points for this component)
Daily drinking
Less than $20 \%$ daily drinking for males: 1 point
Less than $10 \%$ daily drinking for females: 1 point
Frequency of getting drunk
Most male drinkers usually get drunk when they are drinking: 2 points
Most males drinkers often get drunk: 1 point
Most female drinkers usually or often get drunk: 1 point
Usual quantity per drinking session
Males: more than $60 \%$ typically consume four or more drinks per session: 2 points
Males: between $40 \%$ and $60 \%$ consume four or more drinks per session: 1 point
Females: more than $50 \%$ consume four or more drinks per session: 2 points
Females: between $35 \%$ and $50 \%$ consume four or more drinks per session: 1 point

Fiesta binge drinking
Males: fiesta drinking commonly occurs: 1 point
Females: fiesta drinking commonly occurs: 1 point
DRINKING WITH MEALS
(Maximum of 4 points for this component)
Males: rarely or never with meals: 2 points
Males: sometimes with meals: 1 point
Females: rarely or never with meals: 2 points
Females: sometimes with meals: 1 point

## Drinking in public places

(Maximum of 2 points for this component)
Males: common and everyday: 1 point
Females: common and everyday: 1 point
Scoring (Possible range: 0-17 POINTS)
Scoring by summation of individual questions: range $0-17$
10-17 points: assign a pattern value of 4
7-9 points: assign a pattern value of 3
4-6 points: assign a pattern value of 2
$0-3$ points: assign a pattern value of 1
Appendix B: Alcohol-attributable fractions (AAFs) for mortality from injuries
Note: AAFs for morbidity from injuries were derived by multiplying the mortality AAFs by two thirds for motor vehicle accidents and by four ninths for all other types of injury.

| AFR-D |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| U/48 | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U150 | Motor vehicle accidents | 0.10 | 0.07 | 0.25 | 0.08 | 0.28 | 0.12 | 0.12 | 0.09 | 0.10 | 0.07 | 0.10 | 0.07 |
| U151 | Poisonings | 0.00 | 0.00 | 0.19 | 0.15 | 0.10 | 0.09 | 0.10 | 0.09 | 0.10 | 0.09 | 0.05 | 0.04 |
| U152 | Falls | 0.00 | 0.00 | 0.14 | 0.09 | 0.14 | 0.09 | 0.14 | 0.09 | 0.11 | 0.05 | 0.07 | 0.02 |
| U153 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U154 | Drownings | 0.00 | 0.00 | 0.17 | 0.16 | 0.21 | 0.20 | 0.21 | 0.20 | 0.16 | 0.16 | 0.16 | 0.16 |
| U155 | Other unintentional injuries | 0.10 | 0.03 | 0.19 | 0.15 | 0.19 | 0.15 | 0.16 | 0.12 | 0.16 | 0.12 | 0.16 | 0.12 |
| U156 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.09 | 0.06 | 0.09 | 0.06 | 0.07 | 0.05 | 0.07 | 0.05 | 0.03 | 0.03 |
| U158 | Homicide | 0.08 | 0.08 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.06 | 0.06 |

Appendix B: Alcohol-attributable fractions (AAFs) for mortality from injuries (continued)

| AFR-E |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| U/48 | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| Ul50 | Motor vehicle accidents | 0.15 | 0.10 | 0.50 | 0.12 | 0.54 | 0.17 | 0.31 | 0.13 | 0.26 | 0.10 | 0.26 | 0.10 |
| U151 | Poisonings | 0.00 | 0.00 | 0.42 | 0.21 | 0.25 | 0.14 | 0.25 | 0.14 | 0.25 | 0.14 | 0.13 | 0.06 |
| U152 | Falls | 0.00 | 0.00 | 0.34 | 0.13 | 0.34 | 0.13 | 0.34 | 0.13 | 0.27 | 0.08 | 0.20 | 0.04 |
| U153 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U154 | Drownings | 0.00 | 0.00 | 0.39 | 0.23 | 0.45 | 0.28 | 0.45 | 0.28 | 0.37 | 0.22 | 0.37 | 0.22 |
| U155 | Other unintentional injuries | 0.15 | 0.04 | 0.42 | 0.21 | 0.42 | 0.21 | 0.36 | 0.17 | 0.36 | 0.17 | 0.36 | 0.17 |
| U156 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.24 | 0.09 | 0.24 | 0.09 | 0.18 | 0.07 | 0.18 | 0.07 | 0.09 | 0.04 |
| U158 | Homicide | 0.12 | 0.12 | 0.40 | 0.25 | 0.40 | 0.25 | 0.40 | 0.25 | 0.40 | 0.25 | 0.40 | 0.25 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.31 | 0.18 | 0.31 | 0.18 | 0.31 | 0.18 | 0.31 | 0.18 | 0.17 | 0.09 |


| AMR-A |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| U148 | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| UI50 | Motor vehicle accidents | 0.18 | 0.12 | 0.38 | 0.14 | 0.42 | 0.20 | 0.21 | 0.16 | 0.17 | 0.12 | 0.17 | 0.12 |
| UI5I | Poisonings | 0.00 | 0.00 | 0.31 | 0.25 | 0.17 | 0.16 | 0.17 | 0.16 | 0.17 | 0.16 | 0.09 | 0.08 |
| U152 | Falls | 0.00 | 0.00 | 0.24 | 0.15 | 0.24 | 0.15 | 0.24 | 0.15 | 0.18 | 0.10 | 0.13 | 0.04 |
| UI53 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| UI54 | Drownings | 0.00 | 0.00 | 0.28 | 0.27 | 0.33 | 0.32 | 0.33 | 0.32 | 0.27 | 0.26 | 0.27 | 0.26 |
| UI55 | Other unintentional injuries | 0.18 | 0.05 | 0.31 | 0.25 | 0.31 | 0.25 | 0.26 | 0.20 | 0.26 | 0.20 | 0.26 | 0.20 |
| U156 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.16 | 0.11 | 0.16 | 0.11 | 0.12 | 0.09 | 0.12 | 0.09 | 0.05 | 0.05 |
| U158 | Homicide | 0.15 | 0.15 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 |
| UI59 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 | 0.11 | 0.11 |

Appendix B: Alcohol-attributable fractions (AAFs) for mortality from injuries (continued)

| AMR-B |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\underline{\text { U/48 }}$ | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U150 | Motor vehicle accidents | 0.18 | 0.12 | 0.56 | 0.14 | 0.60 | 0.20 | 0.36 | 0.16 | 0.30 | 0.12 | 0.30 | 0.12 |
| U151 | Poisonings | 0.00 | 0.00 | 0.48 | 0.25 | 0.30 | 0.16 | 0.30 | 0.16 | 0.30 | 0.16 | 0.16 | 0.08 |
| UI52 | Falls | 0.00 | 0.00 | 0.39 | 0.15 | 0.39 | 0.15 | 0.39 | 0.15 | 0.32 | 0.10 | 0.24 | 0.04 |
| U153 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U154 | Drownings | 0.00 | 0.00 | 0.44 | 0.27 | 0.50 | 0.32 | 0.50 | 0.32 | 0.43 | 0.26 | 0.43 | 0.26 |
| U155 | Other unintentional injuries | 0.18 | 0.06 | 0.48 | 0.25 | 0.48 | 0.25 | 0.42 | 0.21 | 0.42 | 0.21 | 0.42 | 0.21 |
| U156 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.28 | 0.11 | 0.28 | 0.11 | 0.22 | 0.09 | 0.22 | 0.09 | 0.11 | 0.06 |
| U158 | Homicide | 0.15 | 0.15 | 0.45 | 0.29 | 0.45 | 0.29 | 0.45 | 0.29 | 0.45 | 0.29 | 0.45 | 0.29 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.36 | 0.22 | 0.36 | 0.22 | 0.36 | 0.22 | 0.36 | 0.22 | 0.20 | 0.11 |


| AMR-D |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| U/48 | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| Ul50 | Motor vehicle accidents | 0.11 | 0.07 | 0.42 | 0.09 | 0.46 | 0.13 | 0.24 | 0.10 | 0.20 | 0.07 | 0.20 | 0.07 |
| U151 | Poisonings | 0.00 | 0.00 | 0.34 | 0.16 | 0.20 | 0.10 | 0.20 | 0.10 | 0.20 | 0.10 | 0.10 | 0.05 |
| U152 | Falls | 0.00 | 0.00 | 0.27 | 0.09 | 0.27 | 0.09 | 0.27 | 0.09 | 0.21 | 0.06 | 0.15 | 0.03 |
| Ul53 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Ul54 | Drownings | 0.00 | 0.00 | 0.31 | 0.18 | 0.37 | 0.21 | 0.37 | 0.21 | 0.30 | 0.17 | 0.30 | 0.17 |
| U155 | Other unintentional injuries | 0.11 | 0.03 | 0.34 | 0.16 | 0.34 | 0.16 | 0.29 | 0.13 | 0.29 | 0.13 | 0.29 | 0.13 |
| Ul56 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.18 | 0.07 | 0.18 | 0.07 | 0.14 | 0.05 | 0.14 | 0.05 | 0.06 | 0.03 |
| U158 | Homicide | 0.09 | 0.09 | 0.32 | 0.19 | 0.32 | 0.19 | 0.32 | 0.19 | 0.32 | 0.19 | 0.32 | 0.19 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.24 | 0.14 | 0.24 | 0.14 | 0.24 | 0.14 | 0.24 | 0.14 | 0.12 | 0.07 |

Appendix B: Alcohol-attributable fractions (AAFs) for mortality from injuries (continued)

| EMR-B |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| U/48 | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U150 | Motor vehicle accidents | 0.03 | 0.02 | 0.08 | 0.02 | 0.09 | 0.03 | 0.03 | 0.02 | 0.03 | 0.02 | 0.03 | 0.02 |
| U151 | Poisonings | 0.00 | 0.00 | 0.06 | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.01 | 0.01 |
| U152 | Falls | 0.00 | 0.00 | 0.04 | 0.02 | 0.04 | 0.02 | 0.04 | 0.02 | 0.03 | 0.01 | 0.02 | 0.01 |
| U153 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U154 | Drownings | 0.00 | 0.00 | 0.05 | 0.05 | 0.06 | 0.06 | 0.06 | 0.06 | 0.05 | 0.04 | 0.05 | 0.04 |
| U155 | Other unintentional injuries | 0.03 | 0.01 | 0.06 | 0.04 | 0.06 | 0.04 | 0.04 | 0.03 | 0.04 | 0.03 | 0.04 | 0.03 |
| U156 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.03 | 0.02 | 0.03 | 0.02 | 0.02 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 |
| U158 | Homicide | 0.02 | 0.02 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.02 | 0.02 |


| EMR-D |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| U/48 | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| UI50 | Motor vehicle accidents | 0.01 | 0.01 | 0.04 | 0.01 | 0.04 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| UI5I | Poisonings | 0.00 | 0.00 | 0.03 | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 |
| UI52 | Falls | 0.00 | 0.00 | 0.02 | 0.01 | 0.02 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 |
| UI53 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| UI54 | Drownings | 0.00 | 0.00 | 0.02 | 0.02 | 0.03 | 0.03 | 0.03 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 |
| UI55 | Other unintentional injuries | 0.01 | 0.00 | 0.03 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| U156 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 |
| U158 | Homicide | 0.01 | 0.01 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.01 | 0.01 |

Appendix B: Alcohol-attributable fractions (AAFs) for mortality from injuries (continued)

| EUR-A |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\underline{\text { U/48 }}$ | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U150 | Motor vehicle accidents | 0.23 | 0.15 | 0.46 | 0.18 | 0.50 | 0.25 | 0.27 | 0.21 | 0.22 | 0.15 | 0.22 | 0.15 |
| U151 | Poisonings | 0.00 | 0.00 | 0.38 | 0.31 | 0.22 | 0.21 | 0.22 | 0.21 | 0.22 | 0.21 | 0.12 | 0.10 |
| U152 | Falls | 0.00 | 0.00 | 0.30 | 0.20 | 0.30 | 0.20 | 0.30 | 0.20 | 0.24 | 0.13 | 0.17 | 0.06 |
| U153 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U154 | Drownings | 0.00 | 0.00 | 0.35 | 0.33 | 0.40 | 0.39 | 0.40 | 0.39 | 0.33 | 0.32 | 0.33 | 0.32 |
| U155 | Other unintentional injuries | 0.23 | 0.07 | 0.38 | 0.31 | 0.38 | 0.31 | 0.32 | 0.26 | 0.32 | 0.26 | 0.32 | 0.26 |
| U156 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.21 | 0.14 | 0.21 | 0.14 | 0.16 | 0.12 | 0.16 | 0.12 | 0.07 | 0.07 |
| U158 | Homicide | 0.19 | 0.19 | 0.36 | 0.36 | 0.36 | 0.36 | 0.36 | 0.36 | 0.36 | 0.36 | 0.36 | 0.36 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.27 | 0.27 | 0.27 | 0.27 | 0.27 | 0.27 | 0.27 | 0.27 | 0.14 | 0.14 |


| EUR-B |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| U/48 | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| UI50 | Motor vehicle accidents | 0.17 | 0.11 | 0.54 | 0.13 | 0.58 | 0.19 | 0.34 | 0.15 | 0.29 | 0.11 | 0.29 | 0.11 |
| U151 | Poisonings | 0.00 | 0.00 | 0.46 | 0.24 | 0.28 | 0.15 | 0.28 | 0.15 | 0.28 | 0.15 | 0.15 | 0.07 |
| U152 | Falls | 0.00 | 0.00 | 0.37 | 0.14 | 0.37 | 0.14 | 0.37 | 0.14 | 0.30 | 0.09 | 0.22 | 0.04 |
| U153 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U154 | Drownings | 0.00 | 0.00 | 0.42 | 0.26 | 0.48 | 0.31 | 0.48 | 0.31 | 0.41 | 0.25 | 0.41 | 0.25 |
| UI55 | Other unintentional injuries | 0.17 | 0.05 | 0.46 | 0.24 | 0.46 | 0.24 | 0.40 | 0.20 | 0.40 | 0.20 | 0.40 | 0.20 |
| UI56 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.27 | 0.10 | 0.27 | 0.10 | 0.20 | 0.08 | 0.20 | 0.08 | 0.10 | 0.05 |
| U158 | Homicide | 0.14 | 0.14 | 0.44 | 0.28 | 0.44 | 0.28 | 0.44 | 0.28 | 0.44 | 0.28 | 0.44 | 0.28 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.34 | 0.21 | 0.34 | 0.21 | 0.34 | 0.21 | 0.34 | 0.21 | 0.19 | 0.10 |

Appendix B: Alcohol-attributable fractions (AAFs) for mortality from injuries (continued)

| EUR-C |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| U/48 | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U150 | Motor vehicle accidents | 0.32 | 0.23 | 0.71 | 0.27 | 0.74 | 0.36 | 0.52 | 0.29 | 0.46 | 0.23 | 0.46 | 0.23 |
| U151 | Poisonings | 0.00 | 0.00 | 0.64 | 0.42 | 0.45 | 0.30 | 0.45 | 0.30 | 0.45 | 0.30 | 0.27 | 0.16 |
| U152 | Falls | 0.00 | 0.00 | 0.55 | 0.28 | 0.55 | 0.28 | 0.55 | 0.28 | 0.47 | 0.19 | 0.37 | 0.09 |
| U153 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U154 | Drownings | 0.00 | 0.00 | 0.61 | 0.45 | 0.66 | 0.51 | 0.66 | 0.51 | 0.59 | 0.44 | 0.59 | 0.44 |
| U155 | Other unintentional injuries | 0.32 | 0.11 | 0.64 | 0.42 | 0.64 | 0.42 | 0.58 | 0.36 | 0.58 | 0.36 | 0.58 | 0.36 |
| U156 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.43 | 0.21 | 0.43 | 0.21 | 0.35 | 0.18 | 0.35 | 0.18 | 0.19 | 0.11 |
| U158 | Homicide | 0.28 | 0.28 | 0.62 | 0.47 | 0.62 | 0.47 | 0.62 | 0.47 | 0.62 | 0.47 | 0.62 | 0.47 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.52 | 0.38 | 0.52 | 0.38 | 0.52 | 0.38 | 0.52 | 0.38 | 0.33 | 0.21 |


| SEAR-B |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| U148 | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| Ul50 | Motor vehicle accidents | 0.07 | 0.04 | 0.30 | 0.05 | 0.34 | 0.08 | 0.16 | 0.06 | 0.13 | 0.04 | 0.13 | 0.04 |
| UI5I | Poisonings | 0.00 | 0.00 | 0.24 | 0.10 | 0.13 | 0.06 | 0.13 | 0.06 | 0.13 | 0.06 | 0.06 | 0.03 |
| U152 | Falls | 0.00 | 0.00 | 0.18 | 0.06 | 0.18 | 0.06 | 0.18 | 0.06 | 0.14 | 0.04 | 0.09 | 0.02 |
| UI53 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U154 | Drownings | 0.00 | 0.00 | 0.21 | 0.11 | 0.26 | 0.14 | 0.26 | 0.14 | 0.20 | 0.11 | 0.20 | 0.11 |
| UI55 | Other unintentional injuries | 0.07 | 0.02 | 0.24 | 0.10 | 0.24 | 0.10 | 0.19 | 0.08 | 0.19 | 0.08 | 0.19 | 0.08 |
| UI56 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| Ul57 | Self-inflicted injuries | 0.00 | 0.00 | 0.12 | 0.04 | 0.12 | 0.04 | 0.09 | 0.03 | 0.09 | 0.03 | 0.04 | 0.02 |
| Ul58 | Homicide | 0.06 | 0.06 | 0.22 | 0.12 | 0.22 | 0.12 | 0.22 | 0.12 | 0.22 | 0.12 | 0.22 | 0.12 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.16 | 0.09 | 0.16 | 0.09 | 0.16 | 0.09 | 0.16 | 0.09 | 0.08 | 0.04 |

Appendix B: Alcohol-attributable fractions (AAFs) for mortality from injuries (continued)

| SEAR-D |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| U/48 | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U150 | Motor vehicle accidents | 0.05 | 0.03 | 0.22 | 0.04 | 0.25 | 0.05 | 0.11 | 0.04 | 0.09 | 0.03 | 0.09 | 0.03 |
| U151 | Poisonings | 0.00 | 0.00 | 0.17 | 0.07 | 0.09 | 0.04 | 0.09 | 0.04 | 0.09 | 0.04 | 0.04 | 0.02 |
| U152 | Falls | 0.00 | 0.00 | 0.12 | 0.04 | 0.12 | 0.04 | 0.12 | 0.04 | 0.09 | 0.02 | 0.06 | 0.01 |
| U153 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U154 | Drownings | 0.00 | 0.00 | 0.15 | 0.08 | 0.19 | 0.10 | 0.19 | 0.10 | 0.14 | 0.07 | 0.14 | 0.07 |
| U155 | Other unintentional injuries | 0.05 | 0.01 | 0.17 | 0.07 | 0.17 | 0.07 | 0.14 | 0.06 | 0.14 | 0.06 | 0.14 | 0.06 |
| U156 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.08 | 0.03 | 0.08 | 0.03 | 0.06 | 0.02 | 0.06 | 0.02 | 0.03 | 0.01 |
| U158 | Homicide | 0.04 | 0.04 | 0.16 | 0.08 | 0.16 | 0.08 | 0.16 | 0.08 | 0.16 | 0.08 | 0.16 | 0.08 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.11 | 0.06 | 0.11 | 0.06 | 0.11 | 0.06 | 0.11 | 0.06 | 0.05 | 0.03 |


Appendix B: Alcohol-attributable fractions (AAFs) for mortality from injuries (continued)

| WPR-B |  | Age group (years) |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\underline{\text { U/48 }}$ | III. Injuries | 0-15 |  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | $\geq 70$ |  |
|  |  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| U149 | A. Unintentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U150 | Motor vehicle accidents | 0.10 | 0.07 | 0.25 | 0.08 | 0.28 | 0.12 | 0.13 | 0.09 | 0.10 | 0.07 | 0.10 | 0.07 |
| U151 | Poisonings | 0.00 | 0.00 | 0.19 | 0.15 | 0.10 | 0.09 | 0.10 | 0.09 | 0.10 | 0.09 | 0.05 | 0.04 |
| U152 | Falls | 0.00 | 0.00 | 0.14 | 0.09 | 0.14 | 0.09 | 0.14 | 0.09 | 0.11 | 0.06 | 0.07 | 0.02 |
| U153 | Fires | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U154 | Drownings | 0.00 | 0.00 | 0.17 | 0.16 | 0.21 | 0.20 | 0.21 | 0.20 | 0.16 | 0.16 | 0.16 | 0.16 |
| U155 | Other unintentional injuries | 0.10 | 0.03 | 0.19 | 0.15 | 0.19 | 0.15 | 0.16 | 0.12 | 0.16 | 0.12 | 0.16 | 0.12 |
| U156 | B. Intentional injuries |  |  |  |  |  |  |  |  |  |  |  |  |
| U157 | Self-inflicted injuries | 0.00 | 0.00 | 0.09 | 0.06 | 0.09 | 0.06 | 0.07 | 0.05 | 0.07 | 0.05 | 0.03 | 0.03 |
| U158 | Homicide | 0.08 | 0.08 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 |
| U159 | War | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U160 | Other intentional injuries | 0.00 | 0.00 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.06 | 0.06 |

Appendix C: Alcohol-attributable deaths (ooos) in 2000 by disease category, sex and subregion

| Disease category | Sex | Subregions |  |  |  |  |  |  |  |  |  |  |  |  |  | World |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | AFR-D | AFR-E | AMR-A | AMR-B | AMR-D | EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-A | WPR-B | World | Total |
| Maternal and perinatal conditions | Males | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 |
|  | Females | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |  |
| Cancer | Males | 7 | 18 | 10 | 10 | 1 | 0 | 1 | 30 | 6 | 15 | 4 | 7 | 13 | 147 | 269 | 355 |
|  | Females | 3 | 6 | 7 | 7 | 1 | 0 | 0 | 21 | 4 | 9 | 1 | 1 | 5 | 23 | 86 |  |
| Neuro-psychiatric diseases | Males | 5 | 12 | 7 | 14 | 2 | 1 | 1 | 13 | 4 | 8 | 4 | 9 | 1 | 13 | 91 | 111 |
|  | Females | 1 | 3 | 2 | 2 | 0 | 0 | 0 | 4 | 1 | 3 | 0 | 1 | 0 | 2 | 19 |  |
| Cardiovascular diseases | Males | 8 | 16 | -27 | 48 | 4 | 1 | 3 | -38 | 49 | 120 | 6 | 51 | -6 | 157 | 392 | 268 |
|  | Females | 3 | 5 | -40 | 16 | 2 | 0 | 0 | -129 | 11 | 37 | 1 | 0 | -37 | 8 | -124 |  |
| Other noncommunicable diseases | Males | 9 | 14 | 10 | 24 | 3 | 0 | 1 | 27 | 12 | 26 | 4 | 6 | 4 | 54 | 193 | 242 |
|  | Females | 3 | 4 | 2 | 5 | I | 0 | 0 | 9 | 6 | 12 | I | 2 | I | 3 | 49 |  |
| Unintentional injury | Males | 17 | 45 | 19 | 56 | 9 | 3 | 2 | 26 | 22 | 112 | 31 | 60 | 7 | 75 | 484 | 577 |
|  | Females | 4 | 7 | 6 | 5 | 1 | 0 | 1 | 8 | 3 | 18 | 5 | 12 | 3 | 20 | 92 |  |
| Intentional injury | Males | 6 | 21 | 8 | 55 | 3 | 0 | 0 | 7 | 7 | 58 | 4 | 14 | 3 | 19 | 206 | 248 |
|  | Females | 2 | 5 | 2 | 3 | 0 | 0 | 0 | 2 | 1 | 10 | I | 4 | 1 | 10 | 42 |  |
| All alcoholattributable deaths | Males | 53 | 125 | 27 | 207 | 22 | 6 | 8 | 65 | 100 | 338 | 51 | 148 | 23 | 465 | 1638 | 1804 |
|  | Females | 15 | 30 | -22 | 39 | 6 | 1 | 1 | -85 | 25 | 88 | 9 | 21 | -28 | 66 | 166 |  |
| All deaths | Males | 2206 | 3154 | 1342 | 1459 | 290 | 409 | 1750 | 2020 | 1034 | 1878 | 6358 | 616 | 5483 | 55862 | 26629 | 55861 |
|  | Females | 6 | 3001 | 1392 | 1120 | 237 | 287 | 1602 | 2054 | 916 | 1721 | 1022 | 5764 | 519 | 4944 | 29232 |  |
| Deaths attributable to alcohol as a percentage of all deaths | Males | 2.4 | 4.0 | 2.0 | 14.2 | 7.6 | 1.5 | 0.4 | 3.2 | 9.6 | 18.0 | 0.8 | 24.1 | 0.4 | 0.8 | 6.2 | 3.2 |
|  | Females | 0.8 | 1.0 | -1.6 | 3.5 | 2.3 | 0.3 | 0.1 | -4.1 | 2.8 | 5.1 | 0.9 | 0.4 | -5.3 | 1.3 | 0.6 |  |

Appendix D: Alcohol-related disease burden in DALYs (ooos) in 2000 by disease category, sex and subregion

| Disease category | Sex | Subregion |  |  |  |  |  |  |  |  |  |  |  |  |  | World |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | AFR-D | AFR-E | AMR-A | AMR-B | AMR-D | EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-A | WPR-B | World | Total |
| Maternal and perinatal conditions | Males | 9 | 17 | 1 | 15 | 1 | 0 | 1 | 2 | 4 | 2 | 1 | 15 | 0 | 0 | 68 | 123 |
|  | Females | 7 | 13 | 1 | 12 | 1 | 0 | 0 | 2 | 3 | 2 | 1 | 13 | 0 | 0 | 55 |  |
| Cancer | Males | 107 | 260 | 99 | 116 | 12 | 6 | 11 | 282 | 72 | 175 | 58 | 125 | 117 | 1740 | 3180 | 4201 |
|  | Females | 38 | 72 | 79 | 81 | 13 | 1 | 1 | 202 | 45 | 103 | 14 | 17 | 49 | 305 | 1021 |  |
| Neuro-psychiatric diseases | Males | 305 | 828 | 2113 | 2979 | 305 | 20 | 170 | 1867 | 575 | 1509 | 524 | 1500 | 361 | 5033 | 18090 | 21904 |
|  | Females | 31 | 141 | 682 | 792 | 82 | 3 | 9 | 514 | 109 | 398 | 74 | 101 | 160 | 717 | 3814 |  |
| Cardiovascular diseases | Males | 98 | 207 | -174 | 480 | 38 | 14 | 42 | -233 | 449 | 1161 | 86 | 851 | -39 | 1432 | 4411 | 3983 |
|  | Females | 30 | 53 | -256 | 162 | 16 | 1 | 1 | -627 | 87 | 234 | 9 | 5 | -219 | 76 | -428 |  |
| Other noncommunicable diseases | Males | 149 | 252 | 165 | 531 | 55 | 8 | 12 | 380 | 232 | 486 | 94 | 220 | 67 | 1045 | 3695 | 4555 |
|  | Females | 44 | 74 | 25 | 101 | 20 | 2 | 2 | 147 | 96 | 196 | 17 | 69 | 3 | 66 | 860 |  |
| Unintentional injury | Males | 576 | 1425 | 498 | 1815 | 268 | 100 | 80 | 643 | 675 | 2771 | 912 | 1837 | 141 | 2268 | 14008 | 16495 |
|  | Females | 162 | 280 | 119 | 177 | 29 | 11 | 17 | 136 | 81 | 402 | 133 | 359 | 34 | 545 | 2487 |  |
| Intentional injury | Males | 198 | 632 | 222 | 1919 | 110 | 14 | 12 | 162 | 176 | 1439 | 118 | 379 | 62 | 502 | 5945 | 7062 |
|  | Females | 81 | 153 | 53 | 118 | 9 | 3 | 5 | 42 | 25 | 234 | 35 | 110 | 17 | 231 | 1117 |  |
| All alcoholattributable DALYs | Males | 1441 | 3621 | 2925 | 7854 | 789 | 162 | 328 | 3103 | 2183 | 7543 | 1793 | 4927 | 708 | 12020 | 49397 | 58323 |
|  | Females | 393 | 785 | 702 | 1443 | 170 | 22 | 36 | 416 | 446 | 1570 | 284 | 675 | 43 | 1941 | 8926 |  |
| DALYs attributable to alcohol as a percentage of all DALYs | Males | 2.0 | 3.5 | 11.9 | 17.3 | 8.6 | 1.3 | 0.6 | 11.1 | 10.2 | 21.5 | 5.3 | 2.8 | 8.1 | 29.1 | 6.5 | 4.0 |
|  | Females | 0.6 | 0.8 | 3.2 | 4.1 | 2.2 | 0.2 | 0.1 | 1.6 | 2.5 | 6.5 | 1.0 | 0.4 | 0.6 | 1.8 | 1.3 |  |

## Chapter 13

Illicit Drug use<br>Louisa Degenhardt, Wayne Hall, Matthew Warner-Smith and Michael Lynskey

## Summary

Estimating mortality directly attributable to illicit drug use such as overdose death - the most tangible adverse heath effect of illicit drug useis difficult because of variations in the quality and quantity of mortality data. As a result, it is necessary to make indirect estimates, involving estimates of the prevalence of illicit drug use. However, it is difficult to make even indirect estimates because the use of these drugs is illegal, stigmatized and hidden. Nonetheless, efforts must be made to estimate the contribution that illicit drug use makes to the global burden of disease, because it is a pattern of behaviour that has a substantial adverse effect on the health of those who engage in it. In cohort studies of treated drug users the problematic use of illicit drugs has been associated with an increased overall rate of mortality, and with an elevated rate of a number of individual causes of death, four of which were estimated here: AIDS, overdose, suicide and trauma.

Definitions of the variable of interest are difficult because of deficiencies in the data collected by countries on illicit drug use, and by disagreements over what constitutes "problematic" illicit drug use. The definition used here was long-term regular injecting use of opioids, amphetamines or cocaine. Data on the prevalence of problematic illicit drug use were derived from a range of sources that used variable methods of deriving estimates.

A literature search was conducted of all studies that estimated the prevalence of problematic drug use. Available data on prevalence in countries with data were used to estimate the prevalence of problematic illicit drug use for subregions. ${ }^{1}$ A search was also completed for cohort studies of drug users that had estimated mortality due to the four individual causes of death, and to all causes of death. Data on the number of years of follow-up were extracted from each study and a weighted average annual mortality rate was calculated for each of the four causes
of death, and for their sum. A standardized mortality ratio (SMR) was also derived from previous estimates of the excess mortality from all causes attributable to illicit drugs. Estimates were made for some causes by applying an attributable fraction obtained from sources such as the Joint United Nations Programme on HIV/AIDS (UNAIDS) (for HIVrelated deaths) to estimates of total deaths for some causes. The median estimate of a range of estimates was used as the estimate for each subregion. Estimates were limited to persons aged 15-54 years.

In 2000, the median number of global deaths attributed to illicit drugs estimated by summing the four causes of death was 194058 . There were an additional 10000 deaths from overdose above and beyond those coded as drug use disorders (added to unintentional injuries) or when coded drug use disorder deaths were higher than estimated overdose deaths. The median 2000 estimate derived using the all-cause method was 197383 . Both estimates had wide uncertainty intervals around them ( 113494 to 276584 for sum of four causes; and 101751 to 322456 for all-cause estimates). When morbidity attributable to illicit drug use is added to the estimated mortality, this risk factor accounts for $0.8 \%$ of global disability-adjusted life years (DALYs). The distribution of numbers of deaths between subregions varied between the two methods. These variations in the estimates reflect the considerable uncertainty about prevalence of drug use in different subregions and uncertainty about the applicability of mortality data derived in developed countries to mortality among illicit drug users in developing countries.

The current estimates suggest that illicit drug use is a significant cause of premature mortality among young adults. This is an underestimate of total disease burden because: (i) there are deficits in data on mortality attributable to the use of some illicit drug (most notably cannabis and the newer synthetic drugs like MDMA ${ }^{2}$ ); (ii) there are differences across subregions in the quality of data available on the causes of mortality that were included in the current estimates; (iii) there is an absence of data that would permit estimates of some other causes of mortality and morbidity attributable to illicit drug use, such as hepatitis B and hepatitis C and violence. There is a need for better data on: the prevalence of illicit drug use in developed and developing countries, and on the mortality and morbidity attributable to problematic drug use.

## 1. Introduction

The use of legally proscribed psychotropic substances for non-medical purposes appears to be increasing in many parts of the world (Frischer et al. 1994; UNDCP 2000; UNODCCP 2000) but it is difficult to quantify the rate of increase. It is difficult to estimate the prevalence of this behaviour and its adverse health consequences in individual societies because this behaviour is illicit and therefore often hidden. Even estimating mortality related to illicit drug use, the most tangible adverse
heath effect, is difficult for reasons that are discussed below (Thorley et al. 1977). Nonetheless, efforts must be made to estimate the contribution that illicit drug use makes to the global burden of disease because it is a pattern of behaviour that has a substantial adverse effect on the health and well-being of those who engage in it, producing substantial loss of life and disability (Hulse et al. 1999).

The global burden of death and disability attributable to illicit drugs was first estimated by Donoghoe (1996), as part of the Global Burden of Disease (GBD) project (Murray and Lopez 1996). Donoghoe estimated that illicit drug use was responsible for 100000 deaths globally in 1990 , the majority of which ( $62 \%$ ) occurred in developing countries. Murray and Lopez (1996) pointed out that this estimate may be too low because of difficulties in reliably estimating the prevalence of illicit drug use and its adverse health effects. Donoghoe's estimate was based on the attributable fractions of various causes of mortality and morbidity attributed to illicit drug use by English et al. (1995), who reviewed all studies published up to 1993. The great majority of these studies, which were principally cohort studies, were conducted in the United States of America and Europe.

Since these estimates were made, there has been an apparent increase in illicit drug consumption in developed societies (Australian Bureau of Criminal Intelligence 2000; EMCDDA 2000; Frischer et al. 1994; UNODCCP 2000), and increased incidence of HIV contracted as a result of sharing of injecting equipment by illicit drug users in developing societies (Stimson 1993). This suggests that Donoghoe's 1990 estimates are likely to substantially underestimate the contribution that illicit drug use makes to the global burden of disease in 2000.

In this chapter, we have attempted to estimate the burden of disease due to illicit drug use by combining a range of sources of data on the prevalence of use and indicators of outcome. We also outline the definitions of the "exposure" variable used in making estimates, and outline the causes of burden considered in this chapter. As will become clear, estimates made of this cause of burden are difficult to make given: (i) paucity of data on the prevalence of illicit drug use around the world; (ii) the fact that data on causes of death related to illicit drug use are not well-recorded, so it is necessary to rely on indirect estimates derived from inaccurate prevalence estimates; and (iii) an absence of evidence on the risk of mortality and morbidity due to some causes among illicit drug users.

### 1.1 EXPOSURE VARIABLE

## Substances included

Illicit drug use includes the non-medical use of a variety of drugs that are prohibited by international law. These drugs include: amphetaminetype stimulants, ${ }^{3}$ cannabis, ${ }^{4}$ cocaine, ${ }^{5}$ heroin ${ }^{6}$ and other opioids, ${ }^{7}$ and MDMA (ecstasy). In order to estimate mortality and morbidity attrib-
utable to illicit drug use, we need to clearly define what is and is not included in this risk factor.

This chapter will focus on the burden attributable to amphetamines, cocaine and opioids. Other substances that are illegal in most countries, such as ecstasy, solvents and cannabis, have not been included in the present analysis as there is currently insufficient research information to quantify the health risks associated with these drugs. Thus, their exclusion should not be interpreted as meaning that the use of these drugs is safe. Rather, it reflects a paucity of research on the harm caused by their use.

Relationship to dose, frequency and route of ADMINISTRATION
The risk of premature mortality and morbidity from illicit drug use is dependent on dose, frequency and route of administration. Consequently it is necessary to define what is meant by "use" when defining the exposure variable "illicit drug use". The mortality risks of illicit drug consumption increase with increasing frequency and quantity of consumption (Fischer et al. 1997). Simple prevalence estimates of the proportion of the population that have ever used an illicit drug are likely to be associated with a low average risk since a single occasion of use and infrequent use, the most common patterns of use reported in population surveys, are associated with a small increase in mortality. More accurate estimates of the burden of disease attributable to illicit drugs require estimates of the prevalence of the most hazardous patterns of illicit drug use. These are found in highest prevalence among dependent drug users who typically inject drugs daily or near daily over periods of years. This pattern of use exposes users to the highest chance of fatal overdose (Warner-Smith et al. 2001) and of contracting bloodborne viral diseases (Ross et al. 1992).

The World Health Organization (WHO), following the International Classification of Diseases, defines problem drug use as "harmful drug use" and "drug dependence". Harmful drug use is defined by clear evidence that the substance use is responsible for physical (e.g. organ damage) and psychological harm (e.g. drug-induced psychosis). Drug dependence, as defined in the International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10), requires the presence of three or more indicators of drug dependence (WHO 1993). These include: a strong desire to take the substance; impaired control over the use; a withdrawal syndrome on ceasing or reducing use; tolerance to the effects of the drug; requiring larger doses to achieve the desired psychological effect; a disproportionate amount of the user's time is spent obtaining, using and recovering from drug use; and the user continuing to take other drugs despite associated problems. The problems should have been experienced for at least one month at some time during the previous year. The United Nations Drug Control Programme
(UNDCP) identifies "problem drugs" based on "the extent to which use of a certain drug leads to treatment demand, emergency room visits (often due to overdose), drug-related morbidity (including HIV/AIDS, hepatitis etc.), mortality and other drug-related social ills" (UNDCP 2000).

Most prevalence estimates vary with the assumptions made and the methodology employed. Data provided by the UNDCP do not have the same reliability as large-scale household surveys of the type generally conducted in developed countries. Unfortunately the expense of conducting such surveys makes their use in developing countries unfeasible. Even if such surveys were feasible in all countries, it is generally accepted that surveys underestimate harmful illicit drug use (Hall et al. 2000b).

The European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) has invested considerable resources in developing methods for the collection of data on the prevalence of harmful illicit drug use that are both valid and comparable (EMCDDA 1997). While these standards have been developed for use within the European Union the global adoption of such standards may greatly improve estimates of drugrelated harm. The EMCDDA defines "problem drug use" as injecting drug use (IDU) or long duration or regular use of opioids, cocaine or amphetamines (EMCDDA 1999). The EMCDDA definition is the one that we have adopted in estimating mortality attributable to illicit drugs.

ILLICIT DRUGS NOT INCLUDED IN CURRENT ESTIMATES OF MORTALITY

## Cannabis

Cannabis has a high prevalence of use in many developed societies (Hall et al. 1999b) but there is a lack of well-controlled studies showing that its use increases mortality (Hall and Solowij 1998; WHO 1997). For example, we identified two cohort studies that have examined the effects of regular, prolonged cannabis use on risks of cancer. One of these reported no increase in overall cancer rates among cannabis users (although there were slightly increased rates of prostate and cervical cancer) (Sidney et al. 1997b). A case-control study found a doubling of the odds of aerodigestive cancers among heavy users of cannabis (Zhang et al. 1999) but it was difficult to disentangle the effects of cannabis smoking from those of tobacco smoking because many cannabis users also smoked tobacco (Andreasson and Allebeck 1990).

There are two prospective epidemiological studies of mortality among cannabis users. A Swedish study over 15 years of mortality among male military conscripts found an increased risk of premature death among men who had smoked cannabis 50 or more times by age 18 years (Andreasson and Allebeck 1990). Violent and accidental deaths were the major contributor to this excess. However, the association between mortality and cannabis use disappeared after multivariate statistical adjustment for alcohol and other drug use. Sidney et al. (1997a) re-
ported a 10 -year study of mortality in cannabis users among 65171 members of the Kaiser Permanente Medical Care Program aged between 15 and 49 years. The sample comprised $38 \%$ who had never used cannabis, $20 \%$ who had used less than six times, $20 \%$ who were former users, and $22 \%$ who were current cannabis users. Regular cannabis use had a small association with premature mortality (relative risk of 1.3) that was wholly explained by increased AIDS deaths in men, probably because cannabis use was a marker for male homosexual behaviour in this cohort. It is too early to conclude that cannabis use does not increase mortality because the average age at follow-up was only 43 years, and cigarette smoking and alcohol use were only modestly associated with premature mortality. For these reasons, we have not included any estimate of cannabis' effects on overall premature mortality.

Cannabis produces dose-related impairments in cognitive and behavioural functions that may potentially impair driving an automobile or operating machinery (Chait 1992). These impairments are larger and more persistent in difficult tasks involving sustained attention (Chait 1992). The most serious possible consequence of acute cannabis use is a motor vehicle accident if a user drives while intoxicated (Hall et al. 1994).

The effects of recreational doses of cannabis on driving performance in laboratory simulators and standardized driving courses have been reported as similar to blood alcohol concentrations between $0.07 \%$ and $0.10 \%$ (Hall et al. 1994). However, studies of the effects of cannabis on driving under more realistic conditions on roads have found much more modest impairments (Bates and Blakely 1999; Robbe 1994; Smiley 1999). This is probably because cannabis users are more aware of their impairment and less inclined to take risks than alcohol users (Smiley 1999).

Epidemiological studies of motor vehicle accidents have produced equivocal results because most drivers who have cannabinoids in their blood also have high blood alcohol levels (Hall et al. 1994, 2001). Studies with reasonable numbers of persons who have only used cannabis have not found clear evidence of increased culpability in these drivers (Bates and Blakely 1999; Chesher 1995). For these reasons we have not included any estimate of the contribution that cannabis makes to motor vehicle fatalities.

## Other illicit drugs

Estimating the contribution that MDMA (ecstasy), hallucinogenic substances and inhalants make to premature mortality presents similar problems to cannabis (Boot et al. 2000). While there are case reports of deaths associated with MDMA intoxication (Dowling et al. 1987; Henry et al. 1992; Parr et al. 1997) these appear to be rare by comparison with overdose deaths due to opioids and cocaine in developed societies with good mortality data, such as Australia (Ridolfo and Stevenson
2001). The illicit use of pharmaceuticals and anabolic steroids have also been excluded from further analysis because difficulties in measuring (i) the prevalence of their harmful use and (ii) mortality attributable to their use mean that it is not possible to calculate relative risks. Similarly, the failure to include solvents stems largely from a lack of good evidence on the prevalence and extent or harm attributable to their use.

The exposure variable for illicit drug use in this analysis is, therefore, injection or long duration of use of amphetamines, cocaine or opioids. The failure to include cannabis, MDMA, hallucinogens and inhalants in our estimates of burden of disease attributable to illicit drugs reflects our ignorance of their health risks; it does not imply that the use of these drugs is without risk to users.

## Counterfactual exposure distribution

The theoretical minimum counterfactual exposure distribution is zero illicit drug use. There may be countries in the world that can truly claim to have zero illicit drug use but there must be few of these now. Even countries that have the policy goal of achieving a drug-free society, such as Sweden, do not have zero illicit drug use. Arguably, once illicit drug use and dependence have appeared in a society, it is unrealistic to expect to be able to return to a zero level of illicit drug use. It may be reasonable to aim to reduce the prevalence of the most harmful types of illicit drug use and to minimize the harm that their use causes.

One approach to defining a plausible counterfactual exposure would be to use developed countries with the lowest prevalence of illicit drug use as the basis for the estimate. Countries like Finland and Sweden may be suggested as examples. The weakness with this strategy is that illicit drug use trends are dynamic and countries that currently have low rates may show increases in rates of use (as has recently happened in Sweden) as availability of illicit drugs increases and more favourable social attitudes develop towards illicit drug use among young adults.

It is also not clear what are feasible minimum counterfactuals. It is not clear whether prevention programmes, such as school-based and other intervention programmes, can prevent problem drug use (National Research Council 2001). These programmes have been most widely implemented and evaluated in the United States. After reviewing this evidence, the United States National Research Council recently concluded that the
effectiveness of most of these approaches for reducing substance use is unknown... Some prevention approaches are effective at delaying the initiation or reducing the frequency of tobacco, alcohol and marijuana use [but]...the magnitude of these effects are generally small... [and it] is not clear that preventing or reducing the use of gateway substances translates into a reduced use of cocaine or other illegal drugs (pp. 233-234).

These conclusions have been supported by a recent study of the likely impact of the most effective school-based prevention programmes, which concluded that they would have, at best, very modest effects in preventing cocaine use (Caulkins et al. 1999).

There is better evidence that some treatment programmes (e.g. opioid agonist maintenance treatment) can substantially reduce illicit opioid use and premature mortality from drug overdose ${ }^{8}$ among opioid-dependent persons (Warner-Smith et al. 2001). In the case of opioid-dependent persons, one could examine the effects that enrolling $10 \%$, $20 \%, 30 \%$, etc. of persons who were dependent on illicit opioids in opioid maintenance treatment would have on illicit opioid use, overdose deaths and disability produced by illicit opioid dependence. Similar estimates could be made of the expected reduction in HIV/AIDS among injecting drug users from the introduction of needle and syringe exchange and distribution programmes.

### 1.2 Data sources

To provide data on the prevalence and risks of illicit drug use, a series of extensive computer searches using databases listed below was conducted. The specific parameters of these searches are also listed.

## DATABASES SEARCHED

We carried out a citation search of Medline, Psychinfo and Web of Science, a search of reference lists of identified papers, including a literature search provided by English et al. (1995), which covered the literature published prior to 1993.

## SEARCH TERMS

1. Illicit drug, or substance use, or substance abuse, or drug use, or $d r u g$ abuse, or heroin, or opiates, or cocaine, or amphetamine—limited to human studies published in the English language.

## 2. Prevalence

3. Cohort, or case-control
4. Mortality
5. Morbidity
6. Suicide, or accidents, or HIV, or assault

Strategy: combine 1 and $2 ; 1$ and 3 and $4 ; 1$ and 3 and $5 ; 1$ and 3 and 6.

## 2. Risk factor exposure

### 2.1 Prevalence studies

Given the lack of reliable direct estimates of the health consequences of illicit drug use, it was necessary to make indirect estimates of burden. Hence, the first challenge in quantifying the burden of disease attributable to illicit drugs was to determine the prevalence of exposure to this risk factor. Illicit drug use differs from other risk factors in the GBD project in that one of its defining features, its illegality, makes it difficult to quantify. This presents two problems. First, illicit drug-using individuals are "hidden" and are thus difficult to identify. Second, even if all drug users can be located and interviewed, they may attempt to conceal their use of these drugs.

There are no well-tested and widely accepted "gold standard" methods for producing credible estimates of the number of people who make up the "hidden population" of such drug users (Hartnoll 1997). The preferred strategy is to look for convergence in estimates produced by a variety of different methods of estimation (EMCDDA 1997, 1999). These methods are of two broad types, direct and indirect. Direct estimation methods attempt to estimate the number of illicit drug users in representative samples of the population. Indirect estimation methods attempt to use information from known populations of illicit drug users (such as those who have died of opioid overdoses, and those who are in treatment or the criminal justice system) to estimate the size of the hidden population of illicit drug users.

A large number of studies purporting to be prevalence studies do not present credible prevalence data. Prevalence data reported in peerreviewed literature are scarce and often unrepresentative. In addition, the range of methodologies used makes comparisons between studies difficult. For this reason, other sources of data were sought to complement prevalence estimates reported in the peer-reviewed literature.

### 2.2 Prevalence of problematic illicit drug use

For the purposes of estimating global mortality, data collated by the UNDCP (2000) provides a convenient and comprehensive tabulation of the most recent international prevalence data. The aggregated prevalence data for subregions are displayed in Table 13.1. The principal advantage of using UNDCP data is that it provides a readily accessible set of estimates for the majority of countries in the world. The quality of the data collected and reported by the UNDCP varies across countries and regions from high quality national survey data to key informant and indicator data of uncertain validity.

In some cases, prevalence data provided by the UNDCP were supplemented by data from other agencies, such as the EMCDDA, and the Asian Harm Reduction Network (AHRN). In regions where these addi-
tional data were available, they were used in indirect estimation methods as an additional source of prevalence estimates, thus meaning that these regions had additional estimates of causes of mortality.

UNDCP 2000 DATA ON THE PREVALENCE OF 12-MONTH USE AMONG PERSONS AGED >15 YEARS

Table 13.1 shows the population estimates of each of the 14 subregions, as well as the UNDCP-derived prevalence estimates of problematic use of the three substances considered in current estimates. It can be seen that problematic cocaine use is largely restricted to the Americas, the European Union and the developed countries of Oceania. Conversely, opioid abuse appears to be restricted to Asia and eastern and central Europe, as well as the developed countries of Oceania, the European Union and North America. Patterns of use in developing countries appear to reflect proximity to production areas and trafficking routes that supply the drug markets of developed "consumer" countries.

A challenge when estimating the prevalence of illicit drug consumption is to avoid double counting individuals who use more than one substance. There is strong evidence (principally from developed countries) that few drug users use one drug exclusively (Darke and Hall

Table 13.I Prevalence (\%) of problematic illicit drug use in the past I2 months among persons aged >15 years, by subregion (UNDCP-derived estimates of prevalence) ${ }^{\text {a }}$

| Subregion | Population $>15$ years (000s) | Opioids | Cocaine | Amphetamine |
| :--- | :---: | :---: | :---: | :---: |
| AFR-D | 159577 | 0.09 | 0.26 | 0.31 |
| AFR-E | 190152 | 0.01 | 0.05 | 0.12 |
| AMR-A | 255420 | 0.13 | 0.78 | 0.20 |
| AMR-B | 297625 | 0.03 | 0.24 | 0.20 |
| AMR-D | 44658 | 0.07 | 0.43 | 0.11 |
| EMR B | 86853 | 0.55 | - | 0.02 |
| EMR-D | 204039 | 0.41 | - | 0.14 |
| EUR-A | 339446 | 0.11 | 0.18 | 0.24 |
| EUR-B | 161213 | 0.09 | 0.01 | 0.10 |
| EUR-C | 152432 | 0.19 | 0.01 | 0.04 |
| SEAR-B | 206870 | 0.04 | - | 0.10 |
| SEAR-D | 818521 | 0.15 | - | - |
| WPR-A | 129888 | 0.04 | 0.28 | 0.22 |
| WPR-B | 1131503 | 0.02 | - | 0.34 |

[^51]a Some estimates for subregions are based on data from a small number of countries in the subregion.

1995; Topp et al. 1999). Rather, most users nominate a drug of choice but regularly use a wide range of substances (Darke and Hall 1995; Klee et al. 1990). Thus combining estimates of the size of each population will overestimate the size of the drug-using population. Given that many opioid and stimulant users are polydrug users, and that these drugs are the most harmful illicit drugs, the simplest approach to this problem may be to use the prevalence of regular users of opioids and/or stimulants in each country as the prevalence estimate for problem illicit drug use.

In order to address this issue, a range of three prevalence estimates was derived from the above UNDCP data:

- a low estimate, which assumed that $50 \%$ of each of the prevalence estimates was unique and therefore additive;
- a medium estimate which assumed that $75 \%$ of each of the prevalence estimates was unique and therefore additive; and
- a high estimate, which assumed that the prevalence estimates were completely additive (i.e. that those who used opioids were a separate group from those who used cocaine and amphetamines).

In order to estimate the proportion of persons who had used these drugs problematically in the past year, data from the 1997 Australian National Survey of Mental Health and Well-Being were used. This survey was a structured diagnostic interview of a representative sample of Australian adults aged $\geq 18$ years (Hall et al. 1999d; Henderson et al. 2000). It assessed persons who had used opioids and stimulant drugs for symptoms of DSM-IV ${ }^{9}$ defined abuse and dependence. Of those who had reported using these drugs within the past year, $28 \%$ met criteria for DSM-IV abuse or dependence.

In the current calculations, therefore, it was assumed that $28 \%$ of those who had used these drugs within the past year were problematic users of these drugs.

It must be noted that prevalence estimates were not available for all countries in all subregions. In making estimates from UNDCP data, where countries had no reported prevalence estimates, subregional estimates of prevalence were used by deriving a weighted average prevalence rate from the data that were available from countries in the subregion. This weighted average rate was used in making subregional estimates. Some subregions therefore had estimates based upon only some countries within the subregion, which may make these estimates less representative:

- AFR-D: prevalence estimates based upon estimates provided for Cameroon, Chad, Ghana, Mauritius, Nigeria, Senegal and Sierra Leone;
- AFR-E: prevalence estimates based upon estimates provided for Ethiopia, Kenya, Namibia, Rwanda, South Africa, Uganda, the United Republic of Tanzania and Zimbabwe;
- AMR-D: prevalence estimates based upon estimates for Bolivia, Ecuador and Peru;
- EMR-D: prevalence estimates based upon estimates provided for Egypt, Morocco and Pakistan; and
- WPR-B: prevalence estimates based upon estimates provided for China, the Lao People's Democratic Republic, Malaysia, the Philippines, the Republic of Korea and Viet Nam.


## EMCDDA ESTIMATES OF "PROBLEM DRUG USERS"

These estimates were used to derive alternative estimates of prevalence in countries from EUR-A: Austria, Belgium, Croatia, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Norway, Spain, Sweden and the United Kingdom.

A weighted prevalence rate was derived from these estimates for the whole of EUR-A. The EMCDDA produced low and high estimates of "problem drug users" for countries in the European Union using a variety of estimation methods including capture-recapture and backprojection methods. Both these estimates were used to make lower and upper estimates of prevalence using these data. A median estimate was also calculated when making median estimates for each of the four major causes of mortality, and for all-cause mortality.

## AHRN estimates of IDU in the Asian region

Numbers from the AHRN were used to make estimates of the prevalence of IDU in these countries (IDU in this case was taken to represent the prevalence of "problem drug use" used in the other two estimates). Weighted prevalence estimates of IDU prevalence were only made in the subregion they were classified under if two or more countries reported (see www.ahrn.net). The countries were as follows:

- SEAR-B: Indonesia (no estimate made);
- SEAR-D: Myanmar (no estimate made);
- WPR-A: Japan, Singapore;
- WPR-B: Cambodia, China, Malaysia, Mongolia, the Philippines, the Republic of Korea and Viet Nam.

Table 13.2 shows the prevalence estimates produced from the EMCDDA and AHRN sources. Comparison with estimates in Table 13.1 reveals that the estimates are fairly similar.

In the current chapter, we have used all available estimates of prevalence to make a range of estimates of each cause of mortality. Hence, for

Table I3.2 Alternative estimates of prevalence of problematic drug use in three subregions

| Subregion | EMCDDA low <br> estimate (\%) | EMCDDA high <br> estimate (\%) | AHRN <br> $(\%)$ |
| :--- | :---: | :---: | :---: |
| EUR-A | 0.2 | 0.4 | NA |
| WPR-A | NA | NA | 0.3 |
| WPR-B | NA | NA | 0.01 |

NA Not applicable.
example, EUR-A has two sources of prevalence estimates: EMCDDA estimates (low and high) and UNDCP estimates (low, median and high). This approach was taken so as to make estimates based on as much of the available data as possible.

## 3. Health outcomes

### 3.1 Premature mortality

The major causes of premature death among illicit drug users are relatively directly related to their patterns of drug use. Evidence for these causes comes from studies of premature mortality among cohorts of illicit drug users who have been treated in Europe and North America. (It must be remembered that there is a range of issues surrounding the use of such cohort studies in deriving global estimates of mortality rates, which are discussed in section 6.2.)

Notwithstanding these issues, illicit drug users have elevated rates of four main causes of premature death by comparison with age peers who do not use illicit drugs, namely, drug overdose, HIV/AIDS, suicide and trauma.

## Overdose

"Overdose" refers to two ICD-10 classifications of cause of death: (i) accidental or intentional fatal poisoning caused by specific drugs, and (ii) poisoning deaths occurring among dependent drug users that are attributed to drug dependence. Despite the conceptual simplicity of drug overdose deaths it has been difficult to quantify the number of such deaths with any precision, even in developed countries, for reasons that are discussed below.

## HIV/AIDS

The connection between illicit drug use and HIV/AIDS largely arises from injection as the route of drug administration via drug users sharing
contaminated injecting equipment. This means that it is necessary to establish the prevalence of injecting drug use, rather than harmful drug use per se, in order to calculate the proportion of incident HIV cases that can be attributed to harmful drug use. This can be accomplished by extrapolating from data on the prevalence of injecting drug use among persons who are illicit drug users as indicated in studies in the peerreviewed literature. It can also be estimated by the proportion of HIV/AIDS cases that are attributed to IDU in each country. One issue that exists concerns a lack of data from some countries on the prevalence of AIDS cases that are attributable to IDU. In the current study, we have only used UNAIDS estimates of mortality attributable to injecting drug use.

## Suicide

Suicide is a cause of death in the ICD-10 but, as with overdose deaths, the reliability with which this cause of death is diagnosed may vary between countries depending on a number of variables. Cultural variations in attitudes towards suicide may influence coroners' and mortality registrars' willingness to classify a death as intentional (Domino and Lenaars 1989; Domino and Takahashi 1991).

## Trauma

Trauma includes homicide, motor vehicle accidents and other forms of accidental death. It is likely that this will be underestimated since few cohort studies report mortality rates from all forms of trauma and it is difficult to calculate attributable fractions for these causes because many trauma deaths in drug users may not be recognized as being drug-related.

## ALL-CAUSE MORTALITY

Several studies have calculated standardized mortality ratios (SMRs) for problem drug users. These studies indicated that problem drug users have substantially increased mortality rates, with typical estimates suggesting that they are approximately 13 times more likely to die than their peers (English et al. 1995; Hulse et al. 1999).

### 3.2 Likely sources of morbidity attributable to illicit drug use

Premature death is the most serious adverse health outcome experienced by problem drug users; it is also the best-studied health outcome in this population. Nevertheless, the contribution that illicit drug use makes to the burden of disease is not exhausted by premature death.

First, each of the major causes of premature mortality probably causes substantial morbidity. Second, drug dependence, which is highly prevalent among problem drug users, is also a cause of disability. Third, evidence suggests that the prevalence of hepatitis $B$ and hepatitis $C$ viruses (HBV and HCV) is high among injecting drug users (Alter et al. 1990;

Anderson et al. 1994; Levine et al. 1994; MacDonald et al. 1996, 2000). Both of these viruses are associated with substantial morbidity and premature death due to the sequelae of chronic infection (Alter et al. 1990; MacDonald et al. 1996, 2000).

However, there is little good evidence that allows quantification of the morbidity related to the use of illicit drugs, and it is not possible to make estimates of the burden of disease caused by morbidity resulting from illicit drug use. This means that current estimates of the burden of disease attributable to illicit drug use significantly underestimate the total burden of illicit drug use. An American analysis of the economic costs of drug use (National Institute on Drug Abuse 1992) revealed that problematic illicit drug use cost the United States an estimated US\$98 billion in 1992; of this, the figure for the impact of premature deaths was US\$ 14.6 billion. Health care expenditure cost an estimated US\$ 9.9 billion, and impaired productivity resulting from drug-related morbidity cost an additional US\$ 14.2 billion-clearly, costs from mortality attributable to drug use in the United States were a fraction of those attributable to morbidity. This means that our estimates are likely to underestimate the global burden of disease attributable to illicit drugs. Future estimates of the global burden of disease would be substantially improved by research into the total morbidity that is attributable to illicit drug use.

Outlined below are some of the major outcomes of illicit drug use that may be significant sources of morbidity. Future research is required to better document the nature and extent of morbidity attributable to the use of illicit drugs. In the absence of such data, we assumed that for each of the four causes included, the population attributable fractions for mortality could also be used to estimate the proportion of morbidity explained by that cause. Equivalently, we assumed that if illicit drug use causes a certain proportion of AIDS mortality, it does so by increasing its incidence, also increasing AIDS-related morbidity by a similar proportion. While uncertain, this assumption is closer to underlying mechanisms than assuming that illicit drug use contributed to none of AIDS-related morbidity. This approach cannot be used to estimate morbidity due to conditions that do not cause deaths, such as, neuropsychological impairment.

## NON-FATAL OVERDOSE

The prevalence of non-fatal overdose is not well studied in problem drug users apart from among opioid users in some developed societies, where non-fatal overdose is a common event (Darke et al. 1996; Gossop et al. 1996; Warner-Smith et al. 2001). An unknown proportion of these cases requires acute medical treatment and hospitalization and some of these may develop persistent medical sequelae as a result of non-fatal overdoses, such as cognitive impairment (Darke et al. 2000) and other medical problems (Warner-Smith et al. 2001). There are no good esti-
mates of the prevalence of these outcomes that would permit an estimate to be made of their contribution to the burden of disease. It should be a research priority to obtain better estimates of the prevalence of these forms of morbidity in future studies of illicit drug users so that the contribution that illicit drug use makes to the burden of disease can be better understood.

## AIDS

In developed societies, the widespread availability of anti-retroviral drugs has extended the life expectancy of persons living with HIV/AIDS (Donoghoe and Wodak 1998), with the result that HIV/AIDS may become a chronic condition. However, in these countries we rarely have data on the proportion of treated cases who acquired their infections as a result of IDU. We have assumed that the proportion of treated HIV/AIDS cases that are attributable to IDU is the same as the proportion of AIDS-related deaths that are attributed to IDU.

## Hepatitis B and C

Many injecting drug users in developed countries are infected with HCV. In Australian needle and syringe attendees, for example, the prevalence of HCV infection is estimated at between $50 \%$ and $60 \%$ (National Centre in HIV Epidemiology and Clinical Research 1998). Chronic infection has been estimated to occur in $75 \%$ of infections, and $3-11 \%$ of chronic HCV carriers will develop liver cirrhosis within 20 years. Given the large number of injecting drug users infected with HCV , and the more protracted complications arising from this infection, the net health and economic cost of HCV transmitted by injecting drug use may be as high as, or considerably higher than, those of HIV. Data on the prevalence of this infection among injecting drug users in developed countries is limited; it is non-existent in many developing countries.

Similarly, the prevalence of HBV has been documented as quite high among injecting drug users in developed countries. There is, however, a lack of good evidence on (i) the prevalence of HBV among illicit drug users; (ii) the risk of premature mortality caused by this disease; and (iii) the extent of morbidity that it causes. HBV has therefore not been included in current estimates for these reasons. It would be desirable in future to include estimates of morbidity and disability that HBV and HCV cause among illicit drug users.

## Attempted suicide

Recent studies in Norway (Rossow and Lauritzen 1999) and Australia (Darke and Ross 2000) have found high rates of self-reported suicide attempts among problematic opioid users (Darke and Ross 2000; Rossow and Lauritzen 1999). Survivors of such suicide attempts may require psychiatric and medical treatment and some suffer from medical sequelae. As with non-fatal overdose, there are no data that permit the
morbidity attributable to this cause among problem drug users, but estimates could be obtained by applying the same attributable fraction for suicide deaths to morbidity caused by attempted suicides.

## TRAUMA

In developed societies there are approximately 20 cases of severe injury for every death caused by a motor vehicle accident (MVA) (English et al. 1995). We could assume the same is true for problem drug users if we had a credible estimate of the proportion of motor vehicle fatalities that were attributable to problem drug use. In the absence of this data we estimated the proportion of MVA morbidity attributable to illicit drugs by applying the same attributable fraction for MVA deaths to morbidity caused by motor vehicle accidents.

## Psychiatric disorder

Studies of treated populations of opioid-dependent persons have found a high prevalence of major depression and anxiety disorders (Darke and Ross 1997). It is difficult to sort out cause and effect from these crosssectional data so it is unclear in what proportion of these cases psychiatric disorders preceded and contributed to the development of problem drug use or vice versa. Nor is it clear to what extent pre-existing psychiatric disorders have been exacerbated by problem illicit drug use or vice versa. It is accordingly difficult to estimate what proportion of these disorders are attributable to problem illicit drug use. For these reasons such estimates have not been included in this chapter. Better understanding of the causal relationships between the two is a priority for future research.

### 3.3 Causality

The main evidence for believing that illicit drug use is a cause of premature death, morbidity and disability comes from cohort studies and cross-sectional studies of illicit drug users.

## Mortality

The cohort studies have identified a number of causes of mortality that are more prevalent among problem illicit drug users than their peers, indicating an association between harmful illicit drug use and these causes of mortality. They have rarely been well controlled for potential confounders, such as social disadvantage, which is common among illicit drug users. English et al. (1995) have argued that the mortality excess among illicit drug users is too large to be wholly accounted for by social disadvantage. Moreover, there are good reasons for believing that the relationship is causal in the case of deaths caused by overdose and bloodborne virus infection. The major illicit drugs are known to have adverse effects in overdose that can be fatal. Opioids, for example, produce respiratory depression that can cause death, and this is especially likely to
occur if opioids are used in combination with other central nervous system depressant drugs such as alcohol and benzodiazepines (Darke and Zador 1996; Warner-Smith et al. 2001). Stimulant drugs, such as cocaine and amphetamines, can cause fatal cardiac arrhythmias and strokes (Goldfrank and Hoffman 1993; Platt 1997), which are very rare causes of death in young adults who do not use these drugs. Similarly, the viruses that cause HIV/AIDS, HBV and HCV infections are efficiently spread by contaminated blood in shared injection equipment (Donoghoe and Wodak 1998; MacDonald et al. 1996).

The case for illicit drug use being a contributory cause of suicide is less direct. Depression is a risk factor for suicide and it occurs at higher rates among illicit drug users. Intoxicating drugs like alcohol and opioids, and dependence on these drugs, have been shown in case-control and prospective studies to be risk factors for suicide (Beautrais et al. 1998, 1999). Opioid-dependent persons in treatment report very high rates of attempted suicide (Darke and Ross 2000).

The case for a causal connection between illicit drug use and trauma deaths is less direct still. Driving while intoxicated by alcohol is a well-known risk factor for fatal motor vehicle crashes (English et al. 1995) and the heavy use of alcohol is common among illicit drug users (Darke and Hall 1995; Darke and Ross 1997; Gossop et al. 1998). Opioids are also intoxicating substances that adversely affect driving, although they are much less commonly found in persons killed in fatal car crashes.

## Morbidity

The case for a causal connection between illicit drug use and morbidity caused by drug overdose, HIV/AIDS and HCV, suicide and trauma are the same as for mortality. To these must be added psychiatric disorders and drug dependence. By definition, drug dependence is caused by regular illicit drug use and most regular illicit drug users are dependent on one or more of the drugs that they regularly use.

The causal relationship between psychiatric disorder and illicit drug use is less clear. The two are associated in the general population and this is not attributable to confounding by social and demographic variables (Degenhardt et al. 2001). The direction of the causal relationship is less certain. Conduct disorders, depression and anxiety disorders that develop in early adolescence may predispose young adults to become dependent on illicit drugs. These disorders may also arise as a result of the adverse effects that illicit drug dependence has on the lives of those affected by it, or the rigours of regular illicit drug use may prolong preexisting depressive and anxiety disorders that may have resolved in its absence.

## 4. Risk factor-disease relationship

## $4.1 \quad$ Outcome studies

INCLUSION CRITERIA
The following inclusion criteria were used:

- cohort studies on the use of opioids, cocaine or amphetamines and mortality;
- studies in which SMRs were reported. SMRs are the ratio of observed numbers of deaths in the cohort to the expected number of deaths in people of the same age and sex distribution in the general population; and
- studies in which crude mortality rates (CMRs) could be derived from the available data in the article.


## Exclusion criteria

The following exclusion criteria were used:

- multiple reports of same data set;
- subsets of a cohort; and
- reviews, commentaries, letters and abstracts.

Table 13.3 shows those studies that were not included in the present analyses. CMRs were derived from data on the number of deaths, period of follow-up and number of participants. Where person-years were not calculated by the authors, persons lost to follow-up were assumed to be alive at the end of study period and included in our calculation of personyears observation (to maintain consistency with studies that did not report numbers lost to follow-up). Following previous research (Hulse et al. 1999) it was assumed that persons dying during the period of

Table 13.3 Outcome studies that were not included

| Reference | Reason for exclusion |
| :--- | :--- |
| Vaillant (I966) | Data are a subset of Vaillant (1973) |
| Watterson et al. (1975) | Series of cross-sections, not longitudinal |
| Thorley et al. (1977) | Data are a subset of Wille (198I) |
| Wiepert et al. (1978) | Poorly defined cohort |
| Ghodse et al. (1985) | Study of death register, not a predefined cohort |
| Selwyn et al. (1989) | No cohort defined |
| Frischer et al. (1993) | Retrospective |
| Fischer et al. (1999) | No mortality reported |

follow-up died in the middle of the period (when estimating the personyears at risk). CMRs are unadjusted, expressed as per cent mortality per annum.

### 4.2 Quantification of risk

In determining the risks associated with harmful illicit drug use, it is necessary to rely on the results of cohort studies that have conducted longterm follow-up of individuals identified as using illicit drugs. English et al. (1995) identified a total of 13 such studies investigating mortality associated with illicit opioid use up to 1993 (Barr et al. 1984; Bewley et al. 1968; Cherubin et al. 1972; Engstrom et al. 1991; Frischer et al. 1993; Ghodse et al. 1985; Haastrup and Jepson 1984; Hser et al. 1993; Joe et al. 1982; Perucci et al. 1991; Thorsen and Haarstrup 1975; Vaillant 1973). Through extensive literature searches we identified a further 16 studies that have been published since 1993, excluding studies which used previously published data (Capelhorn et al. 1996; Eskild et al. 1993; Friedman et al. 1996; Fugelstad et al. 1995, 1997; Galli and Musicco 1994; Goedert et al. 1995; Goldstein and Herrera 1995; Keenan et al. 1993; McAnulty et al. 1995; Oppenheimer et al. 1994; Orti et al. 1996; Robertson et al. 1994; van Haastrecht et al. 1996; Wahren et al. 1997; Zaccarelli et al. 1994). The studies summarized in Tables 13.4-13.8 were all studies identified that followed up cohorts of problem or injecting drug users.

The general limitations of cohort studies have been discussed elsewhere (Dart 1995; Feldman 1993; Freeman 1996). The particular limitations of the cohort studies that are most relevant to this project are, first, that these studies were conducted exclusively in developed countries (principally the United States with 11 studies, western Europe with 22 studies and the Western Pacific with two studies). Second, with one exception (McAnulty et al. 1995) these studies drew their samples from people receiving treatment for drug-related problems. Third, the majority of the studies have been done on opioid users, usually injectors. There is much less data on mortality among problem stimulant users. Finally, the majority of cohort studies were conducted in the pre-AIDS era. These limitations will be discussed in more detail later in this chapter.

In studies of all-cause mortality (Table 13.8), a total of 152432 subjects were included, which involved a total of 1035574 person-years of observation, during which time 11633 deaths were recorded. The weighted average all-cause mortality rate was $1.12 \%$ per annum. Pooled crude death rates from the specific causes of death identified by English et al. (1995) were calculated from data reported in these studies (see Tables 13.4-13.7).
Table I3.4 Included outcome studies that examined rates of mortality due to AIDS among problematic drug users

| Site | Reference | $n$ | Population studied | Follow-up (years) | Lost | Personyears | Drug used | Crude mortality rate per 1000 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Australia, Sydney | Capelhorn et al. (1996) | 296 | Methadone | 13 | 39 | 3484 | Heroin | 0 |
| Italy | Goedert et al. (1995) | 4962 | Treatment | 3.88 | - | 21130 | $\ldots$ | 0.71 |
| Italy, Milan | Galli and Musicco (1994) | 2432 | Treatment | 6.7 | - | 16415 | Methadone (94\%) | 0.88 |
| Italy, Rome | Davoli et al. (1997) | 3955 | Treatment | 4 | 198 | 15820 | IDU | 1.06 |
| Italy, Rome | Perucci et al. (1991) | 4200 | Methadone treatment | 8 | Nil | 33600 | Opioids | 0.05 |
| Italy, Rome | Zaccarelli et al. (1994) | 2431 | Treatment | 3.2 | - | 7872 | IDU | 1.13 |
| Netherlands, Amsterdam | Mientjes et al. (1992) | 390 | Methadone treatment | 2.2 | - | 810 | Opioids | 0.37 |
| Netherlands, Amsterdam | van Haastrecht et al. (1996) | 632 | HIV- methadone and STD clinic patients | 4.4 | 18 | 2781 | $\ldots$ | 0.43 |
| New Zealand, Wellington | Dukes et al. (1992) | 997 | Treatment | 9.1 | - | 9073 | Injecting opioid users | Nil |
| Norway, Oslo | Eskild et al. (1993) | 1009 | HIV test centre clients | 3.67 | - | 3706 | IDU | 0.11 |
| Spain, Catalonia | Orti et al. (1996) | 15711 | Hospital emergency departments and treatment | 2.8 | - | 43717 | Opioids | 1.08 |
| Spain, Catalonia | Sanchez-Carbonell and Seus (2000) | 135 | Treatment | 10.5 | - | 1418 | Heroin | 1.48 |
| Sweden | Gronbladh et al. (1990) | 368 | Methadone and untreated | 5-11 | - | 3283 | Heroin | 0 |
| Sweden, Gothenburg | Benson and Holmberg (1984) | 618 | Drug using conscripts, rehab. \& psych. patients, drug using welfare recipients | 10 | - | 5789 | Cannabis, solvents, LSD, stimulants (opioids rare) | 0 |
| Sweden, Lund | Tunving (1988) | 524 | Persons in treatment for opioid, amphetamine, and both opioid and amphetamine use | 10 | 0 | 5240 | Opioids, amphetamines | 0 |

Table I3.4 Included outcome studies that examined rates of mortality due to AIDS among problematic drug users (continued)

| Site | Reference | n | Population studied | Follow-up (years) | Lost | Personyears | Drug used | Crude mortality rate per 1000 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sweden, Stockholm | Fugelstad et al. (1995) | 472 | HIV+ | 3.8 | - | 1793 | $\ldots$ | 0.39 |
| Sweden, Stockholm | Fugelstad et al. (1997) | 1640 | Drug-related hospitalization | 8 | Nil | 13120 | $14 \%$ heroin; $35 \%$ amphetamine; $23 \%$ polydrug | 0.14 |
| United Kingdom, England, London | Oppenheimer et al. (1994) | 128 | Treatment | 22 | 7 | 2816 | Heroin | Nil |
| United Kingdom, Scotland, Edinburgh | Bucknall and Robertson (1986) | 184 | Heroin users attending a general practice | 4 | 4 | 720 | Heroin | 0 |
| United Kingdom, Scotland, Edinburgh | Robertson et al. (1994) | 203 | GP | 10 | 17 | 2030 | IDU | 0.79 |
| USA, CA | Hser et al. (1993) | 581 | Males in compulsory treatment | 24 | 35 | 13064 | "Narcotics" | 0 |
| USA, CT, New Haven | Musto and Ramos (1981) | 91 | Under treatment for drug abuse when recruited | 52 | - | 4732 | Morphine | 0 |
| USA, New York | Friedman et al. (1996) | 858 | Drug and alcohol dependants on welfare | 8 | - | 6864 | Drugs and alcohol | 1.23 |
| USA, OR, Portland | McAnulty et al. (1995) | 1769 | Not in treatment | 1.78 | - | 3149 | IDU | 0 |
| USA, PA, Philadelphia | Zanis and Woody (1998) | 507 | Methadone | 1 | 5 | 507 | Heroin | 0 |
| Netherlands, Amsterdam; USA, MD, Baltimore | van Ameijden et al. (1999) | 2809 | Treatment shelters and community agencies | 6-9 | - | 15107 | IDU | 0 |

[^52]
### 4.3 OvERVIEW OF METHODS OF ESTIMATING THE MORTALITY burden of illicit drugs

Methods for estimating mortality attributable to harmful illicit drug use can be "direct" or "indirect". Direct methods count the number of deaths attributed to illicit drug use by applying attributable fractions to ICD-classified causes of death in national mortality registers. Indirect methods involve estimating mortality by multiplying measures of mortality risk (e.g. relative risk) by the prevalence of the risk factor in the population.

## DIRECT METHODS

The first method, which requires the greatest amount of data, uses the attributable fraction of mortality attributed to harmful illicit drug use calculated for a population for which direct measures of specific cause mortality data are available. This attributable fraction is then used to extrapolate the mortality attributable to harmful illicit drug use in another population.

This method has the advantage of excluding deaths in those exposed that are not due to the risk factor. The source of mortality data to use with this method is the All-Cause Mortality Database compiled by WHO. Attributable fractions for illicit drug use (which have been calculated in countries where direct estimates have been made) can be applied to these data.

In some countries direct measures of mortality are available from mortality registers. This is straightforward, in principle, for deaths caused by overdose, which has an attributable fraction of 1 . Aside from individual country mortality registers, other sources of directly measured mortality data include HIV/AIDS surveillance data available from agencies such as UNAIDS and the United States Census Bureau.

The difficulties involved in applying this method are exemplified by the case of "overdose" deaths. This is the only cause of death that is wholly attributable to harmful illicit drug use so all mortality due to this cause must be the result of the risk factor. It is the cause of death that should be the most easily quantified. However, the great many difficulties inherent in assigning any particular case to this cause of death have been well documented (Advisory Council on the Misuse of Drugs 2000; Danish National Board of Health 1997; WHO 1998).

In most United Nations Member States, cause of death is classified according to ICD-10 codes, which specify whether the cause of death was intentional poisoning (suicide), unintentional poisoning or dependence. Despite the existence of ICD-10 criteria for classification of cause of death, countries differ in the way that deaths are registered and causes of death are classified (Danish National Board of Health 1997; WHO 1998). For example there is one European country in which: ". . . it is well known that about $90 \%$ of drug-related deaths are coded
with the code for unknown cause of death" (p. 51) (Sanchez-Carbonell and Seus 2000).

A recent report by the Home Office has been critical of the system for recording drug-related deaths in the United Kingdom (Advisory Council on the Misuse of Drugs 2000). It noted that deaths may not be classified as drug-related if they are not referred to the coroner (as may happen when a certifying doctor is unaware that the deceased was a drug user) or the death is due to an indirect effect of harmful drug use, such as a viral infection. There also appears to be a great deal of variation between individual coroners in their preparedness to record deaths as drug related. The report notes that: "there are coroners working in areas of known high drug prevalence who never certify a death as related to drug misuse" (p. 80).

Other sources of variation identified in the British report were that neither post-mortem nor toxicological analysis are formally required for suspected drug-related deaths; that the verdicts available to the coroner are not mutually exclusive; that coroners do not have the necessary skills to distinguish between the verdicts available to them, most notably "dependence on drugs" and "non-dependent abuse of drugs"; and that there is no requirement of the coroner to identify the drugs involved (Advisory Council on the Misuse of Drugs 2000).

There are also variations between countries in how much information is gathered about the circumstances or cause of death (Danish National Board of Health 1997; WHO 1998). In Australia, for example, autopsy is routinely conducted on all suspected overdose deaths, making forensic and toxicological data the basis for the classification of cause of death. This, however, is a far from universal practice. In the United States, only $20 \%$ of drug-related deaths are subject to autopsy (WHO 1998). Similarly, the immediate cause of death is recorded in death registers but contributing factors may or may not (Danish National Board of Health 1997; WHO 1998). This can cause large differences in rates of drugrelated deaths based on death register data.

For causes other than overdose, where the attributable fraction is less than 1, the difficulties involved in attributing a death to illicit drug use are compounded. In addition to the caveats discussed above, the simple fact that there is a complete absence of such data in the majority of countries in the world necessitates the use of indirect methods to estimate mortality attributable to harmful drug use.

## INDIRECT METHODS

Indirect methods of estimating mortality can be used when directly recorded data are unavailable or unreliable. The estimates provided by these methods can be validated against direct methods in countries where reliable mortality data are available. For the vast majority of countries in the world, indirect methods provide the only indicator of the extent
of the health consequences of harmful illicit drug use, because of the absence of epidemiological data on drug-related mortality. Three indirect methods can be used to estimate the burden of mortality attributable to illicit drugs.

The simplest method is to multiply mortality rates in cohort studies by the estimated prevalence of problem illicit drug use in the country. This provides an estimate of deaths caused by illicit drugs for each of the causes of death that we have considered.

## KEY INFORMANT DATA

A final source of data is that which researchers in the drug and alcohol field can provide. The WHO Management of Dependence Project surveyed drug researchers in Member States and asked them to provide estimates of the prevalence of harmful drug use and resultant mortality in their country, using the best available data. The sources range in quality from large-scale population surveys to educated guesses based on clinical experience, but such consultations provide an independent source of estimates against which to check the sources outlined above.

Only 15 responses were returned and in most cases responses either reported on published data or data whose validity was difficult to evaluate. This source has not been included formally in current estimates. Future attempts to estimate the contribution that illicit drug use makes to global burden of disease may include such data.

Key informants may also be of use to judge the accuracy and validity of estimates of mortality attributable to the different causes of death. Such key informants are extremely invaluable and note has been made in the text of instances in which key informants reported that our estimates were likely to be underestimates.

### 4.4 Methods used for each cause of mortality <br> AIDS

UNAIDS estimates of death related to IDU were used. No upper and lower estimates were obtained.

## DRUG overdose

It should be noted that rates derived from research on opioid overdoses were included in these calculations and separate estimates of the number of persons dying from stimulant-related overdoses have not been made. There is a lack of good data on rates and/or risk of dying from stimulant-related overdoses. However, it is likely that in countries which have a higher prevalence of cocaine use, cocaine-related overdoses may account for a considerable proportion of all fatal drug overdoses. In the United States, for example, the Drug Abuse Warning Network indicated that in 1999, $28 \%$ of single-drug overdoses were due to co-
caine (National Institute on Drug Abuse, personal communication, 2002).

Our approach has been to assume that overdose rates derived from research on opioid overdose may be applicable to stimulant drugs. In estimates made of overdose deaths, rates of drug use (which include opioids, amphetamines and cocaine) were multiplied by the rate of opioid overdoses derived from cohort studies. Hence, it has been assumed that the same rate of overdose deaths applies to these other drugs as it does to opioid drug use. In the estimates derived from "allcause" rates in cohort studies, overdoses due to amphetamines and cocaine use will be included in this rate. Direct estimates made from the WHO Mortality Database included only deaths due to opioids so may be an underestimate.

In this chapter, data used to derive a number of estimates of the number of persons dying from overdoses were derived from the following sources.

## Direct estimates: attributable fractions combined with data from WHO all-cause mortality database

An attributable fraction was derived from EUR-A countries. This involved obtaining estimates of the total number of deaths coded in ICD as attributed to mental disorders and accidental poisoning due to opioids. The median attributable fraction of these countries was 0.1164 . This attributable fraction was applied to all other subregions to enable a direct estimate to be made.

Data were taken from the WHO all-cause mortality database on deaths attributed to mental disorders and accidental poisoning. Not all countries reported such data. In making estimates for subregions when some data were missing from countries in the subregion, an average overdose rate was calculated using the available data, and this rate was used to estimate the total number of trauma deaths in the subregion. It must be noted that many countries in some subregions did not report data in the WHO database, and some subregions had no countries that reported such data; the following subregions had estimates made from only few countries in the subregion, or had no estimates made:

- AFR-D: Mauritius;
- AFR-E: no estimate;
- AMR-B: Argentina, Belize, Brazil, Costa Rica and El Salvador;
- AMR-D: no estimate;
- EMR-B: Kuwait;
- EMR-D: no estimate;
- EUR-C: Armenia, Azerbaijan, Bulgaria, Kyrgyzstan, Poland, Romania and Slovakia;
- SEAR-B: Thailand;
- SEAR-D: the Democratic People's Republic of Korea;
- WPR-B: no estimate.


## Indirect estimates: cohort-derived mortality rate

A weighted average mortality rate was calculated from cohort studies (see Table 13.5 for included studies). The average was $0.43 \%$ per annum (to 2 decimal places). The upper and lower $95 \%$ confidence intervals of this rate were $0.25 \%$ per annum and $0.64 \%$ per annum.

## Suicide

Suicide was considered as a cause of mortality among illicit opioid users only. Hence, when making indirect estimates of deaths, only rates of opioid use were considered in analyses. Median estimates are reported. Where only one source of data was available in estimating a cause of mortality, then the median of this was used, and the lowest and highest estimate used for the range. If more than one source of estimates was available for a subregion, then the lowest and highest median estimates were used as the range.

## Direct estimates: attributable fractions combined with data from WHO all-cause mortality database

An attributable fraction of 0.09 was used to calculate the proportion of all suicides that were among opioid users. This attributable fraction was derived from an Australian study reported by English et al. (1995). Data were taken from the WHO all-cause mortality database on the number of deaths due to suicide by country, for persons aged $>15$ years. The year for which data were available varied. For those countries that had more than one year of data, the most recent year's data were used.

Not all countries reported such data. In making estimates for subregions in which some data were missing from countries in the subregion, an average suicide rate was calculated using the available data. This rate was used to estimate the total number of suicide deaths in the subregion. It must be noted that many countries in some subregions did not report data in the WHO database, and in some subregions there were no countries that reported such data. The subregions in which estimates were made from only few countries, or which provided no estimates are as follows:

- AFR-D: Mauritius;
- AFR-E: no estimate;
- AMR-B: Argentina, Belize, Brazil and Costa Rica;
- AMR-D: no estimate;
Table 13.5 Included outcome studies that examined rates of mortality due to overdose among problematic drug users

| Site | Reference | $n$ | Population studied | Follow-up (years) | Lost | Personyears | Drug | Crude mortality rate per 1000 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Australia, Sydney | Capelhorn et al. (1996) | 296 | Methadone | 13 | 39 | 3484 | Heroin | 0.66 |
| Denmark, Copenhagen | Haastrup and Jepson (1984) | 300 | Treatment | 7 | 19 | 1967 | Morphine | 1.37 |
| Italy | Goedert et al. (1995) | 4962 | Treatment | 3.88 | - | 21130 | $\ldots$ | 0.30 |
| Italy, Milan | Galli and Musicco (1994) | 2432 | Treatment | 6.7 | - | 16415 | Methadone (94\%) | 0.92 |
| Italy, Rome | Davoli et al. (1997) | 3955 | Treatment | 4 | 198 | 15820 | IDU | 0.58 |
| Italy, Rome | Perucci et al. (1991) | 4200 | Methadone | 8 | Nil | 33600 | Opioids | 0.24 |
| Italy, Rome | Zaccarelli et al. (1994) | 2431 | Treatment | 3.2 | - | 7872 | IDU | 0.55 |
| Netherlands, Amsterdam | van Haastrecht et al. (1996) | 632 | HIV- methadone and STD clinic patients | 4.4 | 18 | 2781 | $\ldots$ | 0.56 |
| New Zealand, Wellington | Dukes et al. (1992) | 997 | Treatment | 9.1 | - | 9073 | Injecting opioid users | 0.25 |
| Norway, Oslo | Eskild et al. (1993) | 1009 | HIV test centre clients | 3.67 | - | 3706 | IDU | 1.56 |
| Spain, Catalonia | Orti et al. (1996) | 15711 | Hospital ER and treatment | 2.8 | - | 43717 | Opioids | 1.09 |
| Sweden | Gronbladh et al. (1990) | 368 | Methadone and untreated | 5-11 | - | 3283 | Heroin | 1.74 |
| Sweden, Gothenburg | Benson and Holmberg (1984) | 618 | Drug-using conscripts, rehab. \& psych. patients, drug-using welfare recipients | 10 | - | 5789 | Cannabis, solvents, LSD, stimulants (opioids rare) | $\begin{gathered} 0.05 \\ \text { (poisoning) } \end{gathered}$ |
| Sweden, Lund | Tunving (1988) | 524 | Treatment: opioid, amphetamine, and both opioid and amphetamine | 10 | 0 | 5240 | Opioids, amphetamines | 0.67 |


| Sweden, Stockholm | Engstrom et al. (1991) | 1630 | Drug-related hospitalization | 12 | - | 19560 | 41\% cocaine/ amphetamine; 12\% heroin; 16\% polydrug; 3I\% other | 0.16 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sweden, Stockholm | Fugelstad et al. (1995) | 472 | HIV+ | 3.8 | - | 1793 | ... | 2.29 |
| Sweden, Stockholm | Fugelstad et al. (1997) | 1640 | Drug-related hospitalization | 8 | Nil | 13120 | 14\% heroin; 35\% amphetamine; 23\% polydrug | 0.82 |
| Sweden, Stockholm | Wahren et al. (1997) | 1494 | Hospitalized for drug dependence | 22 | - | 32868 | 57\% stimulants; $39 \%$ opioids | 0.09 |
| United Kingdom | Ghodse et al. (1998) | 92802 | "Drug addicts" notified to Home Office | 27 | - | 687673 | 65\% opioids | 0.38 |
| United Kingdom, England, London | Oppenheimer et al. (1994) | 128 | Treatment | 22 | 7 | 2816 | Heroin | 0.64 |
| United Kingdom, England, London | Wille (1981) | 128 | Treatment ( $R \times$ heroin) | 10 | - | 1280 | Heroin | 0.86 |
| United Kingdom, Scotland, Edinburgh | Bucknall and Robertson (1986) | 184 | Heroin users attending a general practice | 4 | 4 | 720 | Heroin | 0.55 |
| United Kingdom, Scotland, Edinburgh | Robertson et al. (1994) | 203 | GP | 10 | 17 | 2030 | IDU | 0.74 |
| USA, CA | Hser et al. (1993) | 581 | Males in compulsory treatment | 24 | 35 | 13064 | "Narcotics" | 0.40 |
| USA, CT, New Haven | Musto and Ramos (1981) | 91 | Treatment | 52 | - | 4732 | Morphine | 0 |

Table 13.5 Included outcome studies that examined rates of mortality due to overdose among problematic drug users

| Site | Reference | $n$ | Population studied | Follow-up (years) | Lost | Personyears | Drug | Crude mortality rate per 1000 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| USA, NM, Albuquerque | Goldstein and Herrera (1995) | 1013 | Methadone | 22 | 243 | $\begin{gathered} 22286 \\ (16940 \\ \text { excl. lost) } \end{gathered}$ | $\ldots$ | 0.53 (0.70) |
| USA, New York | Concool et al. (1979) | 1156 | Treatment (84\% methadone) | 7 | 102 | 8092 | Heroin | 0.07 |
| USA, New York | Friedman et al. (1996) | 858 | Drug and alcohol dependents on welfare | 8 | - | 6864 | Drugs and alcohol | 0.12 |
| USA, New York | Vaillant (1973) | 100 | Treatment, male | 20 | 17 | 1660 | Narcotics | 0.35 |
| USA, OR, Portland | McAnulty et al. (1995) | 1769 | Not in treatment | 1.78 | - | 3149 | IDU | 0.41 |
| USA, PA, Philadelphia | Zanis and Woody (1998) | 507 | Methadone | 1 | 5 | 507 | Heroin | 1.18 |
| USA, 18 treatment agencies | Joe and Simpson (1987) | 697 | Treatment | 6 | 142 | 3330 | Opioids | 0.75 |
| USA, 34 treatment agencies | Joe et al. (1982) | 3324 | Treatment | 4 | - | 11710 | Opioids | 0.59 |
| Netherlands, Amsterdam; USA, MD, Baltimore | van Ameijden et al. (1999) | 2809 | Treatment shelters and community agencies | 6-9 | - | 15107 | IDU | 0.50 |

[^53]- EMR-B: Kuwait;
- EMR-D: no estimate;
- SEAR-B: Thailand;
- SEAR-D: the Democratic People's Republic of Korea;
- WPR-B: no estimate.


## Indirect estimates: cohort-derived crude mortality rates

The crude mortality rate due to suicide from cohort studies was also estimated (see Table 13.6 for included studies). The weighted average rate of death per annum due to suicide was $0.24 \%$ (shown here to 2 decimal places). In order to make a range of estimates around this average rate, the standard error of the rate was calculated and $95 \%$ confidence intervals constructed around the rate. These were used as the lower and upper ranges of the mortality rates due to suicide: these were $0.15 \%$ per annum and $0.33 \%$ per annum, respectively.

## Trauma

It must be noted that there are significant problems with estimates of rates/attributable fractions due to trauma, since cohort studies reported different sorts of trauma, and different numbers of causes. In the attributable fraction method of calculation, only road traffic accidents were used to calculate the number attributable to illicit drug use as it is unclear the extent to which homicides or other trauma deaths are due to illicit drug use.

Median estimates are reported. Where only one source of data was available in estimating a cause of mortality, then the median of this estimate was used, and the lowest and highest estimates were used for the range. If more than one source of estimates was available for a subregion, then the lowest and highest median estimates were used as the range.

## Direct estimates: attributable fraction combined with data from WHO all-cause mortality database

Data were taken from the WHO all-cause mortality database on the number of deaths due to motor vehicle or other road traffic accidents (ICD-9 codes E470-E474 and E479) by country. Not all countries reported such data. In making estimates for subregions when data were missing from some countries, an average trauma rate was calculated using the available data. This rate was used for estimating the total number of trauma deaths in the subregion. It must be noted that many countries in some subregions did not report data in the WHO database, and some subregions had no countries that reported such data. In the following subregions estimates were made from only few countries, or no estimates were made:
Table I3.6 Included outcome studies that examined rates of mortality due to suicide among problematic drug users

| Site | Reference | $n$ | Population studied | Follow-up (years) | Lost | Personyears | Drug | Crude mortality rate (per 1000) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Australia, Sydney | Capelhorn et al. (1996) | 296 | Methadone | 13 | 39 | 3484 | Heroin | 0.14 |
| Denmark, Copenhagen | Haastrup and Jepson (1984) | 300 | Treatment | 7 | 19 | 1967 | Morphine | 0.61 |
| Italy, Milan | Galli and Musicco (1994) | 2432 | Treatment | 6.7 | - | 16415 | Methadone (94\%) | 0.06 |
| Italy, Rome | Perucci et al. (1991) | 4200 | Methadone | 8 | Nil | 33600 | Opioids | 0.03 |
| Italy, Rome | Zaccarelli et al. (1994) | 2431 | Treatment | 3.2 | - | 7872 | IDU | 0.04 |
| Netherlands, Amsterdam | van Haastrecht et al. (1996) | 632 | HIV- methadone and STD clinic patients | 4.4 | 18 | 2781 | $\ldots$ | 0.36 |
| New Zealand, Wellington | Dukes et al. (1992) | 997 | Treatment | 9.1 | - | 9073 | Injecting opioid | 0.09 |
| Norway, Oslo | Eskild et al. (1993) | 1009 | HIV test centre clients | 3.67 | - | 3706 | IDU | 0.24 |
| Sweden | Gronbladh et al. (1990) | 368 | Methadone and untreated | 5-11 | - | 3283 | Heroin | 0.09 |
| Sweden, Gothenburg | Benson and Holmberg (1984) | 618 | Drug-using conscripts, rehab. \& psych. patients, drug-using welfare recipients | 10 | - | 5789 | Cannabis, solvents, LSD, stimulants (opioids rare) | 0.29 |
| Sweden, Lund | Tunving (1988) | 524 | Treatment: opioid, amphetamine, and both opioid and amphetamine | 10 | 0 | 5240 | Opioids, amphetamines | 0.33 |
| Sweden, Stockholm | Engstrom et al. (1991) | 1630 | Drug-related hospitalization | 12 | - | 19560 | 41\% cocaine/ amphetamine; 12\% heroin; 16\% polydrug; 3I\% other | 0.79 |


| Sweden, Stockholm | Fugelstad et al. (1995) | 472 | HIV+ | 3.8 | - | 1793 | ... | 0.50 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sweden, Stockholm | Fugelstad et al. (1997) | 1640 | Drug-related hospitalization | 8 | Nil | 13120 | 14\% heroin; 35\% amphetamine; 23\% polydrug | 0.22 |
| Sweden, Stockholm | Wahren et al. (1997) | 1494 | Hospitalized for drug dependence | 22 | - | 32868 | 57\% stimulants; 39\% opioids | 0.30 |
| United Kingdom, England, London | Oppenheimer et al. (1994) | 128 | Treatment | 22 | 7 | 2816 | Heroin | 0.07 |
| United Kingdom, England, London | Wille (1981) | 128 | Treatment ( Rx heroin) | 10 | - | 1280 | Heroin | 0.08 |
| United Kingdom, Scotland, Edinburgh | Bucknall and Robertson (1986) | 184 | Heroin users attending a general practice | 4 | 4 | 720 | Heroin | 0.14 |
| USA, CT, New Haven | Musto and Ramos (1981) | 91 | Treatment | 52 | - | 4732 | Morphine | 0.02 |
| USA, NM, Albuquerque | Goldstein and Herrera (1995) | 1013 | Methadone | 22 | 243 | $\begin{aligned} & 22286 \\ & \text { (16940 } \\ & \text { excl. lost) } \end{aligned}$ | ... | 0.05 (0.07) |
| USA, New York | Vaillant (1973) | 100 | Treatment, male | 20 | 17 | 1660 | Narcotics | 0.20 |
| USA, PA, Philadelphia | Zanis and Woody (1998) | 507 | Methadone | 1 | 5 | 507 | Heroin | 0 |
| Netherlands, Amsterdam; USA, MD, Baltimore | van Ameijden et al. (1999) | 2809 | Treatment shelters and community agencies | 6-9 | - | 15107 | IDU | 0.59 |
| - No data. |  |  |  |  |  |  |  |  |

- AFR-D: no estimate;
- AFR-E: no estimate;
- AMR-B: Argentina, Belize, Brazil, Costa Rica, El Salvador and Paraguay;
- AMR-D: no estimate;
- EMR-B: Kuwait;
- EMR-D: no estimate;
- SEAR-B: Thailand;
- SEAR-D: no estimate;
- WPR-B: the Philippines and the Republic of Korea.

The attributable fraction derived by Ridolfo and Stevenson (2001) of 0.015 was used to calculate direct estimates of trauma due to illicit drugs.

## Indirect estimates: cohort-derived crude mortality rates

Indirect estimates of the number of road traffic accident deaths due to illicit drug use were also made using pooled estimates of the rates of death due to trauma from cohort studies (Table 13.7). Rates of traumatic injury were also high in this group: the weighted average rate of death per annum due to trauma was $0.35 \%$. In order to make a range of estimates around this average rate, the standard error of the rate was calculated and $95 \%$ confidence intervals constructed around the rate. These were used as the lower and upper ranges of the mortality rates due to trauma: these were $0.23 \%$ per annum and $0.46 \%$ per annum, respectively (shown here only to 2 decimal places).

## ALL-CAUSE MORTALITY

Median estimates are reported. Where only one source of data was available in estimating a cause of mortality, then the median of this was used, and the lowest and highest estimate used for the range. If more than one source of estimates was available for a subregion, then the lowest and highest median estimates were used as the range.

## Direct estimates: attributable fractions combined with data from WHO all-cause mortality database

Data were taken from the WHO all-cause mortality database on the total number of deaths by country, for persons aged between 15 and 54 years. The year for which data were available varied so the data from the most recent year were used in those countries that had more than one year of data.

This age group (15-54 years) was chosen as the age group within which excess mortality rates would occur among problem illicit drug
Table I3.7 Included outcome studies that examined rates of mortality due to trauma among problematic drug users

| Site | Reference | $n$ | Population studied | Follow-up (years) | Lost | Personyears | Drug | Crude mortality rate (per 1000) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Australia, Sydney | Capelhorn et al. (1996) | 296 | Methadone | 13 | 39 | 3484 | Heroin | 0.23 |
| Italy | Goedert et al. (1995) | 4962 | Treatment | 3.88 | - | 21130 | $\ldots$ | 1.8 |
| Italy, Milan | Galli and Musicco (1994) | 2432 | Treatment | 6.7 | - | 16415 | Methadone (94\%) | 0.15 |
| Italy, Rome | Davoli et al. (1997) | 3955 | Treatment | 4 | 198 | 15820 | IDU | 0.25 |
| Italy, Rome | Perucci et al. (1991) | 4200 | Methadone | 8 | Nil | 33600 | Opioids | 0.15 |
| Italy, Rome | Zaccarelli et al. (1994) | 2431 | Treatment | 3.2 | - | 7872 | IDU | 0.20 |
| Netherlands, Amsterdam | van Haastrecht et al. (1996) | 632 | HIV- methadone and STD clinic patients | 4.4 | 18 | 2781 | $\ldots$ | 0.18 |
| New Zealand, Wellington | Dukes et al. (1992) | 997 | Treatment | 9.1 | - | 9073 | Injecting opioid users | 0.12 |
| Norway, Oslo | Eskild et al. (1993) | 1009 | HIV test centre clients | 3.67 | - | 3706 | IDU | 0.22 |
| Spain, Catalonia | Orti et al. (1996) | 15711 | Hospital ER and treatment | 2.8 | - | 43717 | Opioids | 0.37 |
| Sweden | Gronbladh et al. (1990) | 368 | Methadone and untreated | 05-11 | - | 3283 | Heroin | 0.06 |
| Sweden, Gothenburg | Benson and Holmberg (1984) | 618 | Drug-using conscripts, rehab. \& psych. patients, drug-using welfare recipients | 10 | - | 5789 | Cannabis, solvents, LSD, stimulants (opioids rare) | 0.07 |
| Sweden, Lund | Tunving (1988) | 524 | Treatment: opioid, amphetamine, and both opioid and amphetamine | 10 | 0 | 5240 | Opioids, amphetamines | 0.15 |

Table 13.7 Included outcome studies that examined rates of mortality due to trauma among problematic drug users (continued)

| Site | Reference | $n$ | Population studied | Follow-up (years) | Lost | Personyears | Drug | Crude mortality rate (per 1000) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sweden, Stockholm | Engstrom et al. (1991) | 1630 | Drug-related hospitalization | 12 | - | 19560 | 4I\% cocaine/ amphetamine; 12\% heroin; 16\% polydrug; 31\% other | 0.35 |
| Sweden, Stockholm | Fugelstad et al. (1995) | 472 | HIV+ | 3.8 | - | 1793 | $\ldots$ | 0.11 |
| Sweden, Stockholm | Fugelstad et al. (1997) | 1640 | Drug-related hospitalization | 8 | Nil | 13120 | 14\% heroin; $35 \%$ amphetamine; $23 \%$ polydrug | 0.21 |
| Sweden, Stockholm | Wahren et al. (1997) | 1494 | Hospitalized for drug dependence | 22 | - | 32868 | 57\% stimulants; 39\% opioids | 0.22 |
| United Kingdom, England, London | Oppenheimer et al. (1994) | 128 | Treatment | 22 | 7 | 2816 | Heroin | 0.14 |
| United Kingdom, England, London | Wille (198I) | 128 | Treatment ( $R \times$ heroin) | 10 | - | 1280 | Heroin | 0.16 |


| United Kingdom, Scotland, Edinburgh | Bucknall and Robertson (1986) | 184 | Heroin users attending a general practice | 4 | 4 | 720 | Heroin | 0.14 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| USA, CA | Hser et al. (1993) | 581 | Males in compulsory treatment | 24 | 35 | 13064 | "Narcotics" | 0.35 (incl. suicide) |
| USA, CT, New Haven | Musto and Ramos (1981) | 91 | Treatment | 52 | - | 4732 | Morphine | 0.08 |
| USA, NM, Albuquerque | Goldstein and Herrera (1995) | 1013 | Methadone | 22 | 243 | $\begin{gathered} 22286 \\ (16940 \\ \text { excl. lost) } \end{gathered}$ | $\ldots$ | 0.27 (0.35) |
| USA, New York | Vaillant (1973) | 100 | Treatment, male | 20 | 17 | 1660 | Narcotics | 0.1 |
| USA, New York | Concool et al. (1979) | 1156 | Treatment (84\% methadone) | 7 | 102 | 8092 | Heroin | 0.26 |
| USA, OR, Portland | McAnulty et al. (1995) | 1769 | Not in treatment | 1.78 | - | 3149 | IDU | 0.16 |
| USA, PA, Philadelphia | Zanis and Woody (1998) | 507 | Methadone | 1 | 5 | 507 | Heroin | 0 |
| USA, 18 treatment agencies | Joe and Simpson (1987) | 697 | Treatment | 6 | 142 | 3330 | Opioids | 0.45 |
| USA, 34 treatment agencies | Joe et al. (1982) | 3324 | Treatment | 4 | - | 11710 | Opioids | 0.38 |
| Netherlands, Amsterdam; USA, MD, Baltimore | van Ameijden et al. (1999) | 2809 | Treatment shelters and community agencies | 6-9 | - | 15107 | IDU | 0.25 |

[^54]users compared to non-users. After calculating mortality rates among this age group, an SMR of 13 was used to calculate the rate of all-cause mortality death among problematic illicit drug users. This was taken from previous studies estimating the excess rates of mortality in this group (English et al. 1995; Hulse et al. 1999). It was assumed for these calculations that the resulting rate of death applied to all illicit drug users. This will underestimate the mortality rate among the minority of illicit drug users who are older than 54 years.

Some countries did not have any death data included in the WHO database. For these countries no individual estimates were made. However, in making calculations for subregions, a weighted average of the all-cause mortality rates was calculated using the data from countries that were included. It was assumed that the countries for which no data were available had the average rate of the other countries in the subregion from which they came. Some subregions, however, had no countries which had appropriate estimates. Those subregions in which there were few or no countries for which estimates could be made were as follows:

- AFR-D: Mauritius;
- AFR-E: South Africa;
- AMR-B: Argentina, Belize, Brazil, Costa Rica, El Salvador and Paraguay;
- AMR-D: no estimate;
- EMR-B: Kuwait;
- EMR-D: no estimate;
- SEAR-B: Thailand;
- SEAR-D: no estimate;
- WPR-B: the Philippines and the Republic of Korea.


## Indirect estimates: cohort-derived crude mortality rate

Crude all-cause mortality rates were also derived from cohort studies included in this project (see Table 13.8). A weighted average all-cause mortality rate was calculated ( $1.12 \%$ per annum), with a $95 \%$ CI of the average rate estimated as between $0.78 \%$ per annum and $1.46 \%$ per annum.
Table I3.8 Included outcome studies that examined rates of all-cause mortality among problematic drug users

| Site | Reference | $n$ | Population studied | Follow-up (years) | Lost | Personyears | Drug | Standardized mortality ratio | No. of deaths | Crude mortality rate (per 1000) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Australia, Sydney | Capelhorn et al. (1996) | 296 | Methadone | 13 | 39 | 3484 | Heroin | ... | 42 | 1.21 |
| Denmark, Copenhagen | Haastrup and Jepson (1984) | 300 | Treatment | 7 | 19 | 1967 | Morphine | 20 | 47 | 2.40 |
| Denmark, Copenhagen | Haastrup and Jepson (1988) | 300 | First time entrants to treatment | 11 | 30 | 2970 | Opioids | $\ldots$ | 78 | 2.63 |
| Denmark, Copenhagen | Segest et al. (1990) | 169 | Methadone | 8 | - | 1352 | Opioids | $\ldots$ | 39 | 2.88 |
| Ireland | Keenan et al. (1993) | 45 | Pregnant on methadone | 6 | - | 270 | Opioids | $\ldots$ | 7 | 2.59 |
| Italy | Goedert et al. (1995) | 4962 | Treatment | 3.88 | - | 21130 | $\ldots$ | 18 | 332 | 1.57 |
| Italy, Milan | Galli and Musicco (1994) | 2432 | Treatment | 6.7 | - | 16415 | Methadone (94\%) | 13.5 | 413 | 2.52 |
| Italy, Rome | Davoli et al. (1997) | 3955 | Treatment | 4 | 198 | 15820 | IDU | 21.2 (Males) 38.5 (Females) | 387 | 2.45 |
| Italy, Rome | Perucci et al. (1991) | 4200 | Methadone | 8 | Nil | 33600 | Opioids | 10.1 | 239 | 0.71 |
| Italy, Rome | Zaccarelli et al. (1994) | 2431 | Treatment | 3.2 | - | 7872 | IDU | (Males) 30.3 HIV+ II.I HIV(Females) 19.4 HIV+ 4.9 HIV- | 181 | 2.30 |

Table I3.8 Included outcome studies that examined rates of all-cause mortality among problematic drug users (continued)

| Site | Reference | $n$ | Population studied | Follow-up (years) | Lost | Personyears | Drug | Standardized mortality ratio | No. of deaths | Crude mortality rate (per 1000) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Netherlands, Amsterdam | Mientjes et al. (1992) | 390 | Methadone | 2.2 | - | 810 | Opioids | $\ldots$ | 29 | 3.58 |
| Netherlands, Amsterdam | van Haastrecht et al. (I996) | 632 | HIV- methadone and STD clinic patients | 4.4 | 18 | 2781 | $\ldots$ | $\ldots$ | 72 | 2.59 |
| New Zealand, Wellington | Dukes et al. (1992) | 997 | Treatment | 9.1 | - | 9073 | Injecting opioid users | 2.44 | 67 | 0.74 |
| Norway, Oslo | Eskild et al. (1993) | 1009 | HIV test centre clients | 3.67 | - | 3706 | IDU | $\ldots$ | 87 | 2.35 |
| Spain, Catalonia | Orti et al. (1996) | 15711 | Hospital ER and treatment | 2.8 | - | 43717 | Opioids | $\ldots$ | 1315 | 3.01 |
| Spain, Catalonia | Sanchez-Carbonell and Seus (2000) | 135 | Treatment | 10.5 | - | 1418 | Heroin | 28.5 | 41 | 3.4 |
| Sweden | Gronbladh et al. (1990) | 368 | Methadone and untreated | 5-11 | - | 3283 | Heroin | 63 street <br> 4 methadone | $\begin{aligned} & 96 \\ & 26 \end{aligned}$ | $\begin{aligned} & 2.92 \\ & 0.45 \end{aligned}$ |
| Sweden, Gothenburg | Benson and Holmberg (1984) | 618 | Drug-using conscripts, rehab. \& psych. patients, drug-using welfare recipients | 10 | - | 5789 | Cannabis, solvents, LSD, stimulants (opioids rare) | $\ldots$ |  |  |
| Sweden, Lund | Tunving (1988) | 524 | Treatment: opioid, amphetamine, and both opioid and amphetamine | 10 | 0 | 5240 | Opioids, amphetamines | ```5.4 (opioids) 2.5 (amphetamines)``` | 62 | 1.18 |
| Sweden, Stockholm | Engstrom et al. (1991) | 1630 | Drug-related hospitalization | 12 | - | 19560 | 41\% cocaine/ amphetamine; 12\% heroin; 16\% polydrug; $31 \%$ other | 5.3 (18.3 for opioid users; 9.0 for cocaine/ amphetamine) | 446 | 2.3 |


| $\stackrel{\sim}{\infty}$ | $\stackrel{\sim}{\square}$ | $\xrightarrow[\sim]{\sim}$ | N | N | ก | $\stackrel{n}{n}$ | $\stackrel{\infty}{+}$ | N | $\stackrel{\checkmark}{-}$ | $\stackrel{\sim}{\text { ¢ }}$ | $\stackrel{\square}{\infty}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | $\stackrel{ \pm}{\sim}$ | $\bar{\sim}$ | $\infty$ | $\frac{\circ}{m}$ | $\underline{\sim}$ | F | の | $\wedge$ | 안 | ㄲ | 안 |

Table 13.8 Included outcome studies that examined rates of all-cause mortality among problematic drug users (continued)

| Site | Reference | $n$ | Population studied | Follow-up (years) | Lost | Personyears | Drug | Standardized mortality ratio | No. of deaths | Crude mortality rate (per 1000) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| USA, NM. <br> Albuquerque | Goldstein and Herrera (1995) | 1013 | Methadone | 22 | 243 | $\begin{array}{r} 22286 \\ \text { (I6940 } \\ \text { xcl. lost) } \end{array}$ | $\ldots$ | 4.0 (Males) <br> 6.8 (Females) | 348 | 1.56 (2.05) |
| USA, New York | Concool et al. (1979) | 1156 | Treatment (84\% methadone) | 7 | 102 | 8092 | Heroin | 1.5 | 45 | 0.56 |
| USA, New York | Friedman et al. (1996) | 858 | Drug and alcohol dependents on welfare | 8 | - | 6864 | Drugs and alcohol | $\ldots$ | 183 | 2.67 |
| USA, New York | Vaillant (1973) | 100 | Treatment, male | 20 | 17 | 1660 | Narcotics | $\ldots$ | 23 | 1.15 |
| USA, OR, Portland | McAnulty et al. (1995) | 1769 | Not in treatment | 1.78 | - | 3149 | IDU | 8.3 |  | 1.05 |
| USA, PA, Philadelphia | Zanis and Woody (1998) | 507 | Methadone | 1 | 5 | 507 | Heroin | ... | 13 | 2.56 |
| USA, 18 treatment agencies | Joe and Simpson (1987) | 697 | Treatment | 6 | 142 | 3330 | Opioids | 6.9 | 52 | 1.56 |
| USA, 34 treatment agencies | Joe et al. (1982) | 3324 | Treatment | 4 | - | 11710 | Opioids | $\begin{aligned} & 14 \text { (<21 years) } \\ & 10 \text { (21-30 years) } \\ & 4 \text { (>30 years) } \end{aligned}$ | 179 | 1.52 |
| Netherlands, Amsterdam; USA, MD, Baltimore | van Ameijden et al. (1999) | 2809 | Treatment shelters and community agencies | 6-9 | - | 15107 | IDU | ... | 264 | 1.75 |

## 5. Estimated mortality attributable to ILLICIT DRUG USE, 2000

### 5.1 Burden of mortality attributable to specific causes

Table 13.9 shows the median indirect estimates of the number of deaths attributed to illicit drug use in 2000 for each of the 14 subregions (Table 13.10 also shows low and high range estimates around these medians).

## AIDS

The second largest individual cause of death was AIDS, with a global median estimate of 59000 deaths. The largest proportion of these deaths was estimated to have occurred in WPR-B (17000). The other two subregions in which the greatest number of deaths from AIDS related to illicit drug use were EMR-D (11000) and SEAR-B (11000).

There is some indication that the estimates for some subregions may be too low. The United States Centers for Disease Control and Prevention reported that in 1999, 5932 AIDS-related deaths occurred in the United States that were attributed to IDU (see http://www.cdc.gov/hiv/ stats/hasr1202.htm). Similarly, reports from UNAIDS experts indicate that estimates for EUR-C may also be too low, with reports that in the Ukraine alone, approximately 3440 deaths occurred due to AIDS in 1999 (UNAIDS, personal communication, 2001). While some reviewers commented that South-East Asian estimates were higher than they expected, recent work has indicated that the number of AIDS deaths in Thailand (one of the countries in this region) was higher than previously estimated (A. Lopez, personal communication, 2001). Recent work in the SouthEast Asia Region is consistent with the possibility that AIDS-related deaths have been underestimated in this region (Reid and Costigan 2002).

## Overdose

Opioid overdose was the next largest cause of death among illicit drug users, with a median estimate of 69152 deaths globally. The two subregions that accounted for the largest number of opioid overdose deaths were SEAR-D (22 989) and EMR-D (12 852), followed by EUR-C and AMR-A.

The estimates may be too low for some subregions. For example, WPR-A, which includes Australia, had a median estimate of 825 deaths, with a high and low estimate of 954 and 696, respectively. Data from the Australian Bureau of Statistics indicates that in 2000, a total of 737 deaths occurred among persons aged 15-44 years (National Drug and Alcohol Research Centre 2000).

## Suicide

Suicide among opioid users was estimated to account for 32216 deaths in 2000. SEAR-D accounted for the greatest proportion of these deaths

Table 13.9 Median indirect estimates of mortality attributed to illicit drug use, by subregion

| Subregion | AIDS | Opioid overdose | Suicide | Trauma |
| :--- | ---: | ---: | ---: | ---: |
| AFR-D | 0 | 1891 | 1191 | 2768 |
| AFR-E | 0 | 407 | 64 | 922 |
| AMR-A | 4000 | 6397 | 2034 | 4057 |
| AMR-B | 5000 | 1845 | 922 | 2342 |
| AMR-D | 0 | 498 | 78 | 716 |
| EMR-B | 0 | 3881 | 673 | 813 |
| EMR-D | 11000 | 12852 | 2015 | 2954 |
| EUR-A | 0 | 5527 | 2355 | 3387 |
| EUR-B | 1000 | 6895 | 1465 | 651 |
| EUR-C | 9000 | 955 | 4156 | 830 |
| SEAR-B | 7000 | 22989 | 576 | 797 |
| SEAR-D | 8000 | 2909 | 1251 | 3128 |
| WPR-A | 17000 | 69152 | 456 | 1028 |
| WPR-B | 59000 |  | 32216 | 9295 |
| Total (median) |  |  |  | 33689 |

Note: There were an additional 10000 deaths from overdose above and beyond those coded as drug use disorders (added to unintentional injuries) or when coded drug use disorder deaths were higher than estimated overdose deaths.
(14 982), with EUR-C accounting for the next largest. A similar number of deaths were estimated to be due to trauma (34 184). WPR- B had the largest numbers of deaths due to this cause (9295), followed by AMRA (4057).

### 5.2 All-CAUSE mortality attributable to illicit drugs

Table 13.11 compares two methods of calculating the total mortality attributable to illicit drug use: (i) adding the above four causes; and (ii) using estimates of "all-cause" mortality derived from cohort studies and attributable fractions. There are some reassuring similarities between the two sources, and some noteworthy discrepancies.

Overall, the global estimates were remarkably similar ("all-cause" estimate 197383 vs "sum" estimate 194058). Of note was the fact that the subregions that had discrepant estimates were largely developing subregions, and not the subregions from which the majority of cohort studies and attributable fractions had been derived. One of the subregions that accounted for the slightly lower estimates using the all-cause mortality method was SEAR-D, whose all-cause estimate (around 11000 ) was only $23 \%$ of its sum estimate. In general, however, for most other subregions, estimates were within close range of each other, or the all-cause estimates were higher. The overall rate of death per 1000

Table 13.10 Mortality range attributable to illicit drug use, by subregion

| Subregion | AIDS | Opioid overdose | Suicide | Trauma | All-cause |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | $0^{\text {a }}$ | 1891 | 1191 | 2768 | 19046 |
| Low | 0 | 1526 | 354 | 1235 | 9754 |
| High | 0 | 2256 | 2028 | 4807 | 28338 |
| AFR-E | 0 | 407 | 64 | 922 | 8286 |
| Low | 0 | 246 | 40 | 412 | 3251 |
| High | 0 | 609 | 87 | 1602 | 13321 |
| AMR-A | 4000 | 6397 | 2034 | 4057 | 40356 |
| Low | 4000 | 5144 | 806 | 718 | 23186 |
| High | 4000 | 7649 | 3261 | 7397 | 54647 |
| AMR-B | 5000 | 1845 | 922 | 2342 | 18425 |
| Low | 5000 | 1530 | 240 | 985 | 13034 |
| High | 5000 | 2159 | 1604 | 3699 | 23817 |
| AMR-D | 0 | 498 | 78 | 716 | 2522 |
| Low | 0 | 300 | 49 | 319 | 1070 |
| High | 0 | 744 | 107 | 1243 | 3985 |
| EUR-A | 0 | 5527 | 2355 | 3387 | 16453 |
| Low | 0 | 2791 | 866 | 712 | 11026 |
| High | 0 | 9108 | 4481 | 4690 | 19533 |
| EUR-B | 1000 | 1281 | 1465 | 651 | 5794 |
| Low | 1000 | 214 | 336 | 473 | 2923 |
| High | 1000 | 2348 | 2595 | 829 | 8665 |
| EUR-C | 3000 | 6895 | 4156 | 830 | 10709 |
| Low | 3000 | 4507 | 707 | 674 | 3474 |
| High | 3000 | 9284 | 7605 | 986 | 17944 |
| EMRB | 0 | 3881 | 673 | 813 | 5012 |
| Low | 0 | 431 | 196 | 317 | 4612 |
| High | 0 | 7332 | 1149 | 1309 | 5412 |
| EMRD | 11000 | 12852 | 2015 | 2954 | 10411 |
| Low | 11000 | 7757 | 1271 | 1319 | 4416 |
| High | 11000 | 19212 | 2759 | 5131 | 16454 |
| SEAR-B | 11000 | 955 | 576 | 797 | 5688 |
| Low | 11000 | 581 | 208 | 745 | 2625 |
| High | 11000 | 1330 | 943 | 849 | 8751 |
| SEAR-D | 7000 | 22989 | 14982 | 3128 | 11024 |
| Low | 7000 | 1824 | 2059 | 1396 | 4676 |
| High | 7000 | 44154 | 27105 | 5434 | 17423 |
| WPR-A | 0 | 825 | 1251 | 1028 | 9916 |
| Low | 0 | 696 | 109 | 246 | 6375 |
| High | 0 | 954 | 2394 | 1809 | 13457 |
| WPR-B | 17000 | 2909 | 456 | 9295 | 33741 |
| Low | 17000 | 1756 | 288 | 8111 | 11329 |
| High | 17000 | 3439 | 624 | 10479 | 90709 |
| Total median | 59000 | 69152 | 32216 | 33689 | 197383 |
| Low | 59000 | 29303 | 8330 | 17622 | 101751 |
| High | 59000 | 110577 | 56742 | 50264 | 322456 |

[^55]Table I3.II Estimates of total mortality attributed to illicit drug use, by subregion

| Subregion | Population (000s) $>15$ years | Sum of four causes of mortality ${ }^{2}$ | Population mortality rate (per 1000) | All-cause mortality ${ }^{\text {b }}$ | Population mortality rate (per 1000) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | 159577 | 5850 | 0.04 | 19046 | 0.12 |
| AFR-E | 190152 | 1393 | 0.01 | 8286 | 0.04 |
| AMR-A | 255420 | 16488 | 0.06 | 40356 | 0.16 |
| AMR-B | 297625 | 10109 | 0.03 | 18425 | 0.06 |
| AMR-D | 44658 | 1292 | 0.03 | 2522 | 0.06 |
| EMR-B | 86853 | 5367 | 0.06 | 5012 | 0.06 |
| EMR-D | 204039 | 28821 | 0.14 | 10411 | 0.05 |
| EUR-A | 339446 | 11269 | 0.03 | 16453 | 0.05 |
| EUR-B | 161213 | 4397 | 0.03 | 5794 | 0.04 |
| EUR-C | 152432 | 14881 | 0.10 | 10709 | 0.07 |
| SEAR-B | 206870 | 13328 | 0.06 | 5688 | 0.03 |
| SEAR-D | 818521 | 48099 | 0.06 | 11024 | 0.01 |
| WPR-A | 129888 | 3104 | 0.02 | 9916 | 0.08 |
| WPR-B | 1 131503 | 29660 | 0.03 | 33741 | 0.03 |
| World | 4178197 | 194058 | 0.05 | 197383 | 0.05 |

a Sum of the median estimates of the following four causes: AIDS, opioid overdose, suicide via opioids and trauma.
b Median estimates of all-cause mortality derived from SMR analyses and pooled CMRs.
persons aged $\geq 15$ years due to illicit drug use, on a global level, was estimated at 0.5 per annum. The highest all-cause mortality rate was estimated to have occurred in AMR-A ( 0.16 per 1000 persons aged $\geq 15$ years), followed by AFR-D ( 0.12 per 1000 persons aged $\geq 15$ years). The lowest rates using all-cause estimates occurred within SEAR-D (0.01 per $1000)$, WPR-B (0.03) and SEAR-B (0.03).

The discrepancies between the two sources of estimates for developed societies suggested that in general (with the exception of AMR-A), the consistency between the two was reasonable. If anything, the all-cause method produced a higher estimate, which is consistent with the fact that the "four cause" method does not include an exhaustive list of all possible causes of death.

In some developing subregions (such as SEAR-D) there was marked discrepancy between the two sources of estimates. This could be due to higher rates of AIDS-related deaths among injecting drug users in these subregions, which were not adequately assessed by using the all-cause method (in which some cohort studies were carried out before AIDS became an issue).

Table I3.12 Proportion of causes of death attributed to illicit drug use among males, by subregion

| Subregion | Proportion among males |
| :--- | :---: |
| AFR-D | 0.89 |
| AFR-E | 0.83 |
| AMR-A | 0.56 |
| AMR-B | 0.63 |
| AMR-D | 0.72 |
| EMR-B | 0.85 |
| EMR-D | 0.82 |
| EUR-A | 0.59 |
| EUR-B | 0.64 |
| EUR-C | 0.79 |
| SEAR-B | 0.96 |
| SEAR-D | 0.83 |
| WPR-A | 0.93 |
| WPR-B | 0.79 |

### 5.3 Age and sex breakdowns

Our ability to make reliable and valid estimates of the age and sex breakdowns of deaths attributable to illicit drug use is extremely limited. Not only are estimates of the prevalence of drug use according to these characteristics limited (or absent) in many countries, it is also the case that evidence on the characteristics of persons dying from the causes examined here are limited. The estimates made below have been made with reference to limited data on the age and sex breakdowns of persons dying from AIDS, overdose, trauma and suicide.

## SEX

We made the following estimates of the sex breakdown. This was completed by using estimates of the proportion of tobacco users who were males in each of the 14 subregions. These had been calculated in each subregion from the smoking risk factor for the GBD project. Table 13.12 shows the estimates for proportion of deaths among males in each of the subregions. Table 13.13 shows the resulting numbers of deaths attributable to illicit drugs by subregion and sex. Table 13.14 provides estimates of the total DALYs attributable to illicit drugs by subregion and sex.

## Age groups

Similarly to the sex breakdowns, the age breakdowns are based on limited data concerning the age distribution of persons dying from the

Table I3.13 Number of deaths attributed to illicit drug use, by subregion and sex

| Subregion | Sum of four causes of mortalty ${ }^{2}$ | All-cause mortality ${ }^{\text {b }}$ |
| :---: | :---: | :---: |
| AFR-D |  |  |
| Males | 5207 | 16951 |
| Females | 643 | 2095 |
| AFR-E |  |  |
| Males | 1156 | 6877 |
| Females | 237 | 1409 |
| AMR-A |  |  |
| Males | 9233 | 22599 |
| Females | 7255 | 17757 |
| AMR-B |  |  |
| Males | 6369 | 11608 |
| Females | 3740 | 6817 |
| AMR-D |  |  |
| Males | 930 | 1816 |
| Females | 362 | 706 |
| EMR-B |  |  |
| Males | 4562 | 4260 |
| Females | 805 | 752 |
| EMR-D |  |  |
| Males | 23633 | 8537 |
| Females | 5188 | 1874 |
| EUR-A |  |  |
| Males | 6649 | 9707 |
| Females | 4620 | 6746 |
| EUR-B |  |  |
| Males | 2814 | 3708 |
| Females | 1583 | 2086 |
| EUR-C |  |  |
| Males | 11756 | 8460 |
| Females | 31251 | 2249 |
| SEAR-B |  |  |
| Males | 12795 | 5460 |
| Females | 533 | 228 |
| SEAR-D |  |  |
| Males | 39922 | 9150 |
| Females | 8177 | 1874 |
| WPR-A |  |  |
| Males | 2887 | 9222 |
| Females | 217 | 694 |
| WPR-B |  |  |
| Males | 23431 | 26655 |
| Females | 6229 | 7086 |
| World |  |  |
| Males | 149425 | 145012 |
| Females | 44633 | 52371 |

[^56]Table I3.14 Burden of disease (000s of DALYs) attributed to illicit drug use in the subregions, by sex

| Subregion | Males | Females |
| :--- | ---: | ---: |
| AFR-D | 428 | 134 |
| AFR-E | 460 | 150 |
| AMR-A | 594 | 185 |
| AMR-B | 586 | 13 |
| AMR-D | 193 | 59 |
| EMR-B | 376 | 64 |
| EMR-D | 478 | 109 |
| EUR-A | 599 | 172 |
| EUR-B | 130 | 39 |
| EUR-C | 340 | 102 |
| SEAR-B | 95 | 22 |
| SEAR-D | 703 | 116 |
| WPR-A | 173 | 76 |
| WPR-B | 256 | 55 |
| World | 5402 | 1477 |

four main causes of death considered here. It was assumed that no persons aged $<15$ years and no persons aged $>54$ years were problematic users of illicit drugs; the 15-54-year age group has typically been found to contain the vast majority of problematic illicit drug users (Anthony and Helzer 1991).

It was assumed that two-thirds of overdose deaths occurred among the $25-44$-year age group, with on-sixth each occurring in the 15-24and 45-54-year age groups, in line with previous research suggesting the bulk of deaths occur in such a pattern (Hall et al. 1999c, 2000c). Deaths related to illicit drug use that were due to trauma were assumed to be disproportionately distributed among younger age groups, with smaller proportions among those aged 35-54 years (see Table 13.15). With the knowledge that AIDS-related deaths usually occur years after contracting HIV, we assigned the deaths the same age distribution as the total HIV/AIDS deaths in each subregion.

## 6. Discussion

### 6.1 Methodological caveats

A number of potential sources of inaccuracy need be acknowledged in our estimates. First, there are a number of factors that determine the proportion of cases of any particular cause of mortality that are attribut-

Table 13.15 Estimated distribution of causes of death attributed to illicit drug use, by age

|  | Age (years) |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| Cause of death | $<15$ | $15-24$ | $25-34$ | $35-44$ | $45-54$ | $>54$ |
| Overdose | 0 | $1 / 6$ | $1 / 3$ | $1 / 3$ | $1 / 6$ | 0 |
| Suicide | 0 | $1 / 3$ | $1 / 3$ | $1 / 6$ | $1 / 6$ | 0 |
| Trauma | 0 | $1 / 3$ | $1 / 3$ | $1 / 6$ | $1 / 6$ | 0 |
| Total $^{2}$ | 0 | 0.13 | 0.34 | 0.29 | 0.24 | 0 |

a Weighted average of the three causes.
able to harmful illicit drug use. These include environmental, cultural or behavioural factors, which are also likely to interact. The risk of contracting HIV/AIDS through injecting drug use, for example, is greatly reduced by providing sterile injecting equipment, and the use of such equipment will be affected by attitudes towards needle sharing. In countries with needle and syringe programmes the attributable fraction of HIV due to injecting drug use is likely to be relatively small compared to similar countries that do not have needle and syringe programmes, even assuming a similar prevalence of other risk factors for HIV transmission in both countries (Hurley et al. 1997). An illustration of this was provided by Lurie and Drucker (1997) who assessed the impact of needle and syringe programmes on the development of the HIV epidemic in Australia and the United States. They estimated that between 10000 and 25000 HIV infections in the United States could have been prevented if needle exchange programmes were implemented as they had been in Australia.

Second, the availability of drug treatment programmes, medical care and a host of other factors that differ between otherwise similar countries may produce differences in the attributable fractions in those countries. For example, van Ameijden et al. (1999) compared mortality in cohorts of heroin users in Amsterdam and Baltimore. They found Amsterdam drug users had an overdose/suicide mortality rate approximately twice that of their counterparts in Baltimore. This was despite the fact that a greater proportion of users in Amsterdam were in methadone maintenance treatment, which has been shown to reduce the risk of overdose. This finding contrasts with a previous finding of the same research group, which attributed lower mortality rates from infectious disease in Amsterdam to drug users having better access to primary health care in Amsterdam than in New York (Mientjes et al. 1992). The variation in mortality rates that result from differences in the complex interactions
of determinants of mortality makes comparisons of cohort studies conducted in different countries problematic.

Third, attributable fractions can only be reliably calculated in countries that collect accurate mortality data. These data are most likely to be found in developed countries vs developing countries (Muller 1982). Caution is required in applying fractions estimated in developed countries to developing ones.

### 6.2 Limitations of cohort studies

As mentioned earlier, the cohort studies of problem illicit drug users have a number of major limitations when used for the purpose of estimating the contribution of problem illicit drug use to the global burden of disease.

## TREATMENT POPULATIONS

The vast majority of cohort studies of mortality among illicit drug users have included people seeking treatment for problem drug use. A small number of studies have compared mortality of drug users while in and out of treatment (Capelhorn et al. 1996; Fugelstad et al. 1995; Gronbladh et al. 1990; Sanchez-Carbonell and Seus 2000; Zanis and Woody 1998). These studies have found that the relative risk of death while in treatment varied from less than 0.2 to 0.8 , with a mean of approximately 0.4 . These studies can be used to produce more accurate estimates of mortality by applying different mortality rates for proportions of users who are and are not in treatment.

## ILLICIT DRUGS USED

Injecting opioid users are over-represented in the cohort studies by comparison with cocaine and other stimulant users. The few studies that report separate data on problem illicit opioid and stimulant use suggest that mortality is higher among opioid users (Engstrom et al. 1991), probably because of the greater risk of fatal overdose from opioids. Stimulant users, by contrast, may be at higher risk of contracting diseases from bloodborne viruses such as hepatitis $B$ and $C$ from sharing injection equipment because they inject at a high frequency when bingeing on their drug of choice (Bux et al. 1995; Chaisson et al. 1989). They may also be more likely to engage in sex for drugs (Chiasson et al. 1991; Darke et al. 1995; Edlin et al. 1994).

## Extrapolation across subregions

Applying direct measures of mortality from cohort studies in developed countries to populations in developing countries is problematic. Developing countries generally have all-cause mortality rates that are significantly higher than the developed countries in which most cohort studies are conducted (WHO 2001). Thus it may be that there is less of a dif-
ferential in mortality rates between the general population and problem drug users in developing countries. Applying the relative risks from developed countries to developing ones may therefore overestimate the mortality attributable to illicit drug use in the latter.

## HIV/AIDS

The majority of cohort studies identified for this project were conducted before the HIV/AIDS epidemic began to affect mortality among injecting drug users. Changes in the epidemiology of HIV and other drugrelated conditions since these studies were conducted may reduce the validity of using prevalence or incidence data to predict mortality. In some developed nations, for example, the incidence of HIV and AIDS may be declining but the large number of prevalent cases may still produce a high burden of mortality (CDC 2001; UNAIDS 2001). Conversely, countries that are still in the early stages of the epidemic may have a high incidence of HIV/AIDS cases that have not yet begun to contribute to mortality. In either case mortality estimates based on the number of incident cases may be inaccurate, for very different reasons. However, in the absence of better data on this issue, UNAIDS data on the number of AIDS deaths in the year 2000 have been used to estimate mortality, since there are significant problems with making estimates from incident cases.

Despite the limitations of cohort studies, they present the most robust epidemiological evidence on the relationship between problem illicit drug use and mortality. When quantifying the burden of mortality attributed to illicit drugs, therefore, cohort studies provide the best basis on which to estimate risk and identify mortality outcomes.

In terms of estimating risk, as we have described above, the use of annual mortality rates derived from studies of illicit drug users in developed countries may underestimate mortality in developing countries. By contrast, applying SMRs from the cohort studies to developed societies may overestimate the mortality rate of drug users in developing countries (which already have higher mortality rates in general), since it is probable that the higher the general mortality rate in any given country, the lower will be the SMR for illicit drug users in that country (Muller 1982). We have used UNAIDS estimates in the current study.

Other data sources can be used to validate estimates of risk derived from cohort studies in some developed societies. In populations where reliable mortality data are collected the attributable fraction of mortality due to a range of conditions that may be related to problem illicit drug use can be calculated. These fractions can then be applied to estimates of mortality in other countries using the WHO all-cause mortality database. The main weaknesses of this method are: that it does not take into account variations in the prevalence of the risk factor; it assumes homogeneity between the population from which the attributable fraction was derived and the population to which it is being applied;
and that cohort studies are representative of the population at risk. It is nonetheless an independent method of calculating mortality that can be used to check estimates of mortality derived by multiplying measures of risk by prevalence estimates.

### 6.3 Summary and conclusions

In summary, in this work we have attempted to estimate the extent of global mortality and morbidity attributable to illicit drug use in 2000. This required estimates of both the global prevalence of problem illicit drug use and the mortality attributable to it. Ideally, such data would include estimation of the numbers of problem, or dependent users, as these are the individuals at greatest risk for drug-related harm. Currently, there are poor data on the prevalence of problem illicit use in many developing countries and there is no consensus on the definition and operationalization of "problem drug use". UNDCP data, supplemented by other sources, provide the best available data, although these have major limitations.

Similarly, there is a considerable amount of data from cohort studies of individuals identified as problem illicit drug users that can be used to estimate the relative risks of death among this group. Unfortunately, most of these studies have been conducted in developed countries on problem opioid users and many were conducted prior to the AIDS pandemic among injecting drug users. A priority for future research must be to assess mortality among illicit drug users in developing countries and, in particular, to examine the extent to which the findings of studies conducted in developed countries are applicable to developing countries.

Furthermore, much of this research has been based on samples of people entering treatment for drug-related problems. Further work is needed to quantify mortality among problem drug users who are not in treatment. Nonetheless, it is clear from the existing cohort studies that problem illicit drug users have a greatly elevated risk of premature death from drug overdose, HIV/AIDS, suicide and trauma.

By comparison with the extensive literature on the health effects of tobacco and alcohol use, very little is known about the adverse health effects of illicit drug use. This situation reflects at least three factors: the recent history of illicit drug use in many countries; the low prevalence of its use in the population compared to alcohol and tobacco; and the fact that its illicit nature encourages users to conceal or deny their drug use, hence inhibiting research on its effects on mortality and morbidity.

In 2000, the median of the two methods of estimating the number of global deaths attributed to illicit drugs was 195721 . Estimates produced by both methods had wide uncertainty intervals around them (113494 for sum of four causes; and 101751 to 322456 for all-cause estimates). When morbidity attributable to illicit drug use is added to the estimated mortality this risk factor accounts for $0.8 \%$ of global DALYs. The distribution of DALYs between subregions varied, reflecting variations in
drug use and death rates, with considerable uncertainty about the applicability of mortality data derived in developed countries to mortality among illicit drug users in developing countries.

Nonetheless, the current estimates suggest that illicit drug use is a significant cause of premature mortality among young adults in the developed and developing world. Our estimate is certainly an underestimate of total disease burden because: (i) there are deficits in data on mortality attributable to the use of some illicit drugs (most notably cannabis and the newer synthetic drugs like MDMA); (ii) there are differences across subregions in the quality of data available on the causes of mortality that were included in the current estimates; and (iii) there is an absence of data that would permit estimates of some other causes of mortality and morbidity attributable to illicit drug use, such as hepatitis B and C and violence.

Given public concerns about the effects of illicit drug use, and indications of a worldwide increase in the production and use of illicit drugs, better research must be done on the adverse health effects of their use. With that in mind we include a list of research priorities.

### 6.4 Research priorities

- There is a need for more rigorously designed prospective studies of mortality and morbidity among problem illicit drug users in developing countries, especially ones which have high rates of HIV/AIDS infection among injecting drug users, and which have experienced substantial increases in rates of such problem drug use in recent years.
- There is also a need for cohort studies of injecting drug users who are not in treatment, since there is evidence that rates of mortality are higher among this group.
- There is a need for better studies of morbidity attributable to nonfatal overdoses, bloodborne viral diseases such as hepatitis B and C, suicide attempts, and trauma among problem illicit drug users in both developed and developing countries.
- There is a global need for better surveillance systems to collect data on key drug-related consequences.
- There is a need for better prevalence estimates of problem illicit drug use in developed and developing countries, especially where there are indications of increased illicit drug use because of proximity to source countries.
- Other specific data collection needs include the following:
- improving the comparability and quality of data on the prevalence of drug use across regions;
- improving methods for accurately estimating the prevalence of, and burden associated with, illicit drug use among non-institutionalized populations;
- improving estimates of the number of "problem drug users" per se, to take into account polydrug use using consistent definitions of "problem drug use";
- obtaining mortality and morbidity data from developing regions from which more accurate methods to estimate the burden of illicit drug use can be derived for those regions;
- improving data on drug use among psychiatric patients, and data on psychiatric morbidity among drug users;
- developing more comparable and accurate mortality data by improving the consistency of procedures used to identify and register drug-related deaths across subregions, for both developed and developing subregions;
- systematic monitoring of mortality by drug type to provide data on mortality associated with non-opioid drugs, especially in the context of developing countries and countries with high HIV prevalence; and
- measurement of the coverage and nature of services in place to reduce burden, especially those aimed at reducing the transmission of bloodborne viruses.


## 7. PROJECTIONS OF ILLICIT DRUG-RELATED HARM

There are a number of indications that rates of illicit drug use and illicit drug-related harm have risen in the past decade. First, developed countries with reasonable mortality data have shown steady increases in drugrelated deaths, especially drug overdose deaths, over the past decade, for example, in Australia (Hall et al. 1999a); Spain (de la Fuente et al. 1995); and the United Kingdom (Hall et al. 2000c). Estimates derived from back projections of both overdose deaths and treatment entry in Australia have shown an increase in estimated number of dependent opioid users (Hall et al. 2000a, 2000b). Second, illicit drug use and drug-related harm such as overdoses and HIV/AIDS have been reported in an increasing number of countries where it was previously rare, such as in eastern Europe, the former Soviet Union, Asia and Africa (UNAIDS 2001; UNDCP 2000; UNODCCP 2000).

Despite indications that drug-related harm is increasing, it is difficult to predict future patterns of illicit drug use and drug-related harm for the following reasons.

First, there is a lack of good time series data on the prevalence of illicit drug use and data on drug-related harm may not be comparable over
time, even in countries with good mortality data systems, because of changes in classification systems (such as successive iterations of the ICD classification system), and because of improvements (and deterioration) in the quality of data that are collected.

Second, although the general trend has been for drug-related deaths to increase during the 1990s, there have also been a number of countries in which drug-related deaths have fallen sharply, often after various policy initiatives have been introduced. In France (Lepere et al. 2001) and Switzerland, for example, drug-related deaths have fallen markedly in the later half of the 1990s, probably in response to a marked expansion of opioid substitution treatment in both countries. In the past two years, Australia has also seen a substantial drop in opioid overdose deaths, after a steep rise from the early 1990s until 1999 (Degenhardt 2002). In this case, some of the early decrease may have been attributable to expanded treatment and educational initiatives to reduce overdose among opioid users. Since the beginning of 2001, the major driver of reduced overdose deaths in Australia has been a substantial drop in the availability of heroin (Weatherburn et al. 2001).

Third, there have been changes in the scale of illicit drug production and in the choice of drugs for illicit manufacture. For example, restrictions on opioid production in Afghanistan in the late 1990s may have reduced heroin supply to Europe, while the supply of cocaine and amphetamine type stimulants (ATS) increased (UNODCCP 2000). The net effect of these changes on drug-related harm is difficult to predict because the effects of ATS on mortality are less well studied and understood than that of opioids (Darke et al. 2000a).

### 7.1 Options for projection

A conservative option may be to assume that: (i) the current global problem will remain at about the current level, but that (ii) the distribution of burden will shift between developed and developing countries with declines in drug-related deaths in developed countries (resulting from expanded opioid substitution treatment) being offset by increases in drug-related deaths in developing societies.

A less conservative option would be to assume that in developed countries, rates of opioid use and opioid-related deaths will continue to rise, but at a slower rate (e.g. $50 \%$ ) than that observed during the 1990s, because of expanded treatment availability. In developing countries, the rate of increase could be projected to follow the same pattern and magnitude as that observed in developed countries like Australia, which have reasonable time series data on trends in opioid-related deaths over the period when illicit opioid use was first introduced and spread (Hall et al. 2000b).

Substantial uncertainty intervals would need to be placed around these estimates to indicate our ignorance of underlying trends and to empha-
size the need to undertake better epidemiological research to improve our estimates of the global burden of disease attributable to illicit drug use.

## Notes

1 See preface for an explanation of this term.
2 MDMA: 3,4 methylenedioxymethamphetamine, a synthetic drug that is used as a stimulant.

3 Amphetamine-type stimulant (ATS): one of a class of sympathomimetic amines with powerful stimulant action on the central nervous system.
4 Cannabis: a generic term for psychoactive preparations (e.g. marijuana, hashish and hash oil) derived from the cannabis sativa plant.
5 Cocaine: an alkaloid central nervous system stimulant drug that is derived from the coca plant.
6 Heroin: an opioid drug derived from the opium poppy.
7 Opioids: generic term applied to derivatives from the opium poppy, their synthetic analogues, and compounds synthesized in the body, which act upon the opioid receptors in the brain. They have the capacity to relieve pain and produce a sense of euphoria, as well as cause stupor, coma and respiratory depression.

8 Drug overdose: the use of any drug in such an amount that acute adverse physical or mental effects are produced. Overdose in this chapter refers to cases in which death is the outcome.

9 DSM-IV: American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, 4th edition.

## References

Advisory Council on the Misuse of Drugs (2000) Reducing drug related deaths. The Stationery Office, London.

Alter M, Hadler SC, Margolis H et al. (1990) The changing epidemiology of hepatitis B in the United States. Journal of the American Medical Association, 263:1218-1222.

Anderson B, Bodsworth N, Rohrsheim R, Donovan B (1994) Hepatitis B virus infection and vaccination status of high risk people in Sydney: 1982 and 1992. Medical Journal of Australia, 161:368-371.
Andreasson S, Allebeck P (1990) Cannabis and mortality among young men: a longitudinal study of Swedish conscripts. Scandinavian Journal of Social Medicine, 18:9-15.
Anthony JC, Helzer J (1991) Syndromes of drug abuse and dependence. In: Psychiatric Disorders in America. Robins LN, Regier DA, eds. The Free Press, New York.

Anthony JC, Warner L, Kessler R (1994) Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: basic find-
ings from the National Comorbidity Survey. Experimental and Clinical Psychopharmacology, 2:244-268.

Australian Bureau of Criminal Intelligence (2000) Australian illicit drug report 1998-1999. Australian Bureau of Criminal Intelligence Publications, Canberra.

Barr HL, Antes D, Ottenberg DJ, Rosen A (1984) Mortality of treated alcoholics and drug addicts: the benefits of abstinence. Journal of Studies on Alcohol, 45:440-452.
Bates MN, Blakely TA (1999) Role of cannabis in motor vehicle crashes. Epidemiologic Reviews, 21:222-232.
Beautrais AL, Joyce PR, Mulder RT (1998) Psychiatric illness in a New Zealand sample of young people making serious suicide attempts. New Zealand Medical Journal, 111:44-48.

Beautrais AL, Joyce PR, Mulder RT (1999) Cannabis abuse and serious suicide attempts. Addiction, 94:1155-1164.

Benson G, Holmberg MB (1984) Drug-related mortality in young people. Acta Psychiatrica Scandinavica, 70:525-534.

Bewley T, Ben-Arie O (1968) Morbidity and mortality from heroin dependence. 2. Study of 100 consecutive patients. British Medical Journal, 1:727-730.

Bewley T, Ben-Arie O, James IP (1968) Morbidity and mortality from heroin dependence. 1. Survey of heroin addicts known to Home Office. British Medical Journal, 1:725-726.
Boot BP, McGregor LS, Hall W (2000) MDMA (Ecstasy) neurotoxicity: assessing and communicating the risks. The Lancet, 355:1818-1821.
Bucknall ABV, Robertson JR (1986) Deaths of heroin users in a general practise population. Journal of the Royal College of General Practitioners, 36:120-122.

Bux D, Lamb R, Iguchi M (1995) Cocaine use and HIV risk behaviour in methadone maintenance patients. Drug and Alcohol Dependence, 37:29-35.

Capelhorn JRM, Dalton MSYN, Halder F, Petrenas A-M, Nisbet JG (1996) Methadone maintenance and addicts risk of fatal heroin overdose. Substance Use and Misuse, 31:177-196.

Caulkins J, Rydell C, Everingham S (1999) An ounce of prevention, a pound of uncertainty. The cost effectiveness of school-based drug prevention programs. Drug Policy Research Center, Santa Monica, CA.
CDC (2001) Centre for Disease Control, National Centre for HIV, STD and TB Prevention. Surveillance reports. Retrieved, 2001, from http://www.cdc.gov/ hiv/stats/hasr1201.htm.
Chaisson RE, Osmond D, Brodie B, Sande MA, Moss AR (1989) Cocaine use and HIV infection in intravenous drug users in San Francisco. Journal of the American Medical Association, 261:561-565.
Chait LD, Pierri J (1992) Effects of smoked marijuana on human performance: a critical review. In: Marijuana/cannabinoids: neurobiology and neurophysiology. Murphy L, Bartke A, eds. CRC Press, Boca Raton, FL.

Cherubin C, McCusker J, Baden M, Kavaler F, Amsel Z (1972) The epidemiology of death in narcotic addicts. American Journal of Epidemiology, 96:11-22.

Chesher G (1995) Cannabis and road safety: an outline of research studies to examine the effects of cannabis on driving skills and actual driving performance. In: The effects of drugs (other than alcohol) on road safety. Parliament of Victoria Road Safety Committee, ed. Road Safety Committee, Parliament of Victoria, Melbourne.

Chiasson M, Stoneburner R, Hildebrandt D, Ewing W, Telzak E, Jaffe H (1991) Heterosexual transmission of HIV-1 associated with the use of smokable freebase cocaine (crack). AIDS, 5:1121-1126.

Concool B, Smith H, Stimmel B (1979) Mortality rates of persons entering methadone maintenance: a seven-year study. American Journal of Drug and Alcohol Abuse, 6:345-353.

Danish National Board of Health (1997) Drug related death in Europe: quality and comparability of data on drug-related deaths: Danish National Board of Health-EMCDDA Working Group, Copenhagen.
Darke S, Hall W (1995) Levels and correlates of polydrug use among heroin users and regular amphetamine users. Drug and Alcohol Dependence, 39:231-235.

Darke S, Ross J (1997) Polydrug dependence and psychiatric comorbidity among heroin injectors. Drug and Alcohol Dependence, 48:135-141.

Darke S, Ross J (2000) The relationship between suicide and overdose among methadone maintenance patients. (NDARC Technical Report No. 100.) National Drug and Alcohol Research Centre, University of New South Wales, Sydney.
Darke S, Ross J, Cohen J, Hando J, Hall W (1995) Injecting and sexual risk-taking behaviour among regular amphetamine users. AIDS Care, 7:19-26.

Darke S, Ross J, Hall W (1996) Overdose among heroin users in Sydney, Australia. I. Prevalence and correlates of non-fatal overdose. Addiction, 91:405-411.

Darke S, Ross J, Hando J, Hall W, Degenhardt L (2000a) Illicit drug use in Australia: epidemiology, use patterns and associated harm. Australian Government Publishing Service, Canberra.

Darke S, Sims J, McDonald S, Wicks W (2000b) Cognitive impairment among methadone maintenance patients. Addiction, 95:687-695.

Darke S, Zador D (1996) Fatal heroin "overdose": a review. Addiction, 91:1765-1772.

Dart JK (1995) The use of epidemiological techniques to assess risk: the epidemiology of microbial keratitis. Eye, 9:679-683.
Davoli M, Perucci C, Rapiti E et al. (1997) A persistent rise in mortality among injecting drug users in Rome, 1980 through 1992. American Journal of Public Health, 87:851-854.
de la Fuente L, Barrio G, Vicente J, Bravo M, Santacreu J (1995) The impact of drug related deaths on mortality among young adults in Madrid. American Journal of Public Health, 85:102-105.

Degenhardt L (2002) Opioid overdose deaths in Australia, 2001 edition. Sydney: National Drug and Alcohol Research Centre, University of New South Wales. Available at http://ndarc.med.unsw.edu.au/ndarc.nsf/website/DrugInfo.factsheets.
Degenhardt L, Hall W, Lynskey M (2001) The relationship between cannabis use, depression and anxiety among Australian adults: findings from the National Survey of Mental Health and Well-Being. Social Psychiatry and Psychiatric Epidemiology, 36:219-227.

Domino G, Lenaars AA (1989) Attitudes toward suicide: a comparison of Canadian and U.S. college students. Suicide and Life-Threatening Behaviour, 19:345-359.

Domino G, Takahashi Y (1991) Attitudes towards suicide in Japanese and American medical students. Suicide and Life-Threatening Behaviour, 21:345-359.
Donoghoe M (1996) Illicit drugs. In: Quantifying global health risks: the burden of disease attributable to selected risk factors. Murray CJL, Lopez AD, eds. Harvard University Press, Cambridge, MA.
Donoghoe M, Wodak A (1998) Health and social consequences of injecting drug use. In: Drug injecting and HIV infection: Global dimensions and local responses. Stimson $G$ et al., eds. UCL Press, London.
Dowling GP, McDonough E, Bost RO (1987) Eve and ecstasy. A report of five deaths associated with the use of MDEA and MDMA. Journal of the American Medical Association, 257:1615-1617.

Dukes PD, Robinson GM, Robinson BJ (1992) Mortality of intravenous drug users: attenders of the Wellington Drug Clinic, 1972-89. Drug and Alcohol Review, 11:197-201.
Edlin B, Irwin K, Faraque S, McCoy C, Word C (1994) Intersecting epidemicscrack cocaine use and HIV infection among inner city young adults. Multicenter Crack Cocaine and HIV Infection Study Team. New England Journal of Medicine, 331:1422-1427.

EMCDDA (1997) Estimating the prevalence of problem drug use in Europe (Scientific Monograph No. 1). European Monitoring Centre for Drugs and Drug Addiction, Luxembourg.

EMCDDA (1999) Study to obtain comparable national estimates of problem drug use prevalence for all EU member states. European Monitoring Centre for Drugs and Drug Addiction, Lisbon.
EMCDDA (2000) Annual report on the state of the drugs problem in the European Union. European Monitoring Centre for Drugs and Drug Addiction, Lisbon.

English D, Holman C, Milne E et al. (1995) The quantification of drug caused morbidity and mortality in Australia, 1995. Commonwealth Department of Human Services and Health, Canberra.

Engstrom A, Adamson C, Allebeck P, Rydberg U (1991) Mortality in patients with substance abuse: a follow-up in Stockholm County, 1973-1984. The International Journal of the Addictions, 26:91-106.

Eskild A, Magnus P, Samuelson SO, Sohlberg C, Kittelsen P (1993) Differences in mortality rates and causes of death between HIV positive and HIV negative intravenous drug users. International Journal of Epidemiology, 22:315-319.

Feldman U (1993) Epidemiologic assessment of risks of adverse reactions associated with intermittent exposure. Biometrics, 49:419-428.
Fischer B, Gliksman L, Rehm J, Daniel N, Medved W (1999) Comparing opiate users in methadone treatment with untreated opiate users: results of a followup study with a Toronto opiate user cohort. Canadian Journal of Public Health, 90:299-305.

Fischer B, Kendall P, Rehm J, Room R (1997) Charting WHO-goals for licit and illicit drugs for the year 2000: are we "on track"? Public Health, 111:271-275.

Freeman, J (1996) Quantitative epidemiology. Infection control and hospital epidemiology, 17:249-255.
Friedman LN, Williams MT, Singh TP, Frieden TR (1996) Tuberculosis, AIDS, and death among substance abusers on welfare in New York city. The New England Journal of Medicine, 334:828-833.
Frischer M, Bloor M, Goldberg D, Clark J, Green S, McKeganey N (1993) Mortality among injecting drug users: a critical reappraisal. Journal of Epidemiology and Community Medicine, 47:59-63.

Frischer M, Green ST, Goldberg D (1994) Substance abuse related mortality: a worldwide review. United Nations International Drug Control Programme, Vienna.
Fugelstad A, Annell A, Rajs J, Agren G (1997) Mortality and causes and manner of death among drug addicts in Stockholm during the period 1981-1992. Acta Psychiatrica Scandinavica, 96:169-175.
Fugelstad A, Rajs J, Bottiger M, de Verdier MG (1995) Mortality among HIVinfected inravenous drug addicts in relation to methadone treatment. Addiction, 90:711-716.
Galli M, Musicco M (1994) Mortality of intravenous drug users living in Milan, Italy: role of HIV-1 infection. AIDS, 8:1457-1463.
Ghodse AH, Oyefeso A, Kilpatrick B (1998) Mortality of drug addicts in the United Kingdom 1967-1993. International Journal of Epidemiology, 27:473-478.

Ghodse AH, Sheehan M, Taylor C, Edwards G (1985) Deaths of drug addicts in the United Kingdom 1967-81. British Medical Journal, 290:425-428.

Goedert JJ, Pizza G, Gritti FM et al. (1995) Mortality among drug users in the AIDS era. International Journal of Epidemiology, 24:1204-1210.
Goldfrank L, Hoffman R (1993) The cardiovascular effects of cocaine-update. In: Acute cocaine intoxication: current methods of treatment. (Research

Monograph No. 123). Sorer H, ed. National Institute on Drug Abuse, Rockville, MD.

Goldstein A, Herrera J (1995) Heroin addicts and methadone treatment in Alburquerque: a 22 year follow-up. Drug and Alcohol Dependence, 40:139-150.

Gossop M, Griffiths P, Powis B, Williamson S, Strang J (1996) Frequency of nonfatal heroin overdose: survey of heroin users recruited in non-clinical settings. British Medical Journal, 313:402.

Gossop M, Marsden J, Stewart D (1998) NTORS at one year. Department of Health, London.
Gronbladh L, Ohund LS, Gunne LM (1990) Mortality in heroin addiction: impact of methadone treatment. Acta Psychiatrica Scandinavica, 82:223-227.
Haastrup S, Jepson PW (1984) Seven year follow-up of 300 young drug abusers. Acta Psychiatrica Scandinavica, 70:503-509.
Haastrup S, Jepson PW (1988) Eleven year follow-up of 300 young opioid addicts. Acta Psychiatrica Scandinavica, 77:22-26.
Hall W, Degenhardt L, Lynskey M (1999a) Opioid overdose mortality in Australia, 1964-1997: birth cohort trends. Medical Journal of Australia, 171:34-37.

Hall W, Degenhardt L, Lynskey M (2001) The health and psychological consequences of cannabis use (NCADA Monograph No. 44). Australian Publishing Service, Canberra.
Hall W, Johnston L, Donnelly N (1999b) Epidemiology of cannabis use and its consequences. In: The health effects of cannabis. Kalant H et al., eds. Centre for Addiction and Mental Health, Toronto.
Hall W, Lynskey M, Degenhardt L (1999c) Heroin use in Australia: its impact upon public health and public order (NDARC Monograph No. 42). National Drug and Alcohol Research Centre, University of New South Wales, Sydney.
Hall W, Lynskey M, Degenhardt L (2000c) Trends in opiate-related deaths in the United Kingdom and Australia, 1985-1995. Drug and Alcohol Dependence, 57:247-254.

Hall W, Ross J, Lynskey M, Law M, Degenhardt L (2000a) How many dependent heroin users are there in Australia? Medical Journal of Australia, 173:528-531.
Hall W, Ross J, Lynskey M, Law M, Degenhardt L (2000b) How many dependent heroin users are there in Australia? (NDARC Monograph No. 44). National Drug and Alcohol Research Centre, University of New South Wales, Sydney.
Hall W, Solowij N (1998) Adverse effects of cannabis. The Lancet, 352: 1611-1616.

Hall W, Solowij N, Lemon J (1994) The health and psychological consequences of cannabis use (NDARC Monograph No. 25). Australian Publishing Service, Canberra.

Hall W, Teesson M, Lynskey M, Degenhardt L (1999d) The 12-month prevalence of substance use and ICD-10 substance use disorders in Australian
adults: findings from the National Survey of Mental Health and Well-Being. Addiction, 94:1541-1550.

Hartnoll R (1997) Cross-validating at local level. In: Estimating the prevalence of problem drug use in Europe. European Monitoring Centre for Drugs and Drug Addiction, ed. Office for Official Publications of the European Communities, Luxembourg.
Henderson S, Andrews G, Hall W (2000) Australia's mental health: an overview of the general population survey. Australian and New Zealand Journal of Psychiatry, 34:197-205.
Henry JA, Jeffreys KJ, Dawling S (1992) Toxicity and deaths from 3,4 methylenedioxymethamphetamine ("ecstasy"). The Lancet, 340:284-287.
Hser Y, Anglin D, Powers K (1993) A 24-year follow-up of Californian narcotics addicts. Archives of General Psychiatry, 50:577-584.
Hulse G, English D, Milne E, Holman C (1999) The quantification of mortality resulting form the regular use of illicit opiates. Addiction, 94:221-230.

Hurley S, Jolley D, Kaldor J (1997) Effectiveness of needle-exchange programmes for prevention of HIV infection. The Lancet, 349:1797-1800.

Joe GW, Lehman W, Simpson DD (1982) Addicts death rates during a four-year post-treatment follow-up. American Journal of Public Health, 72:703-709.

Joe GW, Simpson DD (1987) Mortality rates among opioid addicts in a longitudinal study. American Journal of Public Health, 77:347-348.
Keenan E, Dorman A, O’Connor J (1993) Six year follow-up of forty five pregnant opiate addicts. Irish Journal of Medical Science, 162:444-445.
Klee H, Faugier J, Hayes C, Boulton T, Morris J (1990) AIDS-related risk behaviour, polydrug use and temazepam. British Journal of Addiction, 85:1125-1132.

Lepere B, Gourarier L, Sanchez M et al. (2001) Reduction in the number of lethal heroin overdoses in France since 1994. Focus on substitution treatments. Annales de Medecine Interne, 152:S5-12.

Levine S, Vlahov D, Nelson K (1994) Epidemiology of hepatitis B virus: seroprevalence, risk factors and viral interactions. Epidemiologic Reviews, 16:418-436.

Lurie P, Drucker E (1997) An opportunity lost: HIV infections associated with a lack of needle-exchange programmes in the USA. The Lancet, 349:604-608.
MacDonald M, Crofts N, Kaldor J (1996) Transmission of hepatitis C virus: rates, routes, and cofactors. Epidemiologic Reviews, 18:137-148.
MacDonald M, Wodak A, Dolan K, van Beek I, Cunningham P, Kaldor J (2000) Hepatitis C virus antibody prevalence among injecting drug users at selected needle and syringe programs in Australia, 1995-1997. Medical Journal of Australia, 172:57-61.

Mathers C, Vos T, Stevenson C (1999) The Burden of Disease and Injury in Australia. Australian Institute of Health and Welfare, Canberra.

McAnulty JM, Tesselaar H, Fleming DW (1995) Mortality among injection drug users identified as "out of treatment". American Journal of Public Health, 85:119-120.

Mientjes GH, van Ameijden EK, van den Hoek AJAR, Coutinho RA (1992) Increasing morbidity without rise in non-AIDS mortality among HIV-infected intravenous drug users in Amsterdam. AIDS, 6:207-212.
Muller M (1982) The Health of Nations: a North-South investigation. Faber and Faber, London.
Murray CJL, Lopez AD, eds. (1996) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Global burden of Disease and Injury, Vol. 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Musto DF, Ramos MR (1981) A follow-up study of the New Haven morphine maintenance clinic of 1920. The New England Journal of Medicine, 304:1071-1077.

National Centre in HIV Epidemiology and Clinical Research (1998) HIV/AIDS and related diseases in Australia: Annual Surveillance Report 1998. National Centre in HIV Epidemiology and Clinical Research, Sydney.
National Drug and Alcohol Research Centre (2000) 1999 Australian Bureau of Statistics data on opioid overdose deaths. National Drug and Alcohol Research Centre, University of New South Wales. Retrieved from http://www.med.unsw.edu.au/ndarc.
National Institute on Drug Abuse (1992) The economic costs of alcohol and drug abuse in the United States-1992. Available at http://165.112.78.61/EconomicCosts/Index.html.
National Research Council (2001) Informing America's policy on illegal drugs: what we don't know keeps hurting us. National Academy Press, Washington, DC.

Oppenheimer E, Tobutt C, Taylor C, Andrew T (1994) Death and survival in a cohort of heroin addicts from London clinics: a 22 -year follow-up study. Addiction, 89:1299-1308.
Orti RM, Domingo-Salvany A, Munoz A, MacFarlane D, Suelves JM, Anto JM (1996) Mortality trends in a cohort of opiate addicts, Catalonia, Spain. International Journal of Epidemiology, 25:545-553.

Parr MJA, Heather ML, Botterill P (1997) Hyponatraemia and death after "ecstacy" ingestion. Medical Journal of Australia, 166:136-141.

Perucci CA, Davoli M, Rapiti E, Abeni DD, Forastiere F (1991) Mortality of intravenous drug users in Rome: a cohort study. American Journal of Public Health, 81:1307-1310.
Platt JJ (1997) Cocaine addiction: theory, research, and treatment. Harvard University Press, Cambridge, MA.
Reid G, Costigan G (2002) Revisiting "The bidden epidemic": a situation assessment of drug use in Asia in the context of HIV/AIDS. The MacFarlane Burnet Institute for Medical Research and Public Health, Melbourne.

Ridolfo B, Stevenson C (2001) The quantification of drug-caused mortality and morbidity in Australia, 1998. Australian Institute of Health and Welfare, Canberra.

Robbe HWJ (1994) Influence of marijuana on driving. Institute for Human Psychopharmacology, University of Limberg, Maastricht.
Robertson JR, Ronald PM, Raab GM, Ross AJ, Parpia T (1994) Deaths, HIV infection, abstinence and other outcomes in a cohort of injecting drug users followed up for 10 years. British Medical Journal, 309:369-372.
Ross M, Wodak A, Loxley W, Stowe A, Drury M (1992) Staying negative. Summary of the results of the Australian National AIDS and Injecting Drug Use Study. ANAIDUS, Sydney.

Rossow I, Lauritzen $G$ (1999) Balancing on the edge of death: suicide attempts and life-threatening overdoses among drug addicts. Addiction, 94:209-219.

Sanchez-Carbonell X, Seus L (2000) Ten-year survival analysis of a cohort of heroin addicts in Catalonia: the EMETYST project. Addiction, 95:941-948.

Segest E, Mygind O, Bay H (1990) The influence of prolonged stable methadone maintenance treatment on mortality and employment: an 8 -year follow-up. The International Journal of the Addictions, 25:53-63.
Selwyn PA, Hartel D, Wasserman W, Drucker E (1989) Impact of the AIDS epidemic on morbidity and mortality among intravenous drug users in a New York methadone maintenance program. American Journal of Public Health, 79:1358-1362.

Sidney S, Beck JE, Tekawa IS, Quesenberry CP Jr., Friedman GD (1997a) Marijuana use and mortality. American Journal of Public Health, 87: 585-590.

Sidney S, Quesenberry CP, Friedman GD, Tekawa IS (1997b) Marijuana use and cancer incidence (California, United States). Cancer Causes and Control, 8:722-728.

Smiley A (1999) Marijuana: on road and driving simulator studies. In: The health effects of cannabis. Kalant H et al., eds. Addiction Research Foundation, Toronto.

Stimson G (1993) The global diffusion of injecting drug use: implications for human immunodeficiency virus infection. Bulletin on Narcotics, 45:3-17.
Thorley A, Oppenheimer E, Stimson GV (1977) Clinic attendance and opiate prescription status of heroin addicts over a six-year period. British Journal of Psychiatry, 130:565-569.

Thorsen K, Haarstrup S (1975) Follow-up of 97 young non-psychotic male opiate abusers: relationship between achieving abstinence, age and duration of abuse. Acta Psychiatrica Scandinavica, 51:289-296.

Topp L, Hando J, Dillon P, Solowij N (1999) Ecstasy use in Australia: patterns of use and associated harm. Drug and Alcohol Dependence, 55:105-115.

Tunving K (1988) Fatal outcome in drug addiction. Acta Psychiatrica Scandinavica, 77:551-566.
UNAIDS (2001) Joint United Nations Programme on HIV/AIDS. Retrieved, 2001, from http://www.unaids.org.

UNDCP (2000) Global illicit drug trends 2000. United Nations International Drug Control Programme, Vienna.

UNODCCP (2000) World drug report 2000. United Nations Office for Drug Control and Crime Prevention. Oxford University Press, Oxford.

Vaillant GE (1966) Twelve-year follow-up of New York narcotic addicts. II. The natural history of a chronic disease. The New England Journal of Medicine, 275:1282-1288.

Vaillant GE (1973) A 20-year follow-up of New York narcotic addicts. Archives of General Psychiatry, 29:237-241.
van Ameijden EJC, Krol A, Vlahov D, Flynn C, van Haastrecht HJA, Coutinho RA (1999) Pre-AIDS mortality and morbidity among injection drug users in Amsterdam and Baltimore: an ecological comparison. Substance Use and Misuse, 34:845-865.
van Haastrecht HJA, van Ameijden EJC, van den Hoek JAR, Mientjes GHC, Bax JS, Coutinho RA (1996) Predictors of mortality in the Amsterdam cohort of human immunodeficiency virus (HIV)-positive and HIV-negative drug users. American Journal of Epidemiology, 143:380-391.
Wahren CA, Brandt L, Allebeck P (1997) Has mortality in drug addicts increased? A comparison between two hospitalized cohorts in Stockholm. International Journal of Epidemiology, 26:1219-1226.
Warner-Smith M, Darke S, Lynskey M, Hall W (2001) Heroin overdose: causes and consequences. Addiction, 96:1113-1125.
Watterson O, Simpson DD, Sells SB (1975) Death rates and causes of death among opioid addicts in community drug treatment programs during 1970-1973. American Journal of Drug and Alcohol Abuse, 2:99-111.

Weatherburn D, Jones C, Freeman K, Makkai T (2001) The Australian heroin drought and its implications for drug policy. (Crime and Justice Bulletin No. 59). New South Wales Bureau of Crime Statistics and Research, Sydney.

WHO (1993) The ICD-10 Classification of mental and behavioural disordersdiagnostic criteria for research. World Health Organization, Geneva.
WHO (1997) Cannabis: a bealth perspective and research agenda. World Health Organization, Geneva.
WHO (1998) Opioid overdose. Trends, risk factors, interventions and priorities for action. World Health Organization, Programme on Substance Abuse, Geneva.

WHO (2001) World Health Organization All-Cause Mortality Database. Retrieved, 2001, from http://www-nt.who.int/whosis/statistics/menu.cfm.
Wiepert GD, Bewley TH, d'Orban PT (1978) Outcomes for 575 British opiate addicts entering treatment between 1968 and 1975. Bulletin on Narcotics, 30:21-32.

Wille R (1981) Ten-year follow-up of a representative sample of London heroin addicts: clinic attendance, abstinence and mortality. British Journal of Addiction, 76:259-266.

Zaccarelli M, Gattari P, Rezza G et al. (1994) Impact of HIV infection on nonAIDS mortality among Italian injecting drug users. AIDS, 8:345-350.
Zanis DA, Woody GE (1998) One-year mortality rates following methadone treatment discharge. Drug and Alcohol Dependence, 52:257-260.

Zhang Z-F, Morgenstern H, Spitz M et al. (1999) Marijuana use and increased risk of squamous cell carcinoma of the head and neck. Cancer Epidemiology, Biomarkers and Prevention, 8:1071-1078.

# Chapter I4 

Unsafe sex<br>Emma Slaymaker, Neff Walker, Basia Zaba and<br>Martine Collumbien

## Summary

The risk factor "unsafe sex" has been defined here as sex between a susceptible person and a partner who has a sexually transmitted infection (STI), without taking measures to prevent infection. Unsafe sex cannot therefore be defined a priori (because sex is only unsafe with respect to the context in which it occurs), or measured directly from reported behaviours. A set of behaviours was defined as "risky sex" and the prevalence of various behaviours was estimated for 57 countries. The prevalence of risky sex as defined here is given by the proportion of the population who have had sex in the last year with a non-co-resident partner, and who did not use a condom on the last occasion with that partner. For the comparative risk assessment (CRA) estimates, the prevalence of risky sex between men and women was the primary focus.

The main outcome considered was infection with HIV, which is responsible for the majority of the burden of mortality and morbidity associated with STIs. Infections with Chlamydia trachomatis (chlamydia), Neisseria gonorrhoeae (gonorrhoea), human papillomavirus (HPV) and Treponema pallidum (the causative agent of syphilis; hereafter referred to as "syphilis") were considered in less detail because the information available for these infections is inadequate for detailed analysis.

Infection with HIV/AIDS is the fourth leading cause of mortality in the world. Currently, most ( 29.4 million) of the 42 million people globally who are infected with HIV are concentrated in Africa, but epidemics elsewhere in the world are growing rapidly. Prevalence is increasing most swiftly in eastern Europe and central Asia (UNAIDS/WHO 2002). Most of the infections prevalent in 2001 were acquired through heterosexual sexual intercourse. Most people infected with HIV do not know they are infected, making prevention and control difficult. The other STIs included in the burden estimates, C. trachomatis, N. gonorrhoeae, HPV
and T. pallidum (syphilis), cause morbidity in all regions of the world. Infection with some of these agents can lead to infertility (e.g. C. trachomatis) or cancer (HPV), and an acute STI may enhance the transmission of and susceptibility to HIV.

To estimate the prevalence of sexual risk behaviours, suitable studies were located and, where possible, the data produced by these studies were analysed to create a set of standard indicators for different aspects of sexual behaviour. The prevalence of different sexual behaviours and characteristics varies greatly between countries and between subregions. ${ }^{1}$ The levels of risk behaviour did not vary in a predictable manner, and variations in reported behaviour at the aggregate level do not correspond to differences in HIV prevalences. A literature search was also carried out to identify reported risk factors for HIV infection and estimates of the risk associated with each factor. Since the outcomes are infections transmitted from person to person, the relative risk of infection changes with the prevalence of the infection, and changes in prevalence affect incidence.

Two different approaches were used to estimate the avoidable burden of HIV/AIDS attributable to unsafe sex. For countries in sub-Saharan Africa where the prevalence of HIV/AIDS in adults is high and the epidemic is largely driven by heterosexual sex, a mathematical projection model (the Epidemic Projection Package [EPP]) was used to estimate how many infections were attributable to unsafe sex, and how many were potentially avoidable. For countries where adult prevalence is lower and the spread of HIV/AIDS is confined to specific subgroups, a different approach was used whereby current estimates of HIV/AIDS and projections were based on estimates of sub-epidemics related to the mode of transmission (e.g. injecting drug use, men who have sex with men, heterosexual transmission). For these countries the risk associated with unsafe sex was the percentage of all infections that were sexually acquired. The other STIs were assumed to be entirely the result of unsafe sex and therefore $100 \%$ of the burden caused by these STIs is avoidable.

The modelling exercise suggests that there would not have been an HIV epidemic in Africa had there never been any sexual transmission since of the cases of HIV infection prevalent in 2001, >99\% were associated with a sexually acquired infection at some point in the chain of transmission. In the rest of the world, the estimated percentage of the HIV infections prevalent in 2001 that were attributable to unsafe sex ranged from $25 \%$ in eastern Europe (EUR-C) to $95 \%$ in parts of Latin America (AMR-D). Using these estimates, the mortality attributable to unsafe sex ranged from 4000 deaths in EUR-C to 1632000 in AFR-E. Globally, 2444000 deaths and 75783000 disability-adjusted life years (DALYs) were attributable to this risk factor. If unsafe sex were to cease, most parts of the world would see a substantial drop in the number of new HIV infections.

## 1. Introduction

A variety of infectious agents can be transmitted through sexual contact, including HIV, chlamydia, gonorrhoea, HPV and syphilis). While having sex (which in this chapter refers to vaginal sexual intercourse, unless otherwise stated) may place a person at risk of being infected by one or more of these agents, it is difficult to assess the magnitude of this risk. Sex can only be defined as "safe" or "unsafe" if something is known about the context in which it takes place and with whom. Having sex does not place a person at risk of contracting a disease unless that person's partner has an infection, which they can transmit. Therefore, unlike many other risk factors, which are independent of the situation in the broader population, or with respect to other individuals, unsafe sex cannot be uniquely defined by the set of actions of an at-risk individual. Rather, a definition must be based on an analysis of the individual's actions in light of the background prevalence of disease. The principal health outcome considered in this chapter is the number of adults aged 15-49 years who become infected with HIV as a consequence of unsafe sex, and the number of these infections that is potentially avoidable. Infections in children resulting from vertical transmission were not included since these are not caused directly by unsafe sex, but by infection via the mother, together with a lack of pre- and postnatal treatment of mother and child. The other STIs (chlamydia, gonorrhoea, HPV and syphilis) were considered separately and in less detail because they contribute to a lesser degree to the burden of disease and mortality, and because of the limited amount of information available regarding the prevalence and current transmission dynamics of these infections in different subregions. A thorough review of the epidemiology and importance of these STIs was given in previous burden of disease work (Rowley and Berkley 1998).

The relationship between the risk factor unsafe sex and the disease outcomes, which contribute to the global burden of disease, cannot be described using the standard epidemiological tradition of constant, extrapolable hazards. This is owing to the fact that the outcomes considered in this chapter all relate to infections which are transmitted from person to person. The relative risk of being infected by any one of these diseases is therefore dependent on the prevalence of the disease.

### 1.1 Definitions of unsafe sex

In this chapter, STIs were the only negative outcomes of sexual contact considered. Other potentially deleterious outcomes, such as an unwanted pregnancy or the psychological consequences of sexual violence, are considered elsewhere in the CRA (see chapters 15 and 23). It is important to define the group of people who share a common risk factor for contracting an STI in order to be able to carry out a risk assessment. The risk factor has been called "unsafe sex", but this term does not
immediately suggest a clearly defined characteristic of either an individual or a population that can be used to determine how many people are affected by the risk factor. "Safe" sex has previously been defined as follows:

> Consensual sexual contact with a partner who is not infected with any sexually transmitted pathogens and involving the use of appropriate contraceptives to prevent pregnancy unless the couple is intentionally attempting to have a child. (Berkley 1998)

This definition is not useful for the purposes of this chapter, since many of the ways in which the above definition can be negated would not put an individual at risk of acquiring an STI. For example, sex with an uninfected partner without using contraception does not pose a risk of infection and nor would sex with an infected partner if a condom was used properly.

Therefore, before defining unsafe sex we must first consider what type of classification would be suitable to describe the degree of risk experienced by an individual or population. The risk of contracting an STI depends both on the individual and on the population. Individual behaviours determine whether or not it is possible for infection to occur. The prevalence of infection in the population determines whether or not the individual becomes exposed to an infectious agent. Therefore an ideal measurement of this risk would include both individual and population characteristics.

If it were possible to measure this risk at the individual level, a gradation across the population would be observed. Risk gradation suggests the possibility that a continuous index of risk could be constructed by combining several factors. However, many if not most behavioural factors do not retain a simple dose-response relationship when considered in combination with others. For example, consider a person who is not infected with HIV at a particular instant in time. The frequency of this person's sexual relations with a regular partner could show a dose-response relationship relative to the risk of acquiring infection, but only if the partner were infected. Similarly, the rate at which this person acquired new partners could also show a dose-response relationship, but only if each partner were infected. Past partner history would not be relevant, unless the individual had contracted another STI which could enhance the dose-response relationships between risk of infection and both coital frequency and partner acquisition rate. The conditionality of these interactions makes it practically impossible to quantify and construct a continuous measure of risk.

Individuals must therefore be categorized into static groups based on average levels of risk. This can be done so as to allow for the important effect of STI prevalence if the definitions of risk are based on probability of contact with cases, rather than on reportable behaviours. These
definitions will be valid at an instant in time, or for a very short time period; it is important to realize that such distinctions may be very short-lived since sexual networks are dynamic. The following definitions provide a way of thinking about the true distinctions.

## Unsafe Sex

Unsafe sex occurs if a susceptible person has sex with at least one partner who has an STI, without taking measures to prevent infection. Susceptible people are not yet infected, either because the infectious agent has not been successfully transmitted, or because the agent has been transmitted but infection has not yet been established. Such susceptible people form the group which is truly exposed to infection and they are at a very high risk of becoming infected. For intervention and prevention purposes, this group is not as important as that defined below, because it is too late to prevent the members of this group from being exposed to infection. However, this group is the most relevant in terms of predicting the number of new cases of STIs.

## Hazardous sex

The group of people engaging in hazardous sex are those susceptible persons who either engage in unprotected sex but who have not yet encountered a partner who has an STI, or who have had sex with at least one partner who has an STI, but have taken measures to prevent transmission. These people have the potential to be exposed to infection, either by encountering an infected partner, or if the measures taken to prevent transmission are ineffective (e.g. condom failure). This group of people is important for prevention efforts; a change in the size of this group has the most potential to change the number of new infections occurring in the future.

These two definitions ("unsafe" and "hazardous" sex) would provide a way to allocate people to risk groups if membership of the two groups could be measured. However, there is currently no way by which this can be measured and so it is necessary to use a definition based on reportable behaviours, i.e. "risky sex", as a proxy.

## RISKY SEX

The people who have risky sex share a certain set of behaviours; these can be different in different epidemic situations, but are likely to include having many sexual partners and not using condoms. This classification is based on individual reports and will include infected as well as susceptible people, because infection status is not known from such reports. Ideally, we seek to identify reportable behaviours so that the group identified as having risky sex would include those having unsafe sex and those having hazardous sex but exclude others (those having "safe" sex or no sex).

Figure 14.1(a) shows the relationship between groups of infected and susceptible people in terms of those who have unprotected sex, those who have sex with an infected partner, and those who report a "risk behaviour". The reported risk behaviour is the measurable component of an individual's sexual behaviour.

From this figure it is possible to see why classifying people into risk groups on the basis of reported behaviours is not necessarily a good measure of exposure to infection. Some people will be wrongly classified as at risk because they report risk behaviours, but actually they are already infected. Others will be wrongly classified because they report behaviours which have not exposed them to infection, as they did not behave in exactly the way they reported. Some people will be wrongly classified as not at risk because although they have had a sexual contact which could potentially have led to infection, they did not report this behaviour. This could be deliberate, because they do not wish to admit to "undesirable" behaviour, or unintentional, because the behaviour has been forgotten. People will also be misclassified if the risk behaviour they are asked to report is not the best predictor of the actual risk experienced. For example, in a population of married women this could happen if risk were classified on the basis of reporting sex with nonmarital partners, but the main source of infection was in fact the women's husbands.

Figure 14.1(b) shows where the groups defined above (unsafe sex and hazardous sex) fall in this schematic. The black section shows the group having unsafe sex: the susceptible people who have unprotected sex with a partner who has an STI. By definition, this group falls wholly within the group of susceptibles and includes some of the people who report risk behaviours and some of the people who are classified as having hazardous sex. The group which has hazardous sex is shown in the light grey and dark grey sections and is composed of those who either have unprotected sex or who have protected sex with a partner who has an STI. This includes people who do report risk behaviours and some who do not. Again, by definition, this group includes only susceptible persons. Some members of the group who report risky behaviour, shown in white, are not included in either the group having hazardous sex or the group having unsafe sex.

### 1.2 Estimating levels of risky sex in a population

If there were no STIs, then there would be no unsafe and/or risky sex. In areas where there is a high prevalence of STIs, a larger number of sexual behaviour patterns will be dangerous than in places where very few people are infected with a sexually transmitted pathogen. A pragmatic definition of a specific behaviour, or group of behaviours, (e.g. sex without a condom) as "risky" can be useful, providing it is understood that the degree of risk associated with this behaviour will not be the same in different populations, or at different times in the same population.

Figure I4.I Venn diagram illustrating the relationship between three ways of defining unsafe sexual behaviour
(a) Components of risk behaviour

(b) Correspondence between the components of risk behaviours and definitions of unsafe, risky and hazardous sex


With this caveat in mind, the question arises as to how best to describe populations with different levels of risk.

Aggregate measures of sexual behaviour will inevitably be less informative than more local measures. However, even country-level indicators cannot capture the more subtle variations in sexual mixing patterns, such as partnership concurrency. The level of risk attached to a particular behaviour changes with the prevalence of the infection; if prevalence is high, there are more infected people in the population and so a susceptible person has a greater chance of choosing an infected person as their next sexual partner. In each subregion, the prevalence of STIs and
of certain behaviours varies between the countries. Within the different countries, STI prevalence and sexual behaviour can vary between urban and rural areas, age groups and sexes, socioeconomic classes, religious groups, between people of different sexual orientation and according to other factors such as proximity to transport links and health services. Personality and physiology play a significant role in determining a person's sexual behaviour and, in the case of the latter, susceptibility to infection. The impact of these determinants cannot be measured at the population level, but they are of great importance in determining how many people are exposed to STIs and how many people become infected.

The effect of heterogeneity in sexual behaviour on the ability to measure the level of dangerous exposure is more subtle. When sexual behaviour is measured in a survey, data are only collected regarding the respondents' behaviour. However, the behaviour of the sexual partner is as important a predictor of risk as the behaviour of the respondent. A respondent who has a large number of sexual partners is probably at a high risk of contracting an STI. However, if all of these partners have never had sex with anybody else, the respondent is perfectly safe. Therefore in a population where people vary greatly in the number and frequency of their sexual contacts, a one-sided measure of "sexual behaviour" is difficult to interpret. It is known that in most populations men and women have very different patterns of sexual behaviour. Most populations also have a subset of both men and women who are distinguished by high levels of sexual activity. Both of these imbalances make it difficult to quantify risk based on reported behavioural data from surveys.

The level of risk, to oneself and one's partner, is illustrated assuming different patterns of partnership and condom use in three different epidemic situations in Figure 14.2. The epidemic states correspond to those defined in WHO/UNAIDS (2000). A low-level epidemic is one in which, although HIV infection may have been present in the population for some time, it has not spread outside defined groups at a high risk of infection, and the prevalence among these groups has not exceeded $5 \%$. A concentrated epidemic is one in which HIV infection has spread within defined groups and prevalence has exceeded $5 \%$ in at least one of these groups, but prevalence among pregnant women in urban areas remains below $1 \%$. A generalized epidemic is one in which HIV infection has spread throughout the general population, as indicated by a prevalence of infection of greater than $1 \%$ among pregnant women.

A partnership is mutually monogamous if both partners only have sex with each other for the duration of the relationship. Lifetime mutual monogamy is always safe, regardless of the prevalence of STIs in the population. One-sided lifetime monogamy is safe for one of the partners in this type of relationship: individuals who have sex with a partner who has never had sex with anyone else do not place themselves at risk of infection from this partner, but may themselves present a risk to this partner if they have also had sex with other people. Serial monogamy is

Figure 14.2 Risk matrices: the level of risk to an individual and their partner is illustrated assuming different behavioural patterns in different epidemic situations

Low-level epidemic


Generalized epidemic-moderate

| Condom use | No. of partners | Relationships of riskiest partner |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  |  | Lifetime monogamo | Serially monogamous | Concurrent <br> s partnerships |
| Always | 1 |  |  |  |
| Always | Many |  |  |  |
| Sometimes | 1 |  |  |  |
| Sometimes | Many |  |  |  |
| Never | 1 |  |  |  |
| Never | Many |  |  |  |

Concentrated epidemic

| Condom use | No. of partners | Relationships of riskiest partner |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  |  | Lifetime monogamous | Serially monogamous | Concurrent partnerships |
| Always | 1 |  |  |  |
| Always | Many |  |  |  |
| Sometimes | 1 |  |  |  |
| Sometimes | Many |  |  |  |
| Never | 1 |  |  |  |
| Never | Many |  |  |  |

Generalized epidemic-severe


Generalized epidemic-explosive

| Condom use | No. of partners | Relationships of riskiest partner |  |
| :---: | :---: | :---: | :---: |
|  |  | Lifetime monogamous | Serially Concurrent monogamous partnerships |
| Always | 1 |  |  |
| Always | Many |  |  |
| Sometimes | 1 |  |  |
| Sometimes | Many |  |  |
| Never | 1 |  |  |
| Never | Many |  |  |

## Key to risk levels


defined as a succession of monogamous relationships. These relationships are monogamous from the individual's standpoint, but no assumptions can be made about the behaviour of the partner. Partnerships of this sort may last for days or years. The frequency with which partnerships are dissolved and reformed will affect the risk of acquiring an STI and this will also be affected by the prevalence of STIs in the population. If an individual has sexual partnerships that overlap, such partnerships are said to be concurrent. STIs can be spread more easily if people have sex with several partners within a short space of time. Therefore although having concurrent partnerships is associated with the greatest risk of contracting an STI, serial monogamy with very short intervals between successive partners also places the partners at high risk.

## 2. DATA SOURCES

The data used to calculate levels of risky sexual behaviour came from general population surveys designed to be nationally representative. More than 300 surveys were identified that could potentially have been used in this analysis. Many of these had been carried out under the auspices of the Demographic and Health Surveys (DHS) programme conducted by Macro International. The focus of these surveys was family formation and fertility, and only more recently have questions on sexual behaviour been incorporated. Most DHS data are from African countries, but some surveys have been carried out in South America and Asia. South American countries are also covered by the Centers for Disease Control and Prevention (CDC) Reproductive Health Surveys (RHS) which asked questions about sexual behaviour. CDC has also carried out some surveys in Asian and eastern European countries. Other organizations, such as Population Services International (PSI), also carry out surveys which provide suitable information.

Most of the established market economy countries have carried out their own surveys of sexual behaviour, many of which date from the late 1980s and early 1990s, a time when policy-makers began to be concerned about the potential for the spread of HIV in these populations. For example, the data from the United Kingdom of Great Britain and Northern Ireland used in this analysis date from 1990; the survey was repeated in 2000 but the data were not yet available. The problem of standardization is greater for established market economy countries' surveys because they have been carried out by many different organizations, each of which sought different information to address different concerns.

### 2.1 Search strategy

The scientific literature was searched for information on the prevalence of different sexual behaviours and the relationship between risky sex and STIs. Information dating from after 1990 was used wherever possible.

Identifying survey data sets and or reports which incorporated information on sexual behaviour was not straightforward since these terms are not indexed in the major bibliographic databases. Therefore the use of a formal search strategy alone would not have been adequate. Suitable surveys were located in several ways:

STEP 1: WEB SITES OF SURVEYING ORGANIZATIONS
Organizations that carry out surveys that include information on sexual behaviours provide lists of these on their web sites; this was the first source of information for the majority of surveys. These organizations are:

- Demographic and Health Surveys (DHS), Macro International and Measure, USA (http://www.measuredhs.com) and (http://www. measureprogram.org)
- Reproductive Health Surveys (RHS) carried out by CDC, Atlanta, USA (http://www.cdc.gov/nccdphp/drh/gp_surveys.htm)
- Population Services International (PSI), USA (http://www.psi.org/)
- Family Health International (FHI), USA (http://www.fhi.org/)
- Global programme on AIDS (GPA) listed on http://www.unaids.org/ publications/documents/epidemiology/determinants/Survey_Sexual_ Behaviour.doc


## Step 2: SEARCH OF PUbLISHED MATERIALS

Medline, Popline and Web of Science databases were searched for appropriate publications. Other databases providing qualitative information (such as Psychinfo) were not used because quantitative information was considered more important for this work.

## Step 3: Contact with other researchers

This turned out to be the most efficient strategy because researchers involved with one survey frequently knew of other existing surveys.

## Step 4: Internet search using Google

The Google Internet search engine was used, the principle search terms employed being the names of authors of surveys known to have been carried out and the names of institutes likely to have been involved in suitable surveys. It is not useful to carry out Internet searches using keywords related to sex.

## Search terms

- Popline

Search terms used were "sex behaviour", "condoms, male" "condoms, female" "population" "HIV infections". This yielded 709 references, of which 75 were selected.

- Medline

Search terms employed were sexual behaviour, risk, ratio, odds, changes, sexual behaviour, incidence or prevalence, change*, reduction or lower or decline, HIV.

- Web of Science

Search terms used were sexu* and country name. If a search term returned a large number of hits, it was narrowed by adding "risk".

All the databases were searched for information from countries where there was no DHS, CDC, PSI or national (state) survey available.

### 2.2 Prevalence studies: HIV and other sexually TRANSMITTED INFECTIONS

Data on the prevalence of HIV are generally from national surveillance systems. In most countries, women who attend antenatal care clinics (ANC) are tested for HIV anonymously and these data are taken to be representative of the general population in these countries. Other sources of surveillance data include blood donors, STI clinic patients and military recruits. Only ANC clinic prevalence data were used in this work. These data are collected by the United States Bureau of the Census and at the Joint United Nations Programme on HIV/AIDS (UNAIDS) in Geneva, from which the information is disseminated. The quality, coverage, history and competence of national surveillance schemes vary enormously (Walker et al. 2001). Consequently, prevalence data from some areas are more reliable than from others and more recent estimates are generally more reliable than older ones.

WHO collects the available STI prevalence data on a regular basis. However there is a lack of time-series data, which means that mathematical projection models cannot be used to make projections of future prevalence.

### 2.3 Prevalence studies: sexual behaviour

As described above, it is not clear which types of behaviours best define the group of people who are at risk of contracting an STI. Therefore information was collected on all behaviours which might be important in defining this group.

## Measuring sexual behaviour

## Target population

Many surveys of sexual behaviour have focused on high-risk groups within a population. In countries with concentrated epidemics, most STIs occur within these groups, which are often composed of people such as commercial sex workers or men who have sex with men. Unfortunately, the size of these groups relative to the total population is rarely known.

Information from surveys of groups at a high risk of infection cannot be extrapolated to the general population without an accurate estimate of the overall size of the group. General population surveys are unlikely to find a representative sample of members of groups at a high risk of infection and therefore underestimate the prevalence of risk-associated behaviour in a population. Data from groups at a high risk of infection have not been used directly in this work because sufficient information is rarely available to be able to use these data in the context of aggregate national estimates of risk behaviour. Therefore the estimates of the level of exposure could be too low in countries where STIs occur mainly within groups at a high risk of infection.

## Methods of data collection

Sexual behaviour surveys are a fairly recent activity and the best methods for obtaining the required information have not been established. The most appropriate reference period for information on the number and characteristics of sexual partners is not known. There is no reliable method for comparing data collected for different periods of time. For example, somebody who reports having had one partner in the last month has not necessarily had twelve in the last year, but may well have had more than one partner in this time. Asking people for information from a longer time period will introduce a recall bias. This bias could be a problem because people might be more likely to recall partners of longer duration than those with whom contact is more short-term. This could lead to underestimates of the number of more risky sexual contacts.

If sexual behaviour patterns are changing over time, a cross-sectional survey will not give a good estimate of the cumulative lifetime exposure to risk, because the risk exposure of the youngest age groups at the time of the survey will not reflect the risk that the older age groups experienced when they were young. Ideally, current state measurements should therefore be supported by life course measures, even if we have to rely on recall data for the latter.

Surveys which only collect information on the respondent's own behaviour will misclassify some individuals with respect to their risk of acquiring an STI. They will systematically underestimate risk because people who are at a low risk because of their own behaviour could be put at risk by the behaviour of their partner. If people who are at a low risk always chose low-risk partners, then surveys could accurately estimate the proportion of those who are at risk. There is evidence from selected DHS with a couple subsamples that this is not the case, and that there is substantial misclassification of women as at a low risk based on their own behaviour, but who are in fact at risk through their husbands. This is illustrated in Figure 14.3 and supported by the results of other studies (e.g. Rwanda and Kenya, Chao et al. 1994; Hunter et al. 1994).

Figure 14.3 The proportion of married couples in which at least one partner reports having had extramarital sex during the last year, by country


Source: Data from selected DHS.

The accuracy of survey instruments in correctly evaluating people's behaviour is unknown. In studies where cross-sectional household surveys have been validated with in-depth interviews, it has been found that people tend to under-report "undesirable" behaviours (Konings et al. 1995). The age, sex and personal characteristics of the interviewer may also influence the reporting of sensitive information (Malamba et al. 1994), which is likely to include the behaviours of interest. Many surveys find that the number of partners reported by men greatly exceeds the number reported by women. Two factors contribute to this: general population surveys may fail to include the few women who have a large number of partners, and women may consistently under-report how many partners they have had (Glynn et al. 2001).

In choosing which data to use for the risk assessment, the first criterion was that the survey sample should be representative of the general population. Some surveys, mainly those with a demographic focus, only interviewed women, and some were concerned only with ever-married women (women who are currently married or who have been married at some stage in their lives). The latter samples were generally carried out in countries where it is not possible to discuss sexual behaviour
openly and consequently they provide only a limited amount of information. The age range of persons included in the surveys also varied. If a survey was limited to a narrow population in terms of sex, age or marital status, it was only analysed in the absence of a suitable alternative. This was the case for several countries. The surveys used, the populations covered and the data sources for each are listed in Appendix A. The type of information available and the number of countries and subregions covered are listed in Table 14.1.

The surveys referred to in Table 14.1 were carried out between 1989 and 2001. In general, the most recent survey available was used for each country. Individual-level data were required to calculate values for most of the sexual behaviour indicators because these were not usually given in a suitable format in published reports. In some cases, the data used to calculate the estimates presented in this chapter did not come from the most recent survey because such data were not available at the time of writing.

There were very few countries for which more than one survey was suitable for calculating sexual behaviour indicators. If more than one eligible survey for a country existed, the survey providing the most information was used first. Data from different surveys were not combined when calculating any one indicator for a particular country, but the full set of indicators for a country were not always derived from the same survey. The estimates for each indicator were rated according to how directly each indicator could be calculated from the information elicited by the survey questions and the number of assumptions which had to be made in the calculation. In cases where two estimates were available for one indicator, the estimate that was considered better was used.

The responses received in a survey may have been influenced by the manner in which the questions were phrased. The data presented here were derived from responses to several differently-phrased questions and this may have distorted the results. Within subregions this should not be of concern, beyond increasing the uncertainty of the measurement, as it seems unlikely that this error should vary with respect to exposure to STIs. However, a bias may well be introduced when making comparisons between subregions with different styles of questions because questionnaire styles are generally more similar within subregions.

Most general population data only cover heterosexual behaviour. Those surveys which discuss sex between men are generally carried out only among men who have sex with men and the number of these individuals in a population is rarely known. Therefore the behavioural measures collected for this analysis focussed entirely on heterosexual sex. For some subregions where sex between men plays a key role in the epidemic, this is an important omission. However, data are rarely available on behaviour in homosexual men in the subregions with the greatest burden of STIs, and the focus of this chapter has been largely dictated by the epidemic in these countries. As will be explained in more detail below,
Table I4.I Indicators of sexual behaviour and the number of countries and subregions for which relevant data are available, by

| Indicator of sexual behaviour | Denominator | Numerator ${ }^{\text {a }}$ | Information available ( $n$ ) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Countries |  | Subregions ${ }^{\text {c }}$ |  |
|  |  |  | Female | Male | Female | Male |
| Ever had sex | Everyone | Number who say they have ever had sex | 63 | 42 | 13 | 9 |
| Sexually active in the last year | Everyone who has ever had sex | Number who had sex in the last year | 59 | 40 | 12 | 6 |
| Higher-risk sex in the last year | All who have had sex in the last year | Sex with non-co-resident partner in the last year | 47 | 34 | 10 | 8 |
| Condom use last time had higher-risk sex | All who have had higher-risk sex in the last year | People who used a condom last time had higher-risk sex | 34 | 30 | 7 | 8 |
| Men who had sex with a CSW in the last year | All men | Men who had sex with a CSW in the last year | NA | 41 | NA | 10 |
| Condom use last time had commercial sex | Men who report having had commercial sex in the last year | Men who used a condom last time they had commercial sex | NA | 23 | NA | 6 |
| Young people ${ }^{\text {b }}$ having premarital sex in last year | All young people who have never had a co-resident partner (i.e. currently single) | Never had a co-resident partner and had sex in the last year | 53 | 37 | 9 | 7 |
| Condom use last time had premarital sex | All young, single and sexually active people | Young, single, sexually active and used a condom last time had sex | 33 | 29 | 7 | 7 |


| Young people having multiple partnerships in the last year |  | All young people | Young people who report more than one partner in the last year | 31 | 31 | 7 | 7 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Young people's condom use last time had higher-risk sex |  | All young people who had sex within the last year | Young people who used a condom the last higher-risk sex in the last year | 45 | 28 | 12 | 7 |
| Condom use first time had sex |  | All young people who have ever had sex | Young people who used a condom the first time they had sex | 9 | 10 | 3 | 4 |
| Had sex by age 15 years |  | Everyone | First had sex before the age of 15 years | 57 | 39 | 11 | 8 |
| Median age at first sex |  | Everyone | Lifetable median | 65 | 51 | 12 | 11 |
| Condom use last time had marital sex |  | Married people (including co-resident partnerships that are not legal marriages) | Married people who used a condom the last time they had sex with their spouse | 20 | 22 | 6 | 6 |
| Extramarital sex in the last year |  | All married people | Married people who had sex with a non-co-resident partner during the last year | 25 | 25 | 7 | 7 |
| $\geq 2$ non-marital partners in last year |  | All people who have had sex in the last year | Number who report $\geq 2$ partners, with whom they do not live, during the last year | 18 | 18 | 5 | 6 |
| Number of partners |  | Everyone | Mean and median number | 27 | 30 | 7 | 7 |
| CSW Commercial sex worker. |  |  |  |  |  |  |  |
| NA | Not applicable. |  |  |  |  |  |  |
| ${ }^{\text {a }}$ | Only people who can contrib | to the denominator are included in the numera |  |  |  |  |  |
| b | Young people are defined as p | ple aged 15-24 years inclusive. |  |  |  |  |  |
| c | Subregions for which at least | country-level estimate was available. |  |  |  |  |  |

whilst homosexual men are not included in the exposure estimates, they are included in the estimates of attributable and avoidable infections as a result of the modelling approach taken.

## Flow of data

Having identified a suitable survey, the questionnaire (if available) was assessed to ensure that the data would be suitable for inclusion in this analysis. If suitable, the data were obtained and converted (if necessary) for analysis using Stata version 7.0. Variables were created for as many of the standard indicators (those listed in Table 14.1) as was possible for each survey. These were then used to calculate the weighted numbers of people in each category, and the results were exported to a Microsoft Access database.

## Survey design issues

There is no standard survey questionnaire. Even those carried out by the same organization, such as DHS, differ slightly from country to country and from year to year. DHS use a standard questionnaire for each survey round, but countries do not necessarily use all of, or only, the standard questions in their surveys. The standard questionnaire for the round four DHS has departed from the previous standard in the AIDS module and now asks about the previous three partners, in contrast to the prior rounds which asked about marital and non-marital partners. Other surveys have differently-worded questions and a different structure and order of questions. Therefore the data had to be standardized in some way.

Two major problems emerged while trying to compile the responses to different questionnaires to allow comparison. First, the reference period for questions on sexual behaviour varied. The majority of surveys asked about behaviour in the year prior to the survey but a few used different timescales. It is difficult to relate the responses to questions with one reference period to those with another reference period and therefore some of the data could not be used to calculate the standard indicators. Second, questions relating to condom use followed two styles. One style asked about condom use on the last occasion (with a particular partner). The other asks whether condoms were always, sometimes, or never used (with a particular partner). The latter question is impossible to compare between different surveys since it would be necessary either to quantify "sometimes" or to get an estimate of consistency of condom use with different partners. A significant amount of data on condom use could not be included here for this reason. Work has been done on methods for comparing responses to different types of questionnaire design; however, to do this effectively for this analysis would have required many assumptions to be made, and would thus have introduced another possible source of error.

## Standardization of questionnaires

Given the differences in question wording and questionnaire structure, it was not possible to define a set of rules for this process. Table 14.2 shows some of the questions used in constructing the same indicator for different countries.

### 2.4 Outcome studies: sexual behaviour and HIV/AIDS

## Estimating the relative risk of HiV infection in exposed vs nonEXPOSED PEOPLE

The relative risk or odds ratio for various indicators of sexual behaviour has been assessed in a number of general population studies listed in Table 14.3. The accuracy of these estimates is influenced by the following factors.

## Methodological issues

The time at which a person became infected is an important piece of information because it is their behaviour at around that time which is the most relevant when estimating relative risk. People do not usually know that they are infected, let alone when this occurred, so behaviour is seldom measured for the relevant period of time. This could reduce the chances of detecting a real association. Studies which attempt to find risk factors for STIs, in particular HIV, face problems because of cultural unease about discussing STIs. Other problems include a lack of laboratory resources and expertise in geographical areas with high prevalence, as well as the ethical issues involved in serological testing.

The studies which estimate the risk associated with particular behaviours are mostly cross-sectional. If sexual behaviour patterns are changing over time then these surveys, which look for patterns of association between estimates of exposure and prevalence, could produce misleading results. The behaviour reported by HIV-positive people who have been infected for some time, and whose behaviour has changed between the time of infection and the time of the survey, will not reflect their behaviour at the time of infection. The degree to which people are misclassified in this way will depend on the stage of the epidemic (because in the early stages more infections are recently acquired) and on the reference period used in the survey.

This effect could be mitigated if life course measures were also considered. Comparison between life course and the more recent measures could show if behaviours have changed. Some indicators of behaviour are known to correlate with others. For example, age at which the individual first has sex has been shown to correlate with number of extramarital partners later in life (White et al. 2000) and so inconsistencies in this relationship, where this has been previously documented, could point to changing patterns of behaviour.
Table 14.2

| Name of survey | Question asked | Mode |
| :---: | :---: | :---: |
| Number of people who have ever had sex |  |  |
| NEM European Group | Have you ever had sexual intercourse? | FTF |
| NATSAL 1990 (United Kingdom) | How old were you when you first had sexual intercourse with someone of the opposite sex, or hasn't this happened?" | FTF |
| DHS Zambia 1996 | Married: When was the last time you had sexual intercourse with (your husband/the man you are living with)? Not married: When was the last time you had sexual intercourse (if ever). | FTF |
| DHS Kazakhstan 1999 | How old were you when you first had sexual intercourse (if ever)? | FTF |
| PSI Rwanda 2000 | Avez-vous jamais fait l'amour avec une personne de sexe opposé? | FTF |
| Number of people who had sex in the year before the survey |  |  |
| NEM European Group | With how many persons of the opposite sex have you had sex over the last 12 months, even only once? | FTF |
| NATSAL 1990 (United Kingdom) | When, if ever, was the last occasion you had vaginal sexual intercourse with a (man/woman)? | SAQ |
| DHS Zambia 1996 | Married: When was the last time you had sexual intercourse with (your husband/the man you are living with)? Not married: When was the last time you had sexual intercourse (if ever) | FTF |
| DHS Kazakhstan 1999 | When was the last time you had sexual intercourse? | FTF |
| PSI Rwanda 2000 | Quand avez-vous fait l'amour la dernière fois? | FTF |
| Number of men who had sex with a commercial sex worker in the year before the survey |  |  |
| NEM European Group | Have you ever had sex with a person you paid to have sex? If yes: |  |
|  | When was it for the last time? | FTF |
| NATSAL 1990 (United Kingdom) | Have you ever paid money for sex with a woman? If yes: <br> When was the last time you paid money for sex with a woman? | SAQ |
| DHS Zambia 1996 | Have you given or received money, gifts or favours in return for sex at any time in the last 12 months? | FTF |
| DHS Kazakhstan 1999 | Have you ever paid for sex? <br> If yes: <br> How long ago was the last time you paid for sex? | FTF |
| PSI Rwanda 2000 | Au cours des douze derniers mois, avez-vous reçu de l'argent ou des cadeaux en échange des rapports sexuels ou bien avez-vous payé quelqu'un pour faire l'amour avec vous? | FTF |
| Key: FTF, face-to-face; SAQ, self-administered questionnaire. |  |  |

## 3. <br> Estimating levels of sexual Risk behaviour

### 3.1 FACTORS WHICH DETERMINE THE INCIDENCE OF A SEXUALLY TRANSMITTED INFECTION

Worldwide, there is great variation in the prevalence of STIs and in patterns of sexual behaviour, but there is little concordance in the variation between the two. Figure 14.4 shows schematically some of the factors which theoretically determine the incidence of an STI, using the example of HIV. The first box shows societal factors which determine general patterns of sexual behaviour and sexual mixing. The second shows the characteristics which influence whether the sexual contact is potentially infectious, i.e. whether a person is exposed to infection. The third shows the mediating factors, which affect the potential for transmission of infection from the infected partner. The fourth box shows those factors which determine whether or not the contact results in a new infection.

Table 14.3 shows some of the factors which have been found to be associated with HIV infection in the general population in a variety of studies. ${ }^{2}$ The Ugandan samples are from cohort studies, which were designed to elucidate some of these relationships. Table 14.4 shows some of the behavioural changes which have been reported at the same time as observed HIV prevalence has decreased, as has happened in some countries, most noticeably Uganda and Thailand. Changes in HIV prevalence can be attributed to changes in behaviour if incidence has also decreased, but it is difficult to establish if this is the case because prevalence can decline due to excess mortality among people already infected with HIV.

Distal determinants of behaviour

## Age

Age is correlated with whether or not someone is sexually active and the likelihood that their sexual partner is their spouse. In countries where the HIV epidemic is of recent origin, older age groups may have a lower cumulative exposure to infection because most of their past sexual

Figure 14.4 Factors which can influence the incidence of HIV infection

Table 14．3 Factors found to be associated with HIV infection among members of the general adult population．

| $\begin{aligned} & \text { n } \\ & \vdots \\ & \vdots \\ & 0 \end{aligned}$ | ก！ |  |  |  |  | 501 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  | O＋ |  |
|  | uо！วəJu！s！！！ 1 d／s |  |  |  |  | $\bigcirc$ | O＋ | 5 |
|  | ио！ทวəJu！て－ヘSH |  |  | 5 |  | $\bigcirc$ |  |  |
|  |  |  |  |  |  |  | O＋ |  |
|  | әSDəs！p ıəכ｜ | － |  |  |  |  |  |  |
|  | ILS ұиәıиว |  |  |  |  |  |  |  |
|  | ILS Jo K．107s！${ }^{\text {d }}$ |  | 5 | 5 |  |  |  |  |
|  | әиә！¢¢¢ ә！！ |  |  |  |  |  |  |  |
|  | xəs ןDu $\forall$ | $\bigcirc$ |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | әsn mopuo） | $\bigcirc$ |  |  |  |  |  | $\times$ |
|  | su！yous |  |  |  |  |  |  |  |
|  | әsn әи！рวОว |  |  |  |  |  |  |  |
|  | ןочоэр 8и！＞и！， |  |  | 5 |  | $\bigcirc$ |  |  |
|  |  |  |  | 5 | ז\％ | $\bigcirc$ | O＋ | 5 |
|  | flos z．oddns of xəs PDH | O＋ |  |  |  | OH |  |  |
|  | uo！snJsuda poog |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  | 5 | 5 |  |  | $\bigcirc$ |  |  |
|  |  |  | 5 |  |  | OH |  | 5 |
|  | s．aułıDd fo ıəquin |  |  |  |  | $\bigcirc$ | O＋ | 5 |
|  | KıluDd גәм07 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | ｜วлDıд 1 |  |  |  |  |  |  |  |
|  | วร口｜！！＾әр！şno §u！＞＞0M |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | u！！｜snW－uou ：uo！¢̣！｜วy |  |  |  |  |  |  | 5 |
|  | jooyวs fo zno |  |  |  |  |  |  |  |
|  | ио！̣จวnpə әı0W |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | $\bigcirc$ | O＋ |  |
|  |  |  |  |  |  | $\bigcirc$ |  | 5 |
|  |  |  |  |  |  |  |  |  |
|  | 边 |  |  |  |  |  |  |  |
|  |  | 荷 |  | 毞 | 皆 | $\frac{\pi}{4}$ |  |  |



Table 14.4 Changes in behaviour which have been observed concomitantly with a decline in HIV prevalence or incidence

exposure occurred at a time of low prevalence. The association of HIV infection with young age was not seen in all the studies listed in Table 14.3, and where an association was found it was not always in the same direction and was sometimes different for men and women.

Sex
In most cultures, men and women initiate sexual activity at different ages. The typical age difference between partners may be different for men and women. Societies may condone some sexual behaviours for men and not for women. In countries with generalized epidemics, prevalence is usually higher in women than men, especially in younger age groups.

Travel, place of residence, workplace
Factors such as travel, area of residence and occupation or place of work have been measured differently by the various studies. In those studies where these factors were associated with HIV infection (Auvert et al. 2001a; Nunn et al. 1994; Quigley et al. 1997; Seed et al. 1995) they could be acting as proxy measures for potential contact with infected sexual partners. These factors will all influence the number of sexual partners and proportion of available partners who are infected.

## Religion

In the studies which found religion to be associated with HIV infection, the comparison was between Muslims and non-Muslims (Malamba et al. 1994; Nunn et al. 1994; Quigley et al. 1997). There are two characteristics of Muslims which may be relevant to HIV infection: the customary practice of male circumcision and the requirement to abstain from alcohol. The use of alcohol and other mood-altering substances is an independent risk factor in other studies (Auvert et al. 2001b, 2001c; de Gourville et al. 1998; Gregson et al. 2001). The social values incorporated in the Muslim faith may also cause people to have risky sex less frequently.

## Marital status

Married people usually have sex more frequently than unmarried people. In most countries, being sexually active outside of a co-resident (cohabiting) relationship is associated with an increased incidence of HIV infection. Sex between co-resident partners usually carries a lower risk of infection than sex with other types of partner, so prevalence may be lower among married people. This may depend on how much extramarital sex is taking place: the proportion of people who have sex outside marriage varies between countries (see Figure 14.5). In countries where HIV prevalence is high, the surviving partners of people who died of AIDS will tend to have a higher prevalence of HIV infection than the group of people who are still married. However, in some places being currently married has been shown to increase the risk of HIV infection. This may be because married people gain an additional sexual partner at the time of marriage (Auvert et al. 2001b) compared to their nevermarried peers. The increased frequency of (unprotected) sex within marriage may also increase the risk of HIV transmission.

## Contraceptive method mix

Condoms may be used to prevent unwanted pregnancies but many couples choose to use non-barrier methods. In many cultures, condoms are not seen as appropriate for use within marriage or in a long-standing relationship. A pattern commonly seen in developed countries is that initial condom use with a new partner is followed by a switch to oral contraceptives after a few months, e.g. France (Commissariat Général du

Figure 14.5 The proportion of married men and women who report having had sex with someone other than their spouse in the last year, by subregion


Note: The figure shows the range of values, within a subregion, for the national estimates of the proportion of married people who report extramarital sex. The line across the middle of the box represents the median value, the box itself spans the interquartile range and the lines extend to the adjacent values at either end of the interquartile range (only shown where the adjacent values fall outside of this range). Data points which fall outside this range are plotted separately. There are no suitable data for AMR-A, EMR-B and EMR-D, EUR-B and SEAR-D.

Plan: Observatoire régional de santé d'Ile-de-France and Agence Nationale de Recherches sur le SIDA 2001). In some populations, negative attitudes towards condoms may lead to lower levels of use.

## EXPOSURE

## Prevalence of HIV

The proportion of people infected with HIV in the population is the main factor influencing the probability of having sex with somebody who is infectious for HIV.

## Sexual mixing patterns

Partner selection would be described as completely assortative if people always chose partners who were similar to themselves in all the measured respects. However, the way in which people select their sexual partners is usually incompletely assortative, that is, people tend to choose partners who are similar to themselves in most, but not all respects. The
differences may be predictable and some mixing patterns can have significant influence over the spread of STIs. For example, age-mixing in sexual relationships (older men with young women) is thought to be an important factor in accelerating the spread of HIV (Anderson and May 1991).

Traditional STI epidemiology defines "core" and "non-core" groups. The incidence of infection is high in the core groups, and most transmission occurs within these groups. The core group is composed of people who have a large number of sexual contacts compared to the rest of the population. Core groups tend to be small, and as long as infection remains confined to these groups, the population prevalence will remain fairly low. Since mixing patterns are incompletely assortative with respect to frequency of partner change, there will be occasional contacts between members of the core group and the rest of the population. The people involved in these sorts of partnerships are known as "bridge" groups and provide the route by which an infection moves from the core group to the general population. An example of this would be married men who visit commercial sex workers: married men are mostly members of the non-core group, commercial sex workers are members of a core group and the subset of married men who visit the sex workers is the bridge group. Simple measures of partner change and proportions exposed in either group fail to capture variation in density of exposure which arises from non-random mixing (Anderson and Garnett 2000).

## Number of partners

If condom use is not widespread in the population, then having a greater number of sexual partners means being exposed to a greater risk of infection. This is probably not a linear relationship because in many countries a disproportionate number of STIs occurs among the small group of people who have numerous partners. Most of the surveys listed in Table 14.3 found an increasing risk of infection (Auvert et al. 2001b; Chao et al. 1994; Hunter et al. 1994; Quigley et al. 1997; ter Meulen et al. 1992) and seroconversion (the detection of antibodies to HIV in a person who has not previously produced such antibodies, indicating recent infection with HIV) (Gray et al. 2000) associated with increasing numbers of partners. The reference periods were not the same in these surveys so it is not possible to compare the magnitude of the increased risk; this pattern was not clear in all studies. In the Masaka cohort in Uganda, the effect of the number of partners seemed to be modified by age. There was a greatly increased risk associated with having more partners for those aged $<25$ years, but no clear pattern among older people (Malamba et al. 1994). In the Four Cities study, women reporting a greater number of lifetime partners had a significantly increased risk of being infected with HIV in Kisumu (Kenya), Ndola (Zambia) and Yaoundé (Cameroon) but not in Cotonou (Benin) (Auvert et al. 2001b). Only in Ndola (Auvert et al. 2001b) was an increased chance of being HIV-positive observed among men reporting a higher number of lifetime partners.

In Uganda, a reduction in the number of partners does not appear to have been necessary for a decline in prevalence to occur (Asiimwe Okiror et al. 1997; Kamali et al. 2000). In Zambia, localized decreases in the prevalence of HIV among young women attending antenatal care clinics were observed between 1994 and 1998, and the proportion of people reporting large numbers of sexual partners in the same area in coincident general population behavioural surveys was also seen to decline (Fylkesnes et al. 2001).

## Commercial sex

Contact with sex workers, a group that often has a high prevalence of HIV infection, seems mainly to be important outside of Africa. Commercial sex is difficult to define in a meaningful way across cultures because the exchange of money or gifts may generally accompany sex in some cultures, but this may not mean that the woman has a great many partners, or that she is demanding the payment in return for sex.

## Duration of relationships

Sex within a relationship that has been established for a long time is thought to carry a lower risk of HIV infection than sex with a more recently acquired partner. Logically, this would only be the case if both the partners were mutually monogamous throughout the duration of the relationship. It may be that mutually monogamous partnerships last longer than others and that the observed association is a selection effect.

## MEDiating factors

## Male circumcision

In Africa, male circumcision is associated with a lower probability of male HIV infection (Auvert et al. 2001c; Gray et al. 2000; Hunter et al. 1994; Seed et al. 1995; Weiss et al. 2000). There is a plausible biological mechanism for this (Glynn et al. 2001; Royce et al. 1997), although its importance outside of Africa remains to be measured. It is also unclear whether a circumcised, infected man is less likely to transmit the infectious agent to a female partner than an uncircumcised man.

## Sexually transmitted infections

HIV infection is likely to be associated with a history of infection with another STI because these agents share the same mode of transmission. Being infected with an STI indicates that a person has had a sexual contact which could also have led to HIV infection, if their partner was infectious for HIV. However, it has been found that, in addition to providing a marker for this type of contact (Obasi et al. 1999), the presence of another active disease increases the risk of both HIV transmission and infection (Mbopi Keou et al. 2000). In the studies summarized in Table 14.3, relevant information was collected for different diseases in differ-
ent ways. This is because the locally important STIs vary and the setting of the studies imposes restrictions on the information collected. However, in all the studies, having ever had another STI clearly increased the chance of being infected with HIV.

## Condom use

The efficacy of condoms in preventing the transmission of HIV and other STIs has been established (Weller and Davis 2002). However, only one of the studies listed in Table 14.3 (a study carried out among men attending an STI clinic in India (Rodrigues et al. 1995) found a protective association between reported condom use and HIV infection. The reason for this may be that in African countries condom use is actually a marker for risky sex. That is, condoms are only used by those who (rightly) perceive themselves to be at risk of infection. In this case, condom use would only be protective if condoms were properly used at every risky encounter. Condom use would only be revealed as protective in a statistical analysis if this could be confined to those who indulge in risky sex, or if the propensity to have risky sex could be controlled for. If members of groups at a high risk of HIV infection were initially more likely to use condoms, a protective effect would only become apparent as condom use became more widespread in the general population. The availability and acceptability of other methods of contraception might affect the chances of a couple using a condom.

## Sexual practices

Anal sex, both in homosexual male and in heterosexual couples, carries a higher risk of transmission than other practices. It has been suggested that drying the vagina before sex, and having sex during menses also increase the risk of HIV infection in women. However, this has not been clearly demonstrated (Auvert et al. 2001b; Buve et al. 2001a; Malamba et al. 1994).

## Susceptibility

There is a high incidence of HIV infection among young women who have become sexually active at an early age. A partial explanation for this observation may be that young women are particularly vulnerable to HIV infection because the immaturity of the genital tract renders them physiologically susceptible. This is a complex issue, as demonstrated by the results of the Four Cities study, which showed that the high prevalence of HIV infection among young women was not fully explained by behavioural factors (Glynn et al. 2001).

In Europe, transmission from males to females has been observed to be more efficient than vice versa (Anonymous 1992) but this was not confirmed in Rakai (Uganda) (Quinn et al. 2000). This pattern of differential transmission probabilities between the sexes is inconsistent in the rest of the world (Mastro and Kitayaporn 1998).

### 3.2 Choice of indicators of potentially hazardous SEXUAL BEHAVIOURS

Sexual behaviour can be summarized by a variety of different measures and, as described above, many of these measures have been found to be associated with HIV infection. However, it is also clear that these associations are not found in all populations, nor are they consistent in direction and magnitude across those populations in which an association has been observed. The most appropriate indicators of potentially hazardous behaviour were judged to be those which have been associated with HIV infection in different settings, and which:

- are available and relevant for all age groups, both sexes and all subregions;
- describe an important aspect of behaviour in all subregions; and
- are associated with the risk of acquiring HIV infection, or with being already infected with HIV.

First, an empirical approach was used to identify this subset. Popula-tion-level estimates are available for many of these behavioural indicators and estimates of HIV prevalence are also available for many countries. However, it is well known that there is no simple relationship between observed HIV prevalence and reported sexual behaviours at the population level. In many African countries with generalized epidemics, the national prevalence estimates are based on fitting a mathematical model of the HIV epidemic to observed HIV prevalence data acquired from among women attending antenatal care clinics. An estimate of the model parameter representing the fraction of the population that is at risk of contracting HIV infection was extracted from the model and a regression analysis was conducted to examine the association between this estimate and various indicators of sexual behaviour. The model, known as the Epidemic Projection Package (EPP), is described in detail in section 4.

Estimates for the behavioural indicators were calculated for all countries with data, as described in Table 14.1. Suitable model fits were available for 16 countries, and a complete set of indicators and model fits were available for nine countries. Each country contributed an urban and a rural estimate, bringing the sample size for the regression analysis to 18.

All analyses were carried out in Stata version 7.0. Correlation coefficients were calculated for each combination of model parameter and behavioural indicator. The results of these correlations governed which behavioural indicators were included in a linear regression model. This model also included another parameter, which describes the force of infection, as an instrumental variable. It was not possible to quantify a relationship between the behavioural data and the model parameter using this method. The analysis was hampered by the small sample size
and the associations that did emerge as statistically significant were not easy to interpret. Some indicators whose effects would be expected to be similar (such as age at first sex and the proportion of the population who had ever had sex), when included in the same regression model produced opposing coefficients. A robust analysis would require a much larger sample size than was available, given the large number of behavioural indicators and the high degree of correlation between these indicators.

The failure of our work, and that of other groups, to find a suitable quantification suggests that there may be no single relationship between any one measure of sexual behaviour at the aggregate level and the incidence of HIV infection in the general population. Rather, data at the level of individuals and their partners before infection may be required. The choice of which indicators to present was therefore governed by which indicators were commonly found to be associated with HIV infection in observational studies and the measures recommended by UNAIDS (2000), even if the nature of the association with HIV infection was not clear.

### 3.3 Prevalence of potentially hazardous sexual behaviours

Different sexual behaviour patterns are summarized here by three measures: lifetable median age at first sex; mean number of sexual partners in the last year; and the proportion of adults in the subregion who have had sex with a non-co-resident partner within the year preceding the survey, and who did not use a condom the last time they had sex with this partner. All the indicators were calculated for individual countries and the subregional estimates were created by weighting these estimates by the total population size of the country relative to the subregion. The number of countries and sample size used for each estimate are given in Table 14.5.

No subregions were completely described and there were no data at all for some subregions. Values had to be estimated for the missing categories by extrapolation of the results from other subregions; this was based primarily on the values of the available estimates. If no estimates were available for a subregion, the values were extrapolated from the subregion with the most similar proportion of people currently married (Figure 14.6). Throughout the results, extrapolated estimates are shown in the shaded cells as explained in the footnotes of the tables.

## Median age at first Sex

The median age at which people first had sex is presented in Table 14.6. This was calculated from the reported age at first sexual intercourse, or current age for people who have not yet had sex. Lifetable techniques were used to calculate this measure to allow for the inclusion of those people who had not yet had sex. The age of sexual debut is important because it affects the duration of exposure to STIs. There is evidence that
Table 14.5 Numbers of people and countries on which the estimates for the behavioural indicators considered were based, by

| Subregion ${ }^{\text {a }}$ | Ever had sex |  |  |  |  |  | Had sex in the last year |  |  |  |  |  | Sex with non-co-resident partner in last year ("higher-risk" sex) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Females |  |  | Males |  |  | Females |  |  | Males |  |  | Females |  |  | Males |  |  | $N$ countrie |
|  | 15-29 | 30-44 | 45-59 | 15-29 | 30-44 | 45-59 | 15-29 | 30-44 | 45-59 | 15-29 | 30-44 | 45-59 | 15-29 | 30-44 | 45-59 | 15-29 | 30-44 | 45-59 |  |
| AFR-D | $\begin{array}{r} 52570 \\ 14 \end{array}$ | $32091$ | $6679$ $14$ | $\begin{array}{r} 13912 \\ 12 \end{array}$ | $\begin{array}{r} 9512 \\ 12 \end{array}$ | $\begin{array}{r} 5351 \\ 12 \end{array}$ | $\begin{array}{r} 36747 \\ 13 \end{array}$ | $\begin{array}{r} 28471 \\ 13 \end{array}$ | $\begin{array}{r} 6212 \\ 13 \end{array}$ | $8286$ | $8317$ | $4627$ | $\begin{array}{r} 20020 \\ 7 \end{array}$ | $\begin{array}{r} 14323 \\ 7 \end{array}$ | $2664$ | $\begin{array}{r} 6388 \\ 8 \end{array}$ | $\begin{array}{r} 6426 \\ 8 \end{array}$ | $\begin{array}{r} 3431 \\ 8 \end{array}$ | 26 |
| AFR-E | $\begin{array}{r} 52502 \\ 13 \end{array}$ | $\begin{array}{r} 28 \quad 165 \\ 14 \end{array}$ | $\begin{array}{r} 6220 \\ \quad 14 \end{array}$ | $\begin{array}{r} 11645 \\ 10 \end{array}$ | $\begin{array}{r} 6613 \\ 10 \end{array}$ | $\begin{array}{r} 3150 \\ 10 \end{array}$ | $\begin{array}{r} 36076 \\ 13 \end{array}$ | $\begin{array}{r} 26360 \\ 13 \end{array}$ | $\begin{array}{r} 5773 \\ 13 \end{array}$ | $\begin{array}{r} 8304 \\ 9 \end{array}$ | $\begin{array}{r} 6372 \\ 9 \end{array}$ | $\begin{array}{r} 3036 \\ 9 \end{array}$ | $\begin{array}{r} 20807 \\ 8 \end{array}$ | $\begin{array}{r} 13954 \\ 8 \end{array}$ | $\begin{array}{r} 2713 \\ 8 \end{array}$ | $\begin{array}{r} 6374 \\ 8 \end{array}$ | $\begin{array}{r} 5035 \\ 8 \end{array}$ | $\begin{array}{r} 2291 \\ 8 \end{array}$ | 20 |
| AMR-B | $\begin{array}{r} 22461 \\ 5 \end{array}$ | $\begin{array}{r} 16133 \\ 5 \end{array}$ | $\begin{array}{r} 3589 \\ 5 \end{array}$ | $\begin{array}{r} 4900 \\ 3 \end{array}$ | $\begin{array}{r} 2889 \\ 3 \end{array}$ | $\begin{array}{r} 1632 \\ 3 \end{array}$ | $\begin{array}{r} 14256 \\ 5 \end{array}$ | $\begin{array}{r} 15501 \\ 5 \end{array}$ | $\begin{array}{r} 3475 \\ 5 \end{array}$ | $\begin{array}{r} 4384 \\ 3 \end{array}$ | $\begin{array}{r} 2878 \\ 3 \end{array}$ | $\begin{array}{r} 1622 \\ 3 \end{array}$ | $\begin{array}{r} 6424 \\ 2 \end{array}$ | $\begin{array}{r} 6998 \\ 2 \end{array}$ | $\begin{array}{r} 1363 \\ 2 \end{array}$ | $\begin{array}{r} 3452 \\ 3 \end{array}$ | $\begin{array}{r} 2747 \\ 3 \end{array}$ | $\begin{array}{r} 1487 \\ 3 \end{array}$ | 26 |
| AMR-D | $\begin{array}{r} 35205 \\ 5 \end{array}$ | $\begin{array}{r} 23372 \\ 5 \end{array}$ | $\begin{array}{r} 5290 \\ 5 \end{array}$ | $\begin{array}{r} 3789 \\ 3 \end{array}$ | $\begin{array}{r} 2557 \\ 3 \end{array}$ | $\begin{array}{r} 1402 \\ 3 \end{array}$ | $\begin{array}{r} 19544 \\ 5 \end{array}$ | $\begin{array}{r} 21937 \\ 5 \end{array}$ | $\begin{array}{r} 5065 \\ 5 \end{array}$ | $\begin{array}{r} 2758 \\ 3 \end{array}$ | $\begin{array}{r} 2511 \\ 3 \end{array}$ | $\begin{array}{r} 1368 \\ 3 \end{array}$ | $\begin{array}{r} 7058 \\ 1 \end{array}$ | 9003 1 | $\begin{array}{r} 1865 \\ 1 \end{array}$ | $\begin{array}{r} 1461 \\ 2 \end{array}$ | $\begin{array}{r} 1283 \\ 2 \end{array}$ | $\begin{array}{r} 647 \\ 2 \end{array}$ | 6 |
| EUR-A | $\begin{array}{r} 6119 \\ 8 \end{array}$ | $\begin{array}{r} 6975 \\ 8 \end{array}$ | $\begin{array}{r} 88 । \\ 7 \end{array}$ | $\begin{array}{r} 6359 \\ 8 \end{array}$ | $\begin{array}{r} 7090 \\ 8 \end{array}$ | $\begin{array}{r} 922 \\ 7 \end{array}$ | $\begin{array}{r} 5364 \\ 8 \end{array}$ | $\begin{array}{r} 6922 \\ 8 \end{array}$ | $\begin{array}{r} 876 \\ 7 \end{array}$ | $\begin{array}{r} 5603 \\ 8 \end{array}$ | $\begin{array}{r} 6992 \\ 8 \end{array}$ | $\begin{array}{r} 914 \\ 7 \end{array}$ | $\begin{array}{r} 2811 \\ 1 \end{array}$ | 3148 1 | - | $\begin{array}{r} 2817 \\ 1 \end{array}$ | $\begin{array}{r} 3162 \\ 1 \end{array}$ | - | 26 |
| EUR-B | $\begin{array}{r} 4417 \\ 2 \end{array}$ | $\begin{array}{r} 3222 \\ 2 \end{array}$ | $\begin{array}{r} 616 \\ 2 \end{array}$ | - | - | - | $\begin{array}{r} 4362 \\ 2 \end{array}$ | $\begin{array}{r} 3197 \\ 2 \end{array}$ | $\begin{array}{r} 613 \\ 2 \end{array}$ |  | - |  | - | - | - | - | - | - | 16 |
| EUR-C | 475 2 | 2047 | 508 1 | $\begin{gathered} 582 \\ 1 \end{gathered}$ | $\begin{gathered} 566 \\ 1 \end{gathered}$ | $\begin{array}{r} 292 \\ 1 \end{array}$ | $\begin{array}{r} 1219 \\ 1 \end{array}$ | $\begin{array}{r} 2004 \\ 1 \end{array}$ | $\begin{array}{r} 497 \\ 1 \end{array}$ |  | - | - | $1203$ | 1946 | 422 | 385 | $543$ | $\begin{array}{r} 269 \\ 1 \end{array}$ | 9 |
| SEAR-B | - | - | - | - | - | - | - | - | - | - | - | - | 789 | 739 1 | 147 | 546 1 | $498$ | $\begin{array}{r} 82 \\ 1 \end{array}$ | 3 |
| SEAR-D | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 7 |
| WPR-A | 682 2 | 791 | $\begin{array}{r} 365 \\ 2 \end{array}$ | 675 2 | $\begin{array}{r} 639 \\ 2 \end{array}$ | $\begin{array}{r} 321 \\ 2 \end{array}$ | $\begin{array}{r} 655 \\ 2 \end{array}$ | $780$ | $\begin{array}{r} 348 \\ 2 \end{array}$ | 647 2 | $\begin{array}{r} 632 \\ 2 \end{array}$ | 315 2 | 611 | 68 2 | 264 | 616 | 565 2 | $\begin{array}{r} 261 \\ 2 \end{array}$ | 5 |
| WPR-B | 7412 1 | 5355 1 | $\begin{array}{r} 1158 \\ 1 \end{array}$ | - | - | - | 7305 1 | 5250 1 | $\begin{array}{r} 1146 \\ 1 \end{array}$ | - | - | - | - | - | - | - | - | - | 22 |


| Subregion ${ }^{\text {a }}$ | Condom use last higher-risk sex |  |  |  |  |  | Number of partners |  |  |  |  |  | Age at first sex |  |  |  |  |  | $N$ countries ${ }^{\text {b }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Females |  |  | Males |  |  | Females |  |  | Males |  |  | Females |  |  | Males |  |  |
|  | 15-29 | 30-44 | 45-59 | 15-29 | 30-44 | 45-59 | 15-29 | 30-44 | 45-59 | 15-29 | 30-44 | 45-59 | 15-29 | 30-44 | 45-59 | 15-29 | 30-44 | 45-59 |  |
| AFR-D | 3942 | 999 | 102 | 4308 | 1568 | 365 | 23842 | 14041 | 3037 | 11758 | 7151 | 3911 | 43507 | 26158 | 5465 | 16377 | 10519 | 6057 | 26 |
|  | 7 | 7 | 5 | 8 | 8 | 8 | 6 | 6 | 6 | 9 | 9 | 9 | 11 | 11 | 11 | 11 | 11 | 11 |  |
| AFR-E | 4317 | 1414 | 220 | 4178 | 1205 | 315 | 29204 | 15020 | 3239 | 11140 | 6101 | 2948 | 50776 | 27047 | 5581 | 13371 | 7502 | 3516 | 20 |
|  | 7 | 7 | 6 | 7 | 7 | 7 | 8 | 8 | 8 | 8 | 8 | 8 | 10 | 10 | 10 | 7 | 7 | 7 |  |
| AMR-B | 1762 | 863 | 120 | 2571 | 786 | 312 | 11354 | 8766 | 2183 | 2661 | 1856 | 950 | 20490 | 14588 | 3309 | 6582 | 4216 | 2347 | 26 |
|  | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 4 | 4 | 4 | 2 | 2 | 2 |  |
| AMR-D | 680 | 282 | 32 | 600 | 221 | 91 | 14619 | 10665 | 2513 | 1149 | 814 | 429 | 35317 | 23360 | 5192 | 3770 | 2546 | 1427 | 6 |
|  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 5 | 5 | 2 | 2 | 2 |  |
| EUR-A | 806 | 363 | - | 1200 | 419 | - | 5047 | 5601 | 689 | 5305 | 5659 | 731 | 5903 | 6796 | 683 | 4642 | 5487 | 668 | 26 |
|  | 1 | 1 | - | 1 | 1 | - | 10 | 10 | 8 | 10 | 10 |  | 6 | 6 | 5 | 6 | 6 | 5 |  |
| EUR-B | - | - | - | - | - | - | - | - | - | - | - | - | 4389 | 3226 | 643 | 836 | 2108 | 868 | 16 |
|  | - | - | - | - | - | - | - | - | - | - | - | - | 2 | 2 | , | 1 | 1 | 1 |  |
| EUR-C | 240 | 196 | 30 | 237 | 108 | 21 | 2151 | 2130 | 522 | 573 | 555 | 291 | 2097 | 2043 | 515 | 578 | 570 | 290 | 9 |
|  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | , | 1 | 1 | 1 | 1 | 1 |  |
| SEAR-B | - | - | - | - | - | - | 789 | 739 | 147 | 546 | 498 | 82 | 11558 | 15300 | 3603 | 546 | 498 | 82 | 3 |
|  | - | - | - | - | - | - | 1 | 1 | 1 | 1 | I | 1 | 2 | 2 | 2 | 1 | I | 1 |  |
| SEAR-D | - | - | - | - | - | - | - | - | - | - | - | - | 4091 | 3177 | 700 | - | - | - | 7 |
|  | - | - | - | - | - | - | - | - | - | - | - | - | , | 1 | 1 | - | - | - |  |
| WPR-A | 26 | 11 | 10 | 35 | 23 | 9 | 144 | 297 | 231 | 120 | 262 | 206 | 171 | 309 | 249 | 675 | 640 | 321 | 5 |
|  | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |  |
| WPR-B | - | - | - | - | - | - | - | - | - | - | - | - | 7333 | 5380 | 1199 | - | - | - | 22 |
|  | - | - | - | - | - | - | - | - | - | - | - | - | 1 | 1 | 1 | - | - | - |  |
| - No data. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| a Upper number for each subregion refers to number of people, lower number refers to number of countries from which data were available <br> b Total number of countries in the subregion. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Figure I4.6 The proportion of people who are currently married, by age and subregion


Note: The lowest proportion of older women who are currently married is found in EUR-A, despite the fact that EUR-A falls in the middle of the range for the two younger age groups. This could be due to a larger proportion of women who never marry, or a higher incidence of marital dissolution in this subregion compared to the others.
young women are more susceptible to HIV infection and that people who start to have sex at a younger age may have more risky behaviour over a lifetime than those who delay the first time they have sex. Values for AMR-A were extrapolated from Australia and New Zealand. These values were used instead of those for the WPR-A subregion as a whole because the latter subregion is very heterogeneous and AMR-A is very similar to Australia and New Zealand for the other indicators, where there are data. The values for the EMR-B and EMR-D were extrapolated from EUR-C.

Table I4.6 The median age at first sex: lifetable estimates

|  | Females |  |  |  | Males |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Subregion | $15-29$ | $30-44$ | $45-59$ |  | $15-29$ | $30-44$ | $45-59$ |
| AFR-D | 17.3 | 16.5 | 17.1 |  | 19.7 | 19.4 | 20.3 |
| AFR-E | 17.5 | 16.2 | 15.9 |  | 18.9 | 18.2 | 19.3 |
| AMR-A | 17.5 | 17.5 | 19.5 |  | 17.5 | 17.5 | 18.5 |
| AMR-B | 18.6 | 19.5 | 20.2 |  | 16.5 | 16.5 | 16.5 |
| AMR-D | 19.4 | 18.4 | 18.4 |  | 17.5 | 18.0 | 18.5 |
| EMR-B | 20.5 | 20.5 | 20.5 |  | 18.5 | 19.5 | 19.5 |
| EMR-D | 20.5 | 20.5 | 20.5 |  | 18.5 | 19.5 | 19.5 |
| EUR-A | 18.5 | 18.6 | 20.5 |  | 17.8 | 17.8 | 18.3 |
| EUR-B | 19.5 | 19.7 | 20.3 |  | $20.3^{\mathrm{c}}$ | $20.8^{\mathrm{c}}$ | $21.3^{\mathrm{c}}$ |
| EUR-C | 20.5 | 20.5 | 20.5 |  | 18.5 | 19.5 | 19.5 |
| SEAR-B | $19.16^{\mathrm{a}}$ | $19.0^{\mathrm{a}}$ | $18.2^{\mathrm{a}}$ |  | 18.5 | 18.5 | 20.5 |
| SEAR-D | $16.5^{\mathrm{b}}$ | $16.5^{\mathrm{b}}$ | $15.5^{\mathrm{b}}$ |  | 18.5 | 18.5 | 20.5 |
| WPR-A | 18.8 | 18.8 | 20.1 |  | 19.0 | 19.0 | 19.6 |
| WPR-B | 23.5 | 21.5 | 21.5 |  | 20.9 | 20.13 | 19.1 |

[^57]
## SEX WITH NON-CO-RESIDENT PARTNERS

The proportion of all people who report having had sex within the last year, with a partner with whom they do not live, and who did not use a condom the last time they had sex with that partner is perhaps the closest measure of unsafe sex that it is feasible to calculate and is our working definition of risky sex. Sex outside of a cohabiting (co-resident) partnership (within the last year) without using a condom is thought to carry a greater risk of HIV infection than marital sex. As shown in Table 14.7, there are striking variations in the levels of this indicator across the subregions, but they do not follow the pattern of HIV prevalence.

## Mean number of partners during the last year

Table 14.8 shows the mean number of sexual partners in the preceding year in the adult population (aged 15-59 years), regardless of the relationship to any of the partners. Again, there are clear differences between the subregions.

Table 14.7 The proportion of the adult population (aged I5-59 years) who report having had sex with a non-co-resident partner in the last year, without using a condom on the last occasion

|  | Females |  |  |  | Males |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
|  | Subregion | $15-29$ | $30-44$ | $45-59^{\text {a }}$ |  | $15-29$ | $30-44$ |
| AFR-D | 0.116 | 0.061 | 0.049 |  | 0.239 | 0.161 | 0.090 |
| AFR-E | 0.108 | 0.075 | 0.067 |  | 0.230 | 0.111 | 0.094 |
| AMR-A | 0.070 | 0.040 | 0.030 |  | 0.090 | 0.090 | 0.070 |
| AMR-B | 0.110 | 0.057 | 0.055 |  | 0.218 | 0.120 | 0.122 |
| AMR-D | 0.016 | 0.013 | 0.005 |  | 0.289 | 0.140 | 0.117 |
| EMR-B | 0.073 | 0.078 | 0.055 |  | 0.216 | 0.099 | 0.099 |
| EMR-D | 0.073 | 0.078 | 0.055 |  | 0.216 | 0.099 | 0.099 |
| EUR-A | 0.212 | 0.074 | 0.074 |  | 0.267 | 0.119 | 0.119 |
| EUR-B | 0.073 | 0.078 | 0.055 |  | 0.216 | 0.099 | 0.099 |
| EUR-C | 0.073 | 0.078 | 0.055 |  | 0.140 | 0.087 | 0.048 |
| SEAR-B | 0.116 | 0.061 | 0.049 |  | 0.239 | 0.161 | 0.090 |
| SEAR-D | 0.116 | 0.061 | 0.049 |  | 0.239 | 0.161 | 0.090 |
| WPR-A | 0.068 | 0.043 | 0.025 |  | 0.091 | 0.087 | 0.066 |
| WPR-B | 0.068 | 0.043 | 0.025 | 0.091 | 0.087 | 0.066 |  |

a It was assumed that survey data for women aged I5-49 years applied to women aged I5-59 years.
Note: Extrapolated estimates are given in the shaded cells.

## 4. RISK FACTOR-DISEASE RELATIONSHIP

### 4.1 HIV

HIV infection is known to be sexually transmitted. Some sexual practices with an HIV-positive partner carry a greater risk of infection than others. In some populations, there are groups of people who can be identified as having a greater likelihood of being infected with HIV. The factors that govern whether a susceptible person chooses one of these people at a higher risk of being infected as a sexual partner will influence their own risk of infection. Sexual behaviour and its determinants are not easy to measure, and can vary in several dimensions, all of which may be pertinent for HIV transmission.

It is hard to model the impact of changes in exposure for an infectious disease with person-to-person transmission because the risk associated with exposure will change with changes in the prevalence of the infection. A sexual contact is only an exposure if one partner is infected with HIV and the other is not, and the likelihood of this occurring will change as the prevalence of infection changes. The social perception of risk may feedback to behaviour and further contribute to change. There-

Table 14.8 The mean number of sexual partners in the last year reported by the adult population (aged 15-59 years)

| Subregion | Females |  |  | Males |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 15-29 | 30-44 | 45-59 ${ }^{\text {a }}$ | 15-29 | 30-44 | 45-59 |
| AFR-D | 0.679 | 0.764 | 0.738 | 1.106 | 1.336 | 1.097 |
| AFR-E | 0.729 | 0.928 | 0.809 | 0.923 | 1.132 | 1.040 |
| AMR-A | 1.433 | 1.104 | 0.834 | 1.797 | 1.421 | 1.116 |
| AMR-B | 0.643 | 0.915 | 0.826 | 1.276 | 1.316 | 1.154 |
| AMR-D | 0.492 | 0.849 | 0.742 | 1.413 | 1.629 | 1.235 |
| EMR-B | 0.576 | 0.915 | 0.774 | 1.125 | 1.153 | 1.010 |
| EMR-D | 0.576 | 0.915 | 0.774 | 1.125 | 1.153 | 1.010 |
| EUR-A | 1.248 | 0.987 | 0.912 | 1.378 | 1.134 | 1.130 |
| EUR-B | 0.576 | 0.915 | 0.774 | 1. 125 | 1.153 | 1.010 |
| EUR-C | 0.576 | 0.915 | 0.774 | I. 125 | 1.153 | 1.010 |
| SEAR-B | 0.649 | 0.842 | 0.755 | 4.007 | 1.941 | 1.469 |
| SEAR-D | 0.649 | 0.842 | 0.755 | 4.007 | 1.941 | 1.469 |
| WPR-A | 1.236 | 1.077 | 0.900 | 1.792 | 1.229 | 1.039 |
| WPR-B | 1.236 | 1.077 | 0.900 | 1.792 | 1.229 | 1.039 |

a It was assumed that survey data for women aged 15-49 years applied to women aged 15-59 years.
Base: All respondents.
Note: Extrapolated estimates are given in the shaded cells.
fore a relative risk measured for a particular population at a particular point in time is meaningless for another place or point in time, unless the overall prevalence, the epidemic maturity and the degree to which infected and susceptible people mix are almost identical.

An alternative way to predict the future prevalence of an infection which is transmitted from person to person is to use a recursive mathematical projection model to account for the increase in incidence caused by the increase in the number of prevalent cases. Simpler approaches, based on a static risk of infection, will not adequately capture the dynamics of infection over a period of time if prevalence is high, because the risk of infection will change as the prevalence of infection changes.

If prevalence is low, a simpler approach can be justified because the error introduced in the estimates of the number of new infections by ignoring changes in prevalence is much smaller. The size of the error that results from using an approach based on a static level of risk in a highprevalence situation will depend on the prevalence of the infection, the speed at which prevalence changes and the period of time considered. To illustrate the scale of errors introduced by ignoring the nonlinearities of epidemic dynamics, we note that in a population with an

HIV prevalence of $20 \%$, with a concurrent infection rate among HIVnegative people of approximately $3.5 \%$ per year, over a five-year period the prevalence could increase by $2 \%$ or fall by $3 \%$ without any changes occurring in risk behaviour, but depending on the maturity of the epidemic at the time when the HIV prevalence of $20 \%$ was reached. Currently, UNAIDS estimates that the prevalence of HIV in nine African countries is in the order of $\geq 20 \%$ among women attending antenatal clinics in urban areas (and in four countries the prevalence of HIV is $>20 \%$ among women attending clinics in rural areas) (UNAIDS/WHO 2002). The estimates of avoidable infections presented here are for a fiveyear period. To determine the range of probable outcomes, it is essential that a suitable mathematical model be used to derive estimates of new infections for these subregions, both for the "business-as-usual" scenario, and to estimate what may happen under the different counterfactual scenarios.

## Methods for estimating HIV prevalence over a period OF TIME

UNAIDS/WHO make country-specific estimates and projections of HIV infection worldwide and the UNAIDS Epidemiology Reference Group has developed a model to make projections of HIV prevalence. The model has been implemented in a program known as the Epidemic Projection Package (EPP) (The UNAIDS Reference Group on Estimates Modelling and Projections 2002). EPP is designed to represent the evolution of generalized epidemics and so its use for prediction is confined to countries in which generalized epidemics have developed. In this chapter, EPP was used to calculate estimates for the two African subregions (AFR-D and AFR-E) (Table 14.9).

## Reasons for using the EPP model

There are a number of models which could have been used for the CRA, but the EPP model, currently used by UNAIDS, was deemed to be the most appropriate. The other available models include deterministic models, such as AVERT (Rehle et al. 1998), but most of these make no allowance for behaviour change. The GOALS model (http://www.futuresgroup.com), developed by WHO and the Futures Group models the impact of interventions concerning behavioural change, primarily from a programme manager's or policy-maker's perspective, with the focus on the cost-effectiveness of different interventions. This model requires a much larger amount of input data than EPP and is not appropriate for longer-term projections. Most of the models designed to explore the effects of different interventions are more complex than EPP. Additional assumptions (such as profiles of commercial sex work) would have been needed for such models to be used, as sufficient data are not always available.

Table 14.9 Countries for which an EPP model fit is available

| Country | EPP fit available | Country | EPP fit available |
| :---: | :---: | :---: | :---: |
| AFR-D |  | AFR-E |  |
| Algeria | - | Botswana | $\checkmark$ |
| Angola | $\checkmark$ | Burundi | $\checkmark$ |
| Benin | $\checkmark$ | Central African Republic | $\checkmark$ |
| Burkina Faso | $\checkmark$ | Congo | $\checkmark$ |
| Cameroon | $\checkmark$ | Côte d'lvoire | $\checkmark$ |
| Cape Verde | - | Democratic Republic of the Congo | $\checkmark$ |
| Chad | $\checkmark$ | Eritrea | - |
| Comoros | - | Ethiopia | $\checkmark$ |
| Equatorial Guinea | $\checkmark$ | Kenya | $\checkmark$ |
| Gabon | $\checkmark$ | Lesotho | $\checkmark$ |
| Gambia | $\checkmark$ | Malawi | $\checkmark$ |
| Ghana | - | Mozambique | $\checkmark$ |
| Guinea | $\checkmark$ | Namibia | $\checkmark$ |
| Guinea-Bissau | $\checkmark$ | Rwanda | $\checkmark$ |
| Liberia | - | South Africa | $\checkmark$ |
| Madagascar | - | Swaziland | $\checkmark$ |
| Mali | $\checkmark$ | Uganda | $\checkmark$ |
| Mauritania | - | United Republic of Tanzania | $\checkmark$ |
| Mauritius | - | Zambia | $\checkmark$ |
| Niger | $\checkmark$ | Zimbabwe | $\checkmark$ |
| Nigeria | $\checkmark$ |  |  |
| Sao Tome and Principe | - |  |  |
| Senegal | $\checkmark$ |  |  |
| Seychelles | - |  |  |
| Sierra Leone | $\checkmark$ |  |  |
| Togo | $\checkmark$ |  |  |

$\checkmark$ Available.

- Not available (insufficient data points to fit the model; no generalized epidemic in the smaller countries).

Stochastic models are also available, the prime example being STDSIM (Korenromp et al. 2000; van der Ploeg et al. 1998), which is a complex model requiring detailed specification of a range of demographic, biological and behavioural inputs. STDSIM is designed to closely model the HIV epidemic in small communities and would not have been suitable for use at the international level, despite the fact that it does explicitly model changes in behaviour. A limitation of all sto-
chastic models is the need for repeated runs to ensure reasonably stable results. To run a stochastic model for the countries with sufficient data in all regions of the world would have taken a prohibitive amount of time.

## Structure of the EPP model

The mathematical model used is fully described elsewhere (The UNAIDS Reference Group on Estimates Modelling and Projections 2002; UNAIDS Epidemiology Reference Group 2001), but is summarized below using a slightly simplified notation. EPP models both epidemiology, with a feedback loop from prevalence to incidence, and demography, with competing mortality risks and population renewal. This is important because AIDS mortality is a significant factor in the course of the epidemic. The model was deliberately kept simple to allow projections to be based on real data. The model is not subdivided by either age or sex.

Figure 14.7 shows how the model divides a population into three groups (infected, susceptible and at-risk, and susceptible and not-at-risk), and how people can move between these groups over time.

People enter either the at-risk or not-at-risk group on their 15th birthday. Exit from the not-at-risk group is by death from a non-AIDS-related cause. Exit from the at-risk group is either through a non-AIDS-related death or by becoming infected with HIV and moving to the infected group.

Entry to the population at age 15 years occurs at a constant rate, based on birth rates and rates of survival to age 15 years observed in the population being modelled. Adjustment is made for the impaired fertility of women infected by HIV and for the vertical transmission of HIV. HIVinfected children are assumed not to survive to age 15 years. Death rates from causes unrelated to HIV infection are assumed to be constant. Deaths resulting from AIDS are governed by a mortality function based on a Weibull distribution, which gives survival times after HIV infection. The Weibull survival function is based on data from observational studies in Uganda and the median survival time is compatible with data from Haiti, Thailand and Uganda.

The EPP model is controlled by four main epidemiological parameters:

| $t_{0}$ | The start year for the epidemic |
| :--- | :--- |
| $s_{0}$ | The initial proportion susceptible |
| $r$ | Proportionality constant of the force of infection |
| $\phi$ | (phi) | | The relative recruitment rate into the susceptible category |
| :--- |

These parameters interact, but their main influence is exercised on the shape and location of different parts of the epidemic curve. These effects are shown in Figure 14.8.
Figure 14.7 Flow of people through the EPP model


Figure 14.8 The effects on HIV prevalence of changes in the main epidemiological parameters in the EPP model


The main demographic parameters governing the model are:

| $m$ |  | Modal survival age after HIV infection (Weibull level <br> parameter) |
| :--- | :--- | :--- |
| $k$ |  | Shape parameter for Weibull survival function |
| $\mu$ | (mu) | Adult mortality rate from non-HIV related causes |
| $\lambda$ | (lamda) | Proportion of non-infected children surviving to age |
|  |  | 15 years |
| $\nu$ | (nu) | Vertical transmission proportion |
| $\beta$ | (beta) | Birth rate for the adult population |
| $\delta$ | (delta) | Low fertility adjustment for HIV-positive adults |

In the mathematical exposition below, the following variables are also used, but as they are either derived from the formal parameters listed above, or treated as constants in the normal use of the model, they are not regarded as formal parameters. These are defined below.

Auxiliary constants:
$\begin{array}{lll}\Delta & \text { (Delta) } & \text { Time increment for differential equations } \\ \varepsilon & \text { (epsilon) } & \text { Initial exogenous force of infection }\end{array}$
Endogenous variables, dependent on formal parameters:
$\theta(t)$ (theta) Force of infection between susceptible and infected, at contact time $t$
$\sigma(x)$ (sigma) Proportion of infecteds surviving x years after infection

Finally, the numbers and proportions of not-at-risk, susceptible and infected persons at time $t$ are denoted as shown below:

| Number | Proportion |  |  |
| :--- | :--- | :--- | :--- |
| $N(t)$ | Not-at-risk population | $n_{t}$ | Not-at-risk proportion |
| $S(t)$ | Susceptible population | $s_{t}$ | Susceptible proportion |
| $I(t)$ | Infected population | $i_{t}$ | Infected proportion $=$ <br> prevalence |
| $P(t)$ | Total population |  |  |
| $F(t)$ | 15-year olds entering <br> population | $f_{t}$ | 15 -year-old proportion <br> susceptible |
|  | $1-f_{t}$ | 15 -year-old proportion <br> not-at-risk |  |

The dynamics of the system are given by the following equations. The number of people aged 15 years entering the adult population at time $t$ is the number of uninfected children born 15 years ago multiplied by the probability of surviving to age 15 .

$$
F(t)=\lambda \beta[N(t-15)+S(t-15)+(1-v) \delta I(t-15)]
$$

The proportion of susceptible people entering the population at time $t$ is a function of the overall current proportion of the adult population that is not at risk, governed by the formal parameters $f$ and $s_{0}$, the initial proportion of susceptible people.

$$
f_{t}=f\left(n_{t}\right)=\frac{\exp \left(\phi\left[n_{t}-1+s_{0}\right]\right)}{\exp \left(\phi\left[n_{t}-1+s_{0}\right]\right)-1+1 / s_{0}}
$$

Note that since at time zero there are no infected persons, $1-n_{0}=s_{0}$, so for any value of $\phi, f_{0}=s_{0}$. Similarly, if $\phi=0$, then the proportion of susceptible 15 -year olds is the same as the initial proportion of those who are susceptible at all times, $f_{t}=s_{0}$. If $\phi<0$, recruitment to the susceptible group declines over time; if $\phi>0$, recruitment increases.

The Weibull function gives the proportion of those infected surviving $x$ years after infection.

$$
\sigma(x)=\exp \left(-\mu x-[x / m]^{k}\right)
$$

The force of infection at the $t$ is given by:

$$
\begin{array}{ll}
\theta(t)=\varepsilon & \text { for } t=0 \\
\theta(t)=r \cdot i_{t} & \text { for } t>0
\end{array}
$$

Having defined these variable components, it is now possible to formulate the change-of-state equations governing transitions between the population classes.

$$
\begin{aligned}
& \frac{\Delta N(t)}{\Delta t}=\left(1-f_{t}\right) F(t)-\mu N(t) \\
& \frac{\Delta S(t)}{\Delta t}=f_{t} F(t)-[\theta(t)+\mu] S(t) \\
& I(t)=\int_{x=0}^{t} \theta(x) S(x) \sigma(t-x) d x
\end{aligned}
$$

The last of these equations is presented as an integral equation rather than a differential, because this is the easiest way to express the fact that the infected population consists of survivors who were infected at a range of times in the past.

## Fitting the EPP model to surveillance data

The four epidemiological parameters ( $t_{0}, s_{0}, r$ and $\phi$ ) were fitted to prevalence data from antenatal clinic surveillance using maximum likelihood fitting. The model was fitted twice for each country, once for the clinics in urban areas and once for those in rural areas.

## Alternative implementation of EPP model

The EPP package is designed for use by national AIDS programme managers, to help validate the UNAIDS estimates and projections. Not all the underlying calculations and parameter estimates that are needed for this chapter are the outputs of the standard EPP package, which makes the epidemic scenarios defined by the counterfactual assumptions difficult to create. Therefore, an alternative implementation of the same mathematical model was created as a spreadsheet using the Microsoft Excel program.

## Current levels and projections of HIV prevalence

## Subregions with a bigh prevalence of HIV

Estimates of the current levels of HIV infection and projections of future levels are necessary to be able to calculate the proportion of these infections that is attributable to unsafe sex and thus the proportion that is potentially avoidable.

The current estimates and projections of HIV prevalence in the African subregions (under the baseline scenario of no behaviour change) were based on fits of the EPP model to antenatal clinic surveillance data. These projections were prepared by UNAIDS/WHO. The parameter estimates from these model fits were used in the spreadsheet version of the model to calculate the future prevalence, incidence and number of infections for each of the countries concerned. Subregional estimates were created by combining these estimates, weighted by the total population of each country. Weighted estimates were used because the EPP model could not be fitted for those countries with insufficient data on prevalence (11
countries). It is important to note the time scales used in making the model-based estimates. The last available prevalence estimate generally exerts more leverage on the fitted curve than do other points, and a more robust fit is generally obtained when more data points are used. Therefore prevalence estimates for 2001 were included where available and 2001 was taken as the base year for all projections. The projections of avoidable infections extend until 2006 because the model is designed to give reasonably accurate short-term predictions over a five-year period.

## Other subregions

For the 146 countries in which the prevalence of HIV/AIDS is low, a different approach was used to model the epidemic. For countries with epidemics that are concentrated in groups with higher-risk behaviour (e.g. men who have sex with men; injecting drug users, sex workers and their clients), a three-step process was followed to produce the current estimates (for the end of 2001) of HIV/AIDS prevalence. First, for each country, groups at the highest risk of acquiring HIV/AIDS were identified and estimates of the size of these groups were made. Next, estimates of point prevalence were made by applying the most recent prevalence rates for these groups to the populations. Finally, prevalence in populations at a lower risk of infection was estimated by allowing for transmission from high to low groups via sexual mixing. This estimate was made in one of the following ways. For countries with data from pregnant women, an adjusted prevalence rate from this group was applied to the number of women of reproductive age (aged 15-49 years) to produce an estimate of the number of women infected via sex with a partner from a group at an increased risk of being infected with HIV. Alternatively, for some countries where the epidemic was more recent and there were no data for populations at a lower risk, assumptions were about the number of infected people at a higher risk who had sexual partners with no other risk of infection. A transmission probability was then applied to produce an estimate of the number of women infected via sex with a partner from a group at a higher risk of being infected with HIV.

Projections of the extent of these epidemics up until 2006 were based on assumptions about saturation levels for each of the groups at a higher risk of infection, the time to saturation, and the spread over time from populations with a high risk to populations with a low risk of being infected with HIV.

For these same 146 countries (excluding countries with a generalized epidemic where EPP was used), trends in prevalence of HIV among groups at a high risk of infection were compiled and compared. Saturation levels for each risk group and time to reach saturation were determined by reviewing available data from countries in the subregion. The particular level of, and time to, saturation were applied to the risk groups
in each country based on current level of prevalence and rate of increase in the groups, and by comparison with saturation levels and rates in neighbouring countries.

Using this approach, we projected low growth for countries with longrunning and relatively stable epidemics (e.g. Brazil, Myanmar, the United Kingdom). For countries with recent epidemics, but rapid rates of growth, the projections show much higher rates of increase (e.g. China, Estonia). For all of these countries, we assumed that there was no general heterosexual transmission except from individuals in groups at a higher risk of infection to their lower-risk sexual partners. These procedures, which have been previously described, gave us projections of adult HIV prevalence over time (Stover et al. 2002).

Estimates of incidence were made by using assumptions about survival (median adult survival for those without highly active antiretroviral therapy-HAART—was nine years), growth of populations and levels of accessibility to treatment with HAART. The specific assumptions and procedures used to translate prevalence into estimates of incidence and mortality have been described in detail elsewhere (Stover et al. 2002; The UNAIDS Reference Group on Estimates Modelling and Projections 2002; Walker et al. 2003).

## Attributable infections and disease burden

In most subregions, some data were available on the probable mode of transmission for at least a sample of prevalent infections. These data have been used to estimate how many infections were sexually acquired in each subregion. The estimated burden due to unsafe medical injections and blood transfusions was taken from chapter 22 and from a WHO/UNAIDS review of blood safety. To calculate the proportion of infections that results from unsafe sex, the numbers of all people who, according to the model, were infected via unsafe blood transfusions, unsafe medical injections (based on the subregional level estimates) or due to injected drug use (based on country-level estimates) were combined to form the group infected via non-sexual transmission. The number of infections remaining, i.e. those acquired via sexual contact (either heterosexual or homosexual), was divided by the total number of infections to give the percentage of infections due to unsafe sex.

However, to estimate how many of the HIV infections prevalent in 2001 were truly attributable to unsafe sex, it is not enough to simply calculate how many infections arose from unsafe sex at a particular point in time. The burden of infections which result from unsafe sex is determined by the total number of cases of sexually transmitted HIV infection that have arisen since the beginning of the epidemic. In countries with low-level epidemics, estimates of attributable infections based on the mode of transmission of prevalent cases and estimates which account for the effects of past sexual transmission will be broadly similar. In
countries where prevalence is high, there will be a greater discrepancy between the two estimates. We calculated additional estimates for countries with a high prevalence by re-running the EPP model using the fitted value of the $s_{0}$ parameter (the initial proportion at risk) reduced to $5 \%$ of its original value. This value was used because it is estimated that $5 \%$ of HIV transmission is due to unsafe injections and blood transfusion in these countries (all in the WHO African Region). This estimate is based on the probable mode of transmission for existing infections. Estimates of HIV prevalence based on this reduced value of $s_{0}$ demonstrate what might have happened had there never been any sexual transmission in these populations. The difference in the number of infected persons estimated in 2001 and the number predicted by the model for 2001, under the altered circumstances, was taken to be the number of infections which were attributable to sexual transmission (see Table 14.12). The results shown for the two African subregions correspond to attributable burden, as defined by the CRA methodology. The results presented in Table 14.12 for the other subregions are an approximation of attributable burden, based only on the exposure of prevalent cases. To obtain better estimates of attributable burden in these subregions, we would need information on the patterns of sexual mixing between the groups at a high risk of infection and the general population for the duration of the epidemic.

The fraction of infections attributable to unsafe sex was applied to the mortality and disease burden (Table 14.13). Prevalent HIV infections are the result of episodes of HIV transmission which occurred over the 15 or so years before measurement. Prevalent AIDS cases and recent deaths will be, on average, the product of transmission patterns from approximately 10 years before measurement (in populations where there is no treatment for AIDS). The estimates for the non-African subregions are based on the assumption that the ratio of sexual to non-sexual transmission has not changed significantly over that time. The model-based estimate for Africa accounts for this possibility. If the ratio of sexual to non-sexual transmission has changed significantly over time, the estimates of attributable disease burden based on the current ratio may be inaccurate.

## AVOIDABLE INFECTIONS

## The counterfactual exposure scenarios

As described earlier, it was not possible to measure relationships between specific behaviours and the risk of acquiring HIV infection in a way that could be generalized to all populations. It may be that consistent relationships of this sort do not exist. Therefore counterfactual exposure scenarios cannot be defined in terms relating to measurable changes in behaviour. Predicting changes based on hypothetical scenarios, which are
not linked to a particular group of behaviours but to corresponding model parameters, is the best possible method for estimating how many future infections are potentially avoidable.

The counterfactuals were defined in a way that could be applied in subregions with both low and high prevalence. The counterfactual scenarios selected relate to decreases in the number of people having unsafe sex as represented by model parameters. The scenarios were chosen to provide a range of estimates based on proportional changes in the size of the at-risk group. The counterfactuals were operationalized differently for countries with low and high prevalence because the methods used to project future HIV prevalence in the two situations require different inputs. Three levels of reduction in unsafe sex were used in the calculation of the avoidable proportion of future infections: $100 \%, 50 \%$ and $10 \%$. These levels were achieved by estimating what would happen if all, $50 \%$ or $10 \%$ of the people who were having unsafe sex immediately ceased doing so. In theory, the intermediate counterfactual scenarios ( $50 \%$ and $10 \%$ reductions) could have been engineered to describe a situation in which those who were having unsafe sex reduced the amount of unsafe sex that they were having. The net effect would be the same because the approach is based on person-time at risk, and assumes that length of exposure is proportional to risk of infection.

## Reversibility

Infection with any of these STIs need only occur once to produce disease. Therefore removing exposure to the STI will automatically reduce the risk of infection with immediate effect and this is demonstrated by the results of the HIV prevalence projections under the different counterfactual scenarios. However, in reality it is unlikely that all exposure could be removed and the spread of infection reversed at a particular point in time. Infectious people will remain in the population even if all risky behaviour ceases. Unless every person with an infection (married and unmarried alike) stopped having sex without a condom (i.e. if there was no unsafe sex) they would continue to infect new people. This is the reason for considering counterfactual scenarios that include partial reduction in unsafe sex, as described above.

## Countries with a bigh prevalence of HIV infection

It is possible to define counterfactuals in terms of changes in the size of the EPP model's at-risk group for the countries in the two African subregions. Reductions were made to the size of this group at the start of 2001, first by moving a specified fraction of the at-risk group to the not-at-risk group, and second, by slowing recruitment to the at-risk group by the same amount. Three reductions in the original size of the at-risk group were made: $10 \%, 50 \%$ and $95 \%$. The greatest reduction (resulting from total cessation of unsafe sex) thus resulted in only $5 \%$ of the
original at-risk group remaining at risk after 2001 and recruitment to this group was cut to $5 \%$ of its former level. The size of this group was not reduced to zero because a certain fraction of HIV-infected people will continue to contract their infection through a non-sexual mode of transmission in a non-generalized epidemic. This proportion is estimated to be $5 \%$ of infections in sub-Saharan Africa. While some people will contract their infection in one way, and transmit it in another, the degree to which this happens cannot be estimated for this work.

The ratio of sexual vs non-sexual transmission among those already infected is known, but the ratio of sexual vs non-sexual exposure among the uninfected is not. Implicit in the use of a $95 \%$ reduction in the atrisk group as the theoretical minimum level of unsafe sex is the assumption that these ratios are the same. This may be incorrect because if a mode of transmission is very efficient (e.g. infected blood transfusion) then the incidence of infection among susceptible people who are exposed in this way may be higher than that among people who are otherwise exposed to HIV infection. If different modes of transmission have different rates of infection, the modes most likely to produce an infection will be over-represented among cases of infected people in comparison with the distribution of the different exposures among uninfected people. If the non-sexual modes of transmission in Africa are significantly more efficient than sexual transmission, then the fraction of the at-risk group which is exposed to HIV infection via sexual transmission may be $>95 \%$. However, the opposite could also be true if unsafe medical injections were the most common form of non-sexual transmission; such injections may be associated with a lower infection rate because the reused syringe does not always come into contact with the body fluids that could potentially transmit the infection. There is no means to assess the relative exposure to the different modes of infection and we must instead rely on the data from HIV-infected people, therefore $95 \%$ is an uncertain assumption.

The changes to the model were made through the $s_{0}$ parameter, and not the $r$ parameter because the latter represents the transmission of infection, and the former describes the fraction of the population that is at risk of infection. Conceptually, transmission can be affected by changes in the level of unsafe sex (e.g. the proportion of sexual acts protected by condoms) but this could not be used satisfactorily to describe a counterfactual scenario. To model a total cessation of unsafe sex, we could not reduce $r$ to zero because this would correspond to a total cessation of all HIV transmission. It is not possible to calculate a value of $r$ which is related to the cessation of sexual transmission only. Two implicit assumptions in this approach are worthy of comment:

- a proportionate relationship between hazardous and unsafe sex: we have assumed that, when we reduce the size of the group of people
who have hazardous sex, the size of the group having unsafe sex will decrease by the same amount.
- random mixing in the at-risk population: the EPP model assumes random mixing, i.e. each person in the population has an equal chance of contacting another member. This assumption gives a good representation of the natural dynamics of a generalized epidemic of an STI. The question arises, in relation to the counterfactual scenarios, of whether random mixing is still a reasonable assumption in relation to the "hard core" of those remaining at risk after the sudden decrease in hazardous sexual behaviour. We would argue that in the case of sub-Saharan Africa, where the alternative modes of transmission are predominantly unsafe medical injections and unsafe blood transfusions, random mixing is still a close approximation. In the case of injecting drug users, one might want to model far more intensive contacts within the group of people at risk than outside of it, but use of injected drugs is not as important in these subregions as it is elsewhere in the world.

The spreadsheet (Excel) implementation of the EPP model was used, after the modifications described below were made in order to include the changes described by the different counterfactual scenarios. Recruitment of new members into the at-risk group was slowed by a specified amount, as defined in the counterfactual, starting in 2001 and continuing until the end of the projection in 2006. The slowing of the recruitment to the at-risk group was achieved by reducing the value of the $s_{0}$ parameter by the specified fraction, with effect from 2001. All the other model parameters remained unchanged. At the start of 2001, the size of the at-risk group was reduced to the fraction of its former size defined in the counterfactual, and the people removed from this group were added to the not-at-risk group. These modifications had the effect of reducing the pool of people who could potentially become infected with HIV, and therefore lowered the number of new cases occurring. Figure 14.9 shows how these modifications affected the projected infections. The dynamic relationship between the at-risk and infected groups remains the same, but the relative sizes of the two groups of susceptible people (at-risk and not-at-risk) are drastically altered and the rates of recruitment to both groups are changed.

Although the number of future infections would be small in the absence of unsafe sex, it was necessary to use a model to estimate the avoidable infections for the African subregions for two reasons. First, one of the counterfactual scenarios involves a reduction of only $10 \%$ in unsafe sex, which means that prevalence and the number of new infections remain high. Second, the relationship between current prevalence and the number of new infections in the future is not linear, even over a five-year period.
Figure 14.9 Effect of the modifications made to the EPP model which were used to calculate the number of new HIV infections occurring under different counterfactual scenarios


Countries with a low prevalence of HIV infection
The counterfactuals for other subregions were again engineered to correspond to situations in which unsafe sex was reduced by $10 \%$, $50 \%$ and $100 \%$ (no unsafe sex). Existing data on the distribution of prevalent HIV infections by mode of transmission was applied to the projections for the countries with a low prevalence of HIV infection. Reductions in unsafe sex were assumed to result in a decreased number of new STIs that were equal in proportion to the reduction in unsafe sex.

### 4.2 Other sexually transmitted infections

Estimation of the relationship between unsafe sex and other STIs (chlamydia, gonorrhoea, syphilis and HPV) is subject to the same constraints as that between unsafe sex and HIV infection. Relative risks of infection with chlamydia, gonorrhoea and syphilis following certain behaviours have been estimated. However, like HIV infection, these relative risks will change as the prevalence of infection changes. This problem is compounded by an even greater lack of information for any of these STIs than for HIV. As a result, we have not attempted to quantify this relationship, and assume that for all these STIs, by definition, all current prevalent infections are attributable to unsafe sex. Therefore, the total burden of disease attributed to these STIs can be considered to arise from unsafe sex. This includes cervical cancer attributable to infection with HPV; recent work suggests that all cases of cervical cancer are attributable to infection with sexually transmitted HPV (Walboomers et al. 1999).

In order to make a reasonable estimate of the future prevalence of these STIs, it is necessary to use a mathematical projection model. In common with that for HIV, such a model would need to be fitted to existing time-series prevalence data to create a projection of the future levels of infection. Since there is no appreciable mortality as a consequence of most of these other STIs, a suitable model would be much simpler than those used for HIV. Cervical cancer due to HPV infection would necessitate a model which accounts for mortality. However, the necessary time-series prevalence data are not available for a sufficient number of countries to make this a viable approach. The methods used to calculate the number of new HIV infections that are potentially avoidable cannot therefore be used for these STIs.

STIs have been virtually eliminated from some populations in the recent past. In the early 1950s, the Chinese government initiated a programme to eradicate sexually transmitted diseases that was successful in the short term. The campaign relied on mass screening to identify and treat people with an STI and also involved the abolition of commercial sex work. The methods used might not be transferable to other cultures, but demonstrate that the problem of STIs can be confronted. It has been
suggested that the incidence of STIs only began to increase after China resumed more open relations with the rest of the world in the early 1980s (Cohen et al. 1996).

With this in mind it seems reasonable to assume that all STIs are avoidable, given appropriate changes in sexual and treatment-seeking behaviour, if these changes are accompanied by the provision of suitable services.

## 5. Results

### 5.1 Prevalence of disease outcomes in 2001

Estimates of the current prevalence of HIV and other STIs were based on reported estimates from HIV surveillance and published studies. These were compiled and used to create subregional prevalence estimates (Tables 14.10 and 14.11). The estimates for the two African subregions were based on EPP model fits to antenatal clinic surveillance data. The prevalence in the other subregions was directly based on reported prevalence according to a variety of empirical sources (U.S. Census Bureau 2001).

Table 14.10 The prevalence of HIV infection in the adult population (aged 15-49 years) by subregion, in 2001

| Subregion | HIV prevalence (\%) |
| :--- | :---: |
| AFR-D | 5.05 |
| AFR-E | 11.97 |
| AMR-A | 0.60 |
| AMR-B | 0.55 |
| AMR-D | 1.93 |
| EMR-B | 0.04 |
| EMR-D | 0.35 |
| EUR-A | 0.28 |
| EUR-B | 0.03 |
| EUR-C | 0.73 |
| SEAR-B | 0.45 |
| SEAR-D | 0.63 |
| WPR-A | 0.04 |
| WPR-B | 0.15 |
| World | 1.20 |

Table I4.II The prevalence of chlamydia, gonorrhoea and syphilis in the adult population (all age groups) by subregion, in 2000

|  | Females |  |  |  | Males |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Chlamydia <br> $(\%)$ | Gonorrhoea <br> $(\%)$ | Syphilis <br> $(\%)$ |  | Chlamydia <br> $(\%)$ | Gonorrhoea <br> $(\%)$ | Syphilis <br> $(\%)$ |
| AFR-D | 0.50 | 0.50 | 0.09 |  | 0.47 | 0.47 | 0.07 |
| AFR-E | 0.27 | 0.29 | 0.07 |  | 0.25 | 0.27 | 0.06 |
| AMR-A | 1.05 | 0.41 | 0.03 |  | 0.89 | 0.36 | 0.03 |
| AMR-B | 0.44 | 0.36 | 0.14 |  | 0.37 | 0.30 | 0.11 |
| AMR-D | 0.42 | 0.32 | 0.14 |  | 0.34 | 0.26 | 0.11 |
| EMR-B | 0.67 | 0.22 | 0.02 |  | 0.48 | 0.16 | 0.02 |
| EMR-D | 0.45 | 0.15 | 0.02 |  | 0.37 | 0.13 | 0.01 |
| EUR-A | 0.16 | 0.03 | 0.00 |  | 0.14 | 0.03 | 0.00 |
| EUR-B | 0.20 | 0.10 | 0.01 |  | 0.20 | 0.10 | 0.01 |
| EUR-C | 0.64 | 0.36 | 0.01 |  | 0.60 | 0.33 | 0.01 |
| SEAR-B | 1.53 | 0.55 | 0.10 |  | 1.15 | 0.42 | 0.08 |
| SEAR-D | 1.98 | 1.49 | 0.17 |  | 1.51 | 1.16 | 0.14 |
| WPR-A | 0.48 | 0.37 | 0.02 |  | 0.61 | 0.49 | 0.02 |
| WPR-B | 0.24 | 0.14 | 0.01 | 0.20 | 0.11 | 0.01 |  |
| World | 0.62 | 0.41 | 0.06 | 0.76 | 0.50 | 0.07 |  |

### 5.2 Attributable infections and disease burden

The subregional estimates of the fractions of all HIV infections that are attributable to unsafe sex are given in Table 14.12. These comprise the percentage of infections prevalent in 2001 that were reportedly acquired through sexual contact. Therefore this fraction is directly attributable to unsafe sex. The feedback between prevalence and incidence has not been taken into account in the estimates for subregions outside Africa: in many of these subregions, the attributable fraction could be considerably higher if it included all infections for which sexual transmission had occurred at any point along the chain of transmission. As described above, the fractions were by definition $100 \%$ for other STIs.

### 5.3 Avoidable infections

The estimates of the fraction of infections which is potentially avoidable are given in the following tables and figures. Figure 14.10 shows the proportion of new infections that may be prevented by different reductions $(100 \%, 50 \%, 10 \%)$ in the level of unsafe sex relative to the number of infections which would be expected to occur if there were no change in sexual behaviour. The height of the bar shows the total proportion that could be avoided if there was no unsafe sex. The proportions within the bar show the reductions that would be seen if unsafe sex was reduced

Table 14.12 The proportion of prevalent HIV infections in adults (aged 15-49 years) that is attributable to unsafe sex, by subregion, in 2001

Subregion
\% of HIV prevalence attributable to unsafe sex
AFR-D $>99$
AFR-E $>99$

AMR-A 72
AMR-B 85

AMR-D 95

## EMR-B

 42EMR-D 85
EUR-A ..... 59
EUR-B ..... 64
EUR-C ..... 25
SEAR-B ..... 73
SEAR-D ..... 78
WPR-A ..... 94
WPR-B ..... 52
World ..... 90

Figure 14.10 The proportion of new HIV infections currently predicted to occur during 2002-2006 that could be prevented by different reductions in the practice of unsafe sex, by subregion


Subregion

Table 14.13 The mortality and burden of disease attributable to sexually transmitted infections, cervical cancer and HIV, by subregion, in 2001

| Subregion | STIS |  | Cervical cancer |  | HIV |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mortality (000s) | $\begin{aligned} & \text { DALYs } \\ & (000 \mathrm{~s}) \end{aligned}$ | Mortality (000s) | $\begin{aligned} & \text { DALYs } \\ & (000 \mathrm{~s}) \end{aligned}$ | Mortality (000s) | $\begin{aligned} & \hline \text { DALYs } \\ & (000 \mathrm{~s}) \end{aligned}$ |
| AFR-D | 43 | 2224 | 21 | 283 | 367 | II 45 \| |
| AFR-E | 58 | 2828 | 37 | 508 | 1632 | 50386 |
| AMR-A | 0 | 73 | 6 | 93 | 11 | 350 |
| AMR-B | 1 | 484 | 19 | 293 | 29 | 978 |
| AMR-D | 1 | 73 | 5 | 74 | 23 | 684 |
| EMR-B | 0 | 135 | 3 | 53 | 0 | 4 |
| EMR-D | 19 | 1146 | 8 | 121 | 45 | 1366 |
| EUR-A | 0 | 80 | 8 | 107 | 4 | 128 |
| EUR-B | 1 | 150 | 7 | 112 | 1 | 28 |
| EUR-C | 0 | 130 | 12 | 163 | 4 | 136 |
| SEAR-B | 2 | 465 | 14 | 248 | 39 | 1222 |
| SEAR-D | 57 | 3891 | 82 | I 323 | 268 | 8204 |
| WPR-A | 0 | 34 | 3 | 35 | 0 | 7 |
| WPR-B | 5 | 582 | 29 | 377 | 21 | 839 |
| World | 188 | 12296 | 254 | 3790 | 2444 | 75783 |

by just $10 \%$ and if it was lowered by a half. These results are also given in Table 14.14.

Figure 14.11 shows how many new infections are predicted to occur in 2002-2006 in each subregion under the different counterfactual scenarios. These results are given in Table 14.15. The greatest changes would be seen in the African subregions, where sexual transmission dominates the epidemic. However in subregions such as WPR-B, which includes China, where a large number of new cases is predicted to occur, the proportion of infections that could be avoided is smaller, because use of injected drugs is a more important mode of transmission in this subregion.

It is important to consider the plausibility of the finding that almost all new HIV infections in Africa could be avoided if unsafe sex were to cease immediately despite the continuation of non-sexual transmissions. Intuitively, it seems unlikely that there would be almost no new HIV infections in the five years following the onset of behaviour change: transmission of the virus via other routes would continue, and it has been estimated that $5 \%$ of the newly-diagnosed infections in Africa in 2000 were acquired through a non-sexual mode of transmission. As discussed above, sexual and non-sexual transmission dynamics

Table 14.14 The predicted cumulative proportion of new HIV infections in adults during 2002-2006 that could be prevented by different reductions in unsafe sex, by subregion

|  | Reduction in unsafe sex |  |  |
| :--- | :---: | :---: | :---: |
| Subregion | $10 \%$ | $50 \%$ | IO0\% <br> (No unsafe sex) |
| AFR-D | 21 | 54 | $>99$ |
| AFR-E | 40 | 71 | $>99$ |
| AMR-A | 7 | 36 | 72 |
| AMR-B | 9 | 43 | 86 |
| AMR-D | 10 | 47 | 96 |
| EMR-B | 4 | 35 | 69 |
| EMR-D | 9 | 44 | 87 |
| EUR-A | 6 | 29 | 59 |
| EUR-B | 6 | 36 | 73 |
| EUR-C | 3 | 17 | 33 |
| SEAR-B | 7 | 36 | 73 |
| SEAR-D | 8 | 39 | 78 |
| WPR-A | 9 | 44 | 94 |
| WPR-B | 5 | 32 | 65 |

cannot be considered in isolation. Even without considering the extent to which these transmission networks are interlinked, the sheer scale of the change to the susceptible population serves to illustrate why it is not implausible that HIV transmission would cease if unsafe sex stopped altogether in the African subregions. Consider, for example, the urban areas of an east African country with a population of eight million where the estimated prevalence of HIV infection among women attending antenatal clinics in 2001 is $11 \%$. This gives a total of 898000 prevalent cases, of which 45000 are thought to be non-sexually acquired. The EPP model fit to the observed prevalence data produces an estimate for the susceptible fraction of the total population of $20 \%$. Therefore, there are 1633000 people who could acquire HIV infection at the start of 2001.

Using the same example, to simulate the immediate and total cessation of unsafe sex, the at-risk group was reduced by $95 \%$, such that only $1 \%$ of the total population would be able to acquire HIV infections (5\% of the original $20 \%$ ), or 16000 people. The 898000 cases are still prevalent but not all prevalent cases are potential sources of a new infection. Some HIV-infected people will not exhibit risky behaviours and so will not have the opportunity to transmit infection. For a new case to arise, the HIV-infected people must have an effective contact (i.e. give a blood
Figure I4.II The total number of new HIV infections in adults predicted to occur during 2002-2006 assuming different reductions in unsafe sex, by subregion


Table 14.15 The total number of new HIV infections in adults predicted to occur during 2002-2006 assuming different reductions in unsafe sex, by subregion

| Subregion | Reduction in unsafe sex |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | No change | 10\% | 50\% | $\begin{gathered} 100 \% \\ \text { (No unsafe sex) } \end{gathered}$ |
| AFR-D | 3420598 | 2691787 | 1562368 | Approx. 0 |
| AFR-E | 9250954 | 5512072 | 2724441 | Approx. 0 |
| AMR-A | 240000 | 223200 | 153000 | 68000 |
| AMR-B | 1350000 | 1228500 | 773000 | 195750 |
| AMR-D | 300000 | 270000 | 158000 | 12500 |
| EMR-B | 245000 | 235200 | 160000 | 75000 |
| EMR-D | 665000 | 605150 | 374000 | 84000 |
| EUR-A | 150000 | 141000 | 106000 | 62000 |
| EUR-B | 451000 | 423940 | 288000 | 124000 |
| EUR-C | 1008000 | 977760 | 840000 | 673000 |
| SEAR-B | 552000 | 513360 | 351000 | 150000 |
| SEAR-D | 3720000 | 3422400 | 2268000 | 815000 |
| WPR-A | 8000 | 7280 | 4500 | 500 |
| WPR-B | 5000000 | 4750000 | 3390000 | 1760000 |
| World | 26360552 | 21001649 | 13152309 | 4019750 |

transfusion or unsafe injection) with one of the 16000 members of the at-risk group. In a population of eight million people, the probability of this happening is now much reduced, thus the number of new infections resulting is very small. The non-linearity in the relationship between changes in unsafe sex and the number of infections avoided results in a large fraction of infections averted by a $10 \%$ reduction in unsafe sex in the African subregions.

## 6. Uncertainty

### 6.1 Exposure

## DATA QUALITY

Most of the behavioural surveys included in this analysis were large probability samples, which were weighted to be representative of the general population by age and sex. There may have been a selection or participation bias in these surveys. Reporting bias is probably inevitable in at least some surveys; people may have under-reported behaviours that are seen as undesirable, especially in the light of education and information campaigns aimed at promoting behavioural change. We have
limited means to assess the existence of such biases, and the assumption implicit in this work is that such biases can be ignored.

It is unclear how well quantitative household surveys measure sensitive information such as sexual behaviour and some surveys will have been designed and implemented better than others. There is little to indicate how good a survey is, apart from the quality of the data and an assessment of the questionnaire. One survey (Sri Lanka 1991 GPA survey) was excluded from the analysis because of poor quality data.

## Methodological issues

In creating the set of standard behavioural indicators, different questions were used as though they were synonymous. If these questions or their translations are not in fact equivalent, the calculated indicators will not measure the same thing in all places. This is quite likely, at least with respect to the questions and indicators which depend on a classification of partner type in different countries.

The aggregation of the country-level data to subregional level is perhaps a cause for concern. For some indicators, the values estimated for countries within a subregion varied by as wide a range as was observed between the countries in different subregions. Once combined at a subregional level, this variation was no longer apparent. In addition, countries for which no data were available did not contribute to the subregional estimate; it is unlikely that the subregional estimate would not change if we did have data for the missing countries. It is plausible that, within a subregion, the countries for which no data are available are systematically different from those countries in which sexual behaviour surveys have been carried out. These differences could be related to behaviour.

Extrapolation of the estimates of the prevalence of sexual risk behaviour to subregions where there were no data was based on comparison of the proportions of the population who were currently married in each subregion. Values from the most similar subregion were substituted for the missing data. In subregions where some data were available, the missing values were taken from the subregion which was most similar according to the available estimates. However, it was clear that the subregions did not vary in a predictable manner and this method of extrapolation introduced some unquantifiable error.

## ERROR

Confidence intervals can be calculated around the point estimates of behavioural indicators for individual countries. Almost all of these intervals are very narrow, mainly because most of the exposure data come from very large DHS. The error that was introduced by aggregating these estimates to the subregional level cannot readily be quantified: error is introduced because countries with no data are assumed to have average values for the subregion.

## Omissions

Having concentrated solely on heterosexual sex, the behavioural review has clearly underestimated the amount of risk in populations where the main mode of HIV transmission is sex between men. The data on the prevalence of sex between men are too scanty to be used in an analysis of this type, and to include only the available data would introduce more uncertainty into these estimates. Infections that result from sex between men are included in the burden estimates, and in the estimates of attributable and avoidable infection, for both low- and high-prevalence subregions.

### 6.2 Outcomes

## MODEL-BASED APPROACH

The accuracy of the estimates of the avoidable burden of HIV infection due to unsafe sex depends initially on the precision of the five-year projections of the HIV/AIDS epidemics. These projections represent only one possible future course of the epidemic. The projections for both countries with a high prevalence of HIV infection (using the EPP model) and countries with a low prevalence (using the saturation approach) must be considered as representing a likely course, not the certain future course, of the epidemic.

Beyond the accuracy of the projections of HIV prevalence under the business-as-usual scenario, there are other sources of potential inaccuracy in the estimates of avoidable infections. The estimated proportion of all infections that are not sexually acquired is central to the calculation of the proportion of avoidable infections in all subregions. The figure of $5 \%$ employed for Africa, though widely used, should be viewed as very uncertain. Information on mode of transmission is derived from reports on the way in which people who have been diagnosed with HIV infection are thought to have acquired the infection. There are limitations to this data. In many places, a diagnosis of HIV infection will not be made before the onset of symptoms. If diagnosis is delayed it may be more difficult to identify the source of infection, especially for those people who have had more than one type of exposure. Late diagnoses or failure to diagnose may introduce another bias because the people who receive a timely diagnosis may have acquired their infection in a different manner from those whose infections are not promptly diagnosed. Subregional data are based on national data that have been aggregated to the subregional level. The different national data may be subject to different biases. Countries for which no data are available have been assumed to have the average proportion of HIV infections for the subregion. This may have distorted the picture still further. The direction of this error may be influenced by the scale and stage of the epidemic, the health care system and the equity of access to health care.

## Africa

The EPP model is based on the assumption that sexual mixing patterns are homogeneous in a population. Therefore, the assumption implicit in the estimates of the numbers of avoidable HIV infections is that the reduction in prevalence of hazardous sexual behaviour is evenly distributed among the population. If declines in the prevalence of hazardous sexual behaviour are concentrated in certain groups, and the remaining risk behaviours (unsafe medical injections, unsafe blood transfusions and injected drug use) are also clustered, then the number of avoidable infections might be lower. If the remaining risk behaviours are evenly distributed throughout the population, the reductions in unsafe sex will have an effect on non-sexual modes of transmission. Infections acquired in one way are not necessarily transmitted in the same way (if they are passed on at all). Therefore sexually acquired cases of HIV infection may act as the source of infection for non-sexually-acquired cases. A reduction in unsafe sex that leads to fewer prevalent cases of HIV infection will therefore also lower the number of new non-sexually-acquired cases.

The assumption of random mixing must be tenable for the EPP model to perform well. This model is intended to give accurate projections of future HIV prevalence in a population with a generalized epidemic. If the modes of transmission that remain after unsafe sex is reduced were to be concentrated among certain groups, the subsequent number of new infections would be higher than that forecast using EPP. Therefore the estimate of the proportion of infections which is avoidable may be too high. However, given the current epidemic situation in the two African subregions, the assumption of random mixing, even in the absence of unsafe sex, may hold true because use of injected drugs is uncommon and unsafe medical injections and blood transfusions are less likely to be concentrated among specific groups.

### 6.3 Limitations

The departures from the standard relative risk methodology and the reasons for this have, for the most part, been fully discussed in the text. However, two further differences remain to be explained. The CRA framework requires that all estimates be presented separately for all age groups and for each sex. The estimates of avoidable infections under the different counterfactual scenarios should be made from 2000 until 2030. Neither has been done for unsafe sex because these extensions would greatly add to the uncertainty of the estimates.

The models used for the prevalence projections are only valid in the short term. To extend them beyond 2006 would require additional assumptions about changes in the availability of treatment and prevention efforts. The effects of treatment on transmission are particularly
hard to predict since treatment will tend to increase the prevalence of infection (by prolonging the survival of infected people), but may also reduce the contagiousness of infected persons. Some of the counterfactuals considered include that there will be a massive reduction in the amount of unprotected sex after 2001. This would inevitably have an impact on fertility, which should in turn lower recruitment to the sexually active population. The methods used to predict HIV prevalence do not account for such changes. In the EPP model, as described in section 4, the recruitment of sexually active adults into the two groups of susceptible people is based on the number of births 15 years earlier and on survival rates to the age of 15 years. Therefore changes in fertility initiated in 2001 would not affect the projections until 2016. The EPP model assumes a constant birth rate that does not change over time and that is the same in the two groups of susceptible people. Although it would be possible to alter the process for implementing the counterfactual scenarios to allow for large future changes in fertility, and the emergence of a dramatic fertility differential, there is no information on which to base these estimates.

Because there is no way to estimate the size of the decline in fertility under the counterfactual scenarios in any subregion, there is no way to accurately model these scenarios beyond the short term. Similar issues are encountered in estimating HIV prevalence by age group and sex. In all subregions, prevalence is different for men and women and varies by age group. The distribution of infections by age and sex varies by epidemic duration and is not necessarily the same in all countries in a subregion, which makes it complicated to establish a subregional breakdown. Estimates of the prevalence of HIV infection among men are not available in most countries but must be inferred from prevalence in pregnant women. Changes in the number of people who are at risk of infection can be expected to change the distribution of new infections by age and sex, but the direction of these changes cannot be anticipated. Therefore, to make estimates of avoidable infections by age group and sex would be to add more uncertainty to the existing estimates.

Finally, the most extreme counterfactual scenario presented above produces some dramatic results for the African subregions. Modelling the complete cessation of unsafe sex implies that, even within marriage, discordant couples would no longer have procreative sex. Such a scenario is artificial and unprecedented: there are no historical examples of a total and sudden cessation of exposure to an infectious disease at the macro level. The reason for including this scenario is that if STIs are eliminated from a population there would be no unsafe sex: in this chapter the counterfactual has been defined in terms of the level of unsafe sex. Under the counterfactual scenario of no unsafe sex, the extraordinarily rapid decline in new infections also produces a discontinuity in prevalence, as the average duration of infection rapidly rises among those

Figure 14.12 HIV prevalence projections for subregions in Africa under two scenarios: no change in current levels of unsafe sex and the total cessation of unsafe sex


Note: The line representing the HIV prevalence following the total cessation of unsafe sex has been interpolated between 2000 and 2001 for the prevalence under the counterfactual scenario.
who are already infected, which in turn implies a rapid increase in mortality, since mortality of the people who are HIV-positive increases with the time since infection. This mortality increase exacerbates the decline in prevalence, with the results shown in Figure 14.12.

In any single EPP counterfactual scenario for a subnational population, the apparent effect of the rapid transfer of a large fraction of the susceptible population from the at-risk group to the not-at-risk group would depend on the timing of the decline in risk relative to the "natural" epidemic peak and the level of saturation (the proportion of infected people among infected and susceptible persons). Three different situations are shown in Figure 14.13, which illustrates the model fits for a population in which HIV prevalence is still rising rapidly in 2001, a second population in which growth in prevalence has stabilized in 2001;

Figure I4.13 EPP projections of the size of the infected, at-risk and not-atrisk groups for three subnational populations under two scenarios: no change in current levels of unsafe sex and total cessation of unsafe sex

|  | No change (business-as-usual) | Total cessation of unsafe sex |
| :---: | :---: | :---: |
| Prevalence <br> still <br> increasing <br> in 2001 |  |  |
| Prevalence peaked around 2001 |  |  |
| Prevalence peaked before 2001 |  |  |
| Key: $\square$, not at risk; $\square$, at risk of infection; $\square$, newly infected; $\square$, already infected. |  |  |

and a third population in which HIV prevalence has begun to decline by 2001. The figure shows the relative contributions of the infected, at-risk and not-at-risk groups for two scenarios: no behaviour change (business-as-usual) and total cessation of unsafe sex.

When the data from populations at these three different epidemic stages are amalgamated at the subregional level, the business-as-usual scenario gives the impression of an epidemic with a much broader prevalence peak than that seen in any one national population. However, for the no-more-unsafe-sex scenario, because the decrease in unsafe sex is assumed to occur in the same calendar year in all places, it produces the artificial-looking declines in the number of new infections, shown in Figure 14.13.

## 7. Discussion and conclusions

Unsafe sex is a difficult exposure to address within the standard epidemiological framework of simple exposure measures and constant relative risks. The problem of relating behaviour patterns to risk of HIV infection is hardly a new one. Other researchers have tried to tackle this in many different ways. The Four Cities study (Buve et al. 2001b; Carael and Holmes 2001; Ferry et al. 2001) compared sexual behaviour in two African cities with a high prevalence of HIV infection and two cities with a relatively low prevalence in order to look for determinants of this heterogeneity. Individual and ecological analyses were carried out. Some behavioural factors were found to be more common in the cities with a high prevalence compared to the cities with a low prevalence of HIV infection. These were: young age at having sex for the first time (for women), young age at first marriage and the existence of a large age difference between spouses. Factors which affect transmission and which were more common in the cities with a high prevalence were herpes simplex virus (HSV-2, genital herpes) infection, trichomoniasis (for women) and lack of male circumcision. Factors that were not more common in the high-prevalence cities were: a high rate of partner change, sex with sex workers, concurrent partnerships, a large age difference between non-spousal partners, gonorrhoea, chlamydial infection, syphilis, dry sex and lack of condom use. The factors found more commonly in the cities with a high prevalence of HIV infection do not seem sufficient to explain the differences in prevalence (Buve et al. 2001a). A comparison of rural populations in Zimbabwe and the United Republic of Tanzania has also failed to find differences in sexual behaviour which could explain the higher HIV prevalence observed in the Zimbabwean population (Boerma et al. 2002). In this light, it is perhaps unsurprising that we have not been able to elucidate a relationship.

Using alternative methods for estimating the attributable disease burden, we found that most of the current burden of disease due to HIV infection is attributable to unsafe sex. If all sexual transmission were to cease, there would be just over 4 million new HIV infections between 2001 and 2006, compared to more than 26 million which are forecast to occur if there is no change in the pattern of transmission. Most of the avoidable infections are concentrated in the African subregions, which
is as expected given the current prevalence of HIV infection in these subregions. The other subregions where sexual transmission is expected to be important in the future are SEAR-D and WPR-B. These two subregions contain some countries which already report broad sexual spread of HIV infection, primarily through sex work (Cambodia, Myanmar) and even in countries where the current epidemic is now driven by injected drug use (Indonesia, China), HIV will spread more broadly from the injecting drug users to their sexual partners. In these countries, the fraction of future infections which would be averted by reductions in unsafe sex is higher than the fraction of current infections which is attributable to unsafe sex.
These findings do not come as a surprise. More important for intervention design and programme evaluation would be to identify which aspects of sexual behaviour contribute most to the spread of HIV in different settings. If this were known, the design and implementation of measures to prevent the spread of HIV infection could be improved. However, even in the absence of this information some measures are known to be effective in preventing HIV infection at the individual level. For example, increasing the levels of condom use can only help to slow the spread of infection.

## Acknowledgements

The following kindly provided data:
Measure DHS+, Macro International Inc. provided the DHS for 63 countries.

Australia: Pilot data were obtained from the Australian Study of Health and Relationships. The data were made available by Anthony Smith and provided by Richard de Visser.

India: Data were obtained from the International Institute for Population Studies (IIPS), Mumbai, courtesy of Ravi K. Verma.

Europe: Data for eight countries (England, France, Germany, Greece, Italy, Norway, Portugal and Switzerland) from the European New Encounter Module (NEM) project were provided by Michel Hubert, on behalf of the NEM group.

France: Data from the 2001 Knowledge, Attitudes, Beliefs and Practices (KABP) survey (Grémy et al. 2001) were made available by Ruth Ferry and provided by Julien Chauveau.

Honduras: Data from the Centers for Disease Control and Prevention Encuesta Nacional de Salud Masculina 1996 survey were provided by Leo Morris.

Rwanda and Madagascar: Data were provided by Population Services International (PSI), courtesy of Dominique Meekers.

## Models

We made use of the Epidemic Projection Package (EPP) and Spectrum. These are both available courtesy of the Futures Group.

We are greatly indebted to Tim Brown at the East West Center, not only for providing us with the source code for EPP but also for a modified version of the model, which made it possible to calculate the initial estimates of the avoidable infections in countries with generalized epidemics.

We are grateful to the UNAIDS/WHO Working Group on Global Surveillance of HIV/AIDS and STIs for making their country-specific models of HIV/AIDS available for use in this exercise.

## Notes

1 See preface for an explanation of this term.
2 The exception is the Indian study which was among people attending a clinic for sexually transmitted infections. The study was included to provide some information on Asia.

## References

Ahmed S, Lutalo T, Wawer M et al. (2001) HIV incidence and sexually transmitted disease prevalence associated with condom use: a population study in Rakai, Uganda. AIDS, 15:2171-2179.
Anderson RM, Garnett GP (2000) Mathematical models of the transmission and control of sexually transmitted diseases. Sexually Transmitted Diseases, 27:636-643.

Anderson RM, May RM (1991) Infectious diseases of humans. Dynamics and control. Oxford University Press, Oxford.
Anonymous (1992) Comparison of female to male and male to female transmission of HIV in 563 stable couples. European study group on heterosexual transmission of HIV. British Medical Journal, 304:809-813.
Asiimwe Okiror G, Opio AA, Musinguzi J et al. (1997) Change in sexual behaviour and decline in HIV infection among young pregnant women in urban Uganda. AIDS, 11:1757-1763.
Auvert B, Ballard R, Campbell C et al. (2001a) HIV infection among youth in a South African mining town is associated with herpes simplex virus-2 seropositivity and sexual behaviour. AIDS, 15:885-898.
Auvert B, Buve A, Ferry B et al. (2001b) Ecological and individual level analysis of risk factors for HIV infection in four urban populations in sub-Saharan Africa with different levels of HIV infection. AIDS, 15:S15-30.

Auvert B, Buve A, Lagarde E et al. (2001c) Male circumcision and HIV infection in four cities in sub-Saharan Africa. AIDS, 15:S31-40.
Berkley S (1998) Unsafe sex as a risk factor. In: Health dimensions of sex and reproduction: the global burden of sexually transmitted diseases, HIV, mater-
nal conditions, perinatal disorders, and congenital anomalies. Global Burden of Disease and Injury, Volume 3. Murray CJL, Lopez AD, eds. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Boerma T, Nyamukapa C, Urassa M, Gregson S (2002) The uneven spread of HIV within Africa: a comparative study of biological, behavioural and contextual factors in rural populations in Tanzania and Zimbabwe. Population Association of America 2002 Annual Meeting. Atlanta, GA.
Buve A, Carael M, Hayes R et al. (2001a) The multicentre study on factors determining the differential spread of HIV in four African cities: summary and conclusions. AIDS, 15:S127-131.

Buve A, Carael M, Hayes R et al. (2001b) Multicentre study on factors determining differences in rate of spread of HIV in sub-Saharan Africa: methods and prevalence of HIV infection. AIDS, 15:S5-14.

Carael M, Holmes KK (2001) Dynamics of HIV epidemics in sub-Saharan Africa: introduction. AIDS, 15:S1-4.
Chao A, Bulterys M, Musanganire F et al. (1994) Risk factors associated with prevalent HIV-1 infection among pregnant women in Rwanda. National University of Rwanda-Johns Hopkins University Aids Research Team. International Journal of Epidemiology, 23:371-380.
Cohen MS, Henderson GE, Aiello P, Zheng H (1996) Successful eradication of sexually transmitted diseases in the People's Republic of China: implications for the 21st century. Journal of Infectious Diseases, 174:S223-229.

Commissariat Général du Plan: Observatoire régional de santé d’Ile-de-France, Agence Nationale de Recherches sur le SIDA (2001) Les connaissances, attitudes, croyances et comportements face au vih/sida en France. Evolutions 1992, 1994, 1998, 2001. Paris.
de Gourville EM, Mabey D, Quigley M, Jack N, Mahabir B (1998) Risk factors for concordant HIV infection in heterosexual couples in Trinidad. International Journal of STD and AIDS, 9:151-157.
del Mar Pujades Rodriguez M, Obasi A, Mosha F et al. (2002) Herpes simplex virus type 2 infection increases HIV incidence: a prospective study in rural Tanzania. AIDS, 16:451-462.

Ferry B, Carael M, Buve A et al. (2001) Comparison of key parameters of sexual behaviour in four African urban populations with different levels of HIV infection. AIDS, 15:S41-50.

Fylkesnes K, Musonda RM, Sichone M, Ndhlovu Z, Tembo F, Monze M (2001) Declining HIV prevalence and risk behaviours in Zambia: evidence from surveillance and population-based surveys. AIDS, 15:907-916.
Glynn JR, Carael M, Auvert B et al. (2001) Why do young women have a much higher prevalence of HIV than young men? A study in Kisumu, Kenya and Ndola, Zambia. AIDS, 15:S51-60.
Gray RH, Kiwanuka N, Quinn TC et al. (2000) Male circumcision and HIV acquisition and transmission: cohort studies in Rakai, Uganda. Rakai Project Team. AIDS, 14:2371-2381.

Gregson S, Mason PR, Garnett GP et al. (2001) A rural HIV epidemic in Zimbabwe? Findings from a population-based survey. International Journal of STD and AIDS, 12:189-196.
Grémy I, al. e (2001) Les connaissances, attitudes, croyances et comportements face au VIH-SIDA en France ; évolutions 1992, 1994, 1998 et 2001. ORS Ile-de-France, Paris.
Hunter DJ, Maggwa BN, Mati JK, Tukei PM, Mbugua S (1994) Sexual behavior, sexually transmitted diseases, male circumcision and risk of HIV infection among women in Nairobi, Kenya. AIDS, 8:93-99.
Kamali A, Carpenter LM, Whitworth JAG, Pool R, Ruberantwari A, Ojwiya A (2000) Seven-year trends in HIV-1 infection rates, and changes in sexual behaviour, among adults in rural Uganda. AIDS, 14:0427-0434.

Kilmarx PH, Supawitkul S, Wankrairoj M et al. (2000) Explosive spread and effective control of human immunodeficiency virus in northernmost Thailand: the epidemic in Chiang Rai province, 1988-99. AIDS, 14:2731-2740.
Konings E, Bantebya G, Carael M, Bagenda D, Mertens T (1995) Validating population surveys for the measurement of HIV/STD prevention indicators. AIDS, 9:375-382.
Korenromp EL, van Vliet C, Bakker R, de Vlas SJ, Habbema JDF (2000) HIV spread and partnership reduction for different patterns of sexual behavioura study with the microsimulation model STDSIM. Mathematical Population Studies, 8:135-173.

Malamba SS, Wagner HU, Maude G et al. (1994) Risk factors for HIV-1 infection in adults in a rural Ugandan community: a case-control study. AIDS, 8:253-257.

Mastro TD, Kitayaporn D (1998) HIV type 1 transmission probabilities: estimates from epidemiological studies. AIDS Research and Human Retroviruses, 14:S223-227.
Mbopi Keou FX, Gresenguet G, Mayaud P et al. (2000) Interactions between herpes simplex virus type 2 and human immunodeficiency virus type 1 infection in African women: Opportunities for intervention. Journal of Infectious Diseases, 182:1090-1096.

Nelson KE, Celentano DD, Eiumtrakol S et al. (1996) Changes in sexual behavior and a decline in HIV infection among young men in Thailand. New England Journal of Medicine, 335:297-303.

Nunn AJ, Kengeya Kayondo JF, Malamba SS, Seeley JA, Mulder DW (1994) Risk factors for HIV-1 infection in adults in a rural Ugandan community: a population study. AIDS, 8:81-86.
Obasi A, Mosha F, Quigley M et al. (1999) Antibody to herpes simplex virus type 2 as a marker of sexual risk behavior in rural Tanzania. Journal of Infectious Diseases, 179:16-24.
Quigley M, Munguti K, Grosskurth H et al. (1997) Sexual behaviour patterns and other risk factors for HIV infection in rural Tanzania: a case-control study. AIDS, 11:237-248.

Quinn TC, Wawer MJ, Sewankambo NK et al. (2000) Viral load and heterosexual transmission of human immunodeficiency virus type 1. New England Journal of Medicine, 342:921-929.

Rehle TM, Saidel TJ, Hassig SE, Boney PD, Gaillard EM, Sokal DC (1998) AVERT: a user-friendly model to estimate the impact of HIV/sexually transmitted disease prevention interventions on HIV transmission. AIDS, 12:S27-35.

Rodrigues JJ, Mehendale SM, Shepherd ME, et al. (1995) Risk factors for HIV infection in people attending clinics for sexually transmitted diseases in India. British Medical Journal, 311:283-286.

Rowley J, Berkley S (1998) Sexually transmitted diseases. In: Health dimensions of sex and reproduction. Murray CJL, Lopez AD, eds. Harvard School of Public Health, Boston, MA.

Royce RA, Sena A, Cates W Jr., Cohen MS (1997) Sexual transmission of HIV. New England Journal of Medicine, 336:1072-1078.

Seed J, Allen S, Mertens T et al. (1995) Male circumcision, sexually transmitted disease, and risk of HIV. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 8:83-90.
Stover J, Walker N, Garnett G et al. (2002) Can we reverse the HIV/AIDS pandemic with an expanded response? Lancet, 360:73-77.
ter Meulen J, Mgaya HN, Chang Claude J et al. (1992) Risk factors for HIV infection in gynaecological inpatients in Dar es Salaam, Tanzania, 1988-1990. East African Medical Journal, 69:688-692.

UNAIDS (2000) National AIDS programmes: A guide to monitoring and evaluation. http://www.unaids.org/publications/documents/epidemiology/ surveillance/JC427-Mon\&Ev-Full-E.pdf (accessed 2002).

The UNAIDS Reference Group on Estimates Modelling and Projections (2002) Improved methods and assumptions for estimation of the HIV/AIDS epidemic and its impact: recommendations of the UNAIDS reference group on estimates, modelling and projections. AIDS, 16:W1-W14.
The UNAIDS Reference Group on Estimates, Models and Projections (2002) Estimating and projecting national HIV/AIDS epidemics. The models and methodology of the UNAIDS approach to estimating and projecting national HIV/AIDS epidemics. http://www.tfgi.com/epp_man.pdf.

UNAIDS Epidemiology Reference Group (2001) Recommended methodology for the estimation and projection of HIV prevalence and AIDS mortality in the short-term. http://www.epidem.org/Publications/Meeting\ summary\% 20and\%20recommendations\%20final.pdf.
UNAIDS/WHO (2001) AIDS epidemic update. 92-9173-132-3. Geneva. http://www.thebody.com/unaids/update1201/contents.html.
UNAIDS/WHO (2002) AIDS epidemic update, Geneva.
U.S. Census Bureau (2001) HIV/AIDS surveillance data base (accessed June 2001). http://www.census.gov/ipc/www/hivaidsd.html.
van der Ploeg CPB, van Vliet C, de Vlas SJ et al. (1998) STDSIM: A microsimulation model for decision support in STD control. Interfaces, 28:84-100.

Walboomers JM, Jacobs MV, Manos MM et al. (1999) Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. Journal of Pathology, 189:12-19.
Walker N, Garcia Calleja JM, Heaton L et al. (2001) Epidemiological analysis of the quality of HIV sero-surveillance in the world: how well do we track the epidemic? AIDS, 15:1545-1554.
Walker N, Stanecki KA, Brown T et al. (2003) Methods and procedures for estimating HIV/AIDS and its impact. The UNAIDS/WHO methods for end of 2001. AIDS, 17:2215-2225.

Wawer MJ, Serwadda D, Gray RN et al. (1997) Trends in HIV-1 prevalence may not reflect trends in incidence in mature epidemics: data from the Rakai population-based cohort, Uganda. AIDS, 11:1023-1030.

Weiss HA, Quigley MA, Hayes RJ (2000) Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis. AIDS, 14:2361-2370.

Weller S, Davis K (2002) Condom effectiveness in reducing heterosexual HIV transmission. In: The Cochrane library issue 3. Update Software, Oxford, http://cochrane.athens.ac.uk.
White R, Cleland J, Carael M (2000) Links between premarital sexual behaviour and extramarital intercourse: a multi-site analysis. AIDS, 14:2323-2331.
WHO/UNAIDS (2000) Guidelines for second generation HIV surveillance. World Health Organization and the Joint United Nations Programme on HIV/AIDS, Geneva.
Appendix A

| Country | Year of survey | Age (years) |  | Marital status |  | Sample size |  | Type of survey | Survey organization | Source of information |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Female | Male | Female | Male | Female | Male |  |  |  |
| AFR-D |  |  |  |  |  |  |  |  |  |  |
| Benin | 1996 | 15-49 | 20-64 | All females | All males | 5491 | 1535 | DHS | DHS | Macro International |
| Burkina Faso | 1999 | 15-49 | 15-59 | All females | All males | 6445 | 2641 | DHS | DHS | Macro International |
| Cameroon | 1998 | 15-49 | 15-59 | All females | All males | 5501 | 2562 | DHS | DHS | Macro International |
| Chad | 1997 | 15-49 | 15-59 | All females | All males | 7454 | 2320 | DHS | DHS | Macro International |
| Comoros | 1996 | 15-49 | 15-64 | All females | All males | 3050 | 795 | DHS | DHS | Macro International |
| Ghana | 1998 | 15-49 | 15-59 | All females | All males | 4843 | 1546 | DHS | DHS | Macro International |
| Guinea | 1999 | 15-49 | 15-59 | All females | All males | 6753 | 1980 | DHS | DHS | Macro International |
| Liberia | 1986 | 15-49 | - | All females | - | 5239 | - | DHS | DHS | Macro International |
| Mali | 1996 | 15-49 | 15-59 | All females | All males | 9704 | 2474 | DHS | DHS | Macro International |
| Niger | 1998 | 15-49 | 15-59 | All females | All males | 7577 | 3542 | DHS | DHS | Macro International |
| Nigeria | 1999 | 10-49 | 15-64 | All females | All males | 7647 | 680 | DHS | DHS | Macro International |
| Senegal | 1997 | 15-49 | $\geq 20$ | All females | All males | 8593 | 4306 | DHS | DHS | Macro International |
| Togo | 1998 | 15-49 | 12-59 | All females | All males | 8569 | 3819 | DHS | DHS | Macro International |
| AFR-E |  |  |  |  |  |  |  |  |  |  |
| Burundi | 1987 | 15-49 | $\geq 20$ | All females | Husbands | 3970 | 542 | DHS | DHS | Macro International |
| Central African Republic | 1994 | 15-49 | 15-59 | All females | All males | 5884 | 1729 | DHS | DHS | Macro International |
| Congo | 1999 | 15-50 | 15-50 | All females | All males | 1181 | 930 | STIS |  | Supplied by PSI |

Surveys used to estimate exposure (continued)

| Country | Year of survey | Age (years) |  | Marital status |  | Sample size |  | Type of survey | Survey organization | Source of information |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Female | Male | Female | Male | Female | Male |  |  |  |
| Côte d'lvoire | 1994 | 15-49 | 15-59 | All females | All males | 8099 | 2552 | DHS | DHS | Macro International |
| Ethiopia | 2000 | 15-49 | 15-59 | All females | All males | 15367 | 2607 | DHS | DHS | Macro International |
| Kenya | 1998 | 15-49 | 15-54 | All females | All males | 7881 | 3407 | DHS | DHS | Macro International |
| Lesotho | 1989 | 15-55 | 15-56 | All | All | 1033 | 549 | KABP/PR | GPA | Supplied by ICP |
| Mozambique | 1997 | 15-49 | 15-59 | All females | All males | 8779 | 2335 | DHS | DHS | Macro International |
| Namibia | 1992 | 15-49 | - | All females | - | 5421 | - | DHS | DHS | Macro International |
| Uganda | 1995 | 20-44 | 15-59 | All females | All males | 1750 | 1356 | In depth | DHS | Macro International |
| United Republic of Tanzania | 1999 | 15-49 | 15-59 | All females | All Men | 4029 | 3542 | Interim | DHS | Macro International |
| Zambia | 1996 | 15-49 | 15-59 | All females | All males | 8021 | 1849 | DHS | DHS | Macro International |
| Zimbabwe | 1999 | 15-49 | 15-54 | All females | All males | 5907 | 2609 | DHS | DHS | Macro International |
| AMR-A |  |  |  |  |  |  |  |  |  |  |
| USA | 1997 | 14-20 | 14-20 | All females | All males | 4039 | 4170 | NLSY | NLS | NLS |
| USA | 2000 | $\geq 15$ | $\geq 15$ | All females | All males | - | - | Current population survey |  | Fields and Casper (2001) |
| USA | 1988 | 18-59 | 18-59 | All females | All males | - | - | Sexual behaviour |  | Laumann et al. (1995) |
| AMR-B |  |  |  |  |  |  |  |  |  |  |
| Brazil | 1996 | 15-49 | 15-59 | All females | All males | 12612 | 2949 | DHS | DHS | Macro International |


| Chile | 1998 | 18-69 | 18-39 | All | All | 3163 | 2244 | National Sexual Behaviour Survey | La Comision Nacional del SIDA (CONASIDA) | Published report |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Colombia | 2000 | 15-49 | - | All females | - | 11585 | - | DHS | DHS | Macro International |
| Dominican Republic | 1996 | 15-49 | 15-64 | All females | All males | 8422 | 2279 | DHS | DHS | Macro International |
| El Salvador | 1985 | 15-49 | - | All females | - | 5207 | - | DHS | DHS | Macro International Macro International |
| Honduras | 1996 |  | 15-59 |  | All | - | 2925 | RHS | CDC | Leo Morris at CDC <br> Atlanta |
| Mexico | 1987 | 15-49 | - | All females | - | 9310 | - | DHS | DHS | Macro International |
| Paraguay | 1990 | 15-49 | - | All females | - | 5827 | - | DHS | DHS | Macro International |
| Trinidad and Tobago | 1987 | 15-49 | - | All females | - | 3806 | - | DHS | DHS | Macro International |
| AMR-D |  |  |  |  |  |  |  |  |  |  |
| Bolivia | 1998 | 15-49 | 15-64 | All females | All males | 11187 | 3780 | DHS | DHS | Macro International |
| Ecuador | 1987 | 15-49 | - | All females | - | 4713 | - | DHS | DHS | Macro International |
| Guatemala | 1999 | 15-49 | - | All females | - | 6021 | - | Interim | DHS | Macro International |
| Haiti | 1994 | 15-49 | 15-59 | All females | All males | 5356 | 1610 | DHS | DHS | Macro International |
| Nicaragua | 1997 | 15-49 | - | All females | - | 13634 | - | DHS | DHS | Macro International |
| Peru | 2000 | 15-49 | - | All females | - | 32000 | - | DHS | DHS | Macro International |
| Peru | 1996 | 15-49 | 15-59 | All females | All males | 28951 | 2487 | DHS | DHS | Macro International |
| EUR-A |  |  |  |  |  |  |  |  |  |  |
| France | 1998 | 18-49 | 18-49 | All | All | 819 | 795 | New Encounter Module |  | NEM European Group |
| France | 2001 | 18-59 | 18-59 | All | All | 1892 | 1429 | KABP | ORS | ORS |
| Germany | 1998 | 15-49 | 15-49 | All | All | 1422 | 1161 | New Encounter Module |  | NEM European Group |
| Greece | 1998 | 15-49 | 15-49 | All | All | 1038 | 962 | New Encounter Module |  | NEM European Group |
| Italy | 1998 | 15-49 | 15-49 | All | All | 1384 | 1219 | New Encounter Module |  | NEM European Group |

Surveys used to estimate exposure (continued)

| Country | Year of survey | Age (years) |  | Marital status |  | Sample size |  | Type of survey | Survey organization | Source of information |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Female | Male | Female | Male | Female | Male |  |  |  |
| Norway | 1997 | 15-49 | 15-49 | All | All | 2122 | 1582 | New Encounter Module |  | NEM European Group |
| Portugal | 1999 | 15-49 | 15-49 | All | All | 360 | 320 | New Encounter Module |  | NEM European Group |
| Spain | 1996 | $\geq 15$ | $\geq 15$ | All females | All males | 4258 | 35730 | National Household Survey. sexual behaviour and condom use re HIV | Aids carepsychological and socio-medical aspects of AIDS/HIV | Castilla et al. (1998) |
| Switzerland | 1997 |  |  | All | All | 1418 | 1359 | New Encounter Module |  | NEM European Group |
| United Kingdom | 1990 | 16-59 | 16-59 | All females | All males | 10758 | 8115 | Sexual attitudes and lifestyles | NATSAL survey | Johnson et al. (1994) |
| EUR-B |  |  |  |  |  |  |  |  |  |  |
| Kyrgyzstan | 1997 | 15-49 | - | All females | - | 3848 | - | DHS | DHS | Macro International |
| Poland | 1991 | 20-49 | 20-49 | All females | All males | 3902 | 3783 | FFS | PAU | United Nations |
| Uzbekistan | 1996 | 15-49 | - | All females | - | 4415 | - | DHS | DHS | Macro International |
| EUR-C |  |  |  |  |  |  |  |  |  |  |
| Kazakhstan | 1999 | 15-49 | 15-59 | All females | All males | 4800 | 1440 | DHS | DHS | Macro International |
| Ukraine | 1999 | 15-44 | - | All females | - | 7128 | - | RHS | CDC | Leo Morris at CDC |
| SEAR-B |  |  |  |  |  |  |  |  |  |  |
| Thailand | 1990 | 15-49 | 15-49 | All | All | 1675 | 1126 | PR | GPA |  |


| SEAR-D |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| India | 1993 | 15-60 | 15-60 | All females | All males | 969 | 836 | Sexual behaviour in |  | Kumar et al. (1997) |
| India | 1999 |  | 18-35 | - | All males | - | 2087 | Delhi Orissa | IIPS | IIPS, courtesy of Ravi K. Verma |
| WPR-A |  |  |  |  |  |  |  |  |  |  |
| Australia | 1999/2001 | 19-59 | 19-59 | All females | All males | 782 | 684 | Sexual behaviour- data from pilot for forthcoming national study | Australian Study of Health and Relationships | Data made available by Anthony Smith, Australian Research Centre in Sex Health \& Society at La Trobe University. Data kindly provided by Richard de Visser |
| New Zealand | 1995 | 18-54 | 18-54 | All females | All males2 | 361 both sexes | - | National telephone survey |  | Paul et al. (1995) |
| Singapore | 1989 | 15-49 | 15-49 | All | All | 1109 | 1006 | PR | GPA | Supplied by ICP (200I) Family Health International |
| WPR-B |  |  |  |  |  |  |  |  |  |  |
| Cambodia | 2000 |  | 15-49 | - | All males | - | 3166 | Household BSS | FHI |  |
| Philippines | 1998 | 15-49 | - | All females | - | 13983 |  | DHS | DHS | Macro International |

Surveys used to estimate exposure (continued)


## Chapter I 5

# Non-USE AND USE OF INEFFECTIVE METHODS OF CONTRACEPTION 

Martine Collumbien, Makeda Gerressu and
John Cleland

## Summary

This chapter estimates the burden of disease attributable to non-use of contraception and use of ineffective methods. The health outcomes include obstetric complications and abortion-related morbidity and mortality associated with unintended pregnancies (unwanted and mistimed). We have presented a model for linking data on contraceptive use and fertility preferences to unwanted births and unsafe abortions as intermediate outcomes, which were then related to the maternal disease burden.

The health outcomes considered were the conditions associated with unsafe abortion and unwanted births. The abortion-related conditions are a separate subcategory and the risk of abortion-related consequences is directly proportional to the risk of an unsafe abortion. The obstetric conditions linked to unwanted births are maternal haemorrhage, maternal sepsis, hypertensive disorders of pregnancy, obstructed labour and other maternal conditions. The burden of these obstetric complications attributable to non-use of contraception was assumed to be proportional to the percentage of unwanted births among all births.

Contraceptive use reduces the risk of unintended conception but does not altogether eliminate it, and failure rates are higher for traditional methods than for modern methods. The categorical variable "contraceptive use" has three levels of exposure: non-use, use of traditional methods and use of modern methods. Non-users experience the highest conception rates. The modern method category was used as the reference category for calculating the relative risk of having an abortion and an unwanted birth.

Not all conceptions lead to an avoidable burden, since many pregnancies are desired. We calculated how many unintended pregnancies are expected in one year by first estimating the proportion of women who would become pregnant and combining this with the probability that the
pregnancy would be unwanted or mistimed, based on current reproductive intentions. The proportion of women becoming pregnant was derived from contraceptive failure rates among modern and traditional method users and biological expectations of the number of conceptions among non-users. Within the non-users, conception rates were applied to the fecund women only, excluding those who would not be exposed to pregnancy for biological or behavioural reasons. Abortion probabilities were applied to determine how many of the mistimed and unwanted pregnancies would end as abortions and unwanted births. Unwanted pregnancies would contribute to both abortion-related burden and the obstetric burden of maternal complications. Mistimed pregnancies only contribute to the abortion-related burden since preventing mistimed births by use of more effective contraception does not avert-only delay-any potential associated obstetric burden.

As theoretical minimum exposure we have simulated the contraceptive distribution which would prevail if all women with a desire to either stop childbearing or postpone the next birth for at least another two years, adopt an effective modern method of contraception. All traditional method users and fecund non-users consist of women who want a birth in the next two years. At this theoretical minimum level, the relative risk of an unwanted birth and abortion becomes zero because only the reference category, modern method users, is at risk of unintended pregnancy. Counterfactual levels of relative risk were calculated to take account of the changing distributions of fertility desires within each exposure category.

Subregional ${ }^{1}$ levels of distribution of contraceptive use and the relative risk levels of abortions and unwanted births were derived by aggregating country estimates based on data from 58 Demographic and Health Surveys (DHS). This source includes data on childbearing intentions and contraceptive use at the time of survey. Average methodspecific and duration-specific failure rates were calculated from 18 countries with DHS calendar data on contraceptive use. In each country, the method-duration-specific failure rates were combined with the method mix and data on duration of use of current methods. Abortion probabilities were derived from the World Health Organization (WHO) estimates of incidence ratios (unsafe abortions per 100 live births).

It was estimated that globally $89 \%$ of the disease burden due to abortion complications is attributable to unprotected sex or use of less effective traditional methods. This amounted to 51000 deaths and 4.4 million disability-adjusted life years (DALYs), with $82 \%$ of the burden falling on women aged $<30$ years. The highest absolute burden is experienced in South Asia ( $35 \%$ of the total abortion burden) while in relative terms women in the two African subregions are the worst effected. The burden of disease attributable to maternal conditions arising from unwanted births was 98000 deaths and 4.5 million DALYs. In contrast to abortion, the largest part of the burden befalls women over $30(74 \%)$ since
a higher proportion of all births is unwanted among older women. For women aged $<30$ years, about $7 \%$ of all births could be averted if all women who wished to stop childbearing used a modern method. This proportion is as high as $40 \%$ for the older age group.

## 1. Introduction

Sexual intercourse contributes positively to health and general well-being in both men and women; it leads to increased intimacy in relationships. Sexual intercourse is also an important risk factor for disease and disability. The most important negative consequence of sex is the risk of contracting a sexually transmitted infection, including HIV, through unprotected intercourse. HIV and sexually transmitted infections (STIs) are discussed in chapter 14. In this chapter we have concentrated on the reproductive consequences of sexual intercourse. Notwithstanding recent developments in assisted reproduction, sexual intercourse is a requirement for reproduction for the overwhelming majority of couples. Motherhood is highly valued in most societies, but each pregnancy and childbirth carries a health risk for the woman, and where obstetric services are poor, maternal mortality is still very high. The most recent estimates show that, of the global 515000 maternal deaths in 1995, more than $99 \%$ occurred in Africa, Asia, Latin America and the Caribbean (Hill et al. 2001).

Reduction of maternal mortality and morbidity can be achieved by more effective treatment of pregnancy-related complications. The disease burden can also be reduced by avoiding pregnancies through adoption of effective contraception. Not all pregnancies and births are intended: many are either mistimed or unwanted at any time. Worldwide it is estimated that about 210 million recognizable pregnancies occur every year (The Alan Guttmacher Institute 1999)—of which about $15 \%$ end in spontaneous miscarriage or stillbirth. Another $22 \%$ are terminated by induced abortion and thus can be classified unambiguously as unintended. The remainder-some 133 million-result in the birth of a baby. Evidence from DHS and similar surveys has suggested that, globally, some $20 \%$ of all births are unintended (The Alan Guttmacher Institute 1999). Adding together unintended births and induced abortions, it may be concluded that about $40 \%$ of all pregnancies are unintended.

This chapter is concerned with estimating the burden of maternal complications and abortions that could be avoided if couples increased their use of effective contraception. We have presented a model for linking data on contraceptive use and fertility preferences to unwanted births and unsafe abortions. DHS data for 58 countries were used to calculate attributable fractions: what proportion of these unwanted births and unsafe abortions could be averted by perfect implementation of fertility preference through increased use of effective contraception. These
intermediate outcomes were then linked to estimates of the burden of maternal complications in pregnancy.

## 2. Risk factor definition and health outcomes

### 2.1 Health outcomes

The health outcomes considered for assessing disease burden due to lack of use of effective contraception are the conditions associated with unsafe abortion and unwanted births. The abortion-related conditions are a subcategory under the Global Burden of Disease (GBD) study's causes of maternal conditions. The main causes of mortality and morbidity associated with unsafe abortion are sepsis, following incomplete removal of the fetus, and perforation of the uterus. Based on the International Statistical Classification of Diseases and Related Health Problems (ICD10), the obstetric conditions other than the abortion-related include maternal haemorrhage, maternal sepsis and obstructed labour. Other complications include hypertensive disorders of pregnancy, and the category of "other maternal conditions". These maternal complications were the causes considered in attributing the burden of disease to unwanted births.

Some other conditions are exacerbated by pregnancy. Indirect obstetric complications result from existing disease (malaria, anaemia, hepatitis, cardiovascular disease, tuberculosis and hypertension) but are aggravated by the physiological effects of pregnancies (AbouZahr and Vaughan 2000). Suicide and violence may be pregnancy related, and other forms of psychological morbidity are associated with childbirth and unintended pregnancies. Since the magnitude and the strength of these relationships are largely unknown, none of these conditions have been included in the burden attributable to non-use of contraception.

The morbidity related to use of contraception has been excluded for this exercise. Those conditions include allergic reactions to barrier methods, intrauterine device (IUD)-associated bleeding, and wounds from surgical procedures. Morbidity associated with systemic contraceptive such as the oral contraceptive pill, includes the impact on cardiovascular and hormonal systems and carcinogenicity (AbouZahr and Vaughan 2000).

The burden of obstetric complications attributable to non-use of (or use of less effective) contraceptive methods is proportional to the percentage of all births that are unwanted. Intergenerational effects of contraceptive use on the health of offspring have not been considered, nor has the burden of perinatal outcomes associated with the delivery of unwanted births been taken into account. It is clear that by averting unwanted pregnancies, a proportion of perinatal deaths can be avoided. However, by averting unwanted births the disease burden throughout
infancy and beyond can be reduced. The potential contribution of contraception to infant survival through better birth spacing is also well known. Babies born within 24 months of an elder sibling are at elevated risk of dying in infancy (Trussell and Pebley 1984). Mistimed births may therefore be associated with higher disease burden in childhood. Because of the conceptual problems of considering health impact in the next generation, the outcomes have been restricted to maternal ones.

### 2.2 INTERMEDIATE OUTCOMES: UNWANTED BIRTHS AND UNSAFE ABORTIONS

In the "perfect contracepting society", all women, or couples, who do not wish to have a baby within the next year or so would use effective contraception. Under these circumstances, a small residue of unintended pregnancies would remain because of contraceptive failure but the overwhelming majority of births would be intended. In the real world, however, large discrepancies exist between reproductive wishes and contraception protection. These discrepancies arise for myriad reasons. In developing countries the main direct cause is lack of any contraceptive precautions despite the desire to delay the next child or have no more children. In the demographic literature, non-use of contraception among women desiring to space or limit childbearing is termed "unmet need" for contraception. Estimates of the prevalence of such unmet need in 55 developing countries in the 1990s ranged from $6 \%$ to $40 \%$ of all currently married women (Westoff 2000). The main underlying causes of unmet need in developing countries include a perception that risk of pregnancy is low, opposition to the use of contraception, stemming from the husband's attitude or religious considerations, and concerns about the safety or side-effects of methods. Lack of knowledge about contraceptive methods, or how to access them, are also important contributory causes in some countries.

In industrialized countries, contraceptive practice tends to be higher than in most developing countries. Nevertheless appreciable discrepancies between reproductive motivation and behaviour are also apparent. Contraceptive failure and irregular use of methods are more important direct causes of unintended pregnancies than in developing countries. For instance, in the United States of America about half of all unintended pregnancies are the result of failure or irregular use (Henshaw 1998). Unanticipated sexual intercourse no doubt represents a further risk factor, particularly for single women.

Attitudes towards becoming pregnant are complex and often ambivalent. Typically, two persons are involved, the woman and her husband or partner, whose views do not necessarily coincide. Attitudes may also change over time, particularly between the time before conception and the time following recognition of the pregnancy. For instance, a couple may have no fixed intention to have a baby but nevertheless be delighted when conception occurs. No unambiguous and generally agreed defini-
tion of unintended pregnancy exists. Rather researchers have used a variety of indirect and direct methods of measurement.

The most commonly available and used measure is that employed by the DHS. In this approach women are asked the following question about recent live births and the current pregnancy (if any): "At the time you became pregnant with (NAME OF CHILD) did you want to become pregnant then, did you want to wait until later, or did you want no more children at all?" This question leads to a three-way classification: births wanted at that time; birth not wanted then but later; birth not wanted at any future time. The latter two categories-mistimed and unwanted births-are often grouped together and defined as unintended births or pregnancies. Some authors use the terms unintended and unplanned interchangeably. However the concept of planning implies active preparation for pregnancy (e.g. cessation of contraceptive use, possible dietary changes, etc.) that makes it inappropriate for the large number of countries where contraception is still uncommon.

Unintended pregnancies may be subdivided into those that are unwanted at any time and those that are mistimed. Both categories may lead to induced abortion although in most settings it can be expected that unwanted pregnancies are more likely to be terminated than mistimed ones (Bankole et al. 1999). In 1995, it was estimated that approximately 26 million legal and 20 million illegal abortions occurred worldwide (Henshaw et al. 1999). The legality and safety of abortion are strongly correlated (Rahman et al. 1998). In the developed world where abortion is generally legal, abortion mortality is as low as 0.2 to 1.2 deaths per 100000 procedures. In non-legal settings, when an unskilled provider, using hazardous techniques, terminates the preg-nancy-often in unsanitary conditions-complications for the woman are likely. Of the 20 million illegal abortions globally, 19 million happen in developing countries (Henshaw et al. 1999). Where abortion is either illegal or highly restrictive, abortion mortality averages 390 deaths per 100000 procedures (but as high as 680 in Africa) (WHO 1998). WHO defines an "unsafe abortion" as a procedure for terminating an unintended pregnancy either by persons lacking the necessary skills or in an environment lacking the minimal medical standards or both (WHO 1998). About a third of unsafe abortions lead to serious complications, and about $13 \%$ of the pregnancy-related deaths worldwide are related to complications of unsafe abortion (The Alan Guttmacher Institute 1999).

When unintended pregnancies are not aborted, and no miscarriage or stillbirth occurs, the pregnancy results in a live birth. As mentioned before, in developing countries, the mortality and morbidity risk associated with complications during childbirth is substantial for any birth, whether intended or unintended.

Part of the total burden of obstetric complications during childbirth can, however, be avoided by preventing the unwanted pregnancies (i.e.
those not wanted at any time) through use of effective contraception (Fortney 1987; Winikoff and Sullivan 1987). Avoiding unwanted pregnancies will reduce maternal mortality in two ways: by reducing the number of pregnancies and by reducing obstetric risk (i.e. the risk per pregnancy). Unwanted births tend to occur when women are relatively old and already have several children. Risks to the mother's health of pregnancy and childbirth are higher at older ages. Hence, the obstetric risk as measured by the maternal mortality ratio (maternal deaths per 100000 live births), is reduced by averting high-risk births based on maternal age and parity but the effect is relatively small (Trussell and Pebley 1984; Winikoff and Sullivan 1987).

In many countries unwanted births are particularly likely to occur to women who have low education, poor nutrition and poor access to health services-all conditions associated with a higher maternal complication rate (Berkley 1998). However, this link between socioeconomic conditions and unwanted births is not universal. In countries with low levels of contraceptive practice, as in much of sub-Saharan Africa, educated women are as likely to report unwanted births as uneducated women (Adetunji 1998). However, these effects of obstetric risk (averting high-risk births), are dwarfed by the impact of reducing the overall incidence of pregnancies through the elimination of unwanted births (Fortney 1987). This elimination will have a huge impact on the maternal mortality rate (maternal deaths per 100000 women of reproductive age), and the lifetime risk of dying in childbirth or pregnancy.

So is it reasonable to assume that delivery complications associated with unwanted births are the same as those associated with wanted births? Apart from considerations of maternal age and socioeconomic status, another possibility is that mothers neglect unintended pregnancies in ways that put the mother herself at greater risk. The evidence is meagre but a recent analysis using five DHS concluded that unintended pregnancies are not selectively discriminated against in terms of obstetric care and thus probably did not represent excess risk to mother's health and survival (Marston and Cleland 2003a). As these five surveys include enquiries from Africa, Asia and Latin America, it is reasonable to generalize results, at least to developing regions. On balance, therefore, it is justifiable to assume that obstetric complications are the same for wanted and unwanted births.

How should mistimed (in distinction to unwanted) births be regarded in relation to delivery complications? Births may be classified as mistimed when the woman is too young and wants to delay the first birth, or when she feels births are too closely spaced or when other conditions are not yet conducive to childbearing. Births to very young women (aged $<18$ years) do carry a higher risk, and delaying some of those may therefore avert some obstetric risk, although the effect will be small (Trussell and Pebley 1984). However, reducing mistimed births by contraceptive practice will have little influence on the incidence of pregnancies as the
births will merely be delayed rather than averted. Such delay or postponement will thus not reduce the burden of delivery complications. As discussed earlier, the potential contribution of contraception to infant survival through better birth spacing is well known, with babies born within 24 months of an elder sibling being at elevated risk of dying in infancy (Trussell and Pebley 1984). In contrast to the strong evidence regarding childhood risks, it is uncertain whether shorter birth intervals are associated with an increased risk of maternal mortality or morbidity. The only two published studies give conflicting results (Conde et al. 2000; Ronsmans and Campbell 1998). It is therefore not justified to regard short intervals as a risk factor for obstetric complications. It may be concluded, therefore, that prevention of mistimed births through contraceptive use will make no contribution to the reduction of delivery complications.

### 2.3 The pathway from exposure to health outcomes

When assessing the 1990 disease burden attributable to unsafe sex, Berkley (1998) estimated the percentage of women with an unmet need for family planning and attributed an equivalent proportion of the obstetric and abortion burden to non-use or inappropriate use of contraception (Berkley 1998). In order to follow the comparative risk assessment (CRA) methodology, we have modelled the outcomes (intermediate outcome in terms of unwanted births and unsafe abortion and health outcome in terms of obstetric and abortion-related burden) from exposure (i.e. contraceptive behaviour).

The definition of exposure has to take into account the fact that contraceptive use reduces the risk of conception but does not altogether eliminate it. The probability of accidental pregnancy while using a method depends on the intrinsic or theoretical effectiveness of the method itself (method failure) and on whether it is used consistently and correctly (user failure). Some methods (e.g. condoms, oral contraceptives, withdrawal and periodic abstinence) are much more prone to user error than other methods (e.g. contraceptive sterilization, intrauterine devices). Withdrawal, or coitus interruptus, and periodic abstinence are distinguished from all other commonly used methods by their exceptionally high failure rates. In the family planning literature these two methods are often given the label "traditional" because they are not the product of advanced techniques of biochemistry or engineering. To capture this variability by method, exposure was divided into three levels: non-use, use of traditional methods and use of modern methods. The pathway from exposure to intermediate and health outcome is depicted in Figure 15.1. The diagram relates to sexually active women (non-virgins), and for clarity the categories of wanted pregnancies and births are omitted.

All three levels of exposure will lead to both unwanted and mistimed pregnancies, but the probability is lowest for the modern method users.

Figure I5.I Pathway from exposure to outcome


Failure rates for traditional methods are higher and no protection carries the highest risk. Modern methods have been used as the theoretical-minimum-risk reference category for calculating the relative risks of having an unintended pregnancy (subdivided into unwanted and mistimed) among traditional method users and non-users.

The next step in the model was to examine the reproductive outcomes of these unintended pregnancies: spontaneous fetal loss and stillbirths, abortions, mistimed and unwanted births. Both unwanted and mistimed pregnancies may end in miscarriage or stillbirth, and such events may cause obstetric complications. No evidence exists to suggest that the probability of miscarriage or stillbirth for unintended pregnancies differs from that of intended pregnancies (i.e. the ratio of unwanted over all stillbirths is the same as the ratio of unwanted births over all births). Therefore, these events have been excluded from the calculations of the attributable risk. Since the total burden of maternal complications is a separate input provided by WHO (independent from our model), and complications due to in utero loss are an integral part of this burden, the attributable burden of maternal complications associated with in utero loss has been accounted for. Both unwanted and mistimed pregnancies may be aborted. In terms of ultimate outcome or disease burden, the risk of abortion-related consequences is directly proportional to the risk of an unsafe abortion. Unwanted pregnancies may be carried to term and the burden of maternal complications will be proportional to the percentage of unwanted births among all births. As discussed earlier, mistimed pregnancies carried to term do not contribute to an attributable burden of disease, because this proportion of disease burden would only have been delayed if the pregnancy were not mistimed.

The model in Figure 15.1 was applied to data from 58 DHS to obtain subregional levels of exposure and the relative risk of having unwanted births and unsafe abortions.

## 3. DATA AND METHODS FOR EXPOSURE AND HAZARD

### 3.1 Data sources

In the analysis, we drew heavily on DHS data on childbearing intentions and contraceptive use at the time of survey. DHS were used for several reasons. First, they are the dominant source of information on fertility intentions and contraceptive use in developing countries, with good representation in all highly populated regions. Most of the most populous developing countries have conducted a recent Demographic and Health Survey: Bangladesh, India, Indonesia and Pakistan in Asia; Brazil and Mexico in Latin America; and Ethiopia and Nigeria in sub-Saharan Africa. Indeed China is the only conspicuous absentee but, as will be shown later, this omission is relatively unimportant because the attributable burden is small in this country.

A further reason for reliance on DHS is that all surveys are nationally representative and executed to a high standard, with abundant technical assistance where needed. Surveys are also highly standardized in content, with the important implication that measures of fertility intentions and contraceptive use are comparable across countries. A final pragmatic reason for using DHS is that well-documented, clean data files are available for public use. Other data, especially the reproductive health surveys by the Centers for Disease Control and Prevention (CDC) were considered, since they cover countries mainly in Latin America and eastern Europe. Because of restricted availability, different age ranges and inadequate detail to make analytically important distinctions, we chose to use DHS data only. Similarly, the use of the National Survey of Family Growth (United States) was briefly considered, but the very small overall burden of maternal complication and the different questions used did not warrant the considerable effort that would have been required.

The surveys used in these calculations were those available at the time of calculation in 2001. Table 15.1 lists the countries by subregion, giving the date of survey, whether the sample was restricted to ever-married women (those who are or have been married) or all women, and the legal status of abortion in each country. The table also indicates what proportion of the total female population aged 15-44 years in the subregion is represented by the country and the weights used for aggregating country-specific data into subregional estimates. Ten of the 58 surveys were done before 1990. Although fertility levels and preferences may have changed considerably in the last decade, these surveys were retained
Table 15.I DHS used in estimating exposure and relative risks

| Subregion | Country | Year | Sample <br> (All or ever-married) | Legal status of abortion ${ }^{\text {a }}$ | Proportion of subregional total of $15-44^{\text {b }}$ <br> (\%) | Weight in subregional estimate ${ }^{\text {b }}$ <br> (\%) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | Burkina Faso | 1999 | All | II | 3.9 | 4.9 |
|  | Benin | 1996 | All | 1 | 2.1 | 2.6 |
|  | Cameroon | 1998 | All | II | 5.0 | 6.3 |
|  | Ghana | 1998 | All | III | 6.9 | 8.6 |
|  | Guinea | 1999 | All | II | 2.5 | 3.1 |
|  | Comoros | 1996 | All | II | 0.2 | 0.3 |
|  | Liberia | 1986 | All | III | 1.1 | 1.4 |
|  | Madagascar | 1997 | All | 1 | 5.3 | 6.7 |
|  | Mali | 1996 | All | 1 | 3.6 | 4.5 |
|  | Nigeria | 1990 | All | 1 | 38.4 | 48.2 |
|  | Niger | 1998 | All | 1 | 3.4 | 4.3 |
|  | Senegal | 1997 | All | 1 | 3.2 | 4.0 |
|  | Chad | 1997 | All | 1 | 2.5 | 3.1 |
|  | Togo | 1998 | All | 1 | 1.5 | 1.9 |
| AFR-E | Botswana | 1988 | All | III | 0.5 | 0.7 |
|  | Burundi | 1987 | All | II | 2.0 | 2.9 |
|  | Central African Republic | 1994 | All | 1 | 1.1 | 1.6 |
|  | Côte d'lvoire | 1994 | All | 1 | 4.4 | 6.3 |
|  | Ethiopia | 2000 | All | II | 17.5 | 25.2 |
|  | Kenya | 1998 | All | 1 | 9.4 | 13.6 |

Table I5.I DHS used in estimating exposure and relative risks (continued)

| Subregion | Country | Year | Sample <br> (All or ever-married) | Legal status of abortion ${ }^{\text {a }}$ | Proportion of subregional total of $15-44^{\text {b }}$ <br> (\%) | Weight in subregional estimate ${ }^{\text {b }}$ <br> (\%) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Malawi | 1992 | All | 1 | 3.1 | 4.5 |
|  | Mozambique | 1997 | All | II | 5.7 | 8.2 |
|  | Namibia | 1992 | All | III | 0.5 | 0.7 |
|  | Rwanda | 1992 | All | II | 2.4 | 3.4 |
|  | United Republic of Tanzania | 1996 | All | 1 | 10.0 | 14.4 |
|  | Uganda | 1995 | All | I | 6.2 | 8.9 |
|  | Zambia | 1996 | All | IV | 2.8 | 4.0 |
|  | Zimbabwe | 1994 | All | 11 | 3.8 | 5.5 |
| AMR-B | Brazil | 1996 | All | 1 | 40.9 | 51.8 |
|  | Colombia | 1995 | All | 1 | 9.9 | 12.5 |
|  | Dominican Republic | 1996 | All | 1 | 1.9 | 2.4 |
|  | El Salvador | 1985 | All | 1 | 1.4 | 1.8 |
|  | Mexico | 1987 | All | 1 | 23.3 | 29.5 |
|  | Paraguay | 1990 | All | 1 | 1.2 | 1.5 |
|  | Trinidad and Tobago | 1987 | All | III | 0.3 | 0.4 |
|  | Bolivia | 1998 | All | II | 11.2 | 11.2 |
| AMR-D | Ecuador | 1987 | All | II | 18.3 | 18.3 |
|  | Guatemala | 1995 | All | 1 | 14.7 | 14.7 |
|  | Haiti | 1994 | All | 1 | 11.3 | 11.3 |


|  | Nicaragua | 1997 | All | 1 | 6.9 | 6.9 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Peru | 1996 | All | II | 37.7 | 37.7 |
| EMR-B | Tunisia | 1988 | EM | V | 7.5 | 100.0 |
| EMR-D | Egypt | 1995 | EM | 1 | 20.0 | 23.7 |
|  | Morocco | 1992 | All | II | 9.1 | 10.7 |
|  | Pakistan | 1990 | EM | II | 42.4 | 50.0 |
|  | Sudan | 1989 | EM | 1 | 8.6 | 10.1 |
|  | Yemen | 1997 | EM | I | 4.6 | 5.5 |
| EUR-B | Kyrgyzstan | 1997 | All | V | 2.2 | 4.7 |
|  | Turkey | 1998 | EM | V | 32.7 | 70.8 |
|  | Uzbekistan | 1996 | All | V | 11.3 | 24.5 |
| EUR-C | Kazakhstan | 1995 | All | V | 6.9 | 100.0 |
| SEAR-B | Indonesia | 1997 | EM | 1 | 71.8 | 71.8 |
|  | Sri Lanka | 1987 | EM | 1 | 6.4 | 6.4 |
|  | Thailand | 1987 | EM | II | 21.8 | 21.8 |
| SEAR-D | Bangladesh | 1997 | EM | 1 | 11.1 | 11.9 |
|  | India | 1993 | EM | IV | 80.4 | 86.1 |
|  | Nepal | 1996 | EM | 1 | 1.8 | 2.0 |
| WPR-B | Philippines | 1998 | All | 1 | 4.8 | 100.0 |

${ }^{\text {a }}$ Legal status of abortion: I, permitted only to save the woman's life or prohibited altogether; II, physical health (also to save the woman's life); III, mental health (also to save the woman's life and physical health); IV, socioeconomic grounds (also to save the woman's life, physical health and mental health); V, without restriction as to reason (Source: Center for Reproductive Rights [formerly Center for Reproductive Law and Policy] 1999).
b These weights represent the fraction of the total population (of women aged 15-44 years) of countries with data.
in the analysis since the differences in fertility levels within subregions are even wider.

DHS directly provide most of the information needed for the calculation of attributable risk ratios. This includes information on sexual activity (needed to define exposure), fertility intentions, type of contraceptive method used and probability of contraceptive failure. These are discussed below. However, DHS have one important defect: most do not collect information on induced abortion and those that do yield severe underestimates. The difficulty of obtaining reliable information on induced abortion is the most intractable problem in the study of human reproduction, perhaps not surprisingly in view of the fact that abortion is both illegal and stigmatized in many societies.

In this chapter, we have used unpublished 1995 national estimates of unsafe abortions compiled by WHO. These estimates were made indirectly from data on hospital admissions for abortion complications, weighted by the proportion of abortions that are thought to result in complications requiring admission. Information from community surveys and to a lesser extent from abortion providers' surveys and mortality studies have also been used to derive best possible estimates (WHO 1998). While this is the main base of data used to calculate abortion probabilities, DHS data have also been used for three central Asian republics (Kazakhstan, Kyrgyzstan and Uzbekistan) where abortion is legal and survey estimates are considered reliable (Westoff et al. 1998). In addition, data on legal rates of abortion were used to make adjustments in some countries (Henshaw et al. 1999).

The only other non-DHS data source used in the calculations were biological in nature. Monthly probabilities of conception among nonusers of contraception (i.e. fecundability) and intrauterine mortality were taken from the published literature (Bongaarts and Potter 1983; Leridon 1977).

### 3.2 Defining exposure and fertility preferences

Women who do not have sexual intercourse are obviously not at risk of complications of abortion or childbirth. Virgins and others need to be distinguished. Virgins are excluded from exposure and they do not affect the relative risk calculations. The proportions of virgins have been given as a separate input for each subregion. In 12 surveys (Table 15.1), mainly in Asia where premarital sex is relatively uncommon, only ever-married women were interviewed. For these countries never-married women have been categorized as virgins. All non-virgins were included in the appropriate category of exposure variable "contraceptive status". A large proportion of non-users is not exposed to risk of pregnancy for either biological or behavioural reasons. As the calculations involved estimating births over a 12 -month period, it was decided to classify women who reported no intercourse in the past 12 months as behaviourally unexposed.

Fertility surveys like the DHS include several questions on fertility intentions: total desired family size, whether more children are wanted, the number of additional children wanted and the intended status of recent births and current pregnancy. Retrospective data on recent births could not be used for our purposes of examining the relationship between unwanted births and contraceptive use, since the women were not asked whether they were using a method of contraception at the time of conception. Instead, we needed to use a forward-looking measure on desirability and timing of any future births. Women's response to the questions "Would you like to have a/another child or would you prefer not to have any (more) children?" and "How long would you like to wait from now before the birth of a/another child?" are considered to be relatively unbiased. Women have no reason to misreport their preference for more children (Bongaarts 1990). Moreover, these future childbearing wishes or intentions are predictive of subsequent childbearing (Westoff 1990).

The data used to link contraception with childbearing intentions are the same that provide the input for calculating the now ubiquitous measure of unmet need for family planning, routinely reported from fertility surveys (Dixon-Mueller et al. 1992; Robey et al. 1996). A woman has an unmet need for birth spacing when she wants to postpone the next birth for at least two years, she is not using contraception and is exposed to risk of conception (i.e. sexually active and menstruating). Similarly, women who want no more children and are not current users are defined to have an unmet need for limiting family size. Unmet need refers to the current status (at the time of interview), but we need to assume a steady state for one year in order to project the fertility implications over the next year for each combined level of exposure and fertility preference.

Since we wanted to estimate the yearly number of expected pregnancies, was this assumption that childbearing intentions stay constant for one year valid? Unless the women are re-interviewed the next year there is no way of knowing the exact fertility implications of the stated preferences and whether these preferences remain constant over a one-year span. Over long periods, childbearing intentions can change substantially, especially in countries progressing through a secular decline in fertility (Freedman et al. 1980). A woman's economic and social circumstances, health status and current marriage/partnership will influence her response about childbearing intentions. As these individual personal circumstances change so may her desire for more children. In the shorter term, however, stability of intentions is reasonably high. For instance, in a prospective study in Peru, aggregate levels of fertility preferences were shown to be consistent over a three-year period (Mensch et al. 1995). Even if circumstances and preferences change for individual women, at the aggregate level the changes are likely to be offset by other women whose life circumstances may change in the opposite direction.

### 3.3 Estimating the expected number of UNINTENDED PREGNANCIES

Referring back to Figure 15.1, the first stage was to estimate how many unintended pregnancies we expected in one year. This was done by first estimating the proportion of women who would become pregnant and combining this with the probability that the pregnancy was unwanted or mistimed. The proportion of women becoming pregnant was based on contraceptive failures among modern and traditional method users and on biological expectations of the number of conceptions among non-users.

## Contraceptive failure

Most enquiries of DHS do not collect information that permit the calculation of contraceptive failure rates. However in a subset of 18 developing countries where levels of contraceptive practice are high, the necessary data have been collected in the form of month-by-month calendars of contraceptive use spanning a 60 -month period prior to date of interview. The type of methods used, dates of starting and ending episodes of use together with main reason for stopping (including failure) are ascertained. Failure rates can be calculated by application of life table techniques to these data. Though this retrospective method of measurement makes heavy demands on the memory of respondents, recall is aided by prior entry into the calendar of live births, ascertained earlier in the interview, and the contraceptive data appear to be of high quality (Curtis and Blanc 1997).

We used an unpublished analysis of failure rates for all 18 countries where calendar data have been collected. Despite considerable intercountry variability, the ranking of methods according to failure rates is clear-cut and accords with other evidence (Trussell 1998). Failure rates are low for methods requiring no memory or skill from users (sterilization, IUD, implant, injectable), intermediate for theoretically effective methods that do require inputs from users (oral contraceptives, condoms) and high for periodic abstinence and withdrawal.

Failure rates not only vary by method and by type of user but also by duration of use. They tend to be higher during the initial period of use and subsequently decline. The reason for this trend concerns selectivity. Inefficient users of a method have a high probability of an early failure. With the passage of time, continuing users are increasingly selected for their proficiency of use and thus failure rates fall. This tendency is more marked for methods requiring skill or memory than for other methods. We have used the mean method-specific and duration-specific failure rates for these 18 countries to calculate country-specific aggregate failure rates for all 58 countries. In the left-hand panel of Table 15.2 we have presented the yearly probabilities of experiencing failure by duration of use for each method. These were calculated from the single-decrement

Table I5.2 Calculation of average method-specific failure rates from yearly probability of failure by duration of use for women aged 30-44 years, all countries combined

|  | Yearly probability of <br> experiencing contraceptive |  |  |  |  | Duration of use in <br> completed years <br> (proportional <br> distribution) | Average <br> failure <br> rate <br> (\%) |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| Contraception method | Ist |  | 2nd | 3rd | $4+$ | 0 | Ist | 2nd | 3rd |

life table estimates of failure. The estimates should be interpreted as the cumulated percentage of couples who would experience failure by the end of the 12th month of use, in the absence of other reasons for stopping use. As can be seen, failure rates generally decline by duration of use, although the second-year failure rates are higher than the first year ones for pill, IUD, withdrawal and "other traditional" methods.

For each of the 58 countries, these method-duration-specific failure rates were then combined with data on duration of use of current methods. The resulting country-specific failure rates for each method were calculated separately for the two age groups. Table 15.2 shows the calculation of these method-specific failure rates for women aged 30-44 years (all countries combined).

The final step was to obtain aggregate failure rates for modern and traditional methods by taking into account the relative contribution of the different modern and traditional methods (method-mix) in each country. The country-method-specific failure rates (calculated as in Table 15.2) have been combined with the relative method mix, as shown in Table 15.3. For women aged 30-44 years, the average failure rate experienced by modern method users is $2.3 \%$, while $15.1 \%$ of the traditional method users will conceive in a year. For the younger age group these failure rates are higher at $4 \%$ and $17.3 \%$, respectively, due to their greater reliance on less effective reversible methods, and a shorter length of time for which current users have been using the methods (calculations not shown).

Table 15.3 Aggregate failure rates, calculated from method-specific failure rates and method mix for women aged 30-44 years, all countries combined

|  | Average method-specific <br> failure (\%) | Method mix <br> relative distribution | Aggregate failure <br> rate (\%) |
| :--- | :---: | :---: | :---: |
| Modern methods |  |  | 2.3 |
| Pill | 5.0 | 0.21 |  |
| IUD | 1.4 | 0.22 |  |
| Injections | 2.2 | 0.11 |  |
| Diaphragm/foam/jelly | 13.3 | 0.01 |  |
| Condom | 7.1 | 0.06 |  |
| Norplant | 0.1 | 0.01 |  |
| Female sterilization | 0.1 | 0.36 |  |
| Male sterilization | 0.3 | 0.02 | 15.1 |
| Traditional methods |  |  |  |
| Periodic abstinence | 14.9 | 0.60 |  |
| Withdrawal | 15.3 | 0.32 |  |
| Other traditional | 15.1 | 0.08 |  |

## Conception rates

The estimation of the expected conceptions among the non-users was based on fecundability, which is defined as the probability of conceiving in a month among fecundable women (Bongaarts and Potter 1983). Fecundable women are those capable of conceiving. Conception refers to recognizable conception signified by the delay of first menses after fertilization. Some non-users have obvious biological or behavioural characteristics that make them temporarily unexposed to the chance of conceiving: the currently pregnant, amenorrhoeic women, women who have not resumed sex since the most recent childbirth and women who reported no sex in the past year. Other non-users, such as those who have reached menopause or know themselves to be infecund, are permanently unexposed to the risk of conception. These women have thus been excluded and the conception rates applied to fecundable women who have been sexually active over the past year.

Fecundability is difficult to assess empirically and it is usually estimated from waiting times to conception. Most reliable estimates are available from measuring the length of interval between marriage to first birth among couples using no contraception. Coital frequency is the dominant behavioural determinant of fecundability and most of the variation of fecundability by age can be attributed to a decline in intercourse (James 1979). Fecundability has been tabulated either by age or by coital frequency, but not by both (Bongaarts and Potter 1983; Leridon 1977).

Without sex there can be no conception but there are other biological requirements: the woman needs to ovulate, and insemination must lead to a successful fertilization, which then has to result in a recognizable conception. Mathematical modelling of age-specific fecundability shows that the ability to conceive is quite constant between ages 25 and 40, while the ability to maintain a pregnancy starts to decline much earlier (Weinstein et al. 1990; Wood and Weinstein 1988). In our calculations intrauterine mortality was accounted for in the second step (Figure 15.1) when we estimated how many of the unintended pregnancies resulted in unwanted births and unsafe abortions.

While biological determinants of fecundability are fairly constant across populations, differences in frequency of sexual intercourse do have a substantial impact on fertility (Brown 2000; Weinstein et al. 1993). Increased frequencies of sexual intercourse raise fecundability, but the relationship is not linear. When coital frequency is low, the chance of conceiving is proportional to the frequency, but, at higher levels of monthly frequency, further increases are minimal (Potter and Millman 1986; Weinstein et al. 1990).

Should age-based or coitus-based estimates of fecundability be used? DHS data allow the calculation of both. Most surveys enquire about the most recent date of last sexual intercourse and frequency in the last month. We have used the data on the most recent date of the last intercourse to infer coital frequency, because it is less prone to recall error and normative responses than the question on coital frequency in the last month (Becker and Begum 1994). It is also available for more surveys included in our calculations. When the individual probability of coitus is constant throughout the month, the interval between two acts of intercourse is the reciprocal of the frequency. It has been shown-mathematically and empirically-that the distributions of the time since last sex and the interval between two acts have the same mean and variance (Leridon 1993). We could therefore estimate the coital frequency from the mean time since last sex. Our calculations were based on women who have been sexually active in the past year. Among these women, a large proportion did not have sex during the last month. Therefore, the average time since last sex was converted into monthly coital frequency for those who did have sex within the last month, while the others were given the value of 0 in order to derive the average monthly frequency among all women sexually active during the past year. Twelve countries lacked data on time since last sex, and for those we imputed frequencies taking the average values estimated from other countries for the three levels of fertility intention. The calculation of mean coital frequencies was done separately for samples of ever-married and samples that included all women. The frequencies for all women typically give lower values, especially among women who want to space their births (hereafter referred to as "spacers"), which include many women who may not have a regular partner, or not live with a partner.

Since the frequency of intercourse varies considerably according to childbearing intentions, with spacers and women who want to limit their births (hereafter referred to as "limiters") having less frequent sex than women who want a birth within the next two years, we have calculated expected pregnancies in two ways. The first uses model estimates of fecundability based on coital frequency (Bongaarts and Potter 1983), allowing different conception rates by fertility intention, while the second uses simple age-specific fecundability estimates, with monthly fecundability declining from 0.25 in the early 20s to 0 by age 45 years (Leridon 1977).

Table 15.4 shows the differentials in monthly coital frequencies by fertility desire and its impact on fecundability for the two age groups. Fecundability was expressed as the number of women expected to become pregnant at the end of one year, and this was contrasted with the age-specific fecundability as calculated from the monthly model estimates (Leridon 1977).

Overall, the mean coital frequencies derived from data on duration since last sexual intercourse seem very low, but reflect the fact that these are based on all women who had sex during the past year (rather than on those in stable cohabiting relationships). The effect of including unmarried, non-cohabiting women is especially evident among the spacers aged 15-29 years, a group which includes $25 \%$ of never-married women, compared with $13 \%$ among the limiters and $7 \%$ among those who desire a child soon. Single women have sex less frequently-and are likely to underreport sexual activity. For the 12 countries with evermarried samples, higher levels of coital frequency result in higher expected fecundability. But in most societies, sexual intercourse and therefore risk of pregnancy is not restricted to marriage. A study among nine African countries has shown that the time spent in marriage is not always a good proxy for sexual activity either, with high levels of inac-

Table 15.4 Mean monthly coital frequency by age and fertility preference for fecund non-users, and estimates of coitusbased and age-based fecundability

|  | 15-29 years | $30-44$ years |
| :--- | :---: | :---: |
| Mean coital frequency |  |  |
| Want children soon | 4.4 | 3.9 |
| Spacers | 2.1 | 3.0 |
| Limiters | 3.5 | 2.0 |
| Coitus-based fecundability (proportion pregnant in a year) |  |  |
| Want children soon | 0.82 | 0.78 |
| Spacers | 0.54 | 0.71 |
| Limiters | 0.72 | 0.54 |
| Age-based fecundability (proportion pregnant in a year) | 0.97 | 0.75 |

tivity recorded by married women, especially in West Africa (Brown 2000). The differential in fecundability according to fertility preference persists when the analysis is restricted to married women only and remains important.

Guided by consistency checks on internal validity of the data (discussed below), the relative risk estimates presented in this chapter were calculated using age-based fecundability estimates. Because of the importance of the effect, however, we will return to this topic in the section on uncertainty of estimates at the end of the chapter.

### 3.4 Combining exposure and fertility intention

The expected number of unwanted pregnancies was calculated separately from mistimed pregnancies. In fact, pregnancies were estimated for all nine combinations of exposure (modern method use, traditional method use, no use) and fertility intention (want birth soon, later, never). As shown schematically in Figure 15.2, contraceptive failures can thus result in pregnancies that are classified as intended. They occur among women who want a birth within the next two years.

More detail is provided in Figure 15.3 showing the calculation of the expected proportion of women having an unwanted pregnancy in the next year for each level of exposure, using data for women aged 30-44 years (all surveys combined).

Among the non-users, $11 \%$ were currently pregnant, $19 \%$ were amenorrhoeic or had not resumed sex since last birth, $21 \%$ were infecund or menopausal and another $10 \%$ had not had sex in the past year, leaving

Figure 15.2 Combining data on exposure and fertility intention to estimate pregnancies


Figure 15.3 Expected proportion of women having an unwanted pregnancy in the next year, by exposure

only $40 \%$ of the non-users exposed to the risk of conception. Forty-five per cent of these fecundable women wanted to limit their families (i.e. have no more children) and, based on their coital frequency, they had a $75 \%$ probability of conceiving in a 12 -month period. The product of these three numbers $(40 \% \times 45 \% \times 75 \%)$ gave the proportion of all non-users who were expected to have an unwanted pregnancy in the next 12 months, namely $13.5 \%$.

Among traditional (and modern) method users there may also be women who are biologically or behaviourally unexposed, but they are implicit in the failure rates (in contrast to conception rates). In other words, the presence of unexposed women among users will depress failure rates and thus obviate the need to take further account of such women. The calculation of the expected number of unwanted births was simply the product of the 12 -month failure probability and the percentage of those who wanted no more children. For all surveys combined, the failure rates for modern and traditional methods were $2.3 \%$ and $15.1 \%$, respectively, and the percentage wanting no more children was $76 \%$ among modern method users and $64 \%$ among traditional method users. The calculations in Figure 15.3 show that a non-user is 7.9 times $(13.5 \% / 1.7 \%)$ more likely than a modern method user to have an unwanted pregnancy in the next year.

Table 15.5 illustrates the calculation of the distribution of expected pregnancies in the next year, according to fertility intention. The righthand panel shows that among the modern method users, $1.7 \%$ will experience an "unwanted" failure, while $0.1 \%$ will have an "intended" failure. Among 100 non-users, 13.5 will have an unwanted pregnancy and there will be 9.6 intended pregnancies. The last row gives the

Table 15.5 Distribution of expected pregnancies estimated from contraceptive use by fertility intention (women 30-44, all surveys combined)

|  | Level of exposure: Contraceptive use (\%) | Fertility intention: ${ }^{\text {a }}$ want birth |  |  | Percentage of women expected to have intended, mistimed and unwanted pregnancies (within level of exposure |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Soon ${ }^{\text {b }}$ | Later | Never | Intended | Mistimed | Unwanted |
| Modern users | 28.2 | 5 | 18 | 76 | 0.1 | 0.4 | 1.7 |
| Traditional users | 8.1 | 11 | 25 | 64 | 1.7 | 3.8 | 9.6 |
| Non-users | 63.8 | 32 | 22 | 46 | 9.6 | 6.5 | 13.5 |
| Total | 100 |  |  |  | 6.3 | 4.6 | 9.9 |

${ }^{\text {a }}$ Per cent distribution with $\%$ wanting birth soon $+\%$ wanting birth later $+\%$ never wanting birth adding to $100 \%$.
b Soon means a birth within the next 2 years.
expected pregnancies, weighted by the contraceptive distribution. In total, $21 \%$ of women aged $30-44$ years are expected to get pregnant at current levels of contraceptive use. According to prevailing fertility intentions, $10 \%$ of women will have an unwanted pregnancy.

### 3.5 Estimating pregnancy outcomes

This section refers to the second step in Figure 15.1, estimating unwanted births and unsafe abortions from unwanted and mistimed pregnancies.

## SPONTANEOUS PREGNANCY LOSS

Before applying abortion probabilities to unwanted and mistimed pregnancies, we allowed for miscarriages and stillbirths. Recognizable intrauterine mortality is lowest in the early twenties (16\%), but reaches double this rate by age 45 years (Bongaarts and Potter 1983). From these tabulated data we calculated the average spontaneous pregnancy loss as $17 \%$ and $27 \%$ for the younger and older women, respectively. Pregnancy loss does not vary by fertility intention, and the expected pregnancies in Table 15.5 need to be reduced by $27 \%$, leaving $15.3 \%$ of women aged 30-44 years pregnant. Thus after accounting for pregnancy loss, 7.2\% of women have an unwanted pregnancy and $3.4 \%$ have a mistimed pregnancy. A proportion of these will be aborted while the rest will result in a live birth.

## Abortion probabilities

Abortion probabilities have been derived by converting the WHO country estimates of incidence ratios (unsafe abortions per 100 live births) to abortion probabilities (ratio of abortions to abortions plus births). For countries without estimates $(n=10)$ we took the WHO
regional abortion estimate (WHO 1998). The average abortion incidence ratio of 17.5 abortions per 100 live births translates into an overall abortion probability per pregnancy of $0.15(=17.5 /(100+17.5))$. Because intended pregnancies are most unlikely to be aborted, this probability needs to be converted to relate to unintended pregnancies only, while keeping the overall abortion ratios constant.

The main reason stated by women for having an abortion is to stop bearing children; following this is the wish to postpone pregnancy (Bankole et al. 1999). Other reasons included disruption of education and the belief that they were too young to have children, especially in Africa where single women are sexually active, all adding to the postponement component of abortion. An in-depth study in Maharashtra, showed that $54 \%$ of aborted pregnancies among married women were defined as unwanted and $42 \%$ as mistimed at the time of conception (Ganatra et al. 2000). In three central Asian republics, the proportion of unwanted pregnancies that were aborted ranged from $74 \%$ to $86 \%$, while two-thirds of all mistimed pregnancies were aborted (Westoff et al. 1998).

Consistent with the limited evidence available, our calculations assumed that the abortion probability of a mistimed pregnancy was half that of an unwanted pregnancy. Though this assumption is essentially arbitrary, it is consistent with the judgements of experts. It can be regarded as reasonable but nevertheless must be viewed with caution.

The relative distribution of mistimed, unwanted and planned pregnancies expected for 100 women (at the current contraceptive prevalence and childbearing intentions) was used to relate abortions to mistimed and unwanted births only. Of the 15.3 pregnancies among $30-44$-year olds (after spontaneous pregnancy loss), 3.4 are expected to be mistimed and 7.2 unwanted. So the abortion probability of unintended pregnancies is calculated as $0.15 \times 15.3 /(7.2+3.4 / 2)=0.25$. This probability implies that one in four unwanted pregnancies will be aborted compared with one in eight of mistimed pregnancies. Although we started from the same incidence ratio for both age groups, the fact that only unintended pregnancies are aborted leads to different abortion probabilities for the two age groups. In the countries with high rates of unsafe abortion and lack of systematic data, we know little about how abortions vary by age. Where official statistics are more complete, there are generally two age patterns of abortion ratios (Bankole et al. 1999). The first takes a Ushape, where abortion is high both among unmarried young women and older women who have reached their desired family size. The second pattern is a steady increase in the abortion ratio with age. As default we have chosen to use the same ratio for both age groups. Where consistency checks indicated negative levels of unwanted births in the younger age group (Guinea, Indonesia and Thailand) or where data on age patterns were available (Kazakhstan, Kyrgyzstan, Uzbekistan) the ratio was adjusted by age. For countries that needed correction, the abortion prob-
ability of (any) pregnancy to women aged 30-44 years was assumed to be three times the probability for pregnancies to younger women (estimate based on the data provided for the latter three countries in the article by Bankole et al. 1999).

We have implicitly assumed that abortion probabilities of unwanted (and mistimed) pregnancies are the same regardless of whether they resulted from method failure or non-use. Common sense suggests that women who act on their intention to prevent an unwanted pregnancy by adoption of contraception are more determined to regulate fertility than other women and thus more likely to seek a termination when pregnant. Very limited empirical evidence on differential abortion probabilities supports this expectation. In Turkey, for instance, $28.5 \%$ of unintended pregnancies in 1998 resulting from non-use were aborted, compared with $38.1 \%$ and $35.2 \%$ of pregnancies resulting from modern and traditional method failure, respectively (Senlet et al. 2000). In Kazakhstan, $51 \%$ of unintended pregnancies among non-users were aborted compared with $67 \%$ of (all) contraceptive failures (Westoff 2000). The inferred differences are relatively small and their generalizability is unknown. Hence, it was decided to apply the same abortion probabilities for all three contraceptive use categories. Nevertheless, since the desire to discontinue childbearing altogether is highest among modern method users, our simulations did result in higher abortion probability for all modern method failures relative to traditional method failures and conceptions among non-users.

## Probability and relative risk of having an abortion

Abortion probabilities have been combined with expected proportions of women who have unwanted and mistimed pregnancies in each exposure category. Table 15.6 provides the worked example of the expected proportion of women aged 30-44 years (all surveys averaged) having an abortion. The expected percentages of women having pregnancies are $27 \%$ lower than in Table 15.5, since they are adjusted for spontaneous pregnancy loss.

The abortion probability of 0.25 for unwanted births results in an expected proportion of $3.1 \%$ of non-users having an abortion in the next

Table 15.6 Probability of having an abortion in each exposure category and the resulting relative risk ratios

|  | Expected \% of women having pregnancies |  |  | Expected \% of women having an abortion | $R R$ of having an abortion |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Intended | Mistimed | Unwanted |  |  |
| Modern users | 0.1 | 0.3 | 1.3 | 0.4 | 1.0 |
| Traditional users | 1.2 | 2.8 | 7.1 | 2.2 | 6.0 |
| Non-users | 7.1 | 4.9 | 9.9 | 3.1 | 8.7 |

year $(0.25 \times 9.9 \%$ [unwanted pregnancies] $+0.25 / 2 \times 4.9 \%$ [mistimed pregnancies]). Following the same logic, only $0.4 \%$ of modern method users are expected to have an abortion. Using the modern method users as the reference category the relative risk was derived as the ratio of expected proportion of women having an abortion among non-users compared with modern method users ( $R R=8.7$ ). Similarly, traditional method users were projected to be six times more likely to have an abortion than modern method users.

Probabilities of women in each exposure category of HAVING AN UNWANTED BIRTH

The proportions of modern and traditional method users and of nonusers who are expected to deliver an unwanted birth were calculated by applying the complement of the abortion probability to the unwanted pregnancies. Mistimed pregnancies that end as live births do not contribute to the burden of maternal outcomes. The worked example in Table 15.7 shows that with an abortion probability of 0.25 , threequarters of the $9.9 \%$ of non-users with an unwanted pregnancy are expected to carry it to term. With $7.4 \%$ of non-users having an unwanted birth in the next year, compared with $0.9 \%$ of modern method users, the relative risk is 7.8 .

Since the WHO abortion estimates relate to unsafe abortions only (which is required for the calculation of relative risk to have an unsafe abortion), we have potentially overestimated the expected number of unwanted births in countries where legal abortions are common. This does not affect the relative risk of unwanted birth, but it does affect the proportion unwanted among all births. As can be seen from Table 15.1, most surveys were done in countries with highly restrictive abortion laws, and therefore most abortions will be unsafe. For the countries with available data on legal abortions (Henshaw et al. 1999), the proportions of unwanted births among all births were calculated using legal abortion rates. The legal abortion ratio was used for Turkey. For Bangladesh and India, the official legal rates are very low and are underestimates of actual procedures performed (Henshaw et al. 1999). In India, abortion

Table I5.7 Probability of having an unwanted birth in each exposure category and the resulting relative risk ratios

|  | Expected \% of women having <br> pregnancies |  |  | Expected \% of women <br> having an | RR of having <br> an unwanted <br> birth |
| :--- | :---: | :---: | :---: | :---: | :---: |
|  | Planned | Mistimed | Unwanted | Uirth |  |
| Modern users | 0.1 | 0.3 | 1.3 | 0.9 | 1.0 |
| Traditional users | 1.2 | 2.8 | 7.1 | 5.3 | 5.6 |
| Non-users | 7.1 | 4.9 | 9.9 | 7.4 | 7.8 |

has been legal for 30 years, but abortion services by authorized facilities are inadequate, especially in rural areas. Many women are even not aware that abortion is legal and resort to abortions from both unskilled and skilled providers (Ganatra et al. 2000). In Bangladesh, with highly restrictive abortion laws, the legal rate refers to menstrual regulation services (manual aspiration evacuation of the uterus without prior confirmation of pregnancy). They are widely available, effectively providing abortion up to eight weeks of a woman's last menstrual period (Rahman et al. 1998). For India and Bangladesh we used the WHO estimates on unsafe abortions.

### 3.6 CONSISTENCY CHECKS

Two basic consistency checks guided the assessment of the plausibility of the data inputs and the method assumptions, and informed the adjustments done for a few individual countries. We compared both the total numbers of projected births and the proportion of unwanted births among all births with retrospective estimates based on the question "At the time you became pregnant with (NAME OF CHILD) did you want to become pregnant then, did you want to wait until later or did you want no more children at all?" Age-specific fertility rates published by DHS allowed the calculation of the yearly number of births in each age group. These rates were averaged for a period of three or five years prior to survey data in order to reduce sampling error. The number of recent births per year should agree with our expected number of births per 1000 women in each age group, after adjusting for the proportions of women who never had sex. Exact matches for each survey are not expected, but systematic patterns of excess or shortfall of births may imply regional biases.

Compared with the current births calculated from the age-specific fertility rates, the projection using age-based fecundability estimates (averaging the outcomes for all surveys) results in a $15 \%$ shortfall of expected births for the younger age group, and $2 \%$ excess for the older age group. What may contribute to the underestimation of expected births? A minor factor may be that we discount miscarriages and stillbirths before calculating abortions. Since some of the aborted pregnancies might have resulted in a spontaneous intrauterine death, we may have underestimated the pregnancies carried to term. A second factor may be abortion probabilities that are too high. While this is unlikely overall, the assumption of constant incidence ratios by age may contribute to the shortfall of expected births in the youngest age group. Continuing fertility decline would also contribute to the discrepancy between projected and retrospective fertility levels. Finally, underreporting of sexual exposure (both overstating virginity and time since last sex) may well be a major cause of the deficit of births to young women.

It should be pointed out that in terms of estimating the relative risk ratios of having an unwanted birth or an abortion, overall numbers of
births are less important than the relative distribution of births across fertility intention. As we indicated before, frequency of sexual intercourse is not the same for women who want a birth soon and those who either want to space or stop childbearing. Moreover, there is no reason to doubt the validity of these relative differences. This is an important factor contributing to the uncertainty around the estimates (discussed later).

This leads us into the second consistency check, the comparison of projected and retrospective estimates of the proportion of births that are unwanted. Compared with the proportion of the most recent births that were reported as unwanted, serious discrepancy was apparent with projected proportions of unwanted births: they were $63 \%$ and $143 \%$ higher for the $15-29$-year olds and the $30-44$-year olds, respectively. We discuss below potential biases that may account for the discrepancy.

The proportion unwanted among most recent births is available if this birth occurred within the last five, or in some surveys, three years. It is derived from the question "At the time you became pregnant with (NAME OF CHILD), did you want to have more children then, did you want to wait until later, or did you want no more children?" The question is meant to draw on the memory of the feelings that were held at the time of conception. Though little evidence exists to evaluate to what extent respondents later report as being wanted those children whose conception was initially unwanted, the most important reason why the "wantedness" of the most recent birth may not agree with our projected number of unwanted births is undoubtedly ex-post rationalization (Bongaarts 1990; Westoff 1981). Panel data in Morocco allowed the comparison of reports on the "wantedness" of 0-2-year-old children in 1992 with reports on the same births three years later (Westoff and Bankole 1998). Whereas $6 \%$ of children reported in 1992 as wanted at the time of conception were later described as unwanted, as many as $62 \%$ of the unwanted pregnancies in the first round were reported as wanted in 1995. The older the child, the more likely reports on the feelings or intentions had changed from unwanted to wanted. A comparison among five DHS confirms the result from Morocco. The percentage unwanted declines as the age of children increased, though the trend was less pronounced than in the Morocco data (Montgomery et al. 1997). Data on change in perception before and after a baby is born are not available. However, in a recent qualitative study in the United Kingdom of Great Britain and Northern Ireland among pregnant women and those who recently aborted, women generally agreed that conceptions which were initially "unplanned" or "unintended" could become "wanted", a term many women associated with the decision to carry the pregnancy to term (Barrett and Wellings 2002). Since we do not know how these concepts are translated and understood by different cultures, the magnitude of the rationalization bias cannot be quantified.

An additional explanation is that the most recent birth refers to a child of lower birth order than the projected next birth. Calculating orderspecific estimates on the "wantedness" of the last child and applying these to the birth order distribution shifted by one child allows assessment of the magnitude of this order effect on fertility preferences. The average effect is surprisingly low: $9 \%$ of women aged $<30$ years report their most recent birth as being unwanted, while this would be raised to only $11 \%$ by shifting the order by one child. For the older women the effect was equally small.

A last potential factor contributing to the discrepancy might be the underestimation of the proportion of unwanted pregnancies that are terminated. By considering unsafe abortions, rather than all abortions, we may have overestimated the expected unwanted pregnancies carried to term. When the relative abortion probability of unwanted over mistimed pregnancies is lower than two, then we could again overestimate the number of unwanted births.

Given the considerable differences between the various measures of preferred fertility and the known bias in underreporting of unwanted births (Bongaarts 1990; Westoff 1981), we have to tolerate a level of inconsistency between projected and retrospective estimates of "wantedness". Important biases other than our model assumptions are operating and this precludes the use of the extent of discrepancy as a guide to make adjustments. Nevertheless, for selected countries with available data on legal rates and age pattern of abortion, adjustments were done (as explained above). We have been guided more by disparity between the age groups, than any discrepancy in expected level of fertility. While the abortion adjustments do not affect relative risk ratios, they do lower the proportion of unwanted among all births, which is used in projecting how much of the obstetric burden can be avoided by reducing unwanted births.

### 3.7 Counterfactual scenario

## Counterfactual distribution of Contraceptive use

The burden of maternal outcomes, including abortion, attributable to lack of effective modern contraception was calculated as the reduction in current burden that would be observed if levels of exposure were reduced to a counterfactual distribution of contraceptive use. Theoreti-cal-minimum-risk distribution of contraception does not mean $100 \%$ modern use, but rather that all women with a desire to either stop or postpone childbearing for at least another two years, adopt an effective modern method of contraception. Perfect implementation of fertility preferences among limiters and spacers obviously results in a higher proportion of modern method users and fewer women using traditional contraception or no contraception at all. All traditional method users and fecund non-users now consist of women who want a birth in the next
two years. This theoretical minimum level of exposure was thus simply calculated by shifting all spacers and limiters into modern method use.

The potential impact fractions were used to estimate the proportional reduction in the total number of unwanted births and unsafe abortions by a change in contraceptive prevalence. Potential impact fractions generally assume that only exposure changes, while the relative risk of the outcome for each level of exposure stays the same. However, as we demonstrated in the previous section, the relative risk of an abortion (or unwanted birth) among the traditional method users not only depends on the failure rates, but also on the fertility preference among these users. The factors determining the relative risk among non-users are conception rates, fertility preferences and the proportion fecund among the nonusers. While both the failure and conception rates remain constant, the relative fertility preferences in each exposure category and the proportion of women who are fecund among non-users will change with a change in contraceptive distribution.

## Counterfactual relative risk ratios

How do we expect the relative risk to vary with a change in exposure? Under the scenario of theoretical-minimum-risk, all women with a desire to stop or space childbearing will adopt modern methods and all expected conceptions in traditional method users and non-users will now be intended pregnancies. Because only the reference group (modern users) is at risk of an unintended pregnancy, the relative risk of unwanted births and abortions will be 0 in other groups. For intermediate levels of shifting the counterfactual relative risk will be between 0 and the current level of relative risk.

In deriving these counterfactual levels of relative risk, the degree of shifting was assumed to be the same for spacers and limiters. The calculation involved computing counterfactual proportions of fecundable women among the non-users, and counterfactual fertility preferences in all three categories of exposure.

For each level of shifting, the counterfactual distribution of contraceptive use was determined by subtracting the number of women using traditional methods and no contraception at all and adding that to those using modern methods. The distribution of non-users was recalculated by keeping the number of infecund/menopausal women and those that were not sexually active in the past year constant; the number of fecundable women changed by subtracting the number of limiters and spacers who shifted to using modern methods; the number of pregnant and amenorrhoeic women changed since more effective contraception implies a reduction in the number of births, with at any one time a smaller proportion of women pregnant or in the amenorrhoeic state. Through a process of iteration we imputed the ratio of births in the counterfactual over current population and assumed that the number of pregnant and amenorrhoeic change to the same extent. These numbers were combined
in the counterfactual proportion of fecund women among non-users. The counterfactual distribution of fertility desires among the three categories of exposure was calculated taking account of the level of shifting and counterfactual proportion of fecundity among the non-users. Table 15.8 compares current with counterfactual exposure and fertility intentions, calculated for women aged 30-44 years, all countries combined, assuming $50 \%$ shifting.

These counterfactual contraceptive distributions, fertility desires and proportion fecund among non-users were then combined to obtain relative risks in the same way as explained in the previous section for the "current" relative risk levels. Table 15.9 contrasts the counterfactual relative risk for the same scenario of $50 \%$ shifting, among women aged $15-29$ years and 30-44 years, with the relative risk levels under current exposure.

The relative risks among non-users initially decline faster with the degree of shifting, compared with levels among traditional method users. When half of all non-users and traditional users who currently have an unmet need for spacing or limiting have adopted a modern method, the relative risk for a non-user to have an unwanted birth has decreased from 7.8 to 5.0. The relative risk for traditional users drops to a lesser extent (from 5.6 to 5.2 ). The steeper decline among the non-users can be explained mainly by the fact that the infecund, the menopausal and the sexually inactive women gradually become a larger proportion of all non-users. The pattern of the counterfactual relative risk levels by degree of shifting varies across countries. The average pattern (for all surveys combined, women aged 30-44 years) shows the most common pattern as depicted in Figures 15.4 and 15.5 of a near-linear decline for

Table 15.8 Current and counterfactual contraceptive distribution and fertility preferences among women aged 30-44 years, all surveys combined

|  | Current exposure and intention |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Contraceptive use | Want birth soon | Want to space | Want to limit |
| Modern | 28.2 | 0.05 | 0.18 | 0.76 |
| Traditional | 8.1 | 0.11 | 0.25 | 0.64 |
| Non-use ${ }^{\text {a }}$ | 63.8 | 0.32 | 0.22 | 0.46 |
| Counterfactual exposure and intention (50\% shifting) |  |  |  |  |
| Modern | 37.4 | 0.04 | 0.22 | 0.74 |
| Traditional | 4.6 | 0.20 | 0.22 | 0.58 |
| Non-use ${ }^{\text {a }}$ | 58.1 | 0.49 | 0.17 | 0.34 |

[^58]Table I5.9 Relative risk ratios for unsafe abortions and unwanted births under current and counterfactual exposure, both age groups, all surveys combined

|  | RR under current exposure | $R R$ under counterfactual exposure <br> assuming $50 \%$ |  |  |
| :--- | :---: | :---: | :---: | :---: |
| Abortions | $15-29$ years | $30-44$ years | $15-29$ years | $30-44$ years |
| Modern | 1.0 | 1.0 | 1.0 | 1.0 |
| Traditional | 3.8 | 6.0 | 3.2 | 5.4 |
| Non-use | 6.2 | 8.7 | 3.8 | 5.4 |
| Unwanted birth |  |  |  |  |
| Modern | 1.0 | 1.0 | 1.0 | 1.0 |
| Traditional | 2.9 | 7.6 | 2.7 | 5.2 |
| Non-use | 4.5 |  | 3.1 | 5.0 |

Figure 15.4 Pattern of change in relative risk of an unwanted birth by level of switching for traditional method users, women aged 30-44 years

non-users, while the relative risks for traditional method users tend to stay relatively constant up to $50-60 \%$ shifting, declining rapidly to 0 thereafter.

For traditional use the relative risk of unwanted births and abortions sometimes increases initially at lower levels of shifting. This seems to happen when more spacers than limiters move into the modern method users, which has little effect on the distribution of fertility preferences among the traditional users, but affects the expected proportion of modern users who will be having relatively more "mistimed failures"

Figure I5.5 Pattern of change in relative risk of an unwanted birth by level of switching for non-users, women aged 30-44 years

rather than "unwanted failures". Thus the expected proportion of modern method users having unwanted births actually decreases slightly. Since the expected proportion of unwanted births among the non-users changes much faster, this pattern of an initial relative risk rise is not detected so much among the non-users.

### 3.8 Aggregating country-specific estimates into subregions

The country level results in terms of contraceptive exposure, relative risks of unsafe abortions and unwanted births were aggregated into subregional averages, weighting each country according to the size of the population (see Table 15.1 for weighting factors). The derived estimate was then taken as subregional average, assuming that the countries were reasonably representative of the whole subregion. Table 15.10 presents the coverage of subregional population by countries with data.

As can be seen from the subregional coverage presented in Table 15.10, the African, Latin American and some of the Asian subregions are well represented. The three subregions where relevant data are totally lacking are AMR-A, EUR-A and WPR-A, which mainly consist of industrialized low-fertility countries with small burdens of maternal mortality or unsafe abortion. The exposure distributions for these subregions were imputed from data provided in the United Nations (UN) report on levels and trends in contraceptive use (UN 1999). As the UN data were for married women aged 15-49 years, we assumed $90 \%$ of total estimate for the age group 15-29 years and $110 \%$ of estimate for 30-44year olds. Since safe abortion is widely available in most of these countries, the relative risk levels were arbitrarily set at 1.5 for unwanted births (both for traditional users and non-users). For abortion we took relative

Table 15.10 Coverage of subregional population by countries with data from DHS

| Subregion | Number of DHS per <br> subregion | Percentage of the subregional population <br> covered by countries with DHS data |
| :--- | :---: | :---: |
| AFR-D | 9 | 80 |
| AFR-E | 14 | 69 |
| AMR-A | 0 | 0 |
| AMR-B | 6 | 79 |
| AMR-D | 6 | 100 |
| EMR-B | 1 | 8 |
| EMR-D | 5 | 85 |
| EUR-A | 0 | 0 |
| EUR-B | 3 | 46 |
| EUR-C | 1 | 7 |
| SEAR-B | 3 | 100 |
| SEAR-D | 3 | 93 |
| WPR-A | 0 | 0 |
| WPR-B | 1 | 5 |

risks of 2 and 3, respectively, for traditional method users and non-users. It matters little what the real risks are since the burden of obstetric and abortion complications is negligible. Subregions that are cause for greater concern are EMR-B, EUR-C and WPR-B. Tunisia is the only country representing EMR-B and Kazakhstan the only one in EUR-C. Most worrying is that WPR-B, a subregion in which $83.5 \%$ of the population live in China, is represented solely by the Philippines. In China, because of the strict anti-natal policies, there will be very few unwanted births and most abortions are legal and safe. Therefore an adjustment was needed for WPR-B, which is discussed in the next section on deriving attributable burden.

### 3.9 Subregional estimates of input parameters

The inputs provided for calculating the attributable burden are the contraceptive distributions and relative risks under the current situation and counterfactual (theoretical minimum) scenario, and the proportion of all births that are unwanted.

We first estimated the proportions of women who were excluded from the analysis because they never had sex and were therefore not exposed to the possibility of pregnancies (i.e. virgins). Table 15.11 gives the proportion of the female population that is presumed to be virgins. Our method overestimates virginity in two ways. In the countries with evermarried samples, single women do not enter the analysis and are assumed to be virgins. Clearly, this assumption is not entirely valid. In the other

Table I5.II Average proportion of women who are virgins, by subregion

| Subregion | $15-29$ years | $30-44$ years |
| :--- | :---: | :---: |
| AFR-D | 23.6 | 0.5 |
| AFR-E | 26.1 | 0.5 |
| AMR-A | 15.0 | 1.5 |
| AMR-B | 39.4 | 4.1 |
| AMR-D | 43.0 | 3.7 |
| EMR-B | 65.1 | 7.4 |
| EMR-D | 48.8 | 4.2 |
| EUR-A | 15.0 | 1.5 |
| EUR-B | 46.5 | 3.1 |
| EUR-C | 42.1 | 2.1 |
| SEAR-B | 45.8 | 5.6 |
| SEAR-D | 27.6 | 1.2 |
| WPR-A | 15.0 | 1.5 |
| WPR-B | 59.8 | 8.7 |

countries virginity may be overstated to the extent that single women underreport sexual activity. The estimates for the three subregions where we lacked data, AMR-A, EUR-A and WPR-A, were informed by data from the British Sex Survey in 1990/91 (Johnson et al. 1994).

Current contraceptive use, averaged for each subregion, is shown in Table 15.12 and the counterfactual scenario in Table 15.13. The second table thus provides the contraceptive status that would have been observed if all women with a current need for spacing and limiting were to adopt a modern method. Note that a small residue of traditional method users remains; these are women who want a child within the next two years. The subregional variation in current contraceptive use is thus explained by variations in desired family size and extent of implementation of these desires through contraception (and abortion). Subregional variations in counterfactual contraceptive use reflect differences in desired family sizes and use of abortion as alternative means to implement fertility intentions.

Tables 15.14 and 15.15 show the relative risk estimates for having an unwanted birth and an unsafe abortion, respectively, under the current regime of contraceptive practice and fertility preferences. The levels for the counterfactual minimum risk have not been provided since they are all 0 , as explained earlier. As can be seen, there is wide subregional variation in relative risks, both for traditional method users and nonusers. The subregion that stands out is SEAR-D, where India represents a large part of the population. Modern contraception in India (and

Table 15.12 Averages of current contraceptive distribution among women who ever had sex, by subregion

|  | I5-29 years |  |  |  | $30-44$ years |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Modern <br> method | Traditional <br> method | Non-use |  | Modern <br> method | Traditional <br> method | Non-use |
| AFR-D | 8.0 | 9.1 | 82.9 |  | 10.7 | 7.7 | 81.6 |
| AFR-E | 13.0 | 5.4 | 81.6 |  | 15.6 | 5.6 | 78.8 |
| AMR-A | 60.6 | 2.6 | 26.8 |  | 74.1 | 3.2 | 32.8 |
| AMR-B | 48.1 | 6.9 | 45.0 |  | 62.1 | 6.8 | 31.1 |
| AMR-D | 27.9 | 11.6 | 60.5 |  | 36.2 | 14.6 | 49.2 |
| EMR-B | 30.4 | 7.7 | 61.9 |  | 44.8 | 10.8 | 44.5 |
| EMR-D | 14.9 | 2.6 | 82.4 |  | 23.5 | 4.1 | 72.4 |
| EUR-A | 62.1 | 5.2 | 22.7 |  | 75.9 | 6.4 | 27.7 |
| EUR-B | 36.0 | 16.6 | 47.4 |  | 46.9 | 21.8 | 31.3 |
| EUR-C | 42.4 | 12.4 | 45.2 |  | 55.5 | 12.2 | 32.3 |
| SEAR-B | 54.4 | 2.6 | 43.0 |  | 55.9 | 4.1 | 40.0 |
| SEAR-D | 23.5 | 4.0 | 72.5 |  | 48.3 | 5.3 | 46.4 |
| WPR-A | 48.7 | 5.0 | 36.3 |  | 59.5 | 6.1 | 44.4 |
| WPR-B | 24.1 | 18.0 | 57.9 |  | 29.5 | 20.3 | 50.2 |

Table 15.13 Averages of counterfactual contraceptive distribution (theoretical minimum) among women who ever had sex, by subregion

| Subregion | 15-29 years |  |  | 30-44 years |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Modern method | Traditional method | Non-use | Modern method | Traditional method | Non-use |
| AFR-D | 43.1 | 2.5 | 54.4 | 42.5 | 1.8 | 55.8 |
| AFR-E | 47.7 | 1.5 | 50.8 | 49.9 | 0.8 | 49.4 |
| AMR-A | 60.6 | 2.6 | 26.8 | 74.1 | 3.2 | 32.8 |
| AMR-B | 84.7 | 1.1 | 14.3 | 84.4 | 0.7 | 14.9 |
| AMR-D | 72.9 | 2.0 | 25.1 | 72.8 | 1.7 | 25.5 |
| EMR-B | 66.2 | 1.4 | 32.5 | 82.6 | 1.0 | 16.4 |
| EMR-D | 48.6 | 0.4 | 51.0 | 62.3 | 0.4 | 37.4 |
| EUR-A | 62.1 | 5.2 | 22.7 | 75.9 | 6.4 | 27.7 |
| EUR-B | 70.3 | 2.7 | 27.0 | 82.4 | 0.8 | 16.8 |
| EUR-C | 76.1 | 2.0 | 21.9 | 79.5 | 0.9 | 19.6 |
| SEAR-B | 71.2 | 0.3 | 28.5 | 73.0 | 0.5 | 26.6 |
| SEAR-D | 60.7 | 0.9 | 38.4 | 74.0 | 0.2 | 25.8 |
| WPR-A | 48.7 | 5.0 | 36.3 | 59.5 | 6.1 | 44.4 |
| WPR-B | 72.7 | 1.2 | 26.1 | 70.0 | 1.0 | 28.9 |

Table 15.14 Average relative risk of having an unwanted birth under the current contraceptive scenario, by subregion

|  | Traditional method users |  |  | Non-users |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Subregion | $15-29$ years | $30-44$ years |  | $15-29$ years |  |
| AFR-D | 2.3 | 3.2 |  | $30-44$ years |  |
| AFR-E | 1.6 | 4.4 |  | 3.8 |  |
| AMR-A | 1.5 | 1.5 | 3.0 | 3.5 |  |
| AMR-B | 3.2 | 7.9 | 1.5 | 4.9 |  |
| AMR-D | 4.1 | 6.4 | 8.3 | 1.5 |  |
| EMR-B | 3.2 | 5.9 | 6.2 | 14.3 |  |
| EMR-D | 3.2 | 5.1 | 3.9 | 9.8 |  |
| EUR-A | 1.5 | 1.5 | 3.4 | 8.4 |  |
| EUR-B | 4.0 | 6.3 | 1.5 | 6.8 |  |
| EUR-C | 5.1 | 5.6 | 5.1 | 1.5 |  |
| SEAR-B | 4.1 | 6.1 | 6.9 | 11.1 |  |
| SEAR-D | 4.6 | 22.2 | 4.3 | 12.7 |  |
| WPR-A | 1.5 | 4.6 | 8.0 |  |  |
| WPR-B | 3.7 |  |  | 1.5 |  |

Table 15.15 Average relative risk of having an unsafe abortion under the current contraceptive scenario, by subregion

|  | Traditional method users |  |  | Non-users |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Subregion | $15-29$ years | $30-44$ years |  | $15-29$ years |  |
| AFR-D | 2.6 | 3.5 |  | $30-44$ years |  |
| AFR-E | 3.0 | 4.7 |  | 3.9 |  |
| AMR-A | 2.0 | 2.0 |  | 4.7 |  |
| AMR-B | 4.1 | 8.4 | 3.0 | 4.2 |  |
| AMR-D | 4.3 | 6.8 | 10.6 | 5.4 |  |
| EMR-B | 4.6 | 6.0 | 7.6 | 3.0 |  |
| EMR-D | 3.9 | 5.2 | 6.4 | 16.2 |  |
| EUR-A | 2.0 | 2.0 | 5.0 | 10.6 |  |
| EUR-B | 4.9 | 6.3 | 3.0 | 10.0 |  |
| EUR-C | 5.0 |  | 7.5 | 7.7 |  |
| SEAR-B | 5.5 | 6.4 | 10.5 | 3.0 |  |
| SEAR-D | 6.6 | 22.7 | 6.4 | 11.5 |  |
| WPR-A | 2.0 | 2.0 | 10.9 | 13.3 |  |
| WPR-B | 4.1 |  | 3.2 | 9.2 |  |

Nepal) is dominated by sterilization, which has a very low failure rate. As very few modern method users become pregnant, the relative risks for the other two categories are raised. This observation has complicated the interpretation of the counterfactual relative risks for abortions, as one of our model assumptions has obviously been violated. Keeping failure rates and contraceptive method mix constant with an increase in contraceptive prevalence has become internally inconsistent with the fertility preferences. Where sterilization is the dominant modern contraceptive method, it is impossible to accommodate the needs of spacers at the current method mix. Increased contraceptive use among spacers in such settings would have to involve uptake of reversible methods but this trend would reduce the average effectiveness of modern method use, because reversible methods have higher failure rates than sterilization. And this, in turn, would reduce the relative risk of an abortion for traditional method users and non-users. The attributable fractions derived from these relative risks of unsafe abortions for these countries with high use of sterilization will thus be a slight overestimate. However, since spacers do not contribute to unwanted births, the relative risks for unwanted births remain unaffected. Only when limiters adopt methods other than sterilization, as method choice widens, would the relative risks for unwanted births be affected.

## 4. Attributable burden

By combining the relative risk values for an unsafe abortion (Table 15.15) with the data on current and counterfactual distributions of contraceptive use (Table 15.12 and Table 15.13), we have derived the attributable fractions for unsafe abortions (Table 15.16). These estimates show that a very large proportion of the disease burden due to abortion complications is attributable to unprotected sex or use of less effective traditional methods. The residual is the "unavoidable" burden associated with modern method failure.

Table 15.17 gives the estimated burden of disease attributable to unsafe abortions by subregion. Both deaths and total DALYs have been presented as well as the relative burden in each subregion, expressed as DALYs per 1000 women, to allow better comparison between the subregions. The estimated burden of abortion attributable to non-use and use of ineffective methods of contraception is 4.4 million DALYs, with $82 \%$ of the burden falling on women aged $<30$ years. South Asia with its large population (SEAR-D) has the highest abortion burden at about $35 \%$ of the total abortion burden in both age groups. Although AFR-D and AFR-E include smaller populations than SEAR-D, women in these subregions have the highest relative burden.

The calculation of attributable burden for unwanted births is slightly more complicated. The attributable fraction for unwanted births could not be applied to the total burden of obstetric complications in

Table 15.16 Estimates of attributable fraction for unsafe abortions, by subregion

|  | Attributable fraction (\%) |  |
| :--- | :---: | :---: |
| Subregion | $15-29$ years | $30-44$ years |
| AFR-D | 88 | 89 |
| AFR-E | 88 | 90 |
| AMR-A | 59 | 59 |
| AMR-B | 85 | 87 |
| AMR-D | 87 | 89 |
| EMR-B | 86 | 85 |
| EMR-D | 90 | 90 |
| EUR-A | 56 | 56 |
| EUR-B | 86 | 85 |
| EUR-C | 87 | 86 |
| SEAR-B | 79 | 84 |
| SEAR-D | 93 | 95 |
| WPR-A | 68 | 71 |
| WPR-B | 85 | 88 |

childbirth, since the many wanted births also contribute to maternal deaths and morbidity. We thus had to restrict the burden to that proportion of all births that was unwanted as simulated from contraceptive failure, conception rates and fertility preferences. These proportions are presented in Table 15.18, together with the unadjusted attributable fractions. These fractions express what proportion of unwanted births can be avoided by perfect implementation of fertility preferences. The proportion of unwanted among all births is much higher for the older age group, as the desire to limit family size is much more prevalent than among younger women. The negative values for women aged 15-29 years in EUR-C and SEAR-B reflect the fact that the abortion probabilities assumed for these women were probably still too high. In calculations the proportion was set to 0 , and in the subregions concerned we may have slightly underestimated the burden of maternal complications in childbirth for the younger age group while overestimating it for older women. Finally, these proportions have been combined with the unadjusted attributable fractions (unwanted births only), to give the attributable fractions of all births that could be averted if all women who wish to stop childbearing used a modern method.

As mentioned earlier, the estimate for WPR-B needed adjusting since China, with very low rates of unwanted births, dominates the subregion. The easiest adjustment procedure was to keep the relative risk and contraceptive prevalence estimates for the Philippines, and adjust the
Table 15.17 Estimated burden of disease attributable to unsafe abortions (due to non-use of contraception), by subregion and age

| Subregion | Women 15-29 years |  |  |  | Women 30-44 years |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Attributable fractions (\%) | Deaths | DALYs | DALYs per 1000 women | Attributable fractions (\%) | Deaths | DALYs | DALYs per 1000 women |
| AFR-D | 88 | 5437 | 604075 | 14944 | 89 | 2671 | 91241 | 4062 |
| AFR-E | 88 | 10133 | 895872 | 18686 | 90 | 3717 | 124403 | 4741 |
| AMR-A | 59 | 12 | 496 | 15 | 59 | 14 | 370 | 10 |
| AMR-B | 85 | 808 | 69473 | 1151 | 87 | 435 | 13608 | 297 |
| AMR-D | 87 | 461 | 42841 | 4197 | 89 | 393 | 11887 | 1843 |
| EMR-B | 86 | 65 | 36873 | 1880 | 85 | 180 | 7894 | 680 |
| EMR-D | 90 | 2976 | 385594 | 8468 | 90 | 3601 | 115620 | 4018 |
| EUR-A | 56 | 8 | 1466 | 37 | 56 | 10 | 363 | 8 |
| EUR-B | 86 | 79 | 16217 | 574 | 85 | 37 | 1980 | 87 |
| EUR-C | 87 | 117 | 21562 | 788 | 86 | 103 | 4014 | 144 |
| SEAR-B | 79 | 1009 | 196086 | 4714 | 84 | 1745 | 62122 | 1952 |
| SEAR-D | 93 | 6445 | 1230514 | 7549 | 95 | 8507 | 310247 | 2647 |
| WPR-A | 68 | 0 | 31 | 2 | 71 | 2 | 49 | 3 |
| WPR-B | 85 | 834 | 86591 | 456 | 88 | 999 | 31689 | 176 |
| World | 89 | 28383 | 3587692 | 4710 | 91 | 22413 | 775486 | 1247 |

Table 15.18 Derivation of attributable fractions for unwanted births, proportions of unwanted births and attributable fraction for all births, by subregion and age

| Subregion | Women 15-29 years |  |  | Women 30-44 years |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Attributable fraction (unwanted) | Unwanted births (\%) | Attributable fraction (all) | Attributable fraction (unwanted) | Unwanted births (\%) | Attributable fraction (all) |
| AFR-D | 84\% | 2.4 | 2\% | 87\% | 23.7 | 21\% |
| AFR-E | 82\% | 6.8 | 6\% | 88\% | 34.6 | 31\% |
| AMR-A | 42\% | 2.0 | 1\% | 42\% | 2.0 | 1\% |
| AMR-B | 81\% | 23.5 | 19\% | 85\% | 58.0 | 49\% |
| AMR-D | 84\% | 27.3 | 23\% | 88\% | 61.7 | 54\% |
| EMR-B | 78\% | 9.5 | 7\% | 83\% | 51.8 | 43\% |
| EMR-D | 85\% | 11.9 | 10\% | 89\% | 53.3 | 47\% |
| EUR-A | 40\% | 2.0 | 1\% | 40\% | 2.0 | 1\% |
| EUR-B | 79\% | 11.6 | 9\% | 85\% | 67.9 | 58\% |
| EUR-C | 82\% | -0.8 | 0\% | 86\% | 38.5 | 33\% |
| SEAR-B | 71\% | -0.7 | 0\% | 82\% | 39.9 | 33\% |
| SEAR-D | 84\% | 9.6 | 8\% | 95\% | 67.2 | 64\% |
| WPR-A | 56\% | 2.0 | 1\% | 56\% | 2.0 | 1\% |
| WPR-B | 83\% | 21.7 | 5\% | 87\% | 53.6 | 12\% |

proportion of all births that are unwanted. The whole subregion has an average total fertility rate (TFR) of 2, and we can therefore calculate that $75 \%$ of the subregional births occur in China (with a TFR of 1.8), all of them assumed as wanted. Of course, this assumption cannot be totally correct nor can it be verified, but it is likely to be a close approximation to the truth because of the strict birth control policies that have been applied in China since 1979.

The burden of disease attributable to unwanted births totals 4.6 million DALYs (see Table 15.19). In contrast to abortion, the largest part of the burden befalls women aged $>30$ years ( $75 \%$ ). It is again the same subregions that are most affected, with those in Africa having the highest relative burden.

## 5. Sources of uncertainty

The calculation of aggregate level attributable fractions inevitably involves numerous uncertainties. One of the most serious concerns is the limited availability of data and the need to extrapolate results for a few countries to an entire subregion. Others relate to the quality of the input data, and to the method for simulating the expected numbers of
Table I5.19 Estimated burden of disease attributable to unwanted births (due to non-use of contraception), by subregion and age

| Subregion | Women 15-29 years |  |  |  | Women 30-44 years |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Attributable fractions (\%) | Deaths | DALYs | DALYs per 1000 women | Attributable fractions (\%) | Deaths | DALYs | DALYs per 1000 women |
| AFR-D | 2 | 938 | 53279 | 1318 | 21 | 7284 | 248587 | 11068 |
| AFR-E | 6 | 3527 | 188279 | 3927 | 31 | 15812 | 523605 | 19957 |
| AMR-A | 1 | 2 | 1092 | 34 | 1 | 2 | 481 | 13 |
| AMR-B | 19 | 1 231 | 139381 | 2308 | 49 | 2512 | 152705 | 3338 |
| AMR-D | 23 | 814 | 60070 | 5884 | 54 | 1916 | 88018 | 13647 |
| EMR-B | 7 | 91 | 18566 | 946 | 43 | 835 | 56151 | 4839 |
| EMR-D | 10 | 2719 | 194874 | 4280 | 47 | 13663 | 513596 | 17849 |
| EUR-A | 1 | 1 | 760 | 19 | 1 | 1 | 479 | 10 |
| EUR-B | 9 | 86 | 20887 | 740 | 58 | 214 | 44212 | 1942 |
| EUR-C | 0 | 0 | 0 | 0 | 33 | 146 | 21774 | 782 |
| SEAR-B | 0 | 0 | 0 | 0 | 33 | 3827 | 138968 | 4368 |
| SEAR-D | 8 | 4491 | 369244 | 2265 | 64 | 36894 | 1443505 | 12314 |
| WPR-A | 1 | 0 | 396 | 25 | 1 | 1 | 270 | 17 |
| WPR-B | 5 | 446 | 86893 | 457 | 12 | 997 | 84396 | 469 |
| World | 7 | 14346 | \| 133721 | 1488 | 40 | 84106 | 3316749 | 5335 |

unwanted births and unsafe abortions. Finally, the counterfactual estimates involved further assumptions.

While it was not possible, with one exception, to quantify the magnitude of the effect caused by the nature of the empirical data and assumptions, we can predict the direction in which they operate: whether they lead to an overestimation or an underestimation of the burden of maternal ill-health attributable to lack of effective contraceptive use. We have briefly reviewed the most important uncertainties and their likely effect on our estimates of relative risk and attributable fractions.

### 5.1 Robustness of data

## CURRENT CONTRACEPTIVE USE

Survey data on current contraceptive use have been routinely collected by means of national surveys for 30 years. Their quality is considered high. Trends over time are plausible and the relationship between contraceptive prevalence and fertility rates is strong. Measurement error, where it exists, is likely to take the form of underreporting. In some societies, clandestine use by women occurs (Biddlecom and Fapohunda 1998; Castle et al. 1999) and this is likely to be concealed in conventional surveys. Moreover, some evidence exists to suggest that women underreport male methods, such as condoms, because of shyness and embarrassment (Koenig et al. 1984). Thus some users may be misclassified as non-users, and exposure may be slightly overestimated. The effect on relative risk estimates of such errors will be in the direction of underestimation but is likely to be small compared with other errors.

## Contraceptive failure rates

Contraceptive failure rates were derived from enquiries of 18 DHS where detailed month-by-month information on contraceptive use episodes had been collected. Although this data source is undoubtedly the most appropriate, two types of uncertainty apply: the accuracy of the information and their representativeness. With regard to accuracy, the estimates (Table 15.2) are in general consistent with evidence from more carefully controlled prospective studies, with the exception of the condom where rates were lower than expected (Trussell 1998). However condoms are not a common method of contraception in most developing countries and any error is of minor significance. There are also doubts about representativeness. Calendar data are collected only in countries with high prevalence of use but the average results have been applied to all 58 countries. The validity of the underlying assumption-that failure rates are unrelated to overall levels of use-is unknown, but, again, the error is unlikely to be serious.

## Induced abortion

In countries where abortion is illegal, or highly restricted and heavily stigmatized, it is impossible to obtain reliable information on incidence by means of conventional direct questioning. Hence, in this chapter, we have had to rely on WHO's indirect estimates of unsafe abortion (WHO 1998). While these estimates are widely accepted and cited at global level, a very considerable band of uncertainty surrounds them. However, no means exist of assessing the possible magnitude, or even direction, of error. In a few countries, the number of legal terminations could be taken into account when estimating the projected number of unwanted births. In yet other countries, legal terminations are carried out but no information was available on their number and therefore no allowance could be made. This gap in data leads to an overestimate of relative risks of unwanted births.

## FERTILITY PREFERENCES

The method used to derive attributable and avoidable burden of disease depends heavily on women's statements about their future fertility desires or intentions. While experts agree that this way of measuring preferences is the least problematic of the several alternatives, interpretation is far from straightforward. For instance, attitudes toward future childbearing may be weakly held and ambivalent. Moreover, the attitude of the spouse, or male partner, is not taken into account. These considerations may partly explain why projected estimates of unwanted and mistimed pregnancies based on the pair of questions "Would you like to have a/another child or would you prefer not to have (more) children?" and "How long would you like to wait from now before the birth of a/another child?" are much larger than retrospective estimates of the "wantedness" of recent births and current pregnancies. While there is good reason-and some empirical evidence-to believe that the retrospective estimates are biased downwards by post facto rationalization, the size of the discrepancy between the prospective and retrospective estimates ( $63 \%$ and $143 \%$ for younger and older age groups, respectively) is a matter of concern. The direction of potential bias is clear. To the extent that preferences for future childbearing do not translate into unwanted and mistimed pregnancies, we will have overestimated relative risks as well as exposure.

## SEXUAL EXPOSURE

The method of estimating attributable burden involved the exclusion of virgins from the calculations and the classification of women who report no sexual intercourse in the past 12 months as behaviourally not at risk of pregnancy. In countries with a strong traditional emphasis on pre-marital chastity for women, it is to be expected that single women will underreport sexual activity, which would lead to an underestimate
of expected birth among non-users, thereby underestimating relative risks.

In countries where DHS field staff interviewed ever-married women only, the single women were implicitly categorized as virgins. Countries with no data on single women are typically those where it would be socially unacceptable to ask young unmarried women about sex and reproduction. Although levels of sexual activity will no doubt be low in these countries, the resulting pregnancies are very likely to be unwanted. Most of them will be terminated, often clandestinely and thus most probably "unsafe". Insofar that the WHO estimates on unsafe abortion include procedures to unmarried girls (as estimates are based to a large extent on hospital admissions for abortion complications), these abortions have been attributed to the married women. However, since the exposure for single women is more skewed towards non-users we will have underestimated the attributable burden.

## FECUNDABILITY

A major dilemma arose in the estimation of projected pregnancies among non-users of contraception. We had to choose biological estimates of the monthly probability of conceiving based on woman's age or on reported coital frequency. The reason for preferring age-based estimates was that they gave a closer fit between expected overall births and observed births in the recent past. The deficit in births was much bigger when estimates of fecundability were based on coital frequency, with expected births $39 \%$ and $11 \%$ lower than recent observed births for the 15-29 and 30-44 group, respectively. At low levels of intercourse, the impact of frequency of intercourse on fecundability is substantial with coitus-based conception rates well below the age-based rates, as shown in Table 15.4. The large shortfall in expected births compared with recent age-specific fertility does cast doubt on the reliability of sexual activity data, which appear to be too low to explain current fertility. Brown (2000) in his comparative study in Africa used the same Bongaarts and Potter model estimates of coitus-dependent fecundability with reported coital frequencies in the last month, and came to the same conclusion. However, in defence of the data he has shown good internal consistency between reported monthly frequency and time since last sex. Of course, at low frequencies, sex could be targeted to coincide with ovulation, increasing the probability of a pregnancy, although this would only affect women who desire a pregnancy soon and who are knowledgeable about the timing of ovulation.

The use of fecundability based on reported coital frequency would make a substantial difference to relative risks, reducing them by about a quarter. The relative risk of having an abortion for non-users would decrease from 8.7 to 6.7 , while the relative risk of an unwanted birth would decrease from 7.8 to 5.7 . This is explained by the fact women aged $30-44$ years who say they want to have no more children have less
sex (two times a month) than those who want to space (three times) or those who desire a child within the next two years (3.9 times).

Whereas coital frequency among contraceptive users also varies with fertility intention, failure rates were kept constant regardless of preference. For modern methods, the most important determinant of failure is imperfect use. However, among perfect users, frequency of intercourse is the most important characteristic determining method failure (Trussell 1995). Traditional method failure is likely to be more dependent on coital frequency. This assumption, though far less important than the choice of age-based rather than coitally-based estimates of fecundability for non-users, will act to bias relative risks upwardly.

### 5.2 Assumptions in the basic model

In addition to concerns about the robustness of the empirical data, we had to make several assumptions in the basic model that links exposure to outcomes. The most important of these were:

- The burden is limited to direct obstetric events.
- Obstetric morbidity and mortality are the same for wanted and unwanted births.
- Abortion probabilities are zero for those who want another child in the next two years and are twice as high for limiters than spacers; probabilities are not affected by the proximate cause (failure vs nonuse) and are the same for the two age groups.


## Choice of outcomes

The crucial dilemma in defining the burden was whether or not to include perinatal mortality, much of which stems from unwanted pregnancies. Expert opinion was divided. The final decision to exclude perinatal mortality was based on the judgement that its inclusion would open up a Pandora's box of other intergenerational effects, going well beyond the perinatal period into infancy and childhood. Beyond the mortality of the unwanted children, short interbirth intervals are known to be a major risk factor for infant mortality and can be prevented by contraceptive use to cause better child spacing.

## Obstetric burden is the same for wanted and unwanted births

The evidence base for judging whether the obstetric burden is the same for wanted and unwanted birth was meagre. To the extent that unwanted births are concentrated among older women of low socioeconomic status, it would have been justifiable to assume a higher risk. However, because births at late maternal ages constitute a small fraction of all births and because the link between socioeconomic status and unwanted childbearing varies between subregions, it was decided, by default, to
assume no difference. The effect of errors in this assumption would be to raise relative risks and the attributable burden.

## Abortion probabilities

Several potential biases stem from assumptions that had to be made about the distribution of abortions by age group and exposure status. We made the simplifying assumption that no pregnancies occurring to women who report the desire for a child in the next two years are aborted. Because life circumstances change, this is no more than a close approximation to the truth and a small upward bias on relative risks is possible. A more important possible form of bias operating in the same direction is the assumption that abortion probabilities for unintended pregnancies are the same for non-users and for users who experience contraceptive failure. The available empirical evidence on this matter was insufficient to propose differing probabilities but it is nevertheless likely that modern method users do have a greater propensity to seek terminations than traditional method users and non-users, in which case relative risks would be overestimated. This bias may be offset to the extent that modern method users are more likely than others to seek safe abortions rather than unsafe, illicit abortions.

A bias operating in the opposite direction arises from the assumption that women who experience an unintended pregnancy when they want no more children are twice as likely to seek an abortion than those who wish to postpone the next pregnancy. Such a differential accords with common sense, at least for married women, and is consistent with the available shreds of evidence, but the size of the assumed difference is essentially arbitrary and may be too high. Because modern method users contain a disproportionately large number of limiters, relative risks may be underestimated.

The age pattern of abortion is known to vary between countries as a reflection of large differences in the proportion of young single women who are exposed to the risk of unintended pregnancy and differences in age at marriage. The simplifying assumption that abortion probabilities were constant by age may have led to an overestimation of abortion in the younger age group and an underestimation among older women. This will not affect the relative risks of unsafe abortion or unwanted birth by contraceptive use status but for the younger age group we may have underestimated the proportion of all births that are unwanted and thus the burden of obstetric complications.

### 5.3 Summary of uncertainty

Table 15.20 attempts to summarize in a necessarily crude manner the possible magnitude and direction of data defects and model assumptions on relative risks. A positive symbol (+) indicates that the effect may be to bias risks upwardly and a negative symbol (-) the opposite. A zero

Table I5.20 Possible effects of data limitations and assumptions on relative risks

|  | RRs <br> Abortion | RRs <br> Unwanted births |
| :--- | :---: | :---: |
| Robustness of empirical data |  |  |
| - Contraceptive use | - | - |
| - Failure rates | - | - |
| - Induced abortion | 0 | + |
| - Fertility preferences | +++ | +++ |
| - Sexual exposure | -- | -- |
| - Fecundability | ++ | +++ |
| Assumptions in the model | 0 |  |
| - Definition of burden | -- | 0 |
| - Obstetric burden is same for unwanted and wanted births | -- |  |
| - Abortion probabilities | - |  |
| - No abortions among those who want birth soon | ++ | 0 |
| - Same for failure and non-use | -- | 0 |
| - Twice as high for limiters than spacers | 0 | 0 |
| - No age pattern | 0 |  |

(0) denotes that the direction of the uncertainty, or possible bias, cannot be established. The number of symbols represents our judgement on the magnitude of the possible bias. As may be seen, positive biases are broadly balanced by negative biases.

Quantifying a range around our estimates is not an easy task, and beyond the scope of this exercise. Varying fecundability by fertility desire alone could lead to a $25 \%$ reduction in relative risk levels. Allowing for other biases that work in the same direction we may set $25 \%$ as a minimum range of uncertainty around the estimates at country level. The extrapolation to subregional level may well introduce the biggest cause of uncertainty. Given these inherent limitations in the data and in the complexity of the various assumptions adopted to apply the methodology, the subregional estimates presented are approximate and reflect actual disease burden in general terms.

## 6. Discussion of attributable burden

The calculation of the burden of disease attributable to non-use of modern contraception methods has required a long and complicated series of steps, mainly arising from the fact that exposure has two dimensions: a behavioural one (use or non-use of contraception) and an attitudinal one (the desire to avoid or delay childbearing). Despite the inevitable degree of uncertainty surrounding estimates, some stemming
from inadequacies of empirical evidence and others from necessary assumptions, the key results make good intuitive sense and certainly provide a reliable basis for setting priorities at global and regional levels.

It is estimated that about 57000 women die each year and that 4.9 million healthy life years (measured in DALYs) are lost because of abortions. Globally, about $90 \%$ of this burden is attributable to non-use of modern contraception. Regional differences in the attributable burden are strikingly large. In east and southern Africa (AFR-E), the estimated annual attributable burden exceeds 18500 DALYs per 1000 women aged 15-29 years, and is also high in West Africa, South Asia and some Middle Eastern countries (AFR-D, SEAR-D and EMR-D). By comparison, it is under 40 per 1000 women aged 15-29 years in the industrialized low-fertility subregions AMR-A, EUR-A and WPR-A. Of course, one reason stems from differences in exposure: variations in the propensity of women who want to delay or avoid pregnancy to use modern contraception. But the more important reason concerns differences in access to legal and safe abortion services. Regions with high attributable burden of abortion-related mortality and morbidity are characterized by restrictive abortion laws, and vice versa. From a public health perspective, both issues (low contraception access and use and restrictive abortions) have important policy implications.

The magnitude of abortion-related mortality and morbidity is dwarfed by the obstetric burden stemming from complications of pregnancy and childbirth. It is estimated that about 415000 women die each year from obstetric causes and that about 25 million healthy life years (measured in DALYs) are lost to these conditions. However, only a minority of these pregnancies are unwanted and hence the proportion of this overall disease burden attributable to non-use of modern methods is much lower than for the abortion-related burden: 7\% among younger women rising to $40 \%$ in older women, among whom the desire to avoid all further childbearing is much more common. For both age groups combined, the estimates suggest that 98000 obstetric deaths, representing nearly $20 \%$ of all such deaths, could be prevented each year if all women who desire no more children were to use modern contraceptives. The attributable burden is thus appreciably larger than the attributable abortion-related burden. Huge subregional differences are again apparent. In five subregions (AFR-D, AFR-E, AMR-D, EMR-D and SEAR-D) over 10000 healthy life years are lost per year per 1000 women aged $30-44$ years. The equivalent figure for the industrialized subregions (AMR-A, EUR-A and WPR-A) is below 20. In addition to access to and the use of modern contraception, these stark contrasts stem largely from variations in the coverage and quality of obstetric services.

## 7. Avoidable burden

There is little time lag between a change in contraception and the effect on burden of maternal complications. Current abortions and unwanted births are due primarily to non-use of contraception in the previous year. Thus, determining avoidable risk is very much like calculating attributable risk, but for "exposure in the future". Since fertility in today's medium- and high-fertility subregions is expected to drop in the future, the risk for each woman of death from an obstetric complication is also expected to decrease. However, the total burden of abortion-related complications and maternal outcomes may continue to increase in the next three decades, because the absolute number of women of reproductive age and the total number of births will continue to increase in the highfertility subregions with the highest burden of maternal mortality.

In calculating counterfactual scenarios and attributable burden, the level of obstetric care and the quality of abortion services available were assumed to remain constant at current levels. Only the numbers of unwanted births and abortions determine the potential decrease in burden by uptake of contraception. However a reduction in unintended pregnancies is not the only pathway to lower levels of disease burden. In industrialized countries, there are still high levels of unintended pregnancies and abortions, but the disease burden associated with these is minimal because of the high quality of obstetric and abortion services. Indeed, the avoidable burden in absolute numbers may change more through a decline in the risk attributed to each pregnancy-by improvements in quality and provision of safe obstetric and abortion servicesthan through a decline in unintended pregnancies resulting from the use of effective contraception. It should be emphasized that improvement in risks related to abortions in many low-income countries requires above all a political will to change restrictive laws. Whatever the future may hold in terms of the risk attached to a single pregnancy or birth, it remains relevant and valid to estimate the proportion of the burden avoidable by increased effective use of contraception to avert unwanted births and abortions.

How should future attributable fractions be calculated and what are the necessary assumptions about fertility decline and levels of exposure? Fertility is expected to vary over the next three decades according to the UN medium-variant projections. Specifically all developing countries are now projected to reach replacement level fertility of 2.1 births per woman in the course of this century. Indeed the next UN projection will assume declines to 1.85 births per woman (UN 2002). These projections are rooted in evidence from the past 100 years that suggests that, once fertility has started to decline, it continues to fall until the achievement of low levels. This process of fertility transition appears to be relatively impervious to socioeconomic development. For instance, fertility has declined under conditions of rapidly improving standards of living, as in
many east Asian countries. It has also declined under conditions of economic stagnation or decline, as in Europe in the 1930s and much of east Africa in the past 20 years. Whatever the underlying forces of change, these fertility declines will be achieved primarily through increased levels of contraceptive use (and perhaps abortion), accompanied in some countries by rising age at marriage.

We thus need to project a future contraceptive distribution based on expected declines in fertility in the next three decades. Cross-country comparisons show that a fall in TFR of one child roughly corresponds to an increase in contraceptive prevalence of 15 percentage points (Ross and Frankenberg 1993). However, inferring contraceptive use from future fertility is complicated by the fact that abortion is an alternative means to regulate fertility. In countries experiencing simultaneous fertility decline and rapid changes in desired family sizes, unwanted births, abortion and contraception levels may all rise in parallel. This counterintuitive trend reflects the fact that in societies where couples want large families of, say, five or six children, exposure to the risk of an unwanted pregnancy is bound to be low. As fertility desires fall, the risk increases. For instance, in a society where couples want two children and women marry at 20 years of age, the desired family size will typically be achieved when the wife is in her mid-20s, leaving her exposed to the potential risk of unwanted pregnancy for the next 20 years. Thus it is not surprising that rapid declines in desired fertility can give rise to situations where increased contraceptive practice is unable to meet the growing need for fertility regulation.

One of the clearest examples is the Republic of Korea. As documented from longitudinal data in this country (Bongaarts and Westoff 2000), early on in the fertility transition, both levels of contraceptive use and the incidence of abortion rose in parallel, which in itself provides evidence for a growing unmet need for contraception. While abortion levels reached a peak and declined, contraceptive prevalence continued to rise, as the Republic of Korea progressed through the fertility transition.

The sequence of events in the Republic of Korea is not inevitable. In countries where effective contraception has not been promoted, and is thus relatively inaccessible, heavy reliance on abortion may persist. This is true in Japan and much of eastern Europe and central Asia (Henshaw et al. 1999). But in countries where abortion is very common, evidence suggests that improved availability of family planning services and wider choice of effective contraception can cause a rapid decline in abortion (Henshaw et al. 1999). This is the trend observed in central Asia, where abortion is being replaced by contraceptive use (Westoff et al. 1998). Thus widely varying patterns of change in population levels of contraception and abortion levels are evident in different populations (Marston and Cleland 2003b).

We therefore needed to make assumptions about the level of unmet need at future expected levels of TFR and contraceptive use. What
change in unmet need is to be expected from recent trends? There has been a steady increasing potential need for limiting births (adding the met need and unmet need), in the 1980s and the 1990s in developing countries (Westoff and Bankole 2000). Whereas the average proportion of women using contraception for limiting births increased faster than the potential need in Latin America, Asia and north Africa, this was not true for sub-Saharan Africa, where the proportion of women with an unmet need for limiting increased during the past two decades (Westoff and Bankole 2000).

As explained before, with rapid rates of social change, the need for contraception can grow faster than contraceptive use itself, resulting in rising incidence of unintended pregnancies, unwanted births and unsafe abortions. This implies that at the same levels of future fertility (TFR), populations could have different levels of unmet need, and that women will make trade-offs between abortion and contraception, depending on the legality, availability and perceived quality of services. However, the prediction of changes in unmet need (reductions in Asia and Latin America vs increases in Africa) made the calculation of avoidable risk too complex to operationalize.

The most effective and practical means of obtaining estimates of future exposure levels was to start from the relationship between fertility desires and TFR, and then to infer future contraceptive levels from the projected fertility desires. Fertility changes because of a decline in fertility desires and/or a better implementation of fertility preferences by couples. Future exposure to the risk factor considered here will depend on both. Therefore, the decrease in the maternal burden of disease (linked to a decline in pregnancies) can be split in two components: a reduction due to a lower desired family size and one due to a better implementation of fertility desires. Subregions with a high burden of maternal complications are going through a fertility transition and the desire for smaller families will continue to increase.

We have thus inferred the change in family size desires from the expected decline in TFR. Whereas the fertility preferences within each category of exposure will shift to higher proportions of women with an intention to stop or postpone childbearing, we can keep the propensity to translate desire into effective contraceptive protection constant. We have assumed that within each level of fertility intention (want children now, spacers, limiters), the relative distribution of non-users, traditional and modern method users stays the same, providing us with an estimate of future exposure. The future levels of contraceptive use implicit in the lower family size desires can be taken as the "business-as-usual" exposure. This reflects a rise in use expected from declines in desired family sizes, but is obviously dependent to a large extent on continued or increased investment in subsidized contraceptive services.

Based on this business-as-usual scenario, the avoidable burden then refers to the burden associated with unwanted births and unsafe abor-
tions that could be avoided through perfect implementation of future fertility preferences, over and above the burden that is avoided by the general trend towards lower fertility and the desire to have smaller family sizes. Simulating business-as-usual exposure levels for different time points in the future also requires associated business-as-usual relative risk levels and proportions unwanted among all births.

The step-by-step derivation of all the input data needed for the calculation of avoidable risk is given, followed by a discussion of the trends in input data. Avoidable risk was calculated for 2001, 2005, 2010, 2020 and 2030. For 2001, the 2000 input levels have been used.

### 7.1 BUSINESS-AS-USUAL EXPOSURE AND OTHER INPUTS TO CALCULATE AVOIDABLE BURDEN

## Establishing the association between fertility desire AND TFR

We have related the levels of fertility intention (want more children soon, spacing and limiting) aggregated across the three levels of exposure (modern methods, traditional methods, fecund non-users) to the TFR in the following way. Using the data from the 58 DHS , the cross-sectional linear associations between each of the current aggregate levels of fertility preference and the TFR were assumed to represent rates of change in the fertility desire for a one-unit change of TFR. Among the nonusers, the proportions of women that are not currently exposed to the risk of pregnancy, either behaviourally or biologically, were also correlated to the TFR. With a fall in fertility there will also be a decline in aggregate levels of women who are pregnant and amenorrhoeic at any point in time; these were regressed as one category against the TFR. Amenorrhoeic women include those who abstain sexually after childbirth. Since both the numbers of women who did not have sex in the past 12 months (other than those who are also amenorrhoeic) and menopausal or infecund women are expected to change little with levels of TFR they were combined and then regressed together against the TFR. The regression coefficients were calculated separately for the 15-29- and 30-44year age groups. Table 15.21 shows the results in terms of the slopes (rate of change along the regression line) and also the R -squared values.

## Projected fertility dechine

The expected decline in fertility was calculated from current TFRs and projected TFR levels using the UN projections (medium-variant) for each time point in future. Table 15.22 shows the subregional average TFRcurrent and projected-and the resulting projected decline at each time point. For 2001, the current levels were kept at calculated 2000 levels.

The sudden drop in fertility from current levels of 4.2 to 2.1 in 2005 in the EMR-B subregion is spurious and explained by the fact that

Table 15.2I Association of fertility intentions by TFR
\% change for I unit change in TFR: regression coefficients (R-squared values)

| Age group <br> (years) | Want birth <br> soon | Want to space | Want to limit | Pregnant and <br> amenorrhoeic | Infecund + no sex <br> in past year |
| :--- | :---: | :---: | :---: | :---: | :---: |
| $15-29$ | $1.42(0.14)$ | $-2.87(0.24)$ | $-6.56(0.60)$ | $7.16(0.69)$ | $0.85(0.12)$ |
| $30-44$ | $2.38(0.30)$ | $1.34(0.16)$ | $-13.4(0.73)$ | $8.39(0.83)$ | $1.28(0.09)$ |

Table I5.22 Current and projected future levels of TFR, by subregion

| Subregion | Current <br> TFR | Projected future TFR UN median variant |  |  |  | Projected change in TFR |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 2005 | 2010 | 2020 | 2030 | 2005 | 2010 | 2020 | 2030 |
| AFR-D | 5.9 | 5.4 | 4.9 | 4.0 | 3.1 | -0.5 | -0.9 | -1.9 | -2.8 |
| AFR-E | 5.7 | 5.5 | 5.1 | 4.2 | 3.4 | -0.2 | -0.6 | -1.5 | -2.3 |
| AMR-A | - | - | - | - | - | - | - | - | - |
| AMR-B | 3.1 | 2.3 | 2.2 | 2.1 | 2.1 | -0.8 | -0.9 | -0.9 | -1.0 |
| AMR-D | 4.1 | 3.1 | 2.8 | 2.4 | 2.2 | -0.9 | -1.2 | -1.7 | -1.9 |
| EMR-B | 4.2 | 2.1 | 2.1 | 2.1 | 2.1 | -2.1 | -2.1 | -2.1 | -2.1 |
| EMR-D | 4.6 | 4.2 | 3.8 | 3.1 | 2.5 | -0.4 | -0.8 | -1.5 | -2.1 |
| EUR-A | - | - | - | - | - | - | - | - | - |
| EUR-B | 2.8 | 2.2 | 2.1 | 2.1 | 2.1 | -0.6 | -0.7 | -0.7 | -0.7 |
| EUR-C | 2.5 | 1.9 | 1.9 | 1.9 | 1.9 | -0.6 | -0.6 | -0.6 | -0.6 |
| SEAR-B | 2.7 | 2.1 | 2.0 | 2.0 | 2.0 | -0.5 | -0.6 | -0.6 | -0.6 |
| SEAR-D | 3.4 | 2.9 | 2.5 | 2.2 | 2.1 | -0.5 | -0.9 | -1.2 | -1.3 |
| WPR-A | - | - | - | - | - | - | - | - | - |
| WPR-B | 3.7 | 3.0 | 2.6 | 2.1 | 2.1 | -0.7 | -1.1 | -1.6 | -1.6 |

Tunisia is the only country represented. Fertility has changed dramatically since the last Demographic and Health Survey in 1988.

## Projected levels of fertility preferences

The calculation involved different steps, explained here in detail for women who want to limit family size, using Ghana for illustrative purposes.

1. The current overall percentage of women who want to limit childbearing: This was calculated from the distribution of fertility preferences within each level of exposure. For example, in Ghana, 11.4\% of women aged 15-29 years are modern method users, $10.1 \%$ use tra-
ditional methods and $78.5 \%$ are not using any contraception at all. Among the modern method users, $11 \%$ want to limit their family size. So $1.2 \% ~(=11.4 \% \times 11 \%)$ of all women aged $15-29$ years are modern method users with a desire to limit family size. Since $7 \%$ of traditional method users want no more children, $0.7 \%$ of all women are traditional method users with desire to limit family size. Among the nonusers $47 \%$ are currently fecund and, among these, $6 \%$ want to stop childbearing. This gives $2.2 \%(=78.5 \% \times 47 \% \times 6 \%)$ fecund nonusers with a desire to limit family size. Adding all limiters together gives an aggregate percentage of $4.1 \% ~(=1.2 \%+0.7 \%+2.2 \%)$ of all 15-29-year olds who want no more children.
2. Projected decline in fertility: Fertility in Ghana is projected by the UN to fall by 0.4 children from a current TFR of 4.4 to 4.0 in 2005.
3. Future overall percentage of women who want to limit childbearing in 2005: The projected fertility decline (0.4) was multiplied by the coefficient representing the change in the percentage of limiters (see Table 15.21: -6.56 ) and added to the current percentage $(4.2 \%)$. So in $2005,4.2+6.56 \times 0.4=6.6 \%$ of $15-29$-year olds are projected to want to limit their family size.
4. Steps 1 to 3 were repeated for the four other variables, using the appropriate coefficients in Table 15.21. By 2005, the percentage of women who want more children is projected to decline from $12.9 \%$ to $12.4 \%$ and women who want to space to increase from $41 \%$ to $42.1 \%$. The percentage pregnant and amenorrhoeic would decline from $29 \%$ to $26.3 \%$, while the percentage of infecund and not sexually active women would change from $13 \%$ to $12.7 \%$.

## Projected business-as-usual levels of contraceptive use

Having calculated the new overall distribution of fertility preferences in the population, we used the relative distribution of modern method users, traditional method users and fecund non-users within each level of fertility desire (current scenario), to estimate our business-as-usual exposure variable. For example, $29 \%$ of the limiters were modern method users, $17 \%$ used traditional methods and $54 \%$ were fecund nonusers. Keeping the propensity to translate fertility intention into contraceptive practice constant, we have $29 \%$ of the $6.6 \%$ limiters aged 15-29 years, i.e. $1.9 \%$ using modern methods in 2005 . Adding this to the $12 \%$ of the $12.4 \%$ women who want a child soon and the $21 \%$ of the aggregate $42.1 \%$ spacers who are using a modern method, we have derived an aggregate percentage of $12.3 \%$ modern method users in 2005 $(=1.9 \%+1.4 \%+8.9 \%)$. In total, $10.7 \%$ are traditional method users and $77.1 \%$ are non-users (this last category includes the $26.3 \%$ pregnant and amenorrhoeic and $12.7 \%$ infecund and not sexually active).

## OTHER INPUT FOR BUSINESS-AS-USUAL SCENARIO

The relative distribution of fertility preference within each exposure level was calculated (e.g. 1.9 of the $12.3 \%$ or $16 \%$ of modern method users want no more children). The relative composition of the non-users in five categories of fecund, pregnant, amenorrhoeic, infecund/menopausal and no sex in the past year was also recalculated, and together with the projected contraceptive use levels used as input into the simulations. From this we derived the business-as-usual levels of the relative risk of having an abortion, the relative risk of having an unwanted birth as well as the proportions of unwanted among all births.

These calculations were done for each of the 58 countries at different time points (2005, 2010, 2020, 2030). For 2001, we used the current (2000) levels of contraceptive use and fertility desire. For each time point in the future we estimated five contraceptive prevalence distributions: business-as-usual exposure, theoretical minimum, and three other counterfactuals, shifting $10 \%, 20 \%$ and $30 \%$ of fecund non-users and traditional method users into the modern use category. For each of these we used the corresponding relative risks for abortions, unwanted births and proportion of unwanted births among all births.

### 7.2 Trends in business-AS-USUAL EXPOSURE AND relative risk levels

We have presented the trend in modern method use as derived through aggregating the country data into subregional estimates from 2001 (which was kept as the same as the 2000 level) to 2030. Table 15.23 shows these trends separately for the two age groups. For the three subregions with missing data, we kept the contraceptive distribution constant at the estimated 2000 level.

As expected from the fertility decline (Table 15.22), the contraceptive levels, reflecting the associated decline in fertility desire, are increasing steadily. As suggested by the regression coefficients (Table 15.21), the increase is more marked for the older age group. Since the expected fall in fertility by 2030 is largest for the two African subregions, the business-as-usual contraceptive levels are predicted to increase most steeply here.

In the business-as-usual scenario, the relative risk levels also varied from the current one, as shown in Tables 15.24 and 15.25 for unwanted births and abortions, respectively. The increase is more marked for nonusers than it is for traditional method users. The increase is explained by the fact that the proportion of limiters increases among all three levels of exposure, but relatively more among the non-users.

Logically, as fertility desires are projected to go down with time, and keeping the propensity to translate desire into contraceptive behaviour at 2000 levels, the proportion of unwanted births among all births increases steadily, as can be seen from Table 15.26.

Table I5.23 Projected trends in the proportion of women using a modern method, business-as-usual scenario

| Subregion | Women 15-29 years |  |  |  |  | Women 30-44 years |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 2001 | 2005 | 2010 | 2020 | 2030 | 2001 | 2005 | 2010 | 2020 | 2030 |
| AFR-D | 8.0 | 9.2 | 10.2 | 12.4 | 14.7 | 10.7 | 13.1 | 14.9 | 18.9 | 22.9 |
| AFR-E | 13.0 | 13.9 | 15.4 | 18.5 | 21.1 | 15.6 | 16.9 | 18.9 | 23.1 | 26.7 |
| AMR-A | 60.6 | 60.6 | 60.6 | 60.6 | 60.6 | 74.1 | 74.1 | 74.1 | 74.1 | 74.1 |
| AMR-B | 48.1 | 52.1 | 52.5 | 53.0 | 53.1 | 62.1 | 68.6 | 69.3 | 70.0 | 70.3 |
| AMR-D | 27.9 | 31.9 | 33.1 | 34.9 | 35.7 | 36.2 | 42.4 | 44.4 | 47.3 | 48.7 |
| EMR-B | 30.4 | 42.7 | 42.7 | 42.7 | 42.7 | 44.8 | 61.5 | 61.5 | 61.5 | 61.5 |
| EMR-D | 14.9 | 17.3 | 18.7 | 20.4 | 21.5 | 23.5 | 26.6 | 29.0 | 32.1 | 34.7 |
| EUR-A | 62.1 | 62.1 | 62.1 | 62.1 | 62.1 | 75.9 | 75.9 | 75.9 | 75.9 | 75.9 |
| EUR-B | 36.0 | 39.4 | 40.0 | 40.0 | 40.0 | 46.9 | 51.2 | 51.9 | 51.9 | 51.9 |
| EUR-C | 42.4 | 45.5 | 45.6 | 45.6 | 45.6 | 55.5 | 59.9 | 60.1 | 60.1 | 60.1 |
| SEAR-B | 54.4 | 58.5 | 59.1 | 59.1 | 59.1 | 55.9 | 60.9 | 61.7 | 61.8 | 61.7 |
| SEAR-D | 23.5 | 26.1 | 28.0 | 30.0 | 30.3 | 48.3 | 53.5 | 56.9 | 60.6 | 61.1 |
| WPR-A | 48.7 | 48.7 | 48.7 | 48.7 | 48.7 | 59.5 | 59.5 | 59.5 | 59.5 | 59.5 |
| WPR-B | 24.1 | 26.8 | 28.5 | 30.3 | 30.3 | 29.5 | 33.6 | 36.3 | 39.1 | 39.1 |

### 7.3 BUSINESS-AS-USUAL AND COUNTERFACTUAL SCENARIOS

The main assumptions underlying the projected exposure and risk are as follows:

- Fertility in developing countries will fall in line with the UN medianvariant projections.
- The cross-sectional relationship between fertility desires and fertility itself can be used to project future changes in fertility desires.
- In the business-as-usual scenario, the propensity to translate fertility desires into contraceptive use will remain unchanged.
- The relative popularity of different contraceptive methods and failure rates will remain unchanged.
- The proportion of infecund and sexually inactive non-users will remain constant.


## Fertility trends

The UN medium-variant fertility projections are widely accepted as a reasonably dependable guide to the future. The UN's past record of successful projection over the short term of 20-30 years is impressive and no specific reason exists to doubt recent projections (Bongaarts and

Table I5.24 Projected trends in relative risk of having an unwanted birth, business-as-usual scenario

| Subregion | Traditional method users 15-29 years |  |  |  |  | Non-users 15-29 years |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 2001 | 2005 | 2010 | 2020 | 2030 | 2001 | 2005 | 2010 | 2020 | 2030 |
| AFR-D | 2.3 | 2.3 | 2.3 | 2.4 | 2.4 | 2.8 | 3.3 | 3.8 | 4.8 | 6.0 |
| AFR-E | 1.6 | 1.6 | 1.7 | 1.8 | 1.9 | 3.0 | 3.2 | 3.7 | 4.9 | 6.0 |
| AMR-A | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| AMR-B | 3.2 | 3.3 | 3.3 | 3.3 | 3.3 | 8.3 | 10.4 | 10.6 | 10.9 | 11.0 |
| AMR-D | 4.1 | 4.1 | 4.2 | 4.2 | 4.2 | 6.2 | 7.9 | 8.5 | 9.4 | 9.9 |
| EMR-B | 3.2 | 3.5 | 3.5 | 3.5 | 3.5 | 3.9 | 7.2 | 7.2 | 7.2 | 7.2 |
| EMR-D | 3.2 | 3.2 | 3.3 | 3.4 | 3.4 | 3.4 | 4.0 | 4.6 | 5.4 | 6.1 |
| EUR-A | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| EUR-B | 4.0 | 4.1 | 4.1 | 4.1 | 4.1 | 5.1 | 6.3 | 6.5 | 6.5 | 6.5 |
| EUR-C | 5.1 | 5.1 | 5.1 | 5.1 | 5.1 | 6.9 | 8.1 | 8.1 | 8.1 | 8.1 |
| SEAR-B | 4.1 | 4.2 | 4.2 | 4.2 | 4.2 | 4.3 | 5.1 | 5.2 | 5.3 | 5.2 |
| SEAR-D | 4.6 | 4.9 | 5.0 | 5.1 | 5.1 | 4.6 | 5.4 | 5.9 | 6.6 | 6.7 |
| WPR-A | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| WPR-B | 3.7 | 3.7 | 3.7 | 3.7 | 3.7 | 5.8 | 6.9 | 7.8 | 8.8 | 8.8 |


|  | Traditional method users 30-44 years |  |  |  |  | Non-users 30-44 years |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 2001 | 2005 | 2010 | 2020 | 2030 | 2001 | 2005 | 2010 | 2020 | 2030 |
| AFR-D | 3.2 | 3.3 | 3.4 | 3.6 | 3.7 | 3.5 | 4.3 | 5.2 | 7.2 | 9.6 |
| AFR-E | 4.4 | 4.4 | 4.6 | 4.8 | 4.9 | 4.9 | 5.4 | 6.4 | 9.1 | 11.7 |
| AMR-A | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| AMR-B | 7.9 | 8.4 | 8.4 | 8.4 | 8.5 | 14.3 | 21.3 | 22.3 | 23.4 | 23.8 |
| AMR-D | 6.4 | 6.8 | 6.9 | 7.2 | 7.3 | 9.8 | 14.1 | 15.8 | 18.9 | 20.9 |
| EMR-B | 5.9 | 6.3 | 6.3 | 6.3 | 6.3 | 8.4 | 20.8 | 20.8 | 20.8 | 20.8 |
| EMR-D | 5.1 | 5.1 | 5.2 | 5.3 | 5.4 | 6.8 | 8.2 | 9.7 | 11.9 | 14.0 |
| EUR-A | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| EUR-B | 6.3 | 6.3 | 6.3 | 6.3 | 6.3 | 11.1 | 15.3 | 16.1 | 16.1 | 16.1 |
| EUR-C | 8.6 | 8.6 | 8.6 | 8.6 | 8.6 | 12.7 | 16.5 | 16.7 | 16.7 | 16.7 |
| SEAR-B | 6.1 | 6.3 | 6.3 | 6.3 | 6.3 | 8.0 | 10.0 | 10.3 | 10.4 | 10.4 |
| SEAR-D | 22.2 | 23.0 | 23.4 | 23.8 | 23.9 | 27.8 | 35.8 | 41.9 | 49.5 | 50.0 |
| WPR-A | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| WPR-B | 5.9 | 6.1 | 6.1 | 6.2 | 6.2 | 7.7 | 10.0 | 11.9 | 14.4 | 14.4 |

Table 15.25 Projected trends in relative risk of having an unsafe abortion, business-as-usual scenario

| Subregion | Traditional method users 15-29 years |  |  |  |  | Non-users 15-29 years |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 2001 | 2005 | 2010 | 2020 | 2030 | 2001 | 2005 | 2010 | 2020 | 2030 |
| AFR-D | 2.6 | 2.6 | 2.7 | 2.6 | 2.6 | 3.9 | 4.3 | 4.8 | 5.8 | 7.1 |
| AFR-E | 3.0 | 3.2 | 3.0 | 2.9 | 2.9 | 4.7 | 5.0 | 5.4 | 6.7 | 8.0 |
| AMR-A | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| AMR-B | 4.1 | 4.2 | 4.2 | 4.2 | 4.2 | 10.6 | 13.1 | 13.4 | 13.7 | 13.8 |
| AMR-D | 4.3 | 4.4 | 4.4 | 4.5 | 4.5 | 7.6 | 9.5 | 10.2 | 11.2 | 11.7 |
| EMR-B | 4.6 | 4.7 | 4.7 | 4.7 | 4.7 | 6.4 | 10.7 | 10.7 | 10.7 | 10.7 |
| EMR-D | 3.9 | 3.9 | 3.9 | 4.0 | 4.0 | 5.0 | 5.7 | 6.3 | 7.3 | 8.1 |
| EUR-A | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| EUR-B | 4.9 | 4.9 | 4.9 | 4.9 | 4.9 | 7.5 | 9.1 | 9.3 | 9.3 | 9.3 |
| EUR-C | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 | 10.5 | 12.1 | 12.1 | 12.1 | 12.1 |
| SEAR-B | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 6.4 | 7.5 | 7.7 | 7.7 | 7.7 |
| SEAR-D | 6.6 | 6.8 | 6.9 | 7.0 | 7.0 | 10.9 | 12.5 | 13.5 | 14.7 | 14.8 |
| WPR-A | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| WPR-B | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 | 6.6 | 7.8 | 8.8 | 9.9 | 9.9 |
|  | Traditional method users 30-44 years |  |  |  |  | Non-users 30-44 years |  |  |  |  |
|  | 2001 | 2005 | 2010 | 2020 | 2030 | 2001 | 2005 | 2010 | 2020 | 2030 |
| AFR-D | 3.5 | 3.6 | 3.7 | 3.8 | 3.8 | 4.2 | 5.0 | 5.7 | 7.7 | 10.2 |
| AFR-E | 4.7 | 4.7 | 4.8 | 4.9 | 5.0 | 5.4 | 5.9 | 7.0 | 9.6 | 12.1 |
| AMR-A | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| AMR-B | 8.4 | 8.8 | 8.8 | 8.8 | 8.9 | 16.2 | 23.3 | 24.3 | 25.4 | 25.9 |
| AMR-D | 6.8 | 7.1 | 7.2 | 7.4 | 7.5 | 10.6 | 14.9 | 16.6 | 19.6 | 21.6 |
| EMR-B | 6.0 | 6.3 | 6.3 | 6.3 | 6.3 | 10.0 | 22.9 | 22.9 | 22.9 | 22.9 |
| EMR-D | 5.2 | 5.2 | 5.3 | 5.3 | 5.4 | 7.7 | 9.1 | 10.5 | 12.7 | 14.7 |
| EUR-A | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| EUR-B | 6.3 | 6.2 | 6.2 | 6.2 | 6.2 | 11.5 | 15.7 | 16.5 | 16.5 | 16.5 |
| EUR-C | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 13.3 | 17.1 | 17.3 | 17.3 | 17.3 |
| SEAR-B | 6.4 | 6.6 | 6.6 | 6.6 | 6.6 | 9.2 | 11.2 | 11.6 | 11.7 | 11.6 |
| SEAR-D | 22.7 | 23.4 | 23.7 | 24.0 | 24.1 | 30.9 | 38.5 | 44.2 | 51.4 | 51.9 |
| WPR-A | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| WPR-B | 6.2 | 6.3 | 6.4 | 6.4 | 6.4 | 8.5 | 10.9 | 12.9 | 15.3 | 15.3 |

Table 15.26 Projected trends in the proportion of births that are unwanted among all births, business-as-usual scenario

| Subregion | Women 15-29 years |  |  |  |  | Women 30-44 years |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 2001 | 2005 | 2010 | 2020 | 2030 | 2001 | 2005 | 2010 | 2020 | 2030 |
| AFR-D | 2.4 | 5.8 | 8.7 | 15.0 | 20.6 | 23.7 | 30.8 | 37.6 | 51.1 | 62.3 |
| AFR-E | 6.8 | 7.2 | 9.8 | 15.9 | 20.9 | 34.6 | 35.2 | 42.3 | 56.7 | 67.5 |
| AMR-A | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 |
| AMR-B | 23.5 | 26.5 | 26.8 | 27.1 | 27.2 | 58.0 | 67.5 | 68.5 | 69.4 | 69.8 |
| AMR-D | 27.3 | 32.0 | 33.4 | 35.2 | 36.1 | 61.7 | 73.5 | 77.0 | 81.9 | 84.2 |
| EMR-B | 11.7 | 20.2 | 20.2 | 20.2 | 20.2 | 53.9 | 78.0 | 78.0 | 78.0 | 78.0 |
| EMR-D | 11.9 | 13.6 | 16.3 | 20.4 | 23.7 | 53.3 | 58.3 | 64.9 | 74.4 | 81.5 |
| EUR-A | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 |
| EUR-B | 18.6 | 21.0 | 21.3 | 21.3 | 21.3 | 72.3 | 78.9 | 79.9 | 79.9 | 79.9 |
| EUR-C | 6.0 | 7.9 | 8.0 | 8.0 | 8.0 | 47.6 | 54.9 | 55.3 | 55.3 | 55.3 |
| SEAR-B | 0.4 | 0.6 | 0.7 | 0.7 | 0.7 | 39.9 | 48.6 | 50.0 | 50.1 | 49.9 |
| SEAR-D | 9.6 | 11.1 | 12.1 | 13.3 | 13.5 | 67.2 | 78.5 | 86.1 | 94.6 | 95.3 |
| WPR-A | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 |
| WPR-B | 21.7 | 25.7 | 28.1 | 30.5 | 30.5 | 53.6 | 63.7 | 70.1 | 76.3 | 76.3 |

Butatao 2000). In our view, fertility projections are therefore not a major source of uncertainty, nor is it possible to conjecture about possible departures from projections and their effect on our estimates.

## LINK BETWEEN FERTILITY AND FERTILITY DESIRES

The relationship between achieved and desired fertility is not straightforward. Nevertheless, over the longer term, they tend to move in parallel. To represent this link we calculated cross-sectional correlations in the 58 study countries between fertility rates and fertility intentions and then assumed that these relationships would remain constant over the next 30 years. While we accept that a degree of uncertainty surrounds this procedure, we doubt whether it represents a serious bias.

## PROPENSITY TO TRANSLATE FERTILITY DESIRES INTO CONTRACEPTIVE PRACTICE

In the business-as-usual scenario it was assumed that the proportion of limiters and spacers who use contraception to achieve their intentions remains constant. This assumption was made in adherence to the busi-ness-as-usual concept that exposure changes with changing fertility desires, while the propensity to translate these into effective contraceptive use remains the same. Yet the empirical record of the past 40 years suggests that it is an artificial and unrealistic assumption. Fertility has declined in the past both because desired fertility has fallen and because
desires have been better implemented (Feyisetan and Casterline 2000). A more realistic representation of business-as-usual, of course, would act to reduce the avoidable burden.

## Method mix and failure rates

Estimation of future scenarios used the assumption that choice of methods-or the method mix-within each country would remain the same, as with failure rates. Several forms of uncertainty may be identified: the development and uptake of newly developed forms of contraception; the possibility of a drift towards more effective existing methods; and greater resort to barrier methods in response to the HIV pandemic. The development and widespread use of radically new methods of contraception seems increasingly unlikely in view of lack of major investment by the pharmaceutical industry (Hagenfeldt 1994). The development of a contraceptive vaccine, for instance, seems an increasingly remote possibility. More plausibly, general shifts from less to more effective methods might occur-and indeed are underway in the countries of the former Soviet Union. With the exception of these latter countries, where in the past, access to modern contraception was severely restricted, little evidence exists to support the view that the contraceptive method mix will change over the next 30 years. Contraception in developing countries, unlike Europe or North America, has always been dominated by effective methods: sterilization, intrauterine devices and hormonal methods. In our view, an increased dominance of these methods is possible but not particularly likely. Any such trend would increase relative risks. Offsetting this might be increased uptake of condoms, the only existing contraceptive method that offers protection against HIV and other STIs. So far, no tendency towards greater use of condoms for family planning (and dual protection) within marriage has been recorded (UN 1999). With regard to contraceptive failure for users of specific methods, the assumption of no change is reasonably robust. We are unaware of any evidence of secular trends in the probability of failure.

## InFECUNDITY AND SEXUAL INACTIVITY AMONG NON-USERS

A final assumption in the estimation of future scenarios was that the projection of infecund and sexually inactive non-users would remain constant. As the general health of adults improves, physiological infecundity is more likely to decline than increase, and prolonged sexual abstinence may become less common for the same reason and because of the erosion of customs of postpartum abstinence in sub-Saharan Africa. Any such trend would probably serve to increase relative risks and increase exposure and the avoidable burden.

Table 15.27 summarizes effects of assumptions on estimates of avoidable risk. A positive symbol (+) indicates that the effect may be to bias risks upwardly and a negative symbol (-) the opposite. A zero (0)

# Table I5.27 Possible effects of assumptions on estimates of avoidable risk 

| Future fertility trends | 0 |
| :--- | :---: |
| Future fertility desires | 0 |
| Translation of desires into contraceptive use | ++ |
| Method mix and failure rates | 0 |
| Infecundity and sexual inactivity | - |

denotes that the direction of the uncertainty, or possible bias, cannot be established.

## Acknowledgements

We are obliged to Iqbal Shah and Elisabeth Aahman for access to the unpublished WHO estimates of unsafe abortion at country level. We are grateful to Mohammed Ali for giving access to the unpublished contraceptive failure rates. We have greatly benefited from the helpful comments by the five anonymous reviewers and the regular encouraging feedback from Majid Ezzati.

## Note

1 See preface for an explanation of this term.

## References

AbouZahr C, Vaughan JP (2000) Assessing the burden of sexual and reproductive ill-health: questions regarding the use of disability-adjusted life years. Bulletin of the World Health Organization, 78:655-666.
Adetunji JA (1998) Unintended childbearing in developing countries: levels, trends, and determinants. Macro International, Calverton, MD.
Bankole A, Singh S, Haas T (1999) Characteristics of women who obtain induced abortion: a worldwide review. International Family Planning Perspectives, 25:68-77.

Barrett G, Wellings K (2002) What is a "planned" pregnancy? Empirical data from a British study. Social Science and Medicine, 55:545-557.

Becker S, Begum S (1994) Reliability study of reporting of days since last sexual intercourse in Matlab, Bangladesh. Journal of Biosocial Science, 26:291299.

Berkley S (1998) Unsafe sex as a risk factor. In: Health dimensions of sex and reproduction: the global burden of sexually transmitted diseases, HIV, maternal conditions, perinatal disorders, and congenital anomalies. Global Burden of Disease and Injury, Vol. 3. Murray CJL, Lopez AD, eds. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Biddlecom AE, Fapohunda BM (1998) Covert contraceptive use: prevalence, motivations and consequences. Studies in Family Planning, 29:360-372.

Bongaarts J (1990) The measurement of wanted fertility. Population Development Review, 6:487-506.

Bongaarts J, Butatao RA (2000) Beyond six billion: forecasting the world's population. National Academy Press, Washington DC.

Bongaarts J, Potter RG (1983) Fertility, biology, and behaviour. an analysis of the proximate determinants. Academic Press, New York.
Bongaarts J, Westoff CF (2000) The potential role of contraception in reducing abortion. Studies in Family Planning, 31:193-202.
Brown MS (2000) Coitus, the proximate determinant of conception: intercountry variance in sub-Saharan Africa. Journal of Biosocial Science, 32:145-159.
Castle S, Konaté MK, Ulin PR, Martin S (1999) Clandestine contraceptive use in urban Mali. Studies in Family Planning, 30:231-248.
Conde Agudelo A, Belizan JM (2000) Maternal morbidity and mortality associated with interpregnancy interval: cross sectional study. British Medical Journal, 321:1255-1259.

Curtis SL, Blanc A (1997) Determinants of contraceptive failure, switching and discontinuation: an analysis of DHS. Macro International, Calverton, MD.

Dixon Mueller R, Germain A (1992) Stalking the elusive "unmet need" for family planning. Studies in Family Planning, 23:330-335.
Fortney JA (1987) The importance of family planning in reducing maternal mortality. Studies in Family Planning, 18:109-114.
Freedman R, Freedman DS, Thornton AD (1980) Changes in fertility expectations and preferences between 1962 and 1977: their relation to final parity. Demography, 17:365-378.
Ganatra B, Hirve S, Walawalkar S, Garda L, Rao V (2000) Induced abortions in rural western Maharashtra: prevalence and patterns. (Paper presented at Ford Foundation conference: Reproductive health in India: new evidence and issues, Feb 28-March 1, 2000, Pune, India.)

Hagenfeldt K (1994) Current status of contraceptive research and development. In: Population-the complex reality. Graham-Smith F, ed. Royal Society, London.

Henshaw SK (1998) Unintended pregnancy in the United States. Family Planning Perspectives, 30:24-29, 46.
Henshaw SK, Singh S, Haas T (1999) Recent trends in abortion rates worldwide. International Family Planning Perspectives, 25:44-48.
Henshaw SK, Singh S, Haas T (1999) The incidence of abortion worldwide. International Family Planning Perspectives, 25:30-38.
James WH (1979) The causes of the decline in fecundability with age. Social Biology, 26:330-334.
Johnson AM, Wadsworth J, Wellings K, Field J (1994) Sexual attitudes and lifestyles. Blackwell Scientific Publications, Oxford.

Koenig MA, Simmons GB, Misra BD (1984) Husband-wife inconsistencies in contraceptive use responses. Population Studies, 38:281-298.

Leridon H (1977) Human fertility: the basic components. University of Chicago Press, Chicago, IL.

Leridon H (1993) La fréquence des rapports sexuels. Données et analyses de cohérence. Population, 5:1381-1408.
Marston C, Cleland J (2003a) Do unintended pregnancies carried to term lead to adverse outcomes for mother and child? An assessment in five developing countries. Population Studies, 57:77-94.
Marston C, Cleland J (2003b) Relationships between contraception and abortion: a review of the evidence. International Family Planning Perspectives 29:6-13.

Mensch BS, Arends Kuenning M, Jain A, Garate MR (1995) Meeting reproductive goals: the impact of the quality of family planning services on unintended pregnancy in Peru. (Research Division Working Paper No. 81). The Population Council, New York.

Montgomery MR, Lloyd CB, Hewett PC, Heuveline P (1997) The consequences of imperfect fertility control for children's survival, health, and schooling. Macro International, Calverton, MD.
Potter RG, Millman SR (1986) Fecundability and the frequency of marital intercourse: new models incorporating the ageing of gametes. Population Studies, 40:159-170.
Rahman A, Katzive L, Henshaw SK (1998) A global review of laws on induced abortion, 1985-1997. International Family Planning Perspectives, 24:5664.

Robey B, Ross J, Bhushan I (1996) Meeting unmet need: new strategies. Population Reports, Series J, Family Planning Programs, 43:1-35.
Ronsmans C, Campbell O (1998) Short birth intervals don't kill women: evidence from Matlab, Bangladesh. Studies in Family Planning, 29:282-290.
Ross JA, Frankenberg E (1993) Findings from two decades of family planning research. Population Council, New York.

The Alan Guttmacher Institute (1999) Sharing responsibility: women, society and abortion worldwide. The Alan Guttmacher Institute, New York.

Trussell J (1995) Contraceptive efficacy. Archives of Dermatology, 131: 1064-1068.

Trussell J (1998) Contraceptive efficacy. In: Contraceptive technology, 17th revised edn. Hatcher R, et al., eds. Ardent Media, New York.
Trussell J, Pebley AR (1984) The potential impact of changes in fertility on infant, child, and maternal mortality. Studies in Family Planning, 15(6/1):267-280.

UN (1999) Levels and trends of contraceptive use as assessed in 1998. United Nations Department of Economic and Social Affairs, Population Division, New York.

UN (2002) United Nations Expert Group meeting on completing the fertility transition. United Nations Department of Economic and Social Affairs, Population Division, New York.

Weinstein M, Wood J, Chang M (1993) Age patterns of fecundability. In: Biomedical and demographic determinants of reproduction. Gray R, Leridon H, Spira A, eds. Clarendon Press, Oxford.
Weinstein M, Wood JW, Stoto MA, Greenfield DD (1990) Components of agespecific fecundability. Population Studies, 44:447-467.
Westoff CF (1981) Unwanted fertility in six developing countries. International Family Planning Perspectives, 7:43-51.
Westoff CF (1990) Reproductive preferences and fertility rates. International Family Planning Perspectives, 16:84-89.
Westoff CF (2000) The substitution of contraception for abortion in Kazakhstan in the 1990s. Macro International, Calverton, MD.

Westoff CF, Bankole A (1998) The time dynamics of unmet need: an example from Morocco. International Family Planning Perspectives, 24:12-14.

Westoff CF, Bankole A (2000) Trends in the demand for family limitation in developing countries. International Family Planning Perspectives, 26:56-62.

Westoff CF, Sharmanov AT, Sullivan M, Croft T (1998) Replacement of abortion by contraception in three Central Asian Republics. The Policy Project and Macro International, Calverton, MD.
WHO (1998) Unsafe abortion. Global and regional estimates of incidence of and mortality due to unsafe abortion with a listing of available country data. (WHO/RHT/MSM/97.16) World Health Organization, Division of Reproductive Health, Geneva.

Winikoff B, Sullivan M (1987) Assessing the role of family planning in reducing maternal mortality. Studies in Family Planning, 18:128-143.

Wood JW, Weinstein M (1988) A model of age-specific fecundability. Population Studies, 42:85-113.

## Chapter 16

# UNSAFE WATER, SANITATION AND HYGIENE 

Annette Prüss-Üstün, David Kay, Lorna Fewtrell and Jamie Bartram

## Summary

The disease burden from unsafe water, sanitation and hygiene (WSH) is estimated at the global level taking into account various disease outcomes, principally diarrhoeal diseases. The risk factor is defined as including multiple factors, namely the ingestion of unsafe water, lack of water linked to inadequate hygiene, poor personal and domestic hygiene and agricultural practices, contact with unsafe water, and inadequate development and management of water resources or water systems.

For estimating disease burden of infectious diarrhoea, exposure scenarios are established according to water supply and sanitation infrastructure, the level of faecal-oral pathogens in the environment and populations assigned to these scenarios. The total burdens from schistosomiasis, trachoma, ascariasis, trichuriasis and hookworm disease are all wholly attributable to unsafe WSH and have been quantified at global level as an additional exercise.

Unsafe WSH is an important determinant in a number of additional diseases, such as malaria, yellow fever, filariasis, dengue, hepatitis A and hepatitis E, typhoid fever, arsenicosis, fluorosis and legionellosis, some of which present a high disease burden at global level.

For infectious diarrhoea, six exposure levels were defined, with the lowest risk level corresponding to an ideal situation where WSH plays no role in disease transmission. Exposure prevalence, in terms of infrastructure, was determined from the Global Water Supply and Sanitation Assessment 2000. This assessment is a synthesis of major international surveys and national census reports covering $89 \%$ of the global population. The parameters considered included access to improved water sources and improved sanitation facilities.

Relative risk estimates were based on reviews and large multi-country studies for areas with high faecal-oral pathogen loads in the environment (i.e. principally in developing countries). The proportion of disease
due to unsafe WSH in regions with low faecal-oral pathogen loads was based on a study analysing the relative importance of etiological agents causing diarrhoeal diseases, supported by evidence from selected studies considered to be of high quality. A low faecal-oral pathogen load in the environment was assumed if sanitation coverage exceeded $98 \%$ (which corresponds to the situation in most developed regions).

For the high faecal-oral pathogen exposure group, Esrey's multicountry study (1996) suggests that a mean reduction in diarrhoea of $37.5 \%$ is possible following the introduction of improved water supply and sanitation in developing country environments. For the low faecal-oral pathogen exposure group, data from the study by Mead et al. (1999) suggested that the proportion of diarrhoeal illness attributable to food in the United States of America was approximately $35 \%$ (excluding those illnesses wholly transmitted by food). We have therefore estimated that approximately $60 \%$ was attributable to unsafe WSH. A review by Huttly et al. (1997) of epidemiological studies on hygiene practices in seven nations identified a median reduction of diarrhoea incidence of $35 \%$.

Selected additional studies have suggested ranges of reductions in diarrhoea incidence that could be achieved by reducing the transmission of faecal-oral pathogens through the implementation of interventions, such as point of use treatment and disinfection of stored water (Quick et al. 1999; Semenza et al. 1998). However, this transition has been poorly documented by exposure-risk information, and we considered it appropriate to examine both optimistic and pessimistic estimates in defining the uncertainty around these values.

The disease burden from unsafe WSH was estimated to have been 1.73 million deaths in the year 2000, and $88 \%$ of the global burden of diarrhoeal disease due to infectious diarrhoeal diseases. In addition, schistosomiasis, trachoma, ascariasis, trichuriasis and hookworm disease are fully attributable to WSH-related factors. Typically, the fraction of diarrhoeal disease attributed to unsafe WSH in developed countries is approximately $60 \%$, whereas in developing countries as much as $85-90 \%$ of diarrhoeal illness can be attributed to unsafe WSH. The major part is borne by children in developing countries.

This estimation of the global disease burden caused by unsafe WSH suggests a significant burden of preventable disease attributable to this cause in developing nations, and a non-negligible burden in developed countries.

## 1. Introduction

The disease burden caused by the risk factor unsafe WSH was estimated at the global level in 1990 (Murray and Lopez 1996a). This original estimate examined WSH in terms of diarrhoeal and selected parasitic diseases, based on the partial attribution of their disease burden to the
risk factor. It was found that worldwide the risk factor accounted for $5.3 \%$ of all deaths and $6.8 \%$ of all disability-adjusted life years (DALYs). Other communicable (e.g. hepatitis A and E, malaria) and noncommunicable diseases (arsenicosis, fluorosis, methaemoglobinaemia) were not considered in that assessment.

### 1.1 Rationale for a composite risk factor

Faecal-oral diseases account for the dominant health outcome of the unsafe WSH risk factor and are the main focus of this chapter. However, not all of them could be included in this estimate (e.g. hepatitis A and E). For infectious diarrhoea, the unsafe WSH risk factor comprises a number of transmission routes mediated by a complex interaction of infrastructure issues, which might affect, for example, microbiological hazards from poor quality drinking water, water availability, microbial risks from inappropriate disposal of faecal wastes and behavioural aspects. The transmission routes interact with the efficiency of interventions such as hygiene within the home, hand-washing and rigorous application of point-of-use treatment within domestic properties. Clearly, in any global assessment, the contribution of each element, together with the plethora of interactions, cannot be precisely quantified in every setting. However, as hazard estimates come from large surveys performed in several countries (Esrey 1996), variations in behaviour and their effects on the transmission of faecal-oral pathogens have, to some extent, been internalized in our estimates.

It is likely that the relationship between faecal-oral pathogen dose and the probability of infection is log-linear for many of the infectious diarrhoeal diseases, reaching a plateau for higher exposures (Briscoe 1984; VanDerslice and Briscoe 1995). Sometimes several component causes (see Figure 16.1) may produce similar infection outcomes. This can mean that the introduction of a single intervention in isolation (e.g. the provision of cleaner water supplies) designed to break an infection pathway may result in a negligible reduction in overall disease burden. This is particularly true in communities where the environmental load of faecal-oral pathogens is high (e.g. a community with low sanitation coverage, faecally contaminated drinking-water supplies, irregular refuse collection and poor hygiene practices). This renders the attribution of a disease fraction to a specific factor particularly difficult and, indeed, potentially misleading to the policy-making community. For this reason, it is necessary to consider WSH as interrelated parts of a single causal web in which cutting one major pathway of transmission may well show no (or minimal) effect on the total disease burden, but may, in other circumstances, provoke a dramatic response. Importantly, however, removing a basic pathway (e.g. by providing safe drinking water or improved sanitation) is likely to be a precondition for the success of subsequent interventions to reduce disease burden.

Figure 16.1 Transmission pathways of faecal-oral disease


Management actions concerning water supply and sanitation often involve water resource management, including the control of insect vectors of disease (such as malaria) and soil-borne helminths (such as ascaris). Similarly, environmental management to control disease vectors impacts directly upon water supply and sanitation. Furthermore, access to improved water sources has a significant impact on exposure to agents of some water-based diseases (such as schistosomiasis) and diseases with water-related insect vectors, and improved sanitation reduces certain vector-borne diseases such as trachoma. These intimate interconnections of exposure pathways and control mechanisms suggest that treating water, including supply and resource management, as an integral part of the risk factor unsafe WSH is rational.

### 1.2 Defintition of risk factors

Unsafe WSH adversely affects health through multiple routes.

1. Transmission through contact with water that contains organisms such as Schistosoma spp.
2. Transmission through vectors proliferating in water ecologies related to dams, irrigation schemes and other water resources projects (e.g. malaria, schistosomiasis, lymphatic filariasis). This should be included although it is currently unclear how or whether it can be quantified.
3. Transmission through the ingestion of water as it occurs during drinking and, to some extent, bathing. This category includes diseases from faecal-oral pathogens, dracunculiasis, arsenicosis, fluorosis, from other toxic chemicals and due to excess proliferation of toxic algae.
4. Transmission caused by poor personal, domestic or agricultural practices, including when personal hygiene is affected by lack of water. This includes person-to-person transmission of faecal-oral pathogens, foodborne transmission of faecal-oral pathogens as a result of poor hygiene or use of contaminated water for irrigation or cleaning. Lack of water is in particular linked to diseases such as trachoma and scabies.
5. Transmission through contaminated aerosols from poorly managed water systems (e.g. legionellosis).

Water-related injuries that could be prevented by appropriate water management were not considered in the current estimate because of the different management approaches to their remediation. They were, however, covered by the World Health Organization (WHO 1998), although their disease burden was not quantified. Many social, geographic and behavioural factors, such as hygiene, the domestic storage and potential contamination of potable water, the use of sanitation facilities, etc. are important determinants of health outcome. This set of factors has complex social and behavioural drivers that are highly heterogeneous both within and between nations. In reality, they would modify the effects of the pathways defined in (i) to (v) above. It is beyond the scope of the current assessment to attempt to quantify the unique impacts of this set of factors in each setting.

Diseases relating to unsafe WSH, and their inclusion in the current estimate, are listed in Table 16.1. This first assessment of disease burden should be considered an initial estimate, which will benefit from refinement as additional information becomes available. Table 16.1 is not exhaustive, as the linkages between water and health are extensive and complex. For example, it is likely that the role of inadequate water for food production, and therefore nutrition, will be particularly important, in addition to the direct impact of infectious diarrhoea on nutrition.

### 1.3 Evidence of causality on infectious diarrhoea

As illustrated in Table 16.1, numerous separate faecal-oral illnesses fall under the "umbrella" of infectious diarrhoea. Their commonality derives from their mode of transmission, in that the source of the pathogen is human (or less commonly, animal) faeces which can cause infection in a new host upon ingestion. The shortest route of transmission is from person-to-person (a hygiene issue), while longer routes include transfer of pathogens to a food crop, as well as to drinking water or recreational

| Disease outcome | Included in current estimate |
| :---: | :---: |
| Infectious diarrhoea, including: cholera, salmonellosis, shigellosis, amoebiasis, other bacterial, protozoal and viral intestinal diseases ${ }^{\text {a }}$ | Yes (acute effects only) |
| Typhoid and paratyphoid fevers | Partly included in estimate for infectious diarrhoea, but would benefit from separate, more precise consideration |
| Hepatitis A | No |
| Hepatitis E | No |
| Fluorosis | No |
| Arsenicosis | No |
| Legionellosis | No |
| Methaemoglobinaemia | No |
| Schistosomiasis ${ }^{\text {a,b }}$ | Yes |
| Trachoma ${ }^{\text {a,b }}$ | Yes |
| Ascariasis ${ }^{\text {a,b }}$ | Yes |
| Trichuriasis ${ }^{\text {a,b }}$ | Yes |
| Hookworm ${ }^{\text {a,b }}$ | Yes |
| Dracunculiasis ${ }^{\text {b }}$ | No (disease close to eradication) |
| Scabies | No |
| Dengue ${ }^{\text {a }}$ | No |
| Filariasis ${ }^{\text {a }}$ | No |
| Malaria ${ }^{\text {a }}$ | No |
| Japanese encephalitis ${ }^{\text {a }}$ | No |
| Onchocerciasis ${ }^{\text {a }}$ | No |
| Yellow fever | No |
| Impetigo | No |
| Drowning ${ }^{\text {a }}$ | No |

[^59]water, as summarized in Figure 16.1. The predominant route will depend upon the survival characteristics of the pathogen as well as local infrastructure and human behaviour. While some of the diseases contained in the group diarrhoeal disease, as defined for the purpose of this project, are relatively mild and self-limited, others may be more severe and cause long-lasting sequelae (Hunter 1997). The disease burden based on these studies has not been taken into account in this estimate.

The fact that faecal-oral pathogens can be spread via the water route is well established (Andersson and Bohan 2001; Esrey et al. 1991; Hunter 1997; Snow 1855). The following sections briefly outline the evidence
for infectious diarrhoea causality in relation to water, sanitation and hygiene. For the most part, studies examining the issue have been intervention studies, which have looked at changes in water supply, excreta disposal or hygiene practices, and assessed the effects on diarrhoea morbidity or mortality rates (generally in young children). Another significant group of investigations comprise case-control studies, particularly following outbreaks suspected to be caused by potable water contamination in developed nations.

## SANitation

Ideally, sanitation (i.e. human excreta management) should result in the isolation or destruction of pathogenic material and, hence, a break in the transmission pathway. In a comprehensive literature review, Esrey et al. (1991) identified 30 studies, from a variety of different countries (including Bangladesh, Brazil, Chile, Guatemala, Kenya, Malaysia and Panama), that examined the impact of sanitation on disease transmission. Twenty-one of those studies reported health improvements (median $22 \%$ reduction in diarrhoea morbidity), with a greater median reduction being seen in the rigorous studies ( $36 \%$ reduction). Several studies have isolated various faecal-oral pathogens from the faeces of sick people and the transmission of such pathogens isolated from infected faeces to human hosts has been shown in numerous studies (e.g. for Shigella [Dupont et al. 1989]). Clearly, the relationship is both plausible and coherent.

## Water

The number of outbreaks of infectious diarrhoea caused by faecal-oral pathogens in developed countries attests to the efficiency of this mode of transmission. In the United States, for example, 14 outbreaks of infectious etiology associated with drinking water were reported for the twoyear period 1997-1998 (Barwick et al. 2000).

In developing countries, it is not only water contaminated at source or during distribution that is an issue, but water stored within the home which may also become contaminated (arguably a hygiene issue). For example, in a literature review, VanDerslice and Briscoe (1993) found 11 observational studies showing that mean coliform levels (an indicator of contamination) were considerably higher in household water containers than in the original source waters.

Numerous epidemiological studies and outbreak investigations have found an association between poor water quality and infectious diarrhoea. In France, water that did not meet microbiological standards was associated with an increased risk of gastroenteritis (RR 1.36, CI 1.24-1.49) (Ferley et al. 1986). In the Philippines, Moe et al. (1991) reported an odds ratio (OR) of 1.92 (CI 1.27-2.91) for diarrhoea following consumption of water contaminated with high levels of Escherichia coli (a faecal indicator bacteria). Mahalanabis et al. (1991)
reported that children with prolonged diarrhoeal illness (more than 14 days) were more likely to drink water from an unprotected water source (OR 1.56, CI 1.18-2.06). Birmingham et al. (1997) conducted an epidemiological investigation to identify sources of infection and risk factors for cholera in Burundi during an epidemic in 1992. Water from Lake Tanganyika was implicated, as a case-control study found that both bathing in the lake (OR 1.6, CI 1.1-2.1) and drinking its water (OR 2.78 , CI $1.0-7.5$ ) were independently related to illness; additionally Vibrio cholerae O1 was isolated from the lake water.

As seen above, the causal relationship between ingesting water of poor sanitary quality and diarrhoeal illness has been observed worldwide, using a variety of techniques and assessing quality in a number of different ways. The biological gradient can be illustrated by increases in infectious diarrhoea morbidity as contamination levels increase, and also as consumption of water from a single contaminated source increases. For example, Njemanze et al. (1999) examined the annual diarrhoeal incidence rate (per 1000 population) in 39 communities in Imo State, Nigeria, in relation to the characteristics (including pollution) of their drinking water source. Sources were classified from A to C with A representing the most desirable sources (with favourable geology, sparse population and clean and unpolluted water). Diarrhoeal incidence rate was found to show a statistically significant increase with a mean of 1.61 for category A, a mean of 6.25 for category B, and a mean of 15.6 for category C.

The relationship between infectious diarrhoea and transmission of pathogens through water is both plausible and coherent. Isolation and enumeration of specific pathogens in water are often not feasible or very imprecise; thus a more common measure of faecal contamination is derived from the use of indicator bacteria. There have been many studies using such indicator species that have demonstrated the faecal contamination of drinking water sources in both developed and developing countries (e.g. Ampofo 1997).

## Hygiene

A number of studies have attempted to examine the role of personal and domestic hygiene, although in many cases some of the "hygiene" measures or interventions could also impact on sanitation, and hygiene interventions may also interact with water quality.

Six studies examined by Esrey et al. (1991) identified reductions in diarrhoea morbidity associated with the uptake of hygiene interventions. These ranged from $14 \%$ to $48 \%$, with a median reduction of $33 \%$. In a more recent review, Huttly et al. (1997) identified a further four studies addressing the impact of improved hygiene. All four studies showed a decrease in diarrhoea, as did a subsequent study of Curtis et al. (2000). These studies were conducted in diverse locations including Bangladesh, Burma, Guatemala and the United States.

The temporal adoption of hygiene measures can be illustrated by the study by Ahmed et al. (1993). This group compared cleanliness and diarrhoea levels in villages with and without hygiene education interventions. Higher adoption rates of the intervention were associated with a better cleanliness state, which was paralleled by a decrease in diarrhoea and malnutrition rates. These differences were found to increase over time as more villagers adopted the intervention.

Alam et al. (1989) studied the effect of four different hygiene measures (source of washing water; presence of faeces in the yard; handwashing before serving food; and hand-washing after defecation). They showed decreasing diarrhoea incidence as the number of adopted hygienic practices increased (4.9 cases per child-year for one practice to 2.6 cases for all four; $P<0.01$ ).

A review by Feachem (1984) documented the presence of pathogens on the hands following toilet activities. In the same review, Feachem also noted a number of studies on hand-washing which demonstrated the almost complete removal (98-100\%) of seeded bacteria.

### 1.4 EVIDENCE OF CAUSALITY ON OTHER OUTCOMES

## Schistosomiasis

Schistosomiasis is caused by infection with trematodes of the Schistosoma species. Transmission of the disease occurs when people come into contact with water containing cercariae (the mobile larval stage of the life cycle), which penetrate the skin. Water is contaminated by infected humans who excrete the schistosome eggs in their faeces or urine (depending upon the Schistosoma species). The final link in the chain of infection is provided by an intermediate snail host, which the parasite needs in order to complete its life cycle. Current knowledge on disease transmission indicates that the disease is fully attributable to unsafe WSH.

Esrey et al. (1991) identified 12 studies that related water and sanitation facilities to the rates of schistosomiasis. Reported decreases in infection rates varied between $59 \%$ and $87 \%$, with the median value of the rigorous studies being a $77 \%$ reduction. Numerous studies, in addition to those identified above, have noted the relationship between contact with contaminated water and high levels of infection with schistosomiasis (Hunter 1997). These have been conducted in various countries and have examined different Schistosoma species. Lima e Costa et al. (1991) found that individuals reporting water contact less than once a week had a smaller excess risk of schistosomiasis than those reporting water contact at least weekly (OR 3.0, CI 1.3-6.6 in comparison to OR 4.3, CI 2.6-7.0).

A number of studies have examined reinfection with schistosomiasis following an intervention programme (such as treatment of infected individuals). In China, Zhaowu et al. (1993) found that reinfection was
associated with the frequency of water contact, the type of water contact and the proximity of residence to snail-infected water. In Brazil, discontinuation of a control programme led to an increased prevalence of schistosomiasis (Coura-Filho et al. 1994). Risk factors for the disease included any form of water contact (OR 2.79, CI 1.19-6.85).

The relationship is plausible and the results of numerous studies are coherent and do not conflict with what is known about the disease. Interventions centring on water and sanitation provision designed to either decrease water contamination or decrease contact with contaminated water have proved to be effective in reducing the rates of schistosomiasis (e.g. Barbosa et al. 1971; Jordan 1972).

## TRACHOMA

Trachoma is a chronic contagious eye disease, which can result in blindness, caused by Chlamydia trachomatis. Transmission occurs by several routes (Dolin et al. 1997), all of which are hygiene related (e.g. direct infection by flies, person-to-person from clothing used to wipe children's faces and by hand-to-face contact). Risk factors for the disease include lack of facial cleanliness, poor access to water supplies, lack of latrines and a high number of flies.

A total of 16 studies were identified by Esrey et al. (1991) which examined the role of WSH on the level of trachoma. The median reduction in trachoma was $50 \%(0-91)$ from all the studies and $27 \%(0-9)$ when considering the rigorous studies. More recently Prüss and Mariotti (2000) identified 39 studies which examined the level of trachoma in relation to environmental causation; they report that relative risks ranged between 1 and 4 . Thirteen of the 16 studies identified by Esrey et al. (1991) reported positive effects, i.e. a water, sanitation or hygiene intervention resulted in lower levels of trachoma. The studies were conducted in a variety of locations including Australia, China, India, Mexico, Mozambique, the Sudan and Tunisia.

Prüss and Mariotti (2000) reported that the biological gradient was verified in most of the studies in which it was investigated, although they also noted that few studies examined this issue. Preventative measures through hygiene education and interventions aimed at reducing fly numbers have both resulted in decreases in trachoma (Emerson et al. 1999; Sutter and Ballard 1983).

## ASCARIASIS

Ascariasis is caused by the large roundworm Ascaris lumbricoides. Eggs are passed in the faeces of an infected person and in poor sanitation conditions may contaminate the soil. Ingestion of infective eggs, from contaminated soil or from uncooked products contaminated with soil or wastewater containing infective eggs, cause the disease. Transmission does not occur from person to person. The knowledge on transmission pathway indicates that the disease is fully attributable to unsafe WSH.

The eggs can survive for months or years in favourable conditions and can, thus, pose an infective hazard for a considerable period of time.

A total of 14 studies examining the level of ascariasis and water and sanitation provision were identified by Esrey et al. (1991). These studies reported reductions between $0-83 \%$, with a median reduction from all the studies of $28 \%$. More recently, Cifuentes (1998) reported big differences in infection between children exposed to untreated wastewater and those exposed to either partially treated wastewater or rainwater irrigation (OR 5.71-13.18, depending upon the age group under consideration). Similar results were reported by Habbari et al. (2000), who showed that Ascaris infection was five times higher in children in the wastewater impacted regions compared to control regions. In Indonesia, Toma et al. (1999) reported a $64 \%$ reduction in Ascaris infection in people who used a latrine compared with those who did not.

A biological gradient is suggested from the results of the four rigorous studies identified by Esrey et al. (1991) where the rate of morbidity reduction was dependent upon the level of sanitation facility. The work of Cifuentes (1998) also indicates a dose-response relationship with children exposed to increasingly contaminated water having increased rates of infection.

The relationship is plausible and the study results are coherent. Eggs have been isolated from faecal samples, soil samples, water samples and hand-washing samples (Jonnalagadda and Bhat 1995). Additional experimental evidence is provided by the studies that have examined the increased use of latrines and noted the parallel decrease in both egg counts in soil and levels of infection (e.g. Arfaa et al. 1977).

## TRICHURIASIS

Trichuriasis is caused by ingestion of the human infectious eggs of the whipworm Trichuris trichiura. The infection is not directly transmissible from person to person. As with other faecal-oral transmitted diseases, the mode of transmission indicates that the disease is fully attributable to unsafe WSH, although the risk factors for trichuriasis in relation to WSH do not seem to have been as well researched as the other illnesses covered here. Studies of prevalence often show an association between Ascaris and Trichuris infection (Anderson et al. 1993; Saldiva et al. 1999; Smith et al. 2001), suggesting similar modes of transmission.

Of the studies that were identified, Henry (1981) found that Trichuris infections decreased by $50 \%$ after water supplies and latrines were installed in a rural area of Saint Lucia. Rajeswari et al. (1994) noted that the prevalence of infection was associated with a number of factors, including socioeconomic status, water supply, sanitary disposal of faeces and family size. Similarly, Narain et al. (2000) found that open field defecation and large family size were independently associated with Trichuris infection.

## HOOKWORM DISEASE

Hookworm infection is caused by Ancylostoma duodenale or Necator americanus, and results from the ingestion or skin penetration of the hookworm larvae that live in the soil. Larvae develop in the soil through the deposit of faeces containing eggs from infected persons. The disease is therefore caused by poor sanitation and hygiene practices. The disease is not transmitted from person to person.

Eleven studies were identified by Esrey et al. (1991) which examined water, sanitation and hookworm infection. From the nine that could be used to calculate a reduction in morbidity, the range was $0-100 \%$, although only one of these was considered to be rigorous. Sorensen et al. (1994) found that the severity of hookworm infection was lower in children coming from communities with good sanitary facilities.

Norhayati et al. (1995) studied the reinfection of children in a hookworm endemic area. In the absence of any interventions the reinfection rate at 4 -months post-treatment was $30 \%$. The authors suggested that long-term strategies incorporating education on personal hygiene, provision of toilets and safe water supply were required to control the rapid reinfection. Humphries et al. (1997) reported that hookworm egg counts were significantly higher in Vietnamese women who used fresh human faeces as a fertilizer in comparison to those who used either treated human faeces or did not use human faeces as a fertilizer.

## 2. Methods

The approach builds on methods presented in Prüss et al. (2002), which are further developed in this estimate.

There is strong evidence that, even in developed nations, there is a considerable burden of disease associated with poor-quality potable water or inappropriate sewage disposal and sanitary control. This was demonstrated by disease outbreaks such as the cryptosporidiosis and Escherichia coli O157 epidemics, which affected Canada, the United States and the United Kingdom of Great Britain and Northern Ireland (Andersson and Bohan 2001; Bouchier 1998; Bruce-Grey-Owen Sound Health Unit 2000). In addition, there is a background of sporadic cases in which unsafe WSH has been implicated (Fewtrell and Delahunty 1995). Hence significant health gain is achievable through further improvement in developed nation WSH conditions. This improved condition represents the theoretical minimum exposure in which no disease transmission would occur through unsafe WSH.

As the five transmission pathways of the various outcomes caused by unsafe WSH are quite different (see section 1.2 of this chapter), two approaches for estimating the disease burden were chosen according to the outcome. The estimates of the burden of infectious diarrhoeal disease caused by unsafe WSH are based on exposure information. The burden of other diseases is entirely due to unsafe WSH.

### 2.1 Estimating exposure for diarrhoeal diseases

For estimating the burden of diarrhoeal disease caused by unsafe WSH, we used a scenario-based approach to define exposure categories. In this approach the risk of diarrhoeal disease is conditioned by a typical exposure or a representative combination of risk factors at commonly encountered levels. Six scenarios ${ }^{1}$ (Table 16.2) were defined on the basis of the following:

## Table 16.2 Exposure scenarios

| Level | Description | Environmental faecal-oral pathogen load |
| :---: | :---: | :---: |
| VI | Population not served with improved water supply and no improved sanitation in countries which are not extensively covered by those services (less than 98\% coverage), and where water supply is not likely to be routinely controlled | Very high |
| $\mathrm{Vb}^{\text {a }}$ | Population having access to improved water supply but not served with improved sanitation in countries which are not extensively covered by those services, and where water supply is not likely to be routinely controlled (less than 98\% coverage) | Very high |
| $\mathrm{Va}^{\text {a }}$ | Population having access to improved sanitation but no improved water supply in countries where less than 98\% of the population is served by water supply and sanitation services, and where water supply is likely not to be routinely controlled | High |
| IV | Population having access to improved water supply and improved sanitation in countries where less than $98 \%$ of the population is served by water supply and sanitation services, and where water supply is likely not to be routinely controlled | High |
| $111{ }^{\text {b }}$ | IV and improved access/quality to drinking water; or IV and improved personal hygiene; or IV and drinking water disinfected at point of use, etc. | High |
| II | Population having access to improved water supply and sanitation services in countries where more than $98 \%$ of the population is served by those services; generally corresponds to regulated water supply and full sanitation coverage, with partial treatment for sewage, and is typical in developed countries | Medium to low |
| 1 | Ideal situation, corresponding to the absence of transmission of diarrhoeal disease through WSH | Very low |
| a Transitions between exposure levels Va and Vb do not generally occur. |  |  |
| Cluster of possible improvements over scenario IV, but not reaching scenario II. |  |  |

- the type of water and sanitation infrastructure; and
- the load of faecal-oral pathogens in the environment based on qualitative assessment of sources and disease circulation in the community.

This choice was based on the absence of comprehensive exposure information at individual level and also the lack of relative risk information relating to individual exposure. Risk information was gathered from the literature to match each of the scenarios.

Scenario I represents the minimum theoretical risk and II the situation typically encountered in developed countries. These two scenarios have very low to medium loads of faecal-oral pathogens, characterized by more than $98 \%$ coverage in improved water supply and sanitation and a regional incidence of diarrhoea of less than 0.3 per person per year (Anonymus 2000; Murray and Lopez 1996b). Scenarios IV-VI are in a high faecal-oral pathogen environment, typical for developing countries. Scenario III represents any intervention that improves on scenario IV, and does not currently occur widely. As such, various transitions can be proposed for scenario III and so it is represented as a cluster of possibilities rather than a specific scenario (see Figure 16.2).

Figure 16.2 Scenarios determining risk of diarrhoeal disease from unsafe WSH


## Data sources and quantification of exposure

The exposure scenarios were selected according to available information on exposure-risk relationships and exposure information from the Global water supply and sanitation assessment 2000 (WHO/UNICEF/ WSSCC 2000). The data on water supply and sanitation coverage provided in this assessment are a compilation of two main sources: household surveys, and to a lesser degree assessment questionaires. Relevant information from available household surveys performed on a large scale was accessed, including:

- Demographic Health Surveys (DHS) performed by Macro International and funded by the United States Agency for International Development;
- United Nations Children's Fund's (UNICEF) Multiple Indicator Cluster Surveys (MICS);
- national census reports; and
- other national sample household surveys.

DHS and MICS are national cluster sample surveys, covering several thousand households in each country. The samples are stratified to ensure they are representative of urban and rural areas of each country. In household surveys, consumers are asked to identify the type of water facility they use from a list of technologies. In estimating coverage for the Global Water Supply and Sanitation Assessment 2000, the types of access to services were categorized into "improved" (e.g. borehole, protected dug well, simple pit latrine) and "not improved" (e.g. unprotected well, vendor-provided water, bucket latrines). In addition, national assessment questionnaires were completed by the relevant national agencies in cooperation with WHO and UNICEF country staff. The resulting country estimate for coverage was then based on linear regressions prepared according to available survey data. In the rare cases where household survey data were not available, the coverage figures adopted were those estimated by a local expert committee, based on national assessments and information provided by the country's water authorities.

The Assessment 2000 provides data for water supply and sanitation for almost every country, with information typically available for more than $90 \%$ of the population in every region. It is the only comprehensive assessment of this kind and is, therefore, the single source used for assessing exposure in this analysis. Overall, the Assessment 2000 represents more than $89 \%$ of the global population. Only the European and Western Pacific Regions contain large data gaps, with information on water supply and sanitation coverage lacking for some large countries. Subregions ${ }^{2}$ with low information coverage include EUR-A ( $25 \%$ information coverage), EUR-B ( $65 \%$ ), EUR-C $(11 \%)$ and WPR-A ( $15 \%$ ). In
each case (based on those countries responding), we considered that the available figures on coverage were likely to be representative of the whole subregion. Countries without information were ascribed subregional coverage rates.

Using data from the Assessment 2000, it is not possible to assess whether those served with improved water corresponded to those with improved sanitation, as only coverage was reported. Reports suggest a strong societal and individual preference for improved water supply over improved sanitation, and this is further supported by the higher levels achieved worldwide for improved water supply when compared to improved sanitation. In apportioning populations among exposure scenarios IV and Vb , we therefore assumed that people with improved water supplies were likely to have access to improved sanitation.
The population of each country was thus assigned to the various scenarios based on the Assessment 2000 as described above, and popula-tion-weighted regional means were calculated. The resulting exposure distribution is represented in Table 16.3. In 2000, the percentage of people served with some form of improved water supply worldwide reached $82 \%$ ( 4.9 billion), and $60 \%$ ( 3.6 billion) had access to improved sanitation facilities. In 2000, one sixth ( 1.1 billion people) of the world's population was still without access to improved water supply and two fifths ( 2.4 billion people) lacked access to improved sanitation.

Table 16.3 Distribution of the population in exposure scenarios, 2000

| Subregion | II (\%) | IV (\%) | Va (\%) | Vb (\%) | VI (\%) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | 0 | 54 | 5 | 6 | 35 |
| AFR-E | 0 | 42 | 10 |  | 38 |
| AMR-A | 99.8 | 0 | 0 | 0 | 0.2 |
| AMR-B | 0 | 76 | 1 | 9 | 14 |
| AMR-D | 0 | 68 | 0 | 7 | 25 |
| EMR-B | 0 | 83 | 5 | 8 | 4 |
| EMR-D | 0 | 66 | 0 | 16 | 18 |
| EUR-A | 100 | 0 | 0 | 0 | 0 |
| EUR-B ${ }^{\text {a }}$ | 0 | 79 | 8 | 1 | 12 |
| EUR-C ${ }^{\text {a }}$ | 0 | 94 | 5 | 0 | 1 |
| SEAR-B | 0 | 70 | 3 | 7 | 19 |
| SEAR-D | 0 | 35 | 0 | 53 | 12 |
| WPR-A | 100 | 0 | 0 | 0 | 0 |
| WPR-B | 0 | 42 | 1 | 33 | 24 |

[^60]Scenario I does not occur on a large scale and, in global terms, is probably negligible, hence its omission from Table 16.3. Scenario III is a poorly characterized series of transition states between IV and II and is not separately accounted for. Such scenarios are nevertheless important concepts in policy development and are therefore retained in the model described in Figure 16.2.

### 2.2 RISK FACTOR-DISEASE RELATIONSHIPS FOR DIARRHOEAL DISEASES

## APPROACH

We selected major reviews, multi-country studies or studies of superior design to quantify the transition between two or more chosen exposure scenarios. This included the review and multi-country study by Esrey (Esrey 1996; Esrey et al. 1991), the reviews by Huttly et al. (1997) and Mead et al. (1999), in conjunction with key literature and high quality studies published since the review papers (Payment et al. 1991, 1997; Quick et al. 1999; Semenza et al. 1998). The majority of this literature was based on intervention studies and surveillance information. The final selection of used studies depended largely on the degree to which the study exposure data could be matched with the chosen exposure scenarios and also the sample size and quality of studies. Brief details on the chosen studies are outlined in Table 16.4.

## Relative risk for exposure scenario II

The ideal situation (scenario I) is the theoretical minimum ( $\mathrm{RR}=1$ ). In scenario II, the pathogen load is mostly transferred from land to water (e.g. in discharge of normally treated sewage, such as biological secondary treatment, to surface water). Such pathogens can potentially pass through potable water treatment systems, which can not guarantee $100 \%$ pathogen elimination in even the most advanced plants used in developed nations. Water contaminated with such pathogens is also used for other purposes such as recreation and irrigation. Hygiene behaviour is still imperfect in scenario II, and small population groups may still be served with poorly regulated community supplied water. In scenario I, the ideal scenario, all this would not occur.

Relative risk for scenario II was based on the review by Mead et al. (1999). Mead et al. assessed the level of all infectious foodborne illness in the United States, using data from a large number of surveys and other sources (including FoodNet, the National Notifiable Disease Surveillance System, the Public Laboratory Information System, the Foodborne Disease Outbreak Surveillance System, the National Hospital Discharge Survey, the National Vital Statistics System and a number of published studies). Based on the literature, they also estimated the percentage of each disease caused by foodborne transmission. This is a very comprehensive study based on more than 400000 diagnosed cases, bringing
Table I6.4 Key studies and reviews

| Reference | Study population | Sample size | Outcome measured/ reported | Reductions | Comments |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Esrey (1996) | Representative populations from Bolivia, Burundi, Ghana, Guatemala, Morocco, Sri Lanka, Togo, Uganda | 16880 | Diarrhoea morbidity nutritional status, child development | $20.8-37.5 \%$ <br> according to type of infrastructure | Detailed examination of effects of incremental improvements in water and sanitation based on survey data |
| Huttly et al. (1997) | Bangladesh, Burma, India, Indonesia, USA for hand-washing; Bangladesh, Guatemala, Thailand, Zaire for various other forms of behaviour | NA | Diarrhoea morbidity | Median reduction 35\% for hand-washing; median 26\% for other hygiene behaviours | Review paper/5 intervention studies on hand-washing and 5 on other hygiene behaviours |
| Mead et al. (1999) | Gastrointestinal illness in the USA population | More than 400000 diagnosed patients | Foodborne illness | Approx. 60\% of gastrointestinal illness due to unsafe WSH ${ }^{\text {a }}$ | Surveillance data |
| Payment et al. (1991) | 606 households in Montreal, Canada | 2408 | Diarrhoea morbidity | 35\% | Water quality intervention |
| Payment et al. (1997) | 1400 families in Montreal, Canada | 5253 | Diarrhoea morbidity | 14-40\% | Water quality intervention |
| Quick et al. (1999) | Two Bolivian communities | 791 | Diarrhoea morbidity | 45\% for all age groups | Water quality intervention |
| Semenza et al. (1998) | Householders in Nukus, Uzbekistan | 1583 | Diarrhoea morbidity | 62-85\% | Water quality intervention |
| NA Not applicable. a Extrapolated from | ly results for the purpose of this analysis. |  |  |  |  |

together numerous different data sources and some assumptions relating to likely levels of underreporting. According to this study, about $35 \%$ of intestinal illness in the United States is foodborne. The level of faecal-oral illness due to unsafe WSH was estimated as $100 \%$ of the cases of infectious diarrhoea, less the percentage due to foodborne transmission. This is probably an underestimate as it is likely that unsafe WSH play a role in some foodborne transmission (e.g. through irrigation of food products with pathogen-contaminated water or via an infected food handler). After deduction of the portion of foodborne transmission and accounting for likely ratios of person-to-person transmission through aerosols of certain viruses (estimated as up to $25 \%$ for rotavirus and astrovirus), the remaining fraction attributable to unsafe WSH is about $60 \%$. This order of magnitude is supported by intervention studies acting on point-of-use treatment of drinking water in Canada (Payment et al. 1991, 1997) and hand-washing in the United States (Black et al. 1981), reporting reductions of $40 \%, 35 \%$ and $48 \%$, respectively. A $60 \%$ reduction in disease corresponds to a relative risk of $2.5(\mathrm{RR}=1 /(1-0.6))$ for exposure scenario II.

## Risk transition between scenarios II and IV

Scenarios II and IV represent high and low environmental pathogen loads. Intervention studies were not available, as it is not possible to transform environments high in pathogen load into environments low in pathogen load; doing so would imply completing the coverage in improved water supply and sanitation in a reasonable time frame and without simultaneous change in other major determinants of health. Therefore, relative risks for scenarios between II and IV were estimated using selected studies.

- Scenario IV and improved drinking-water quality: Quick et al. (1999) examined the level of diarrhoea prevention that could be achieved through point-of-use water treatment along with safe water storage. This study was selected as the intervention strongly reduces the pathway of transmission through drinking water, and "simulates" the reduction that could be achieved by improved drinking water quality and its handling inside the house. The study randomized 791 participants into two groups. The intervention group received a special storage container (preventing hand contact with the stored water) and a supply of disinfectant. The control group not receiving the intervention was similar in terms of demographic characteristics, sanitary conditions and baseline water quality. During the baseline investigations only $5 \%$ of household samples were free of E. coli. During the study period this varied between $0 \%$ and $13 \%$ of the control group (with the median level being between 5000 and 85000 of E. Coli $/ 100 \mathrm{ml}$ ), while the intervention group exceeded $50 \%$ of households at all times, rising to almost $80 \%$ on one occasion (median $E$.
coli counts were zero, throughout). Overall diarrhoea reductions of $44.7 \%$ in the total population and $54.5 \%$ in children have been reported by Quick et al. ( $\mathrm{RR}=1.81$ and 2.20 ). The reduction of $44.7 \%$ was selected as a component in the transition between II and IV in this analysis.

In a randomized intervention study in 240 households (120 with and 120 without access to municipal piped water) with a total population of 1583 in Uzbekistan (Semenza et al. 1998), approximately half of the households without piped water were trained to chlorinate their drinking water within the home and store it in a safe manner. Diarrhoea morbidity was markedly lower in the home-chlorination group (28.8/1000 subjects per month), compared to $75.5 / 1000$ in the piped water group and $179.2 / 1000$ in the no piped water group (i.e. a $62 \%$ reduction in diarrhoea rates for an intervention with home chlorination of drinking water, as compared to those living in areas with access to piped water $[\mathrm{RR}=2.6]$; in individuals without a piped supply, the same intervention achieved a $85 \%$ reduction in disease $[R R=6.7])$. The authors considered that home chlorination of water was unlikely to affect disease transmission via other routes, and suggested that a large fraction of the diarrhoeal pathogens in this area were spread through water.

- Scenario IV and improved personal bygiene: reductions in diarrhoea morbidity have been reviewed by Huttly et al. (1997), and handwashing resulted in a median $35 \%$ reduction in diarrhoea incidence $(\mathrm{RR}=1.5)$. The results of this review outlined possible achievements due to a reduction in the transmission pathway of hygiene, which in itself is conditioned by the pathogen load in the environment.

Risk transition between scenarios IV and VI
The multi-country study conducted by Esrey (1996) provided data to allow calculation of relative risks between scenarios IV, Va, Vb and VI. This study examined whether incremental health effects relating to diarrhoea and nutritional status resulted from incremental improvements in water and sanitation conditions and was based on DHS from eight countries from five different regions (Bolivia, Burundi, Ghana, Guatemala, Morocco, Sri Lanka, Togo, Uganda). DHS included information on diarrhoea prevalence, child weight, child height, child age, source of drinking water and type of sanitation facility. In addition, the survey data were supplemented by field studies that determined current levels of diarrhoea prevalence in children aged 3-36 months. According to this study, a reduction of $20.8 \%$ in diarrhoeal disease rates $(R R=1.26)$ could be observed when progressing from scenario VI to Vb (i.e. when providing an improved water supply), and $37.5 \%(R R=1.6)$ when progressing from VI to Va (i.e. when providing improved sanitation facilities). When progressing from VI to IV (i.e. when providing both an improved water

Table 16.5 Relative risks

|  | Exposure categories or transition between scenarios |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | 11 | III | IV | Va (to IV) | Vb (to IV) ${ }^{\text {a }}$ | $\begin{gathered} \text { VI (to } \\ \text { Vb) } \\ \hline \end{gathered}$ | VI (to Va and IV) |
| Risk reduction ${ }^{\text {b }}$ | NA | 60\% | Various ${ }^{\text {c }}$ | 45\% and 35\% | 0\% | - | 20.8\% | 37.5\% |
| Partial relative risk ${ }^{\text {a }}$ | NA | 2.5 | Various ${ }^{\text {c }}$ | $\begin{aligned} & 1.81 \text { and } \\ & 1.54 \end{aligned}$ | 1.0 | $\begin{aligned} & \text { I. } 60 / \mathrm{I} .26 \\ & =1.27 \end{aligned}$ | 1.26 | 1.60 |
| Absolute relative risks (compared to scenario I) | I | 2.5 | Various ${ }^{\text {c }}$ | 6.9 | 6.9 | 8.7 | 11.0 | 11.0 |

NA Not applicable.

- No data.
abtained by calculating the remaining risk differences between VI to Vb as compared to VI and IV .
b Relative to the scenario below.
c See text
supply and improved sanitation facilities), a reduction of $37.5 \%$ was also achieved. This implies that no further reduction in diarrhoeal disease is achieved when implementing an improved water supply, when improved sanitation is already available. These data are supported by the review of Esrey et al. (1991), which provides similar results for the same types of interventions.

The resulting relative risks are obtained by multiplying the relative risks between each scenario, summarized in Table 16.5.

According to our model, the risks of diarrhoea incidence in developing countries are 2.8 to 4.4 times higher (Table 16.5) than current risks in developed countries. The same order of magnitude of difference in diarrhoea rates was reported by various compilations of health statistics or studies (Esrey 1996; Murray and Lopez 1996b).

### 2.3 Estimating risk factor-Disease relationships for diseases other than diarrhoeal diseases

The World bealth report 2001 (WHO 2001) provided estimates of the burden of additional diseases that are exclusively (or virtually exclusively) caused by unsafe WSH (Table 16.6).

### 2.4 Sources of uncertainty

## METHOD

The method is based on typical scenarios, characterized by a combination of sub-risk factors, which should represent most of the world's situations. Certain population groups may not be captured by any one of these scenarios, but the number of groups is probably small, which may

Table 16.6 Global disease burden caused by selected water-related diseases other than infectious diarrhoea in 2000

| Disease | Deaths (000s) | DALYs (000s) |
| :--- | :---: | :---: |
| Schistosomiasis | 11 | 1713 |
| Trachoma | 0 | 1161 |
| Ascariasis | 6 | 1252 |
| Trichuriasis | 2 | 1640 |
| Hookworm disease | 6 | 1829 |
| Total | 25 | 7595 |
| Source: WHO (200I). |  |  |

be partly internalized in the risk estimates. For example, differences are likely to exist in the specific WSH practices in the various households in the same exposure scenario. The effect of these differences should, however, largely be captured in the large samples on which this study is based. The current study is therefore based on average risks for large population groups within which a variety of individual practices and situations are represented.

## Exposure estimates

The Water Supply and Sanitation Assessment 2000, which reports individual country data, exhibited variable precision between respondents, particularly in relation to rural and tribal populations. A more precise exposure estimate would require actual assessments, such as the water quality of the supply. Such measures are impractical on a large level. The Assessment 2000, however, captures exposure information for a majority of the world's countries and represents a solid source of information. Uncertainty in water supply and sanitation coverage has therefore not introduced major uncertainty into our analysis.

## RISK ESTIMATES

This analysis used large surveys and multi-country studies where available. It is therefore based on risk averages, i.e. the average of risk related to the described scenarios across the world and across an array of situations. While this method may not be suitable for specific local settings, it should provide a reasonable estimate for large regions.

Where no large surveys, reviews or multi-country studies were available (i.e. in part the transition between scenarios IV and II), the use of sentinel studies for "global" application may constitute a significant source of error. Therefore, this analysis has been selective on the basis of study quality and coverage, to ensure maximum transferability.

As much of the described imprecision will remain largely unquantifiable, upper and lower uncertainty boundaries are based on varying the

Table 16.7 Low and high relative risk estimates

|  | Exposure scenario |  |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
|  | I | $I I$ | $I I I$ | $I V$ | $V a$ | $V b$ | VII |
| Lower estimate | I | 2.5 | Variable | 3.8 | 3.8 | 4.9 | 6.1 |
| Best estimate | I | 2.5 | Variable | 6.9 | 6.9 | 8.7 | 11.0 |
| Upper estimate | I | 2.5 | Variable | 10.0 | 10.0 | 12.6 | 16.0 |

estimates of the potentially greatest source of uncertainty-the transition between scenarios IV and II (i.e. from an environment with a high faecal-oral pathogen load to one with a low faecal-oral pathogen load). The lower estimate was based solely on the improvement that can be achieved by implementing personal hygiene measures ( $35 \%$ risk reduction or a RR of 1.54). This is comparable with the best estimate, which is based on a combination of improved water quality and improved hygiene (see Table 16.5). For the upper estimate, the additional risk reduction relating to the provision of continuous piped water supply (i.e. improved access to water) was considered in addition to hygiene improvements. This is represented by a relative risk of 2.6 (from a $62 \%$ risk reduction from the study by Semenza et al. 1998) in addition to that resulting from hygiene improvements. The resulting relative risks for each of the estimates are summarized in Table 16.7.

The same relative risk is assumed for all age groups. As most of these rates have been assessed for children and the largest disease burden also occurs in that age group, the error of applying the same relative risks to adults is probably small. Also, several studies that have assessed both relative risks for children and adults have shown that figures do not generally differ dramatically, although the impact on young children tends to be higher (e.g. Quick et al. 1999).

It should be noted that faecal-oral disease transmission is partly conditioned by the prevalence of the risk factor at community level. For example, protection of drinking water depends on the effective implementation of an intervention by all members of the community, whereas studies have often been performed at individual level, generally resulting in underestimation of the benefits of community-wide interventions.

## SEQUELAE AND DELAYED EFFECTS

Estimation of the burden of disease due to infectious diarrhoea is based upon the acute diarrhoeal episode and associated mortality. Several of the agents of infectious diarrhoea are associated with other health effects, often delayed. These may add significantly to the burden of disease, as is the case of campylobacteriosis, for example. Inadequate evidence was available to reliably estimate the additional burden of disease.

## 3. Results

The attributable fractions, deaths and number of DALYs are listed in the annex tables (see CD-ROM accompanying this book), for the 14 subregions, males, females and eight age groups.

Globally, in the year 2000, almost 1.73 million deaths due to diarrhoeal diseases were attributable to unsafe WSH as defined in the exposure variable used in this work; $68 \%$ of them are children. Most of these deaths, $>99 \%$, occur in developing countries. The attributable fractions of diarrhoeal disease vary between $60 \%$ in developed countries to $85-90 \%$ in developing countries. The difference in disease burden between developed and developing subregions, despite the relative similarities in attributable fractions, is largely due to the lower incidence and case fatality rates of diarrhoeal disease in developed nations. The African subregions alone, together with SEAR-D and EMR-D, bear $88 \%$ of the death burden. The disease burden in males and females is similar. The disease burden from the five other diseases that have been quantified separately is 25000 deaths and 7.6 million DALYs, also concentrated in developing countries. This chapter highlights and confirms the concentration of the burden of disease due to the risk factor unsafe WSH in poor countries and on children- $99.7 \%$ of DALYs and $99.8 \%$ of deaths occur in developing countries, with $80 \%$ of DALYs among children. Globally, $3.1 \%$ of all deaths and $3.7 \%$ of DALYs were attributable to water, sanitation and hygiene, caused by the diseases we could include in this analysis. In the age group 0 to 4 years, these percentages amounted to $11 \%$ of all deaths and $9 \%$ of all DALYs, which shows the importance of this risk factor.

## 4. Projections of future exposure

As the methods for estimating disease burden rely heavily on water supply and sanitation coverage, these are the main parameters that need to be projected for estimating future burden. Progress with water supply and sanitation coverage is affected by factors such as demographic change, income, policies and investments, education, technology, types and management of infrastructure, and involvement of the community and the public and private sectors. In practice, these vary widely within and between countries, making future projections difficult and complex.

To some extent these factors respond to major national and international policy initiatives. The International Drinking-water Supply and Sanitation Decade (1981-1990) established momentum that certainly produced an acceleration of investments from 1981 to 1990 and beyond this period. The Millennium Declaration established the targets of halving the proportion of the population not served with safe water supply by 2015 and improving sanitation for the urban poor. The impact of these historic and future activities on either overall progress with service levels or upon the factors outlined above is difficult to assess. The
proposed coverage forecast method and respective coverage figures generated are presented below.

While efforts are ongoing to develop a model for forecasting improved water supply and sanitation coverage based upon understanding of the factors outlined above, the lack of sufficient data has limited the value of this in preparing future projections. Water supply and sanitation coverage may be predicted by certain distal causes such as income and education; however, in the given time frame a prediction based on past evolution and future demographic changes was preferred. Prediction on the basis of the Human Development Index provided similar results at global level. Global data sets on service coverage generated by WHO in the 1980s were primarily based on country reporting and provided results with limited comparability. More recently, WHO and UNICEF have assessed water supply and sanitation coverage in 1990 and 2000 (WHO/UNICEF/WSSCC 2000), based on household survey data and data by service providers (water agencies, ministries) in the absence of survey data. This shift in methodology provided more reliable and comparable data. The prediction was thus based on the following points.

- It was assumed that the same number of people that acquired coverage between 1990 and 2000 would acquire coverage per decade during the next three decades.
- Population projections from the United Nations Statistics Division (UN 2001) were used.
- For EUR-B and EUR-C, progress in the decade 1990-2000 shows declining trends and does not provide a reasonable basis for projection. Zero change in absolute numbers served was assumed.

It is important to note that this projection assumes that local, national and international efforts as undertaken in the last decades, will continue. The method further assumes no development of approaches or technologies that will enable a shift for part of the population into exposure scenario I, the ideal scenario.

Coverage was projected separately for each subregion. A summary of projected water supply and sanitation coverage is provided in Table 16.8, and a detailed projection per subregion, according to the exposure scenarios used in this analysis, is presented in Table 16.9.

The data presented suggest that the water supply goal and target ${ }^{3}$ adopted in the Millennium Declaration are likely to be achieved globally if a similar effort as compared to that undertaken in the last decade is continued until the year 2015. For certain subregions, however, the target may not be achieved, namely those in the African continent, as well as EMR-B (where coverage is already high and where half of the population are likely to experience an important risk reduction), EURB and EUR-C, under the assumption that past trends will continue.

Table 16.8 Global projection of water supply and sanitation coverage

| Year | Total population (millions) | Access to improved water sources |  |  | Access to improved sanitation |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Population served (millions) | $\begin{gathered} \% \\ \text { served } \end{gathered}$ | Population not served (millions) | Population with access (millions) | \% having access | Population without access (millions) |
| 1990 | 5255 | 4072 | 77 | 1183 | 2582 | 49 | 2673 |
| 2000 | 6057 | 4976 | 82 | 1081 | 3646 | 60 | 2411 |
| 2010 | 6826 | 5894 | 86 | 932 | 4739 | 69 | 2087 |
| 2015 | 7207 | 6353 | 88 | 854 | 5285 | 73 | 1922 |
| 2020 | 7579 | 6802 | 90 | 777 | 5831 | 77 | 1748 |
| 2030 | 8270 | 7681 | 93 | 589 | 6902 | 83 | 1368 |

It is not always possible to see clear trends within scenarios. This is likely to be due to highly variable rates of population growth and movement of populations between scenarios. Greatest health gains are likely to be associated with movement of populations from scenario VI to better circumstances. AFR-D and AFR-E contain the highest proportion of population in scenario VI. The forecasts indicate that the situation is not likely to change dramatically over the next 30 years if the trends of the last decade continue.

SEAR-D and WPR-B present large proportions of their population with fairly good levels of coverage but relatively low proportions of people served with sanitation facilities (scenario Vb ). AMR-A, EUR-A, and WPR-A have reached or will soon reach $100 \%$ coverage (scenario II). SEAR-B and AMR-B are projected to make good progress, tending from exposure scenarios Vb and VI to IV and II. AMR-D, in addition to developing a trend similar to the trend above is likely to make considerable progress towards scenario II. EUR-C should experience a large shift into scenario II (two large countries that were close to full coverage in 1990 will reach such status by 2010).

## Sources of error/SENSITIVITY analysis

Factors such as water scarcity, competition for water resources and the cumulative effects of pollution of water resources are likely to both increase the cost of interventions and reduce their sustainability. This is due to pressures on both the quality and availability of water resources and would suggest that the projections might be optimistic estimates.

## Acknowledgements

We gratefully acknowledge the comments and suggestions provided by Steve Luby and Carlos Corvalan, the participants in the review meeting of the WHO comparative risk assessment in Auckland in December

Table 16.9 Projection of distribution of exposure by scenarios, 2000 to $2030^{\text {a }}$

| Subregion | 11 (\%) |  |  |  | IV (\%) |  |  |  | Va (\%) |  |  |  | Vb (\%) |  |  |  | VI (\%) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Year |  |  |  | Year |  |  |  | Year |  |  |  | Year |  |  |  | Year |  |  |  |
|  | 2000 | 2010 | 2020 | 2030 | 2000 | 2010 | 2020 | 2030 | 2000 | 2010 | 2020 | 2030 | 2000 | 2010 | 2020 | 2030 | 2000 | 2010 | 2020 | 2030 |
| AFR-D ${ }^{\text {b }}$ | 0 | 1 | 1 | 0 | 54 | 55 | 54 | 54 | 5 | 2 | 1 | 1 | 6 | 11 | 12 | 14 | 35 | 32 | 32 | 31 |
| AFR-E | 0 | 16 | 12 | 12 | 42 | 44 | 42 | 48 | 10 | 7 | 4 | 3 | 9 | 7 | 11 | 6 | 38 | 27 | 31 | 31 |
| AMR-A | 99.8 | 100 | 100 | 100 | 0.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.2 | 0 | 0 | 0 |
| AMR-B | 0 | 0 | 4 | 4 | 76 | 81 | 81 | 85 | 1 | 0 | 0 | 1 | 9 | 10 | 10 | 9 | 14 | 8 | 4 | 2 |
| AMR-D | 0 | 0 | 17 | 16 | 68 | 77 | 65 | 68 | 0 | 1 | 1 | 1 | 7 | 6 | 5 | 3 | 25 | 15 | 13 | 11 |
| EMR-B | 0 | 3 | 7 | 53 | 83 | 83 | 79 | 34 | 5 | 5 | 5 | 5 | 8 | 4 | 3 | 1 | 4 | 5 | 6 | 7 |
| EMR-D | 0 | 23 | 22 | 21 | 66 | 53 | 59 | 65 | 0 | 0 | 0 | 0 | 16 | 10 | 6 | 3 | 18 | 13 | 13 | 11 |
| EUR-A | 100 | 100 | 100 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EUR-B | 0 | 7 | 7 | 6 | 79 | 72 | 73 | 74 | 8 | 6 | 6 | 5 | 1 | 1 | 1 | 1 | 12 | 13 | 14 | 14 |
| EUR-C | 0 | 74 | 72 | 70 | 94 | 23 | 24 | 25 | 5 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 3 |
| SEAR-B | 0 | 27 | 20 | 19 | 70 | 45 | 56 | 61 | 3 | 0 | 1 | 0 | 7 | 17 | 17 | 17 | 19 | 11 | 7 | 3 |
| SEAR-D | 0 | 0 | 0 | 0 | 35 | 39 | 47 | 54 | 0 | 0 | 0 | 0 | 53 | 58 | 53 | 46 | 12 | 2 | 0 | 0 |
| WPR-A | 100 | 100 | 100 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| WPR-B | 0 | 0 | 0 | 0 | 42 | 49 | 60 | 67 | 1 | 0 | 1 | 1 | 33 | 31 | 29 | 26 | 24 | 19 | 14 | 8 |

Rounding of percentages may lead to sums slightly different from $100 \%$.
Source: José Hueb, personal communication.

2000, and the five peer reviewers. We also wish to acknowledge the valuable support and contribution of the U.S. Environmental Protection Agency. This chapter has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency.

## Notes

1 The scenarios are equivalent to exposure categories used in other chapters in this book, in the sense that there is increasing risk across scenarios defined based on faecal-oral load. The term scenario is used here, as the shift from one level of faecal-oral load to another may occur due to changes in any of the multiple dimensions of exposure (water, sanitation and hygiene).
2 See the preface for an explanation of this term.
3 To halve the proportion of people not having access to water supply services by 2015 compared to 1990 .

## References

Ahmed NU, Zeitlin MF, Beiser AS, Super CM, Gershoff SN (1993) A longitudinal study of the impact of behavioural change intervention on cleanliness, diarrhoeal morbidity and growth of children in rural Bangladesh. Social Science and Medicine, 37:159-171.

Alam N, Wojtyniak B, Henry FJ, Rahaman MM (1989) Mothers' personal and domestic hygiene and diarrhoea incidence in young children in rural Bangladesh. International Journal of Epidemiology, 18:242-247.
Ampofo JA (1997) A survey of microbial pollution of rural domestic water supply in Ghana. International Journal of Environmental Health Research, 7:121-130.
Anderson TJ, Zizza CA, Leche GM, Scott ME, Solomons NW (1993) The distribution of intestinal helminth infections in a rural village in Guatemala. Memorias do Instituto Oswaldo Cruz, 88:53-65.
Andersson Y, Bohan P (2001) Disease surveillance and waterborne outbreaks. In: Water quality: guidelines, standards and health. Assessment of risk and risk management for water-related infectious disease. Fewtrell L, Bartram J, eds. World Health Organization, Geneva.

Arfaa F, Sahba GH, Farahmandian I, Jalali H (1977) Evaluation of the effect of different methods of control of soil-transmitted helminths in Khuzestan, south-west Iran. American Journal of Tropical Medicine and Hygiene, 26: 230-233.

Barbosa FS, Pinto R, Souza OA (1971) Control of schistosomiasis mansoni in a small north east Brazilian community. Transactions of the Royal Society of Tropical Medicine and Hygiene, 65:206-213.
Barwick RS, Levy DA, Craun GF, Beach MJ, Calderon RL (2000) Surveillance for waterborne-disease outbreaks-United States, 1997-1998. Morbidity and Mortality Weekly Reports, 49:S1-36.

Birmingham ME, Lee LA, Ndayimirije N et al. (1997) Epidemic cholera in Burundi: patterns of transmission in the Great Rift Valley Lake region. The Lancet, 349:981-985.

Black RE, Dykes AC, Anderson AC et al. (1981) Handwashing to prevent diarrhoea in day-care centers. American Journal of Epidemiology, 113:445451.

Bouchier IAD (1998) Report of the Group of Experts on Cryptosporidium in water supplies. Department of the Environment, Transport and the Regions. London.

Bruce-Grey-Owen Sound Health Unit (2000) The investigative report on the Walkerton outbreak of waterborne gastroenteritis. Available at http:// www.publichealthbrucegrey.oc.ca/private/Report/SPReport.htm.
Briscoe J (1984) Intervention studies and the definition of dominant transmission routes. American Journal of Epidemiology, 120:449-455.

Cifuentes E (1998) The epidemiology of enteric infections in agricultural communities exposed to wastewater irrigation: perspectives for risk control. International Journal of Environmental Health Research, 8:203-213.
Coura-Filho P, Rocha RS, Farah MW, da Silva GC, Katz N (1994) Identification of factors and groups at risk of infection with Schistosoma mansoni: a strategy for the implementation of control measures? Revista Instituto de Medicina Tropical de São Paulo, 36:245-253.
Curtis V, Cairncross S, Yonli R (2000) Domestic hygiene and diarrhoeapinpointing the problem. Tropical Medicine and International Health, 5:22-32.

Dolin PJ, Faal H, Johnson GJ, Minassian D et al. (1997) Reduction of trachoma in a sub-Saharan village in absence of a disease control programme. The Lancet, 349:1511-1512.
Dupont HL, Levine MM, Hornick RB, Formal SB (1989) Inoculum size in shigellosis and implications for expected mode of transmission. Journal of Infectious Diseases, 159(6):1126-1128.
Emerson PM, Lindsay SW, Walraven GE et al. (1999) Effect of fly control on trachoma and diarrhoea. The Lancet, 353:1401-1403.
Esrey SA (1996) Water, waste, and well-being: a multicountry study. American Journal of Epidemiology, 143:608-623.
Esrey SA, Potash JB, Roberts L, Shiff C (1991) Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. Bulletin of the World Health Organization, 69:609-621.
Feachem RG (1984) Interventions for the control of diarrhoeal diseases among young children: promotion of personal and domestic hygiene. Bulletin of the World Health Organization, 62:467-476.
Ferley JP, Zmirou D, Collin JF, Charrel M (1986) Etude longitudinale des risques liés à la consommation d'eaux non conformes aux normes bactériologiques. Revue d'Epidémiologie et de Santé Publique, 34:89-99.

Fewtrell L, Delahunty A (1995) The incidence of cryptosporidiosis in comparison with other gastrointestinal illnesses in Blackpool, Wyre and Fylde. Journal of the Chartered Institution of Water and Environmental Management, 9:598-601.

Habbari K, Tifnouti A, Bitton G, Mandil A (2000) Geohelminthic infections associated with raw wastewater reuse for agricultural purposes in BeniMellal, Morocco. Parasitology International, 48:249-254.
Henry FJ (1981) Environmental sanitation infection and nutritional status in infants in rural St. Lucia, West Indies. Transactions of the Royal Society of Tropical Medicine and Hygiene, 75:507-513.

Humphries DL, Stephenson LS, Pearce EJ, The PH, Dan HT, Khanh LT (1997) The use of human faeces for fertilizer is associated with increased intensity of hookworm infection in Vietnamese women. Transactions of the Royal Society of Tropical Medicine and Hygiene, 91:518-520.
Hunter PR (1997) Waterborne disease: epidemiology and ecology. John Wiley \& Sons Ltd., Chichester, England.
Huttly SRA, Morris SS, Pisani V (1997) Prevention of diarrhoea in young children in developing countries. Bulletin of the World Health Organization, 75:163-174.
Jonnalagadda PR, Bhat RV (1995) Parasitic contamination of stored water used for drinking/cooking in Hyderabad. Southest Asian Journal of Tropical Medicine and Public Health, 26:789-794.

Jordan P (1972) Epidemiology and control of schistosomiasis. British Medical Bulletin, 28:55-59.

Lima e Costa MFF, Rocha RS, Leite MLC et al. (1991) A multivariate analysis of socio-demographic factors, water contact patterns and Schistosoma mansoni infection in an endemic area in Brazil. Revista Instituto de Medicina Tropical de São Paulo, 33:58-63.
Mahalanabis D, Alam AN, Rahman N, Hasnat A (1991) Prognostic indicators and risk factors for increased duration of acute diarrhoea and for persistent diarrhoea in children. International Journal of Epidemiology, 20:1064-1072.
Mead PS, Abdel-Moneim M, al-Erian RA, al-Amari OM (1999) Food-related illness and death in the United States. Emerging Infectious Diseases, 5:607-625.

Moe CL, Sobsey MD, Samsa GP, Mesolo V (1991) Bacterial indicators of risk of diarrhoeal disease from drinking-water in the Philippines. Bulletin of the World Health Organization, 69:305-317.
Murray CJL, Lopez AD, eds. (1996a) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, Vol 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.
Murray CJL, Lopez AD, eds. (1996b) Global health statistics: a compendium of incidence, prevalence and mortality estimates for over 200 conditions. Global Burden of Disease and Injury, Vol 2. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Narain K, Rajguru SK, Mahanta J (2000) Prevalence of Trichuris trichuria in relation to socio-economic and behavioural determinants of exposure to infection in rural Assam. Indian Journal of Medical Research, 112:140-146.

Njemanze PC, Anozie J, Ihenacho JO, Russell MJ, Uwaeziozi AB (1999) Application of risk analysis and geographic information system technologies to the prevention of diarrheal diseases in Nigeria. American Journal of Tropical Medicine and Hygiene, 61:356-360.
Norhayati M, Oothuman P, Fatmah MS, Muzain Minudin Y, Zainuddin B (1995) Hookworm infection and reinfection following treatment among Orang Asli children. Medical Journal of Malaysia, 50:314-319.

Payment P, Richardson L, Siemiatycki J, Dewar R, Edwardes M, Franco E (1991) A randomized trial to evaluate the risk of gastrointestinal disease due to consumption of drinking water meeting current microbiological standards. American Journal of Public Health, 81:703-708.
Payment P, Siemiatycki J, Richardson L, Renaud G, Franco E, Prévost M (1997) A prospective epidemiological study of gastrointestinal health effects due to the consumption of drinking water. International Journal of Environmental Health Research, 7:5-31.

Prüss A, Mariotti SP (2000) Preventing trachoma through environmental sanitation: a review of the evidence base. Bulletin of the World Health Organization, 78:258-266.

Prüss A, Kay D, Fewtrell L, Bartram J (2002) Estimating the burden of disease due to water, sanitation and hygiene at global level. Environmental Health Perspectives 110:537-542.
Quick RE, Venczel LV, Mintz ED et al. (1999) Diarrhoea prevention in Bolivia through point-of-use water treatment and safe storage: a promising new strategy. Epidemiology and Infection, 122:83-90.
Rajeswari B, Sinniah B, Hussein H (1994) Socio-economic factors associated with intestinal parasites among children living in Gombak, Malaysia. Asia Pacific Journal of Public Health, 7:21-25.

Saldiva SR, Silveira AS, Philippi ST et al. (1999) Ascaris-Trichuris association and malnutrition in Brazilian children. Paediatric and Perinatal Epidemiology, 13:89-98.

Semenza JC, Roberts L, Henderson A, Bogan J, Rubin CH (1998) Water distribution system and diarrheal disease transmission: a case study in Uzbekistan. American Journal of Tropical Medicine and Hygiene, 59:941-946.

Smith H, Dekaminsky R, Niwas S, Soto R, Jolly P (2001) Prevalence and intensity of infections of Ascaris lumbricoides and Trichuris trichiura and associated socio-demographic variables in four rural Honduran communities. Memorias do Instituto Oswaldo Cruz, 96:303-314.
Snow J (1855) On the mode of cholera communication. John Churchill, London.
Sorensen E, Ismail M, Amarasinghe DK, Hettiarachchi I, Dassenaieke TS (1994) The effect of the availability of latrines on soil-transmitted nematode infections in the plantation sector in Sri Lanka. American Journal of Tropical Medicine and Hygiene, 51:36-39.

Sutter EE, Ballard RC (1983) Community participation in the control of trachoma in Gazankulu. Social Science and Medicine, 17:1813-1817.

Toma T, Miyagi I, Kamimura K et al. (1999) Questionnaire survey and prevalence of intestinal helminthic infections in Barru, Sulawesi, Indonesia. Southeast Asian Journal of Tropical Medicine and Public Health, 30:68-77.
UN (2001) World urbanization prospects. The 2001 revision. Urban and rural areas. (POP/OB/WVP/Rev.2001/1). United Nations, New York.
UN (2000) United Nations Millennium Declaration. (Resolution 55/2, adopted at the Millennium Summit, September 6-8.) United Nations, New York.
VanDerslice J, Briscoe J (1993) All coliforms are not created equal: a comparison of the effects of water source and in-house water contamination on infantile diarrheal disease. Water Resources Research, 29:1983-1995.
VanDerslice J, Briscoe J (1995) Environmental interventions in developing countries: interactions and their implications. American Journal of Epidemiology, 141:135-144.

WHO (1998) Guidelines for safe recreational-water environments: coastal and fresh-waters. Draft for consultation. World Health Organization, Geneva.
WHO (2001) World health report 2001. World Health Organization, Geneva.
WHO/UNICEF/WSSCC (2000) Global water supply and sanitation assessment 2000 report. World Health Organization/United Nations Children's Fund Water Supply and Sanitation Collaborative Council, Geneva.
Zhaowu W, Kaiming B, Liping Y, Guifeng Y, Jinhua Z, Qili L (1993) Factors contributing to reinfection with Schistosomiasis japonica after treatment in the lake region of China. Acta Tropica, 54:83-88.

## Chapter I7

# URBAN AIR POLLUTION 

Aaron J. Cohen, H. Ross Anderson, Bart Ostro, Kiran Dev Pandey, Michal Krzyzanowski, Nino Künzli, Kersten Gutschmidt, C. Arden Pope III, Isabelle Romieu, Jonathan M. Samet and Kirk R. Smith

## Summary

Current scientific evidence, derived largely from studies in North America and Western Europe (NAWE), indicates that urban air pollution, ${ }^{1}$ which is derived largely from combustion sources, causes a spectrum of health effects ranging from eye irritation to death. Recent assessments suggest that the impacts on public health may be considerable. This evidence has increasingly been used by national and international agencies to inform environmental policies, and quantification of the impact of air pollution on public health has gradually become a critical component in policy discussions as governments weigh options for the control of pollution.

Quantifying the magnitude of these health impacts in cities worldwide, however, presents considerable challenges owing to the limited availability of information on both effects on health and on exposures to air pollution in many parts of the world. Man-made urban air pollution is a complex mixture with many toxic components. We have chosen to index this mixture in terms of particulate matter (PM), a component that has been linked consistently with serious health effects, and, importantly, levels of which can be estimated worldwide. Exposure to PM has been associated with a wide range of effects on health, but effects on mortality are arguably the most important, and are also most amenable to global assessment. Our estimates, therefore, consider only mortality. Currently, most epidemiological evidence and data on air quality that could be used for such estimates comes from developed countries. We have had, therefore, to make assumptions concerning factors such as the transferability of risk functions, exposure of the population and their underlying vulnerability to air pollution, while trying to ensure that these assumptions are transparent and that the uncertainty associated with them is assessed through appropriate sensitivity analyses.

In order to provide estimates for all 14 subregions, ${ }^{2}$ models developed by the World Bank were used to estimate ambient concentrations of inhalable particles (particulate matter with an aerodynamic diameter of $<10 \mu \mathrm{~m}, \mathrm{PM}_{10}$ ) for PM in 3211 national capitals and cities with populations of $>100000$ using economic, meteorological and demographic data and the available measurements. To allow the most appropriate epidemiological studies to be used for the estimation of the burden of disease, the estimates for $\mathrm{PM}_{10}$ were converted to estimates of fine particles (particulate matter with an aerodynamic diameter of $<2.5 \mu \mathrm{~m}$, $\mathrm{PM}_{2.5}$ ) using available information on geographic variation in the ratio of $\mathrm{PM}_{2.5}$ to $\mathrm{PM}_{10}$. Population-weighted subregional annual average concentrations of $\mathrm{PM}_{2.5}$ and $\mathrm{PM}_{10}$ were obtained using the population of the cities in the year 2000 .

Our estimates of the burden of disease were based on the contributions of three health outcomes: mortality from cardiopulmonary disease in adults, mortality from lung cancer, and mortality from acute respiratory infections (ARI) in children aged $0-4$ years. Numbers of attributable deaths and years of life lost (YLL) for adults and children (aged 0-4 years) were estimated using risk coefficients from a large cohort study of adults in the United States of America (Pope et al. 2002) and a metaanalytical summary of five time-series studies of mortality in children, respectively. Base-case estimates were calculated assuming that the risk of death increases linearly over a range of annual average concentrations of $\mathrm{PM}_{2.5}$, between a counterfactual (or referent) concentration of $7.5 \mu \mathrm{~g} / \mathrm{m}^{3}$ and a maximum of $50 \mu \mathrm{~g} / \mathrm{m}^{3}$.

The results indicate that the impact of urban air pollution on the burden of disease in the cities of the world is large, but this is likely to be an underestimate of the actual burden, on the basis of an assessment of sources of uncertainty. There is also considerable variation in our estimates among the 14 subregions, with the greatest burden occurring in the more polluted and rapidly growing cities of developing countries. We estimated that air pollution in urban areas worldwide, in terms of concentrations of PM, causes about $3 \%$ of mortality attributable to cardiopulmonary disease in adults, about $5 \%$ of mortality attributable to cancers of the trachea, bronchus and lung, and about $1 \%$ of mortality attributable to ARI in children. This amounts to about 0.80 million premature deaths ( $1.4 \%$ of the global total) and 6.4 million YLL $(0.7 \%$ of the global total). This burden occurs predominantly in developing countries, with $39 \%$ of attributable YLL occurring in WPR-B and $20 \%$ in SEAR-D. The highest proportions of the total burden occurred in WPRB and EUR-B, where urban air pollution caused $0.7-1.0 \%$ of the burden of disease.

We quantified the statistical uncertainty of our base-case estimates by estimating the joint uncertainty in the estimates of annual average concentration of PM and the estimates of the relative risks. Estimates worldwide and for most subregions vary by less than two-fold (50\%
uncertainty interval). Model uncertainty due to assumptions about the shape of the concentration-response function, the choice of counterfactual level for PM, and other factors were assessed in sensitivity analyses. For the most part, the worldwide estimates in each sensitivity case are within the $50 \%$ uncertainty intervals for the base-case estimates. The sensitivity analyses indicate that our base-case estimates were most sensitive to our choice of concentration-response function and theoretical level of minimum exposure.

## 1. Introduction

The potential for serious consequences of exposure to high levels of ambient air pollution was made clear in the mid-20th century, when cities in Europe and the United States experienced episodes of air pollution, such as the infamous London Fog of 1952 and Donora Smog of 1948, that resulted in large numbers of excess deaths and hospital admissions. Subsequent clean air legislation and other regulatory actions led to the reduction of ambient air pollution in many regions of the world, and particularly in the wealthy developed countries of North America and Europe. New epidemiological studies, however, conducted over the last decade, using sensitive designs and methods of analysis, have identified adverse health effects caused by combustion-derived air pollution even at the low ambient concentrations that now generally prevail in cities in North America and western Europe (Health Effects Institute 2001). At the same time, the populations of the rapidly expanding mega-cities of Asia, Africa and Latin America are increasingly exposed to levels of ambient combustion-related pollution that rival and often exceed the levels experienced in developed countries in the first half of the 20th century. Current scientific evidence, derived largely from studies in North America and western Europe, indicates that urban air pollution causes a spectrum of effects on health, ranging from eye irritation to death (Anonymous 1996a, 1996b). Recent assessments suggest that the impacts on public health may be considerable (Brunekreef 1997; Cifuentes et al. 2001; COMEAP 2001; Künzli et al. 2000; Ostro and Chestnut 1998). This evidence has increasingly been used by national and international agencies to inform environmental policies, and quantification of the impact of air pollution on public health has gradually become a critical component in policy discussions as governments weigh options for the control of pollution.

Quantifying the magnitude of the impact of air pollution in cities worldwide, however, presents considerable challenges owing to the limited availability of information on both effects on health and on exposures to air pollution in many parts of the world. Measurements of urban air pollution, when available, are available largely for a nonrepresentative sample of urban areas. Many areas of the world lack measurements of any kind, and these must then be estimated using
statistical models (see below). On the basis of these considerations, we defined the target population for this risk assessment exercise as the residents in the year 2000 of national capital cities and of cities worldwide with populations of $>100000$.

Man-made urban air pollution, which is derived largely from combustion processes, is a complex mixture containing many toxic components. We indexed this mixture in terms of PM, a component that has been consistently linked with serious effects on health, and, importantly, the levels of which can be estimated worldwide. Exposure to PM has been associated with a wide range of effects on health, but its effects on mortality are arguably the most important, and are also most amenable to global assessment. Our estimates, therefore, consider only mortality. Currently, most epidemiological evidence and data on air quality that could be used for such estimates come from developed countries. We have had, therefore, to make assumptions concerning factors such as the transferability of risk functions, exposure of the population and their underlying vulnerability to air pollution, while trying to ensure that these assumptions are transparent and that the uncertainty associated with them is assessed through appropriate sensitivity analyses.

The general framework for estimating the global burden of disease attributable to specific risk factors is described in chapters 1 and 25. Briefly, the approach involves estimating an attributable fraction(s) for each risk factor in each of the 14 subregions of the world. Estimating the attributable fraction for urban air pollution requires several steps. First, the exposure to urban air pollution of the population of each subregion must be estimated. Second, a theoretical minimum level of exposure must be specified. The attributable fraction quantifies the impact of exposure above this theoretical minimum level. Finally, deriving the attributable fraction requires the estimation of the gradient of risk between the theoretical minimum level and the estimated subregional exposure. These risk functions are derived from epidemiological studies for the purposes of estimating the global burden of disease. As discussed below, epidemiological studies generally estimate exposure to air pollution in terms of ambient concentrations, thus, we use the term "concentration-response" (rather than "exposure-response") to describe the risk function.

This chapter describes our approach to estimating the attributable fraction and presents our estimates of the attributable burden of disease caused by urban air pollution. First, we briefly review background information on exposure to air pollution and then describe our choice of the theoretical minimum level and the approach to estimating the exposure to PM of the populations of the world's cities. Next, we review the current information on the effects of air pollution on health and describe our approach to deriving the concentration-response function(s). Finally, we present and discuss our estimates of the attributable burden and their uncertainties.

## 2. Exposure to Urban air pollution from COMBUSTION SOURCES

Combustion of fossil fuels for transportation, power generation, and other human activities produces a complex mixture of pollutants comprising literally thousands of chemical constituents (Derwent 1999; Holman 1999). Exposure to such mixtures is a ubiquitous feature of urban life. The precise characteristics of the mixture in a given locale depend on the relative contributions of the different sources of pollution, such as vehicular traffic and power generation, and on the effects of the local geoclimatic factors. The relative contribution of different combustion sources is a function of economic, social and technological factors, but all mixtures contain certain primary gaseous pollutants, such as sulfur dioxide $\left(\mathrm{SO}_{2}\right)$, nitrogen oxides $\left(\mathrm{NO}_{\mathrm{x}}\right)$ and carbon monoxide (CO), that are emitted directly from combustion sources, as well as secondary pollutants, such as ozone $\left(\mathrm{O}_{3}\right)$, that are formed in the atmosphere from directly-emitted pollutants. The pollutant mixture also contains carcinogens such as benzo( $\alpha$ )pyrene, benzene and 1,3 -butadiene. When petrol contains lead $(\mathrm{Pb})$, as is still the case in many developing countries, this element is a common constituent of the pollution mix, assessed in a separate chapter in this volume (chapter 19).

All combustion processes produce particles, most of which are small enough to be inhaled into the lung either as primary emissions (such as diesel soot), or as secondary particles via atmospheric transformation (such as sulfate particles formed from the burning of fuel containing sulfur). Their concentrations (in micrograms per cubic metre, or $\mu \mathrm{g} / \mathrm{m}^{3}$ ) are generally measured as inhalable and fine particles, $\mathrm{PM}_{10}$ and $\mathrm{PM}_{2.5}$, respectively. ${ }^{3}$ However, the total suspended particle mass (TSP) is still the only particle measurement available in many developing countries (Krzyzanowski and Schwela 1999).

Pollution from the combustion of fossil fuels is largely emitted into the outdoor air, but human exposure occurs both indoors and outdoors (Ozkaynak 1999). An individual's exposure to ambient urban air pollution depends on the relative amounts of time spent indoors and outdoors, the proximity to sources of ambient air pollution, and on the indoor concentration of outdoor pollutants. The indoor concentrations depend on factors such as the circulation of the indoor air and the degree to which constituents of the outdoor combustion mixture penetrate and persist in the indoor environment. Studies conducted largely in Europe and North America have shown that the fine particles generated from combustion outdoors both effectively penetrate and persist in many indoor environments. Gases, such as sulfur dioxide and ozone, may penetrate the indoor environment, but generally do not persist because of their reactivity. In some rural areas of developing countries, indoor cooking on unvented coal- or biomass-burning stoves is the most significant exposure to pollution from combustion sources. The burden of disease caused by
such exposure is addressed in chapter 18. The actual dose delivered to the lung or other organs will further depend on the type of pollutant, the breathing pattern and physical characteristics of the individual that determine the extent and site of deposition.

Governments in many parts of the world monitor ambient concentrations of air pollution as part of regulatory programmes designed to protect public health and the environment (Grant et al. 1999). The most extensive monitoring systems are in the United States and western Europe, where regular monitoring of ambient air quality has been in place since the mid-1970s. The most frequently and routinely monitored air pollutants include sulfur dioxide ( $\mathrm{SO}_{2}$ ), nitrogen oxides ( $\mathrm{NO}_{\mathrm{x}}$, including NO and $\mathrm{NO}_{2}$ ), carbon monoxide ( CO ), ozone $\left(\mathrm{O}_{3}\right)$, lead $(\mathrm{Pb})$, black smoke (BS) or soot, and PM. National monitoring systems also exist in other parts of the world, but access to the data collected by these systems and international standardization of the monitoring methods are limited. The World Health Organization (WHO) Air Management Information System (AMIS) (WHO 2001c) collects the available information, but the reporting from many regions is poor, and for some regions there are no data in the WHO database. The various designs of the networks, differences in monitoring objectives and limited availability of the collected data for the outside users limit access to the information on population exposure in the greater proportion of the world's cities. In some parts of the world (e.g. in most of the countries of the former Soviet Union), the monitoring systems exist but do not provide the data necessary for assessment of the impact on health (Krzyzanowski and Schwela 1999). More details about the data available for this analysis are provided in further sections of this chapter.

These monitoring systems currently provide much of the data on exposure to urban air pollution that have been used in epidemiological research, although some studies establish their own monitoring networks when routinely-collected data are either unavailable or of poor quality, or to measure specific air pollution constituents, such as specific known carcinogens. Typically, monitoring sites are located in the city centre or throughout a given metropolitan area, in order to more accurately reflect the average residential exposure of the population. The data from monitors sited so as to measure emissions from specific sources, such as a local industry or heavy vehicular traffic, are frequently excluded from the data sets, as they may significantly deviate from the average levels of exposure experienced by the population.

Exposure estimates that rely exclusively on data from one or more stationary monitoring sites may provide inaccurate estimates of the short- and/or long-term average personal exposures of study populations (Navidi and Lurmann 1995; Zeger et al. 2000). The direction and magnitude of the errors that will be induced in estimates of the relative risk attributable to exposure to air pollution depend on the precision of the air quality monitoring data (or models used to generate the estimates of
the concentration of pollution), the applicability of one estimate to the entire target population and the correlation of the errors with the health outcome. Generally, such errors will be smaller for pollutants that tend to be uniformly distributed over large urban areas, and that penetrate efficiently indoors, both of these features being the case for fine PM produced by combustion. If the errors in the estimates of exposure are uncorrelated with the risk of the health outcome, then the estimates of relative risk attributable to air pollution will, in most cases, be too low (i.e. biased to the null) (Navidi and Lurmann 1995).

### 2.1 Definition of the air pollution metric for exposure variable

We selected $\mathrm{PM}_{10}$ and $\mathrm{PM}_{2.5}$ as the indicators of exposure to urban air pollution from combustion sources. As noted above, PM is a ubiquitous component of the mixtures emitted into, and formed in, the ambient environment by combustion processes, and indicates the presence of these mixtures in outdoor air. Most importantly, these measures of particulate air pollution have been used in many epidemiological studies from around the world, of both mortality and morbidity of air pollution, and so provide the best overall indicator of exposure for our purposes (see section 3). Although other components of ambient air pollution from combustion sources are associated with these and other effects on health (Anonymous 1996a, 1996b), particulate air pollution has been found to be consistently and independently related to the most serious effects of air pollution, including daily and longer-term average mortality (California Air Resources Board 2002; Health Effects Institute 2001; U.S. Environmental Protection Agency 2002; WHO 2000a, 2003). There is some evidence, although much less than that for PM, linking ozone to premature mortality, particularly during the summer months (Abbey et al. 1999; Health Effects Institute 2000b). However, despite recent progress in developing models to estimate tropospheric (ground-level) ozone on a global scale, it was not currently feasible to derive the subregional estimates that would have been required for this project. In many developing countries, exposure to lead in the ambient air may also be of great consequence, having effects on mortality perhaps via effects on blood pressure. The impacts of lead in outdoor air are dealt with in chapter 19.

PM has been linked to serious effects on health after both short-term exposure (days to weeks), and more prolonged exposure (years), although there remains some uncertainty as to the distribution of induction times with regard to mortality (see below). We chose the annual average concentration(s) of PM as the exposure metric(s) because it corresponds to the time-scales of a priori interest for estimates of attributable and avoidable burden in the Global Burden of Disease (GBD) project, and because it was used to estimate the effects of exposure to PM in the key epidemiological study that provides our estimates of the concentration-response function.

### 2.2 Estimation of annual average concentrations of particulate matter

## Air pollution measurements used in estimating annual AVERAGE CONCENTRATIONS

The availability of measurements of ambient concentrations of PM varies widely across the globe, making estimation of annual average concentrations a considerable challenge (Krzyzanowski and Schwela 1999). To estimate ambient PM concentrations for all 14 subregions, we used a model (Global Model of Ambient Particulates [GMAPS]) recently developed at the World Bank to estimate concentrations of $\mathrm{PM}_{10}$ in cities, on the basis of available measurements of PM at population-oriented monitoring sites (Pandey et al. forthcoming). The model incorporates information on factors such as fuel mix, level of economic development, demographics and weather, in order to predict ambient concentrations of $\mathrm{PM}_{10}$ in urban residential areas. These estimates of $\mathrm{PM}_{10}$ were converted to $\mathrm{PM}_{2.5}$ using available information on geographic variation in the ratio of $\mathrm{PM}_{2.5}$ to $\mathrm{PM}_{10}$. For each PM metric, the population-weighted subregional annual average was derived using the population of each city within each subregion in the year 2000.

The GMAPS model developed at the World Bank can be used to generate estimates of concentrations of $\mathrm{PM}_{10}$ in all world cities with populations of $>100000$, and in national capitals. The estimation model is based on available measurements of $\mathrm{PM}_{10}$ and TSP from population-oriented monitoring stations in cities worldwide for the period 1985 to 1999, retrieved in October 2001. In all cases, data from a monitoring site were included if and only if it was clearly identified as a residential or mixed residential site (see section 2.3 for definition). For instance, city averages reported for many Chinese cities (National Environmental Protection Agency of China 2000) were not included in the model estimation because the location of these sites could not be ascertained.

In principle, the monitoring data used for calculation of annual averages should be collected throughout the year, since seasonal patterns in the data are fairly common. More than $85 \%$ of cities in Europe and the United States collect measurements of PM throughout the year. The representativeness of the data for cities in other parts of the world could not be confirmed. In addition, in many countries where PM was measured throughout the year, it was only measured on every sixth day. The methods for measuring concentrations of PM also varied, both gravimetric and automatic methods (tapered element oscillating microbalance monitors [TEOMS] or beta gauge monitors) being included.

Most of the data on annual average ambient concentrations used in the model come from AMIS (WHO 2001c). This information is submitted to WHO by national environmental agencies and air quality control authorities, which perform these measurements using nationally
approved methods and standards of data quality. The data set contains the annual mean concentration of selected air pollutants, including PM, by monitoring site. Additional data, such as 95 th percentiles of daily means, are also available for some sites. Although WHO requests that all Member States provide data for compilation in the AMIS database, the reported data are still limited because many countries do not have air quality monitoring networks. Additionally, some countries with monitoring networks may not report the data because of poor data quality or limited ability to process and report the data.

The data from AMIS were supplemented with other sources of data on TSP and $\mathrm{PM}_{10}$ from monitoring sites. These included data for European cities collected by WHO/European Centre for Environment and Health (ECEH) for the Health Impact Assessment of Air Pollution (HIAAP) project in 1999 from both national and local environmental agencies (WHO 2001a), data for Canadian cities provided by Environment Canada (www.ec.gc.ca) and statistics Canada (http://www.statcan.ca/english/ads/cansimII/index.htm), and data for cities in the United States from the U.S. Environmental Protection Agency AIRS database (Aerometric Information Retrieval System 2001). Data for Chinese cities were also obtained from the Environmental Quality Reports from China (National Environmental Protection Agency of China 2000), and Mexican cities from the Instituto Nacional de Ecología (INE), SEMARNAP, Mexico (Instituto Nacional de Ecología 2000). Additional data were also obtained from the World Bank URBAIR studies of air pollution in Jakarta and Kathmandu (Grønskei et al. 1997a, 1997b). To limit undue influence of the data from cities in the United States, data used from the United States AIRS database were limited to the years 1996-1999.4

Measured annual average concentrations of $\mathrm{PM}_{10}$ and TSP data from monitoring sites were available for 512 unique locations in 304 cities in 55 countries over the period 1985-1999, and provided 1997 timelocation data points. For some sites and years, data on both TSP and $\mathrm{PM}_{10}$ were available, yielding a total of 2344 individual observations. ${ }^{5}$ The number of cities with measured data on PM from monitoring sites in each subregion and for each year by PM measure is shown in Table 17.1. A total of 304 cities reported either the annual average concentrations of $\mathrm{PM}_{10}$ or TSP for at least 1 year between 1985 and 1999. Of these, 51 cities reported both $\mathrm{PM}_{10}$ and TSP while 165 cities, mostly in North America and western Europe, reported $\mathrm{PM}_{10}$ only, and the remaining 88 cities reported data for TSP only.

Coverage of cities and populations with data from monitoring sites varies significantly across different subregions (Figure 17.1). For instance, data from monitoring were available for fewer than two cities for six of the subregions, AFR-D, AFR-E, AMR-D, EMR-B, EMR-D and SEAR-B. In contrast, data from monitoring sites were available for 218 cities in NAWE, of which 174 report data on $\mathrm{PM}_{10}$. The 304 world cities

Table I7.I Number of cities for which data on particulate matter are available from monitoring sites, by subregion, year and type of particulate matter

|  | PM ${ }_{10}$ or TSP | PM ${ }_{10}$ | TSP |
| :---: | :---: | :---: | :---: |
| Subregion |  |  |  |
| AFR-D | 2 | 0 | 2 |
| AFR-E | I | 0 | I |
| AMR-A | 123 | 118 | 25 |
| AMR-B | 19 | 12 | 12 |
| AMR-D | 2 | 2 | 2 |
| EMR-B | 0 | 0 | 0 |
| EMR-D | 1 | 1 | 0 |
| EUR-A | 95 | 56 | 43 |
| EUR-B | 22 | 7 | 17 |
| EUR-C | 7 | 1 | 7 |
| SEAR-B | 2 | 0 | 2 |
| SEAR-D | 11 | 11 | 10 |
| WPR-A | 5 | 5 | 4 |
| WPR-B | 14 | 3 | 14 |
| World | 304 | 216 | 139 |
| Year |  |  |  |
| 1985 | 28 | 7 | 28 |
| 1986 | 52 | 15 | 50 |
| 1987 | 53 | 9 | 52 |
| 1988 | 47 | 16 | 45 |
| 1989 | 53 | 17 | 51 |
| 1990 | 64 | 20 | 60 |
| 1991 | 63 | 30 | 60 |
| 1992 | 70 | 34 | 67 |
| 1993 | 73 | 41 | 68 |
| 1994 | 78 | 40 | 73 |
| 1995 | 73 | 42 | 68 |
| 1996 | 156 | 132 | 54 |
| 1997 | 144 | 127 | 40 |
| 1998 | 211 | 150 | 81 |
| 1999 | 166 | 143 | 40 |
| 1985-1998 | 267 | 187 | 127 |

Figure I7.I Cities from which data on exposure to $\mathrm{PM}_{10}$ or TSP during 1985-1999 are available from monitoring cites


Source: K. D. Pandey, Personal Communication.
with data from monitoring account for $9 \%$ of the total number of cities with a population of $>100000$ worldwide and have a combined population in the year 2000 of around 559 million, or about $28 \%$ of the global urban population (Table 17.2).

## Global Model of Ambient Particulates (GMAPS)

The GMAPS model econometrically estimates a fixed-effect model of the concentrations of urban ambient PM using the latest available data from WHO and other sources, as outlined above. The estimating Equation 1 focuses on the anthropogenic sources of pollution and the capacity of the natural environment to generate, disperse and dissipate pollutants. ${ }^{6}$ Its determinants include the scale and composition of economic activity, the energy mix, the strength of local regulation of pollution, and geographic and atmospheric conditions that affect the transport of pollutants.

$$
\begin{align*}
& C_{i j k t}=\sum_{k=1}^{K} \beta_{k} Z_{k}+\sum_{f=1}^{F} \beta_{E f} E_{f k t}+\sum_{g=1}^{G 2} \beta_{M g} M_{g i k}+\beta_{R} R_{k t}+\beta_{N} N_{j k t}+\beta_{D} D_{j k} \\
& \quad+\beta_{S c a l e} S_{c a l e_{i k t}}+\beta_{Y} Y_{k t}+\beta_{T} \text { Trend }_{i j k t}+\beta_{Y T} Y_{k t} \text { Trend }_{i j k t} \\
& \quad+\theta_{S} S_{i j k t}+\theta_{\text {scale } S_{i j k t} S_{c a l e}{ }_{j k t}+\theta_{Y} S_{i j k t} Y_{k t}+\theta_{T} S_{i j k t} \text { Trend }_{i j k t}}^{\quad+\theta_{Y T} S_{i j k t} Y_{k t} \text { Trend }_{i j k t}+\sum_{g=1}^{G 1} \theta_{M g} S_{i j k t} M_{g j k}+\varepsilon_{i j k t}} \tag{1}
\end{align*}
$$

Table 17.2 Cities for which measurements of particulate matter are available from monitoring sites, by subregion

| Subregion | Number of cities |  |  | Urban population (000s) in 2000 ${ }^{\text {a }}$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Cities in subregion | Cities with monitoring sites | \% with monitoring sites | Cities in subregion | Cities with monitoring sites | \% with monitoring sites |
| AFR-D | 107 | 2 | 2 | 66960 | 14914 | 22 |
| AFR-E | 105 | 1 | 1 | 68367 | 2388 | 3 |
| AMR-A | 267 | 123 | 46 | 232439 | 178240 | 77 |
| AMR-B | 399 | 19 | 5 | 217159 | 64121 | 30 |
| AMR-D | 47 | 2 | 4 | 29512 | 3291 | 11 |
| EMR-B | 89 | 0 | 0 | 56621 | 0 | 0 |
| EMR-D | 126 | 1 | 1 | 99397 | 8124 | 8 |
| EUR-A | 429 | 95 | 22 | 161808 | 79160 | 49 |
| EUR-B | 182 | 22 | 12 | 81756 | 21494 | 26 |
| EUR-C | 275 | 7 | 3 | 109178 | 7670 | 7 |
| SEAR-B | 68 | 2 | 3 | 53708 | 18793 | 35 |
| SEAR-D | 356 | 11 | 3 | 214175 | 67081 | 31 |
| WPR-A | 242 | 5 | 2 | 100079 | 11459 | 11 |
| WPR-B | 519 | 14 | 3 | 528318 | 81817 | 15 |
| World | 3211 | 304 | 9 | 2019479 | 558553 | 28 |
| a The total urban population is for 3211 cities with populations $>100000$ and national capitals. The total urban population in 2000, including cities of all sizes, is 2.8 billion. |  |  |  |  |  |  |

where
$C_{i j k t}=\log$ of concentration of PM in monitoring station $i$, city $j$, country $k$, at time $t$
$Z_{k}=$ binary variable for country $k$
$E_{f k t}=\log$ of per capita energy consumption of energy source type $f$ for country $k$ at time $\mathrm{t}(\mathrm{f}=1 \ldots F)$
$M_{g j k}=\log$ of meteorological/geographic factor $g$ for city $j$, country $k$ (factors $g=1 \ldots G 1$ affect $\mathrm{PM}_{10}$ concentration in a different way than TSP concentration. Factors $g=G 1+1 \ldots G 2$ do not make a distinction between $\mathrm{PM}_{10}$ and TSP)
$R_{k t}=\log$ of population density of country $k$ at time $t$
$N_{j k t}=\log$ of population of city $j$, country $k$, at time $t$
$D_{j k}=\log$ of local population density in the vicinity of city $j$ in country $k$

Scale $_{j k t}=\log$ of scale of economy (intensity of economic activity) for city $j$, country $k$ at time $t$
$\mathrm{Y}_{\mathrm{kt}}=\log$ of income per capita (1-year lagged 3-year moving average) of country $k$ at time $t$
$\operatorname{Trend}_{\mathrm{ijkt}}=$ time trend $(1985=1,1986=2, \ldots 1999=15)$
$\mathrm{S}_{\mathrm{ijkt}}=$ binary variable for PM type measured at monitoring station $i$, city $j$, country $k$, at time $t,\left(1=\mathrm{TSP}, 0=\mathrm{PM}_{10}\right)$, and
the $\beta_{S}$ and $\theta_{S}$ are the parameters that are estimated by the model.
Equation 1 jointly determines the concentrations of total suspended particulate matter (TSP) and inhalable particulates ( $\mathrm{PM}_{10}$ ) in residential areas. Most cities in developing countries only monitor TSP and not $\mathrm{PM}_{10}$. Adoption of the pooled specification permits use of all available data and provides better information about the concentrations of PM, especially for cities in developing countries. Limiting the estimation sample to $\mathrm{PM}_{10}$ observations is sensible only if knowledge of the concentration of TSP in a city makes no contribution to predicting $\mathrm{PM}_{10}$. Since $\mathrm{PM}_{10}$ comprises the smaller size particles within TSP, this assumption is clearly unreasonable. The pooled specification allows for separate estimation of concentrations of $\mathrm{PM}_{10}$ and TSP for each city by setting the binary variable, $S_{i j k t}$, equal to zero or one, as shown in Equations 2 and 3.

$$
\begin{gather*}
\log \left[P^{\prime 2} 0_{i j k t}\right]=\sum_{k=1}^{K} \beta_{k} Z_{k}+\sum_{f=1}^{F} \beta_{E f} E_{f k t}+\sum_{g=1}^{G 2} \beta_{M g} M_{g j k}+\beta_{R} R_{k t}+\beta_{N} N_{j k t}  \tag{2}\\
\quad+\beta_{D} D_{j k}+\beta_{S c a l e} \text { Scale }_{j k t}+\beta_{Y} Y_{k t}+\beta_{T} \text { Trend }_{i j k t}+\beta_{Y T} Y_{k t} \text { Trend }_{i j k t}
\end{gather*}
$$

$$
\begin{align*}
& \log \left[T S P_{i j k t}\right]=\sum_{k=1}^{K} \beta_{k} Z_{k}+\sum_{f=1}^{F} \beta_{E f} E_{f k t}+\sum_{g=1}^{G 2} \beta_{M g} M_{g j k}+\beta_{R} R_{k t}+\beta_{N} N_{j k t} \tag{3}
\end{align*}
$$

$$
\begin{aligned}
& +\sum_{g=1}^{G 1} \theta_{M g} M_{g j k}
\end{aligned}
$$

To reduce undue influence from extreme values, all of the continuous variables in the model were specified in log form and each exogenous variable in the estimation sample was truncated to the middle $98 \%$ range observed in the estimation sample.

The estimation Equation 1 includes country-specific binary variables, $Z_{k}$, to control for economic, social and natural factors that are not captured by the other explanatory variables. These include differences in the quality of the data on ambient concentration and in collection methods across countries, the degree of regulatory heterogeneity within a country, the relative importance of intercity transport, proximity of and pollution levels in neighbouring cities and the composition of economic activity. The country-specific binary variables measure the average concentration of PM in each country during the 15 -year period 1986-1999, controlling for variations within the country caused by factors accounted for in the remainder of the estimating Equation 1. In contrast, the rest of the estimation model (1) explains the marginal contribution of the included factors to deviations in the ambient concentration in the city from this average.

The primary determinants of the observed variations in the ambient concentrations of PM within a country in the estimation model are:

Energy consumption. The model includes six separate per capita energy consumption categories-coal, oil, natural gas, nuclear, hydroelectric, combustible renewables and wastes-that account for all energy consumed in each country for which data are available from the International Energy Agency's (IEA) Annual Energy Balance database (International Energy Agency 2001a, 2001b). The separate inclusion of each type of energy source accounts for differences in emission factors, variations in economic activity and intensity of fuel use across countries. In addition, the model also includes per capita consumption of petrol and diesel used in the transportation sector, also available from IEA's database, to capture additional detail about one of the most significant contributors to ambient concentrations of PM.

Meteorological and geographic factors. The model includes 22 atmospheric and geographic factors for each city to account for both the dissipative/dispersive capacity of the natural environment and natural sources of particulates, such as desert dust storms, forest fires and sea spray. These include a suite of 18 climatic variables representing the long-
term average climatic conditions related to local atmospheric conditions and transport of PM , consisting of the annual average (average of the monthly data) and seasonal changes (measured as the standard deviation of the monthly data) for the following nine factors: mean temperature, diurnal temperature, mean precipitation, barometric pressure, wind speed, percentage cloud cover and frequency of wet, sunny and frosty days (New et al. 1999). ${ }^{7}$ In addition, two meteorological variables related to energy demand (heating and cooling degree-days) are estimated for each city from the mean monthly temperature. Two topographical variables related to atmospheric transport-distance from the city centre to the nearest point on the coastline, calculated using the geographic information system (GIS), and elevation of the city, derived from a global digital elevation model (USGS 1996)—are also included in the model.

City and national population and national population density. These variables provide measures of the scale and intensity of the pollution problem in each city. The data on population comes from the Demographic yearbook published by the United Nations (UN 2001).

Local population density. The local population density in the vicinity of each city provides a measure of the intensity of pollution. It is estimated from the Gridded Population of the World (version 2), available from the Consortium for International Earth Science Information Network (CIESIN 2000). This data set provides the best available population data for about 120000 administrative units, converted to a regular grid of population counts at a resolution of about 5 km . The local population density in the vicinity of each city is the average population density for all grid cells within a $20-\mathrm{km}$ radius of the city centre.

Local intensity of economic activity. Most cities do not collect data on the amount or composition of economic activity. Instead, the local gross domestic product (GDP) per square kilometre computed as the product of the national per capita GDP and the local population density in the vicinity of each city is used as a proxy for the intensity of economic activity within each city (World Bank 2002).

National income per capita. This variable is used to capture the following national indicators: valuation of the quality of the environment, strength of environmental policy and regulation, the institutional capacity to enforce environmental policies, and the potential use of cleaner fuels along the fuel-use chain as countries develop. It is measured as a 1-year lag of the average of the previous 3 years (World Bank 2002).

Time trends. The model includes two time-trend variables (with $1985=1 \ldots 1999=15$ ) to allow for differential time trends for $\mathrm{PM}_{10}$ and TSP particulate pollution. Both of these variables are in turn interacted with lagged national per capita income to allow trends to vary across countries on the basis of differential valuation and improvements in environmental quality across countries as measured by the level of economic development. These trends measure changes in concentrations of PM
that are caused by factors not already captured in the model, such as technological changes, improvements in knowledge and structural shifts in the composition of economic activity. They do not represent the unconditional aggregate trends in concentrations of PM.

Binary variable to differentiate $P M_{10}$ and TSP. The model includes a binary variable indicating whether PM is measured as TSP or $\mathrm{PM}_{10}$. This binary variable is also allowed to interact in the model with other variables to allow for size class differences in the composition of particulates across cities and countries. It provides a better representation of intercity differences across the world, rather than assuming a uniform relationship across all cities. The $\log$ of the ratio of $\mathrm{PM}_{10}$ to TSP in each city can be estimated by subtracting Equation 3 from 2, as shown in Equation 4. The key determinants of this ratio are the scale of economic activity, differential trends across countries, level of economic development and strength of environmental policy, and the subset of meteorological variables that are directly related to particle size (annual mean and seasonal variations in wind speed, precipitation and frequency of wet days).

$$
\begin{align*}
& \log \left[P M 10_{i j k t}\right]-\log \left[\operatorname{TSP}_{i j k t}\right]=-\theta_{S}-\theta_{\text {Scale }} S_{c a l e_{j k t}}-\theta_{Y} Y_{k t} \\
& \quad-\theta_{T} \text { Trend }_{i j k t}-\theta_{Y T} Y_{k t} \text { Trend }_{i j k t}-\sum_{g=1}^{G 1} \theta_{M g} M_{g j k} \tag{4}
\end{align*}
$$

In order to facilitate predictions for countries not included in the estimation, a secondary model shown in Equation 5 is estimated to explain the average level of ambient PM concentration in each country.

$$
\begin{equation*}
\dot{\beta}_{k}=\sum_{f=1}^{F} \gamma_{E f} \bar{E}_{f k}+\gamma_{R} \overline{\bar{R}}_{k}+\gamma_{R} \bar{Y}_{k}+u_{k} \tag{5}
\end{equation*}
$$

where
$\dot{\beta}_{k}=$ country-specific binary variable coefficient estimated in Equation 1
$\bar{E}_{f k}=\log$ of average per capita energy consumption of energy type $f$ for country $k$ during 1985-1999 ( $f=1 \ldots F$ )
$\bar{R}_{k}=\log$ of average population density of country $k$ during 1985-1999
$\bar{Y}_{k}=\log$ of average national per capita income of country $k$ during 1985-1999 (1-year lagged average of previous 3 years)
This secondary model (5) explains the average level of pollution under reference conditions for a country, on the basis of the scale of the economy, the composition of economic activity as measured by the energy mix, and the strength of local pollution regulations and the institutional capacity for implementing these regulations.

### 2.3 Model outputs

The GMAPS model is designed to obtain the best city-level prediction of concentrations of PM for a wide range of cities on the basis of the limited amount of data from monitoring available, so it focuses on increasing the fit of the model. It is not designed to provide a causal model for ambient concentrations of air pollution. The estimation model (1) explains $88 \%$ of the variation in the observed data from monitoring, indicating a good fit (Pandey et al. forthcoming). The overall correlation between the measured and the predicted data is around 0.9 for both $\mathrm{PM}_{10}$ and TSP observations (see Table 17.3), and is $>0.80$ for all years and for both observations of $\mathrm{PM}_{10}$ and TSP, with the exception of $\mathrm{PM}_{10}$ in 1985. The correlation by subregion is smaller than that over time, ranging between 0.2 and 0.9 for subregions with more than 10 data points. The correlations for subregions with fewer data points are smaller than 0.2 and are less precisely estimated. A negative correlation for EUR-B is driven by a single erroneous observation for Bucharest, Romania, where the observed concentration of $\mathrm{PM}_{10}$ is higher than that of TSP. These results originated from two different monitoring locations; had the model been re-estimated without this particular $\mathrm{PM}_{10}$ observation, the correlation for the subregion would have been 0.32 .

Subregion- and PM type-specific scatter plots of model predictions compared to actual data also show a clustering of points around the solid line drawn at a $45^{\circ}$ angle, indicating that the actual values are close to the predicted values. As would be expected, the predicted values are less extreme than the actual values at both tails, owing to the truncation of all explanatory variables to the middle $98 \%$ range of the estimation sample. F-tests revealed that all of the eight aggregate factors in the model added significant explanatory power to the regression.

The secondary estimation model (5) explains $85 \%$ of the variation in the estimated average level of pollution in a country, indicating that this model provides a good fit. The explanatory power of the secondary model is not as robust to changes in the estimation sample owing to significant uncertainties in the estimated dependent variable.

Out-of-sample predictions were used to validate the model using both statistical and heuristic criteria. The model was re-estimated using subsamples of the data on the basis of different cut-off points for per capita income, to examine the appropriateness of extrapolating from a model primarily based on industrialized cities in North America and western Europe to cities in developing countries. The resulting estimates from the model were used to predict concentrations of $\mathrm{PM}_{10}$ in residential areas in the out-of-sample cities located in developing countries. A second set of estimates was also made comprising income-based subsamples using only the available data on $\mathrm{PM}_{10}$ from monitoring in residential sites. These validation estimates consistently showed that out-of-sample correlations were higher when data on TSP were included in the estimations. Furthermore, the out-of-sample correlations on aggregate ranged

Table 17.3 Correlation between observed concentrations of particulate matter at monitoring sites and predictions by subregion, year and type of particulate matter

|  | $P M_{10}$ or TSP |  | PM ${ }_{10}$ |  | TSP |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | No. of observations | Correlation | No. of observations | Correlation | No. of observations | Correlation |
| Subregion |  |  |  |  |  |  |
| AFR-D | 6 | 0.86 | 0 | NA | 6 | 0.86 |
| AFR-E | 2 | -1.00 | 0 | NA | 2 | -1.00 |
| AMR-A | 1273 | 0.79 | 938 | 0.59 | 335 | 0.67 |
| AMR-B | 361 | 0.80 | 215 | 0.52 | 146 | 0.75 |
| AMR-D | 34 | 0.88 | 18 | 0.31 | 16 | 0.72 |
| EMR-D | I | NA | 1 | NA | 0 | NA |
| EUR-A | 182 | 0.85 | 75 | 0.82 | 107 | 0.73 |
| EUR-B | 63 | 0.83 | 16 | -0.29 | 47 | 0.78 |
| EUR-C | 54 | 0.84 | 1 | NA | 53 | 0.83 |
| SEAR-B | 9 | 0.14 | 0 | NA | 9 | 0.14 |
| SEAR-D | 158 | 0.81 | 65 | 0.69 | 93 | 0.80 |
| WPR-A | 69 | 0.85 | 36 | 0.86 | 33 | 0.20 |
| WPR-B | 132 | 0.92 | 21 | 0.49 | 111 | 0.89 |
| World | 2344 | 0.94 | 1386 | 0.89 | 958 | 0.92 |
| Year |  |  |  |  |  |  |
| 1985 | 35 | 0.95 | 7 | 0.11 | 28 | 0.85 |
| 1986 | 68 | 0.93 | 17 | 0.81 | 51 | 0.92 |
| 1987 | 65 | 0.93 | 11 | 0.94 | 54 | 0.92 |
| 1988 | 70 | 0.93 | 20 | 0.91 | 50 | 0.92 |
| 1989 | 76 | 0.94 | 21 | 0.90 | 55 | 0.94 |
| 1990 | 91 | 0.94 | 24 | 0.90 | 67 | 0.94 |
| 1991 | 101 | 0.96 | 34 | 0.94 | 67 | 0.95 |
| 1992 | 116 | 0.94 | 38 | 0.95 | 78 | 0.93 |
| 1993 | 130 | 0.94 | 49 | 0.95 | 81 | 0.92 |
| 1994 | 138 | 0.94 | 46 | 0.94 | 92 | 0.93 |
| 1995 | 144 | 0.94 | 59 | 0.94 | 85 | 0.93 |
| 1996 | 330 | 0.92 | 259 | 0.88 | 71 | 0.90 |
| 1997 | 298 | 0.89 | 253 | 0.79 | 45 | 0.92 |
| 1998 | 377 | 0.88 | 289 | 0.84 | 88 | 0.84 |
| 1999 | 305 | 0.90 | 259 | 0.83 | 46 | 0.87 |
| All years except 1999 | 2039 | 0.94 | 1127 | 0.90 | 912 | 0.92 |
| NA Not app | cable. |  |  |  |  |  |

between 0.40 and 0.59 , based on the income cut-off used, and lend support to the modelling approach.

Since cities with data from monitoring are not representative of all cities and account for a small fraction of urban residents in developing countries, the following heuristic criteria were also used to evaluate the predictions of the model.

- Comparison of the relative variation of the predictions within countries and between countries relative to the actual data: The model predictions exhibited significant variations both across countries and across cities within a country. The predicted variations within a country were about $60 \%$ of those between countries and were comparable to the corresponding variations in the actual data.
- Number of cities for which predictions were outside the range of the estimation sample: The predictions for $\mathrm{PM}_{10}$ were within the range observed in the actual data. They continued to be within bounds when the same fractions of values are removed from the tails of the estimated and measured data.
- Magnitude of predictions outside the range observed in the estimation sample: Of the 304 cities with data from monitoring, concentrations of $\mathrm{PM}_{10}$ exceeded $200 \mu \mathrm{~g} / \mathrm{m}^{3}$ in three cities and concentrations of TSP exceeded $400 \mu \mathrm{~g} / \mathrm{m}^{3}$ in 10 cities. The predicted concentrations of $\mathrm{PM}_{10}$ exceeded $200 \mu \mathrm{~g} / \mathrm{m}^{3}$ in only four out of 3226 cities.
- Range of the $P M_{10}: T S P$ ratio: The $\mathrm{PM}_{10}$ :TSP ratio predicted by the model is between 0.24 and 0.98 and spans the middle $95 \%$ of the range observed in the actual data. The mean ratio predicted by the model is 0.49 ; the ratio for half of the cities lies between 0.39 and 0.56 .
- Comparison of the uncertainty in estimates for cities, relative to the amount of available information for neighbouring cities: Bootstrap error estimates of the prediction error for the city showed that the confidence intervals were wider for cities with no data from monitoring and are largest in the countries with no data from monitoring.

The robustness of the model was tested using alternative specifications of the model based on the goodness-of-fit of the model and the heuristic criteria outlined above. The alternative models that were considered were:

- Linear model. The linear model provides undue weight to the extreme values in the explanatory variables, resulting in predictions that are orders of magnitude larger than those for cities with data from monitoring.
- Explanatory variable truncation. The model was re-estimated with four different levels of truncation for the explanatory variables: no truncation, truncation to the actual range for the cities with data from monitoring, truncation to the middle $98 \%$ range of all explanatory variables for these cities, and truncation to the middle $90 \%$ range for these cities. Estimates based on the first two of these were sensitive to some of the extreme data points in the estimation sample, resulting in large variations in the predictions. Estimates from the last truncation were rejected because more than one quarter of the observations were truncated, leading to a poorer model fit.
- Energy consumption variables. The model was re-estimated with three alternative measures for the energy consumption variables: energy consumption per area, total per capita energy consumption and share of each energy type in the total energy mix, and the product of national per capita energy consumption by energy type and city population density. The specification per area resulted in predictions that were unstable and orders of magnitude larger than those observed in any city because of truncation of extreme values in countries with missing data on fuels. The second and third specifications resulted in poorer fits with over-predictions for $>100$ cities with values outside of the observed range of concentrations of $\mathrm{PM}_{10}$.
- Income. The model was estimated using income-squared and incomecubed terms to measure the impact of national per capita income. Higher order terms were unstable and resulted in predictions that were orders of magnitude larger than those observed in any city.
- $P M_{10}: T S P$ ratio. A number of different models were estimated from full interactivity of the binary variable $S_{i j k t}$ with all of the continuous variables, to no interactivity with the continuous variables. The full interactivity model was rejected because it predicted physically implausible $\mathrm{PM}_{10}$ :TSP ratios of 2 for a significant number of cities. The limited model with no interactivity was rejected because it over-predicted the results for many cities in the Middle East and North Africa that contain a larger fraction of wind-blown coarser particles. Other models were estimated that incrementally added groups of variables, such as energy type and the other climate variables. These were all rejected using the heuristic criteria outlined above.
- Location of monitoring sites. The sensitivity of the model predictions to the inclusion of mixed residential sites was examined by reestimating the model using only pure residential sites. ${ }^{8}$ Although estimates for some individual cities change in significant ways, predictions at the subregional level, and for most countries, are not statistically different as compared to when mixed sites are included.
- Inclusion of non-residential sites. A more inclusive model, which jointly estimates concentrations of PM in residential and nonresidential sites indicated that most model parameters were relatively stable and that the model predictions for subregional residential concentrations of $\mathrm{PM}_{10}$ were not significantly different for most subregions.
- Additional monitoring data. The sensitivity of the model was tested to the inclusion of additional monitoring data that became available between October 2001 and July 2002. The aggregate PM predictions were not statistically different for all subregions, except for EMR-B where concentrations of PM increase by nearly $50 \%$. This is primarily owing to the inclusion of data for Kuwait City, which is the only city for which data from monitoring sites are available in this subregion.
- Influential data point. The sensitivity of the model predictions to influential data points was examined using bootstrap error techniques. Variations in the predictions based on different subsamples of the data were used to estimate the degree of uncertainty in the model estimates.


## Estimating ambient concentrations of particulate matter in cities

The average subregional ambient concentrations used in this work are estimated from the city-level model predictions for 1999, the latest year for which all of the explanatory variables were available. The estimates of concentrations of $\mathrm{PM}_{10}$ in each city for 1999 were generated using a three-step approach. First, for all cities located in countries with at least one population-oriented monitoring site, the concentration of $\mathrm{PM}_{10}$ was estimated using the GMAPS model, as specified in Equation 1. The concentration of $\mathrm{PM}_{10}$ cannot be estimated using Equation 1 alone for cities in every country, because the average level of pollution in the country as measured by the country binary variable was not available for countries without monitors. Therefore, in the second step, the secondary Equation 5 was used to predict the country coefficient for countries without monitors. These predictions were combined with estimates from Equation 1 that explain variations around the average level to generate 1999 concentrations of $\mathrm{PM}_{10}$ for these cities in these countries.

Finally, for cities with actual data from monitoring, a best estimate of concentration of $\mathrm{PM}_{10}$ in the city in 1999 was generated by incorporating information on concentrations from previous years. Specifically, an average residual for each city was determined by comparing each yearspecific predicted value generated by the estimation model (1) with the actual monitored value for that year and city. This served to adjust the model predictions for local factors that are known but unmeasured in the model, such as the composition of local economic activity. Given the
large year-to-year variations in the available measured data even at the same monitoring station, correcting for the average residual provides a better representation of long-term average factors affecting concentrations of PM in a city than using the actual monitored value for the last year of data from monitoring alone.

## Estimating subregional ambient concentrations of PARTICULATE MATTER

To avoid extrapolating outside the sample frame, all exogenous variables were truncated to the range used in the estimation sample. When necessary, missing explanatory variables for the country were filled in with the median values for economically similar countries located in the same geographic area. For most subregions, data were available for at least $95 \%$ of the cities, accounting for at least $95 \%$ of the population in each subregion. In contrast, data on either fuel, GDP or gross national product (GNP) were missing for $20-30 \%$ of the cities, accounting for $20-30 \%$ of the population for each of the four subregions AFR-D, AFR-E, EMRB and EMR-D. In all, data on either fuel, GDP or GNP were completed in this way for 176 cities worldwide, accounting for $5 \%$ of the total world urban population.

The estimated annual average concentrations of $\mathrm{PM}_{10}$ in urban areas for world cities with populations of $>100000$ and national capitals are shown in Figure 17.2. Each circle on the map represents a city and is shaded according to the estimated concentrations of $\mathrm{PM}_{10}$ in that city. Standards currently in place in North America and western Europe lie

Figure 17.2 Estimated annual average concentrations of $\mathrm{PM}_{10}$ in cities with populations of $>100000$ and in national capitals


[^61]Table 17.4 Population-weighted predicted $\mathrm{PM}_{10}$ and TSP and percentiles of the distribution of estimated concentrations of $P M_{10}$

| Subregion | Predicted point estimate ( $\mu \mathrm{g} / \mathrm{m}^{3}$ ) |  |  | Percentiles of the distribution of estimated PM $10\left(\mu \mathrm{~g} / \mathrm{m}^{3}\right)$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | PM 10 | TSP | PM ${ }_{10} /$ TSP | 5\% | 25\% | 50\% | 75\% | 95\% |
| AFR-D | 68 | 195 | 0.350 | 32 | 43 | 61 | 72 | 84 |
| AFR-E | 39 | 104 | 0.372 | 30 | 35 | 39 | 44 | 58 |
| AMR-A | 25 | 39 | 0.642 | 24 | 25 | 25 | 25 | 25 |
| AMR-B | 37 | 79 | 0.470 | 35 | 36 | 38 | 39 | 42 |
| AMR-D | 51 | 146 | 0.349 | 37 | 43 | 48 | 53 | 58 |
| EMR-B | 40 | 118 | 0.341 | 23 | 30 | 34 | 39 | 48 |
| EMR-D | 110 | 276 | 0.397 | 62 | 78 | 99 | 110 | 127 |
| EUR-A | 26 | 49 | 0.531 | 25 | 26 | 26 | 27 | 28 |
| EUR-B | 48 | 118 | 0.406 | 41 | 44 | 46 | 48 | 50 |
| EUR-C | 31 | 90 | 0.340 | 21 | 25 | 29 | 33 | 38 |
| SEAR-B | 108 | 245 | 0.439 | 39 | 86 | 105 | 129 | 151 |
| SEAR-D | 84 | 206 | 0.409 | 73 | 80 | 84 | 88 | 96 |
| WPR-A | 32 | 50 | 0.646 | 27 | 30 | 32 | 34 | 37 |
| WPR-B | 89 | 221 | 0.403 | 73 | 83 | 89 | 96 | 104 |
| World | 60 | 144 | 0.417 | 51 | 56 | 58 | 62 | 65 |

between $30-60 \mu \mathrm{~g} / \mathrm{m}^{3}$. Therefore, we defined a middle group with concentrations of $\mathrm{PM}_{10}$ in the range of $30-60 \mu \mathrm{~g} / \mathrm{m}^{3}$. Cities with values that fell outside this range were sorted into two groups of cities with higher concentrations and two groups with lower concentrations (thus forming a total of five groups). Worldwide, about $30 \%$ of the urban population live in the less polluted cities while $40 \%$ live in the more polluted cities. The remaining $30 \%$ of people live in cities with concentrations of $\mathrm{PM}_{10}$ in the middle range. However, there are significant regional differences. More than $70 \%$ of the people living in NAWE live in cities with concentrations of less than $30 \mu \mathrm{~g} / \mathrm{m}^{3}$, meeting the most stringent standards. In contrast, more than $70 \%$ of the populations in SEAR-D, WPR-B, EMR-D and SEAR-D live in cities where concentrations exceed even the most lenient standards.

This difference can also be seen in the estimated population-weighted mean concentrations of $\mathrm{PM}_{10}$ for each subregion, which are presented in Table 17.4. These are computed from 1999 estimates of concentrations of $\mathrm{PM}_{10}$ in cities, using the populations of each city in 2000 as weights. We have not directly used data for cities with data from monitoring in computing the subregional averages, to avoid incorporating short-term transitional variations into our exposure estimates. ${ }^{9}$ The mean exposures in the most polluted subregions (EMR-D, SEAR-B, SEAR-D and

WPR-B) are about three times higher than those in the least polluted subregions (AMR-A and EUR-A). The table also shows predictions of population-weighted average concentrations of TSP and the size composition of PM for each subregion. Finer particles account for a larger share of PM in the highly industrialized countries of AMR-A and EURA compared to the other less industrialized subregions.

We quantified the uncertainty in our estimates of the subregionspecific mean concentrations of PM using a bootstrap technique. In this method, the model is re-estimated many times (200 trials) using a randomly repeated sample of the observations used in estimating the model. For each trial, city and population-weighted subregional predictions of PM are generated using the methods described above. The predictions from all trials are sorted from highest to lowest to obtain the percentile distribution of concentrations of $\mathrm{PM}_{10}$ for each subregion and are also shown in Table 17.4. The degree of certainty in the point estimates of concentration of $\mathrm{PM}_{10}$ for each subregion is directly related to the number of observations available from monitoring of PM. For example, the two subregions (AMR-A and EUR-A) with the most frequently monitored cities also have the smallest confidence intervals for $\mathrm{PM}_{10}$ values. In contrast, the five subregions with two or fewer cities that are monitored (AFR-E, AFR-D, EMR-B, EMR-D and SEAR-B) have larger confidence intervals than the other subregions. The width of the confidence intervals for these subregions depends on the geographic and climatic similarity of their cities with monitored data. For example, confidence intervals for AFR-E and EMR-B are about half of those for AFR-D and for EMR-D.

The estimates also show that substantial differences exist in the average concentration of PM within each subregion. The share of the urban population in cities with populations $>100000$ and in national capitals according to estimated concentrations of $\mathrm{PM}_{10}$ is given in Figure 17.3. All cities in AMR-A, EUR-A, EUR-C and WPR-A are estimated to have concentrations of $\mathrm{PM}_{10}$ of $<60 \mu \mathrm{~g} / \mathrm{m}^{3}$. In contrast, $95 \%$ of the urban population in SEAR-B and about $82 \%$ of the urban population in WPRB and EMR-D are exposed to $>60 \mu \mathrm{~g} / \mathrm{m}^{3} \mathrm{PM}_{10}$. We also estimate that a high proportion of the urban population in SEAR-D is exposed to high annual average concentrations of $\mathrm{PM}_{10}$.

Since some of the health outcomes are based on $\mathrm{PM}_{2.5}$, rather than $\mathrm{PM}_{10}$, concentrations for this pollutant had to be estimated. Cityspecific concentration of $\mathrm{PM}_{2.5}$ was estimated as a fixed proportion of $\mathrm{PM}_{10}$. Available measurements indicate that the ratio of $\mathrm{PM}_{2.5}$ to $\mathrm{PM}_{10}$ ranges from 0.5 to 0.8 in many urban areas in developed countries, (California Air Resources Board 2002; U.S. Environmental Protection Agency 2002). Limited evidence suggests that a similar ratio may exist in large cities in other subregions. For example, a recent study from China reports the $\mathrm{PM}_{2.5}: \mathrm{PM}_{10}$ ratio to be in the range of 0.51 to 0.72 in four urban locations (Quian et al. 2001). However, in areas impacted

Figure 17.3 Distribution of the urban population according to estimated concentrations of $\mathrm{PM}_{10}$ in cities with populations of $>100$ and in national capitals, by subregion

by more crustal particles (e.g. arid areas or cities with a significant number of unpaved roads or windy days), the ratios are likely to be much lower. These areas will have a greater proportion of $\mathrm{PM}_{10}$ in the coarse size range of $2.5-10 \mu \mathrm{~m}$. For example, evidence from the Coachella Valley (i.e. the Palm Springs area), an arid range of southern California suggests that the $\mathrm{PM}_{2.5}: \mathrm{PM}_{10}$ ratio is 0.35 (Ostro et al. 1999b). Therefore, we assumed a ratio of 0.5 for our base case and have examined the sensitivity of our results to this assumption. Specifically, for our sensitivity analysis, for cities in AMR-A, EUR-A, EUR-B, EURC and WPR-A (including the United States, Canada, all Europe, Japan, Singapore, Australia and New Zealand), a higher scaling factor of 0.65 was used, assuming relatively more combustion-related particles, while a lower scaling factor of 0.35 was used for cities in all other subregions.

Estimates of the annual average population-weighted concentration of $\mathrm{PM}_{2.5}$ for each subregion were calculated in a similar manner to that for $\mathrm{PM}_{10}$, using the estimated concentration of $\mathrm{PM}_{2.5}$ for the city in 1999 and the population for each city in 2000.

### 2.4 Choice of the theoretical-minimum-Risk exposure

Studies of mortality associated with both short- and long-term exposure to PM, discussed below, have been unable to detect a threshold below
which there is no effect of exposure. For most results presented below, we estimated the burden of disease with respect to a counterfactual concentration of $7.5 \mu \mathrm{~g} / \mathrm{m}^{3} \mathrm{PM}_{2.5}$ ( or $15 \mu \mathrm{~g} / \mathrm{m}^{3} \mathrm{PM}_{10}$ ). This value is close to the lowest concentration observed in the epidemiological study (Pope et al. 2002) from which we derived the concentration-response functions used for the majority of our estimates. This choice avoids extrapolating the concentration-response function(s) below the concentrations actually observed in the epidemiologic studies from which they were derived, although health benefits may well accrue from reductions below those concentrations.

We were aware, however, that for some cities the estimated (and observed) concentrations of PM are lower, e.g. in AMR-A (United States and Canada), and that achieving such concentrations more widely would be not only desirable, but also feasible in some settings (U.S. Environmental Protection Agency 2002). Moreover, previous impact estimates have been sensitive to where this value was set (Künzli et al. 2000). Therefore, we also conducted sensitivity analyses in which the theoretical minimum concentration was halved and doubled (see below).

## 3. Health effects of exposure to urban AIR POLLUTION

The past 10-15 years have seen a rapid increase in research on the health effects of air pollution, and it is now widely accepted that exposure to urban air pollution is associated with a broad range of acute and chronic health effects, ranging from minor physiological disturbances to death from respiratory and cardiovascular disease (Anonymous 1996a, 1996b; Figure 17.4). Recently, a committee of the American Thoracic Society identified effects on respiratory health associated with air pollution, which should be considered adverse, spanning outcomes from death from respiratory diseases to reduced quality of life, and included some irreversible changes in physiological function (American Thoracic Society 2000). In general, the frequency of occurrence of the health outcome is inversely related to its severity, with the consequence that assessing total health impact solely in terms of the most severe, but less common, outcomes, such as mortality, will underestimate the total health burden of air pollution (WHO 2001b).

A large body of epidemiological research, discussed in more detail below, provides evidence that exposure to air pollution is associated with increased mortality and morbidity. The respiratory and cardiovascular systems appear to be the most affected. A growing body of toxicological and clinical evidence currently offers some limited insight into the mechanisms through which exposure to air pollution may produce the effects on respiratory and cardiovascular outcomes observed in epidemiological studies (Anonymous 1996a, 1996b; Health Effects Institute 2002). These mechanisms may involve decrements in pulmonary func-

Figure I7.4 The relative frequencies of health events associated with exposure to air pollution

tion, effects on heart rate variability and inflammatory response. Longterm bioassays and other studies of toxicity provide evidence for the mutagenicity and/or carcinogenicity of some components of urban air pollution, such as emissions from diesel-powered vehicles (Cohen and Nikula 1999; Diesel Working Group 1995).

Air pollution may elicit both acute and chronic biological responses. Acute responses to air pollution in otherwise healthy persons may be confined to reversible physiological adaptations resulting from natural defence mechanisms (e.g. watery eyes, cough or a transient fall in lung function). Acute responses may, however, also increase the severity or duration of an already established respiratory infection or of diseases such as asthma or chronic obstructive lung disease that have already placed the individual in a vulnerable position, and increase the risk of hospital admission or even death. If such vulnerability were temporary, for example, a severe infection of the lower respiratory tract, the individual might have recovered and lived for some time, had it not been for the added factor of exposure to air pollution at the time the individual was most vulnerable because of the infection. On the other hand, if the individual had a terminal chronic condition, such as severe chronic obstructive pulmonary disease or chronic congestive heart failure, exposure to air pollution might advance death by only a short time, this being imminent in any case. There is limited epidemiological evidence to suggest that ambient air pollution may contribute to the development
of diseases such as chronic obstructive pulmonary disease, for which smoking and, in developing countries, indoor air pollution, are also risk factors (Abbey et al. 1999; Pope and Dockery 1999; Tager et al. 1998). Distributions of short-and long-term vulnerability, reflecting the prevalence of acute and chronic cardiorespiratory disease, may well differ across populations worldwide. This will have implications for the transferability of risk functions from studies in populations in NAWE to populations where differences in genetic factors, diet, tobacco smoking, extent of urbanization, distribution of wealth and other factors related to social class, have resulted in different patterns of disease.

For example, recent analyses of two cohorts in the United States (Krewski et al. 2000) showed clearly that the effects of long-term exposure to air pollution on mortality depend on attained educational level, with the largest relative effects observed among the least educated. Recent studies in developing countries have also reported such gradients in the relative risks of mortality (O'Neill et al. 2003). It is not clear what factor(s) might be responsible for these observations (e.g. aspects of occupation or diet), but it is reasonable to expect that they might vary across the globe. Differences in vulnerability to air pollution introduce a source of uncertainty in our estimates that can currently be only partially quantified.

Epidemiological evidence about exposure-response relationships is most directly applicable to the risk assessment of air pollution, because it comes from the direct observation of human populations under relevant conditions (Samet 1999). Epidemiological studies have limitations that are largely a result of their observational nature. These relate to the accurate measurement of exposure, definition of outcomes and interpretation of associations that are observed. Assessing the causality of such associations requires a process of scientific reasoning that considers all evidence, including that from experimental studies (WHO 2000b). While there remain many gaps in our knowledge about the explanations for epidemiological associations, they can provide the best evidence to guide action to reduce the exposure of the population to air pollution, and to undertake health impact assessments, provided the uncertainties are recognized.

The epidemiology of air pollution takes advantage of the fact that concentrations of urban air pollution, and thus human exposure, vary in both time and space. For the most part, current epidemiological research has focused on either one or the other dimension, but infrequently on both within the same population(s). Short-term temporal variation in concentrations of air pollution over days and weeks has been used to estimate effects on daily mortality and morbidity. Spatial variation in long-term average concentrations of air pollution has provided the basis for cross-sectional and cohort studies of long-term exposure.

### 3.1 Studies of short-term exposure

The effects of short-term exposure to air pollution have been extensively studied in time-series studies in which daily rates of health events (e.g. deaths or hospital admissions) in one or more locales are analysed in relation to contemporaneous series of daily concentrations of air pollutants, and other risk factors (e.g. weather) that vary over time periods of months or years. Regression techniques are used to estimate a coefficient that represents the relationship between exposure to pollution and the outcome variable. The usual method of regression models the logarithm of the outcome, and thus arrives at an estimate of the relative risk, a proportional change in the outcome per increment of ambient concentration. There has been a rapid increase in the number of these studies as computing and statistical techniques have improved and as data on outcomes and air pollution have become more extensive and easily accessible from routine sources. It is a strength of these studies that individual cofactors, such as smoking, nutrition, behaviour, genetic factors, etc., are unlikely confounders because they are not generally associated, on a day-to-day basis, with the daily concentration of air pollution. Studies of time series have found associations between concentrations of PM in the air and a large range of outcomes. These have been reviewed by Pope and Dockery (1999) and include daily mortality (all causes, respiratory, cardiovascular), hospitalization for respiratory diseases (all causes, chronic obstructive pulmonary disease, asthma, pneumonia) and for cardiovascular diseases (acute myocardial infarction, congestive cardiac failure). Since this review, associations have also been reported for primary health care visits for disease of the lower respiratory tract, and diseases of the upper respiratory tract of both infective and allergic origin (Hajat et al. 2001, 2002). However, recent methodological studies and re-examination of earlier work indicate that the magnitude of the estimates of relative risk from time-series studies of daily mortality depends on the approach used to model both the temporal pattern of exposure (Braga et al. 2001) and potential confounders that vary with time, such as season and weather (Health Effects Institute 2003).

The acute effects of air pollution have also been studied longitudinally in panel studies, which can provide evidence of physiological effects at an individual level. Small groups, or panels, of individuals are followed over short time intervals, and health outcomes, exposure to air pollution and potential confounders are ascertained for each subject on one or more occasions. Panel studies have generally reported associations of exposure to urban air pollution with increased prevalence of symptoms involving the upper and lower respiratory tract, and increased rates of asthma attacks and medication use. Associations with short-term reduction in lung function and the prevalence of cough symptoms have been reported in studies in the United States (Pope and Dockery 1999), but are not consistently supported by studies in Europe (Roemer et al. 1999).

## Time-Series studies in adults across the world

Studies of time series concerning daily mortality and, to a lesser extent, daily hospital admissions, have been conducted in cities throughout the world. A recent meta-analysis summarized the evidence from $>100$ studies of daily mortality (Stieb et al. 2003). In addition, large studies have now been conducted using uniform methods for assembling and analysing data from multiple cities: APHEA 2 (Air Pollution and Health: A European Approach) (Katsouyanni et al. 1996, 2001) and NMMAPS (National Mortality and Morbidity Air Pollution Study) (Health Effects Institute 2000a, 2000b) in the United States. These multi-city studies have confirmed the findings of earlier studies of individual cities in finding positive associations between daily mortality and hospital admissions and concentrations of PM, and have also attempted to explain the heterogeneity among cities in the relative risks associated with exposure to air pollution. For example, in the APHEA 2 Study, it was found that the effects of PM on mortality were modified by mean concentrations of nitrogen dioxide (Katsouyanni et al. 2001), and in the NMMAPS Study, daily mortality was modified by the long-term average concentrations of $\mathrm{PM}_{10}$. Levy et al. (2000) reported that the effects of $\mathrm{PM}_{10}$ were greater in cities where $\mathrm{PM}_{2.5}$ comprised a higher proportion of $\mathrm{PM}_{10}$.

Most studies of time series are from countries in NAWE, where air pollution is low and decreasing and populations are characterized by western lifestyles and patterns of disease. To examine the epidemiological evidence for other non-NAWE countries, we searched a database of studies of time series and panel studies compiled at St George's Hospital Medical School, for which researchers had systematically ascertained, reviewed, and abstracted results from studies published in the peerreviewed scientific literature (WHO 2003). All studies meeting prespecified quality criteria related to adequacy of confounder control, and which provided estimates of the concentration-response relationship and its statistical precision were included. We classified them by the subregion in which the study was performed and tabulated the results for six outcomes: all-cause mortality, respiratory mortality, cardiovascular mortality, infant and child mortality, reduction in peak expiratory flow rate and cough symptom.

The distribution of time-series and panel studies by outcome and subregion is shown in Tables 17.5(a) and (b). Up to mid-November 2001, the number of studies from AMR-A (71) and EUR-A (75) far exceeded the total for the remaining 12 subregions (42). The next largest contributor was AMR-B (Central and South America), with 18 studies. The table shows the numbers of panel studies presenting usable numerical estimates. Only 14 studies from non-NAWE subregions were identified, compared with 64 from the NAWE subregions. Some of the non-NAWE countries had lifestyles and patterns of disease similar to those in NAWE-those in Australasia, for example.

Table 17.5 Distribution of studies by outcome status and subregion
(a) Selected time-series studies

| Subregion | Outcome status |  |  |  |  | Total (from subregion) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Cause of mortality |  |  | Hospital admissions/ emergency room visits | Other time-series studies |  |
|  | $\begin{gathered} \text { All } \\ \text { causes } \end{gathered}$ | Respiratory disease | Cardiovascular disease |  |  |  |
| AFR-D | 0 | 0 | 0 | 0 | 0 | 0 |
| AFR-E | 0 | 0 | 0 | 0 | 0 | 0 |
| AMR-A | 34 | 11 | 13 | 39 | 2 | 71 |
| AMR-B | 10 | 9 | 6 | 5 | 0 | 18 |
| AMR-D | 0 | 0 | 0 | 0 | 0 | 0 |
| EMR-B | 0 | 0 | 0 | 0 | 0 | 0 |
| EMR-D | 0 | 0 | 0 | 0 | 0 | 0 |
| EUR-A | 44 | 28 | 24 | 28 | 3 | 75 |
| EUR-B | 4 | 3 | 4 | 0 | 0 | 5 |
| EUR-C | 0 | 0 | 0 | 0 | 0 | 0 |
| SEAR-B | 1 | 1 | 1 | 0 | 0 | 1 |
| SEAR-D | 0 | 0 | 0 | 0 | 0 | 0 |
| WPR-A | 4 | 4 | 4 | 2 | 0 | 6 |
| WPR-B | 8 | 3 | 4 | 4 | 0 | 12 |
| World (from outcome status) |  | 113 |  | 78 | 5 | 187 |

(b) Selected panel studies

|  | Outcome status for valid studies |  | Total for selected studies <br> (from subregion) |
| :--- | :---: | :---: | :---: |
| Subregion | 0 | 0 | 0 |
| AFR-D function | Symptoms | 0 |  |
| AFR-E | 0 | 0 | 28 |
| AMR-A | 19 | 17 | 3 |
| AMR-B | 3 | 3 | 0 |
| AMR-D | 0 | 0 | 0 |
| EMR-B | 0 | 0 | 0 |
| EMR-D | 0 | 0 | 38 |
| EUR-A | 32 | 31 | 3 |
| EUR-B | 3 | 2 | 2 |
| EUR-C | 2 | 1 | 0 |
| SEAR-B | 0 | 0 | 1 |
| SEAR-D | 0 | 1 | 4 |
| WPR-A | 4 | 1 | 1 |
| WPR-B | 0 | 1 | 79 |
| World (from | 62 | 57 |  |
| outcome status) |  |  | 0 |

Figures 17.5 and 17.6 show the results for daily mortality and PM, by subregion, for adults and children, respectively. These estimates were scaled to $\mathrm{PM}_{10}\left(\mathrm{PM}_{2.5}=0.6 \times \mathrm{PM}_{10}, \mathrm{BS}=0.5 \times \mathrm{PM}_{10}, \mathrm{TSP}=2 \times \mathrm{PM}_{10}\right)$. These scaling factors were decided after examining a number of co-located measures, but are likely to be variable across the individual cities. Estimates of random effects and fixed effects are shown because there is heterogeneity.

The pooled estimates are shown in Table 17.6. Most of the studies of mortality showed relative risks of $>1.0$ (i.e. a change of $>0.0 \%$ ), with lower $95 \%$ confidence intervals also $>1.0$. However, the studies showing the largest confidence intervals also tended to have the largest effects, this indicating the possibility of publication bias. There was considerable heterogeneity in the estimates. For this reason, the summary estimate for random effects is more appropriate because it takes into account the greater uncertainty. The estimate for random effects was increased and statistically significant for mortality from all causes, respiratory and cardiovascular diseases. It is remarkable that for daily mortality, the pooled estimate for the non-NAWE subregions of $0.5 \%$ increase in daily mortality per $10 \mu \mathrm{~g} / \mathrm{m}^{3}$ increase in PM was very similar to the estimates produced by the APHEA 2 and NMMAPS studies of 0.6 and 0.5 , respectively. A recent meta-analysis of 109 published studies from around the world reports similar estimates (Stieb et al. 2003).

These results indicate that daily mortality is positively associated with short-term exposure to urban air pollution at time-scales in the order of days, in all subregions where this association has been measured. They also suggest that the relative effect of exposure may also be of similar magnitude in different parts of the world.

## Air pollution and reproductive and child health

Six time-series studies of daily mortality report associations between particulate pollution and adverse effects in children, and all of them are from non-NAWE countries. Their estimates are shown in Figure 17.6. Four were from São Paulo (Conceiao et al. 2001; Gouveia and Fletcher 2000; Pereira et al. 1998; Saldiva et al. 1994), one from Mexico City (Loomis et al. 1999) and one from Bangkok (Ostro et al. 1999a). The Bangkok study was of $\mathrm{PM}_{10}$ and daily mortality from all causes in children aged $<6$ years. The study conducted in Mexico City evaluated the impact of daily changes in concentrations of $\mathrm{PM}_{2.5}$ and total mortality in children aged $<1$ year. Three studies in São Paulo (Conceiao et al. 2001; Gouveia and Fletcher 2000; Saldiva et al. 1994), conducted during different periods of time, all reported an association between PM and mortality from respiratory disease in children aged $<5$ years. The study conducted in São Paulo by Pereira et al. (1998) investigated the association of exposure to urban air pollution with intrauterine mortality. Some of the relative risks reported in these studies were $>1.0$, but only the estimate from Mexico City was statistically significant at the $95 \%$ level.
Figure 17.5 Percentage change in mean daily number of non-accidental deaths, in adults, per $10 \mu \mathrm{~g} / \mathrm{m}^{3}$ increase in $\mathrm{PM}_{10}$,

Note: Scaling factors used: $\mathrm{PM}_{2.5}=0.6 \times \mathrm{PM}_{10}, \mathrm{BS}=0.5 \times \mathrm{PM}_{10}, \mathrm{TSP}=2 \times \mathrm{PM}_{10}$.
Figure 17.6 Percentage change in mean daily number of non-accidental deaths, in children, per $10 \mu \mathrm{~g} / \mathrm{m}^{3}$ increase in $\mathrm{PM}_{10}$, by study

Note: Scaling factors used: $\mathrm{PM}_{25}=0.6 \times \mathrm{PM}_{10}, \mathrm{BS}=0.5 \times \mathrm{PM}_{10}, \mathrm{TSP}=2 \times \mathrm{PM}_{10}$.
Pooled estimates of daily mortality from all causes and concentrations of $\mathrm{PM}_{10}$, from time-series studies, excluding
North America and western Europe

|  | All-cause mortality (\% change) | Mortality from respiratory disease (\% change) | Mortality from cardiovascular disease (\% change) | Infant and child mortality (\% change) | Lung function (regression coefficients) | Symptoms (odds ratios) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| No. of studies | 29 | 18 | 18 | $5^{\text {a }}$ | 11 | $10^{\text {b }}$ |
| Heterogeneity | $\mathrm{Q}=\underset{(\mathrm{P}<0.00 \mathrm{I})}{\mathrm{I} 21.62 \text { on } 28 \mathrm{df}}$ | $\begin{aligned} \mathrm{Q}= & 68.86 \text { on } 17 \mathrm{df} \\ & (\mathrm{P}<0.00 \mathrm{I}) \end{aligned}$ | $\begin{gathered} \mathrm{Q}=51.40 \text { on } 17 \mathrm{df} \\ (\mathrm{P}<0.00 \mathrm{I}) \end{gathered}$ | $\begin{gathered} \mathrm{Q}=7.72 \text { on } 4 \mathrm{df} \\ (\mathrm{P}=0.102) \end{gathered}$ | $\begin{aligned} & Q=16.16 \text { on } 10 \mathrm{df} \\ &(P=0.095) \end{aligned}$ | $\begin{gathered} \mathrm{Q}=26.68 \text { on } 9 \mathrm{df} \\ (\mathrm{P}=0.002) \end{gathered}$ |
| RE (95\% CI) | 0.5 (0.4-0.6) | 0.6 (0.2-1.1) | 0.3 (0.1-0.5) | $1.0(-0.9-3.1)$ | -0.024 (-0.146-0.098) | 1.006 (0.990-1.022) |
| FE (95\% CI) | 0.4 (0.3-0.4) | 0.4 (0.2-0.6) | 0.1 (0.1-0.2) | 1.0 (-0.1-2.1) | -0.007 (-0.095-0.081) | 1.006 (0.998-1.013) |
| NMMAPS study 20 USA cities (Health Effects Institute 2000b) | 0.46 (0.27-0.65) | - | - | - | - | - |
| APHEA 2 study 29 European cities (Katsouyanni et al. 2001) | 0.62 (0.4-0.8) | - | - | - | - | - |

Key: RE, random-effects estimate; FE, fixed-effects estimate; $\mathrm{Q}, \chi^{2}$ statistical test for heterogeneity; df, degrees of freedom.
I study (Pereira et al. 1998) not included in meta-analysis due to uncommon outcome.
I study (Awasthi et al. 1996) not included in meta-analysis due to uncommon particle measurement.

Infant mortality from respiratory disease and other adverse perinatal events, such as low birth weight and malformations, have also been associated with more prolonged exposure to air pollution (Bobak and Leon 1999; Wilhelm and Ritz 2003; Woodruff et al. 1997). Woodruff et al. followed a large birth cohort in the United States for one year and estimated the relative risk of mortality associated with residential exposure to $\mathrm{PM}_{10}$ in the first two post-natal months, conditional on a variety of potential confounders, including maternal smoking. They reported an increase in total mortality of $4 \%$ per $10 \mu \mathrm{~g} / \mathrm{m}^{3}$, and $20 \%$ for mortality from respiratory causes. Two studies recently evaluated changes in infant mortality associated with reductions in industrial emissions caused by a recession and mandated reductions in pollution resulting from the United States Clean Air Act Amendments of 1970 (Chay and Greenstone 1999, 2001). Using county-level data, the authors estimated that 4-8 infant deaths per 100000 live births were prevented for each $1 \mu \mathrm{~g} / \mathrm{m}^{3}$ reduction in TSP.

## Studies of acute morbidity

Far fewer studies have been conducted of the association of exposure to urban air pollution with acute morbidity, especially in non-NAWE countries. There were insufficient studies in any one outcome group to allow formal meta-analysis of non-NAWE studies, but most reports suggested a positive association with urban air pollution, consistent with that observed in NAWE countries, especially for hospital admissions (Atkinson et al. 2001; Burnett et al. 1999; Health Effects Institute 2000b). A recent study compared directly the effects of air pollution on hospital admissions in China, Hong Kong Special Administrative Region (Hong Kong SAR) and London. Similar associations were observed for $\mathrm{PM}_{10}$ and gaseous pollutants and hospital admissions for ischaemic heart disease in both locations, and the associations were strongest during seasons of low humidity in both cities, but no association with admissions for cardiac disease was observed in Hong Kong SAR (Wong et al. 2002).

### 3.2 Studies of long-term exposure

COHORT STUDIES OF MORTALITY FROM CHRONIC RESPIRATORY AND CARDIOVASCULAR DISEASE
Cohort studies take advantage of spatial heterogeneity in concentrations of air pollution to compare the incidence of disease and death in populations exposed in the long term to differing levels of pollution. By following large populations for a number of years, cohort studies provide estimates of both attributable numbers of deaths and, more importantly, average reductions in life span attributable to air pollution.

The evidence from cohort studies of populations in Europe and the United States indicates that long-term exposure to urban air pollution
is associated with an increase in total and cardiopulmonary mortality in adults (Dockery et al. 1993; Hoek et al. 2002; Lipfert et al. 2003; McDonnell et al. 2000; Pope et al. 2002). In each of these studies, the effects of potential confounders such as cigarette smoking, occupation and prior medical history were adjusted for in regression analyses. Most studies find the strongest and most consistent associations with exposure to PM , and $\mathrm{PM}_{2.5}$ appears to be more closely associated with mortality than $\mathrm{PM}_{10}$ or TSP (Dockery et al. 1993; Pope et al. 2002). The recently published results of a study conducted in the Netherlands confirm the impacts of long-term exposure to air pollution, and in particular that related to road traffic, on mortality (Hoek et al. 2002).

Unfortunately, the cohort studies provide little information on when exposure to air pollution acts to increase the risk of mortality (i.e. the induction time for mortality attributable to exposure to long-term exposure to air pollution), making it difficult to estimate when the effects of reduction of air pollution might be observed.

Comparable cohort studies have not yet been carried out in developing countries. However, the imposition of restrictions on the sulfur content of fuel for power generation and transportation in Hong Kong SAR, instituted over short time intervals in 1990, provided opportunities for researchers to measure directly the impact of reducing air pollution on long-term average mortality (Hedley et al. 2002). Hedley et al. (2002) documented both changes in ambient air quality subsequent to the imposition of the restrictions, and declines in long-term average rates of mortality from cardiovascular and respiratory diseases associated with those changes. Comparison of changes in mortality in more and less polluted areas of Hong Kong SAR provided limited ability to account for secular changes in other risk factors for mortality that could have produced the observed decrease in mortality following the change in the sulfur content of fuel. A similar study was also published recently by Clancy et al. (2002), who measured decreased long-term average mortality in Dublin after the banning of the sale of bituminous coal in Dublin in 1990.

## The American Cancer Society study

The American Cancer Society (ACS) study (Pope et al. 2002) in the United States is by far the largest cohort study of air pollution and longterm average mortality to date. The ACS study of air pollution and mortality is based in the ACS Cancer Prevention II Study, an on-going prospective cohort of approximately 1.2 million adults from all 50 states (Calle et al. 2002). Friends and neighbours recruited cohort members on behalf of the ACS. Participants were enrolled in 1982, when they were aged $\geq 30$ years, and their mortality has been ascertained through to 1998. Data on a wide range of risk factors for cancer and other chronic diseases were obtained from each participant. The ACS study links the data for approximately 500000 cohort members with data on air pol-
lution from metropolitan areas throughout the United States. The first study of air pollution and mortality in this cohort (Pope et al. 1995) was based on follow-up through to 1990 . That study reported increases in mortality from cardiopulmonary disease for $19.9 \mu \mathrm{~g} / \mathrm{m}^{3}$ fine particulate sulfate (relative risk of $1.26,95 \%$ CI 1.16-1.37), and from lung cancer (relative risk of $1.36,95 \%$ CI 1.11-1.66). These findings were subsequently corroborated in an independent re-analysis (Krewski et al. 2000). A more recent analysis of this cohort extended follow-up through to 1998, and ascertained 40706 deaths from cardiopulmonary disease, and 10749 from lung cancer. Data were analysed using Cox proportional hazards regression models that incorporated both random effects and non-parametric spatial smoothing to adjust for unmeasured factors correlated spatially with air pollution and mortality across the United States. The models also adjusted for age, sex, race, education, marital status, body mass, diet, alcohol consumption, occupational exposures and the duration and intensity of cigarette smoking, all measured via questionnaire at enrolment.

Concentrations of ambient air pollution had, in general, declined across the United States between 1982 and 1998. Measurements of ambient concentrations of fine particulate air pollution $\left(\mathrm{PM}_{2.5}\right)$ in the cities where subjects resided at enrolment were available for periods both briefly preceding enrolment (1979-83) and immediately after follow-up (1999-2000). In separate regression analyses, cohort members were assigned estimates of exposure corresponding to their city-of-residencespecific value for each of those periods, as well as for the average value across the two periods. For a change of $10 \mu \mathrm{~g} / \mathrm{m}^{3}$ in the ambient concentration of $\mathrm{PM}_{2.5}$, the smallest relative increases were observed for the mean concentration of the time period 1979-1983. This estimate was based on data from 61 cities, with a mean concentration of $\mathrm{PM}_{2.5}$ of $21.1 \mu \mathrm{~g} / \mathrm{m}^{3}$, and a range of $10-30 \mu \mathrm{~g} / \mathrm{m}^{3}$. The relative risks for a $10 \mu \mathrm{~g} / \mathrm{m}^{3}$ change in the concentration of ambient $\mathrm{PM}_{2.5}$ were larger when exposure was specified as the average of the ambient concentrations of the two time periods. This may be explained by the fact that the estimates from the earliest periods are more subject to random (and nondifferential) error. However, it also suggests that more recent exposures may be exerting the strongest effects on mortality, an interpretation also offered in the recent re-analysis of the earlier follow-up of the ACS cohort (Krewski et al. 2000). Unfortunately, it was not possible to derive individual time-varying estimates of exposure from the available data (e.g. detailed residence histories were unavailable), precluding direct evaluation of the induction time for mortality attributable to exposure to air pollution.

## Long-term exposure and the incidence of chronic disease

Little evidence is available concerning exposure to air pollution and the incidence of chronic cardiovascular or respiratory disease. One study in
the United States reported an association of long-term exposure to $\mathrm{PM}_{10}$ with the incidence of self-reported physician-diagnosed chronic bronchitis (Abbey et al. 1999). A recent case-control study reported an association between short-term exposure and the incidence of non-fatal myocardial infarction (Peters et al. 2001). Cross-sectional studies have found associations with reduced lung function and increased respiratory symptoms in both adults and children, which might in part represent chronic disease as the result of long-term exposure. Several recent crosssectional studies in large Chinese cities have reported increased prevalence of respiratory symptoms in adults (Qian et al. 2001; Zhang et al. 1999) and elementary school children (Qian et al. 2000; Zhang et al. 2002) exposed to urban air pollution. A cross-sectional study in Delhi observed reductions in pulmonary function in residents of highly polluted areas, but little evidence of increased prevalence of symptoms (Chhabra et al. 2001).

## AIR POLLUTION AND LUNG CANCER

Epidemiological studies over the last 40 years have observed that general ambient air pollution, chiefly composed of the by-products of the incomplete combustion of fossil fuels, is associated with small relative increases in the incidence of lung cancer. The evidence derives from studies of trends in the incidence of lung cancer, studies of occupational groups, comparisons of urban and rural populations, and case-control and cohort studies using diverse exposure metrics. Recent prospective cohort and case-control studies which have controlled for the effects of cigarette smoking, occupation and other risk factors have consistently observed small increases in the relative risk of lung cancer in relation to exposure to particulate air pollution (Abbey et al. 1999; Dockery et al. 1993; Krewski et al. 2000; Pope et al. 2002; Samet and Cohen 1999). A recent Swedish case-control study reported that excess lung cancer was related specifically to exposure to mobile sources of air pollution, with the largest effects observed for exposure occurring 20 years prior to diagnosis (Nyberg et al. 2000).

### 3.3 Choice of outcomes and hazards

## Studies used for hazard estimates

The use of results from time series to estimate the disease burden attributable to urban air pollution is problematic for various reasons. First, data on rates of occurrence, such as hospital admissions, primary health care consultations or asthma exacerbation are not collected in many countries. Thus there is no baseline upon which to develop an estimate of health impact. Mortality is an exception in that data are available from death registration or indirect demographic methods in all subregions.

The application of the time-series concentration-response functions to the assessment of mortality impact, however, is limited. Specifically, it is not possible to use the results of studies of time series to estimate YLL in adults. This is because time-series studies of daily mortality do not in themselves allow the estimation of lost life time, but rather only allow estimation of the number of deaths that have been brought forward by an unspecified amount of time. Recent research has made clear that the time-series estimates reflect deaths that may have been brought forward by as much as several months, rather than simply advancing the time of death in frail people by a few days (Schwartz 2000; Zeger et al. 1999). The design of the time-series study of daily mortality, which requires the control of long-term variation in air pollution, precludes estimation of greater losses (Künzli et al. 2001; Leksell and Rabl 2001; McMichael et al. 1998). Thus, the time-series studies only provide an estimate of the daily number of deaths brought-forward.

Cohort studies include not only people whose deaths were advanced by recent exposure to air pollution, but also those who died from chronic disease caused by long-term exposure (COMEAP 1998; Künzli et al. 2001), thus they provide a more comprehensive estimate of the effects on mortality. Furthermore, because their relative risks can be applied to population life tables, the effects of air pollution on life span can be estimated (Brunekreef 1997; COMEAP 2001; Hurley et al. 2000; Sommer et al. 2000)

The situation may be different for children. In developing countries, the major causes of death, such as acute respiratory disease, are very likely to result from a severe acute infection, which represents a brief window of vulnerability. If the child survives, they might be expected to fully recover and enjoy a full life expectancy. If we assume that death was not otherwise imminent, then these deaths, on average, represent the loss of considerable life years. Under such an assumption, one could use time-series estimates to estimate YLL in children aged $0-4$ years.

In making these estimates, several further considerations should be kept in mind. The first is that the effects of cumulative exposure over several weeks are several times greater than those obtained by using a single day lag and thus underestimate the impacts on health. The second is that other air pollutants in the mixture may be exerting additional effects, may interact with particles or may be confounding the associations of particles. Ozone, for example, is also toxic and while its effects tend to be independent of PM, it also seems to modify the effect of particles on number of hospital admissions (Atkinson et al. 2001).

## Definition and specification of health outcomes

We estimated the burden of disease imposed by mortality from cardiopulmonary disease and lung cancer in adults, and from ARI in children aged 0-4 years. We made this choice despite the fact, discussed above, that other serious health effects of air pollution are well-
documented, and that still others appear from more limited evidence to be of potential concern. Our decision to focus on mortality outcomes was made on the basis of the following considerations.

- Strength of evidence. A large body of research from many parts of the world indicates that ambient air pollution causes increased daily mortality from cardiovascular and respiratory disease. There appear to be comparable effects in the cities of developed and developing countries, on the basis of the limited evidence available. Although no cohort studies of mortality have as yet been conducted in developing countries, the possibility that associations comparable to those observed in the United States would be observed is strengthened by the results of the studies of daily mortality, and the limited results from studies of morbidity.
- Consistent definition of the end-point. Mortality per se is a welldefined event that is registered in most countries. For this reason, epidemiologists have frequently measured the effect of air pollution on total mortality from all natural causes, ascertained from death certificates or other sources of vital statistics. Other outcomes, such as bronchitis and the symptom of wheeze, are subject to very large variations in severity, and without such qualification their impact on health is difficult to assess. The definitions of other possible health outcomes, such as restricted activity days, use of primary health care services, diagnoses and school absences, are likely to vary with national culture and among health care systems.

The cause of death is more problematic because it is not certified medically in many countries, and there are considerable differences between and within countries in terms of diagnostic practice. Nevertheless, we propose to base our estimates primarily on cardiopulmonary, rather than total, mortality. There is strong evidence from both time-series and cohort studies that ambient air pollution specifically increases the risk of death from these causes. Moreover, variations in the relative contribution of non-cardiopulmonary mortality among countries could increase the error in the burden assessment, particularly in countries with lower cardiopulmonary death rates, potentially leading to overestimates of impact. Since considerable cross-coding is likely, we have chosen to use the combined cardiopulmonary group consisting of GBD infectious and chronic respiratory diseases and selected cardiovascular outcomes for adults. In children, death from cardiovascular diseases is rare and the pulmonary group is adequate.

- Availability of baseline occurrence rates. Data on age-specific mortality are collected or estimated using consistent methods for all subregions. This is not the case for some important potential measures of morbidity, such as the frequency of asthma attacks, or measures of
the utilization of health care services outside of Europe and North America.
- Importance of the end-point in terms of health impact. Although the impacts of air pollution on other health end-points must certainly contribute to the global burden of disease, mortality, quantified in terms of either numbers of deaths or reduced survival time, currently plays the most prominent role in impact assessments. We chose these three specific mortality outcomes (mortality from cardiopulmonary causes and lung cancer in adults aged $\geq 30$ years, and mortality from ARI in children aged $0-4$ years) because they allow us to estimate YLL, as discussed above.
- Feasibility within the time constraints of the current work. Given additional time and resources, future efforts might possibly consider, for example, using the evidence from the International Study of Asthma and Allergies in Childhood (ISSAC) (Anonymous 1998), which has data on prevalence from 60 countries as a baseline upon which to estimate the effect of particles on the exacerbation of asthma. Another possibility might be the effect on hospital admissions or primary health care visits.


### 3.4 Developing the concentration-response functions

We derived concentration-response functions for three end-points to produce the estimates of global burden of disease reported in this work: mortality from cardiopulmonary causes and lung cancer in adults aged $\geq 30$ years, and mortality from ARI in children aged $0-4$ years. As discussed earlier, we made no estimates of the impacts of PM on the incidence of disease, so the disability-adjusted life years (DALYs) quantify only YLL.

We assumed a log-linear risk model which leads to the following formula for the relative risk $(R R)$ in a population whose exposure is estimated by an average concentration of pollution $C$ relative to the reference level $C_{0}$ :

$$
\begin{equation*}
R R=\exp \left[\beta\left(C-C_{0}\right)\right] ; \tag{6}
\end{equation*}
$$

where, $C_{0}$, the reference, or theoretical minimum level of exposure, is defined as above and $\beta$ is the estimated effect of PM on the health outcome of interest. We calculated a subregion-specific relative risk for each of the 14 subregions using a population-weighted mean of concentrations for all cities in the subregion calculated as follows:

The subregion-specific relative risk for outcome $i$ in subregion $k$ related to $\mathrm{PM}_{2.5}, R R_{2.5 \mathrm{jk}}$, is:

$$
\begin{equation*}
R R_{2.5 i k}=\exp \left[\beta_{2.5 i} \times\left(C k_{2.5}-7.5\right)\right] \tag{7}
\end{equation*}
$$

where $C_{k 2.5}$ is the subregion-specific population-weighted mean concentration of $\mathrm{PM}_{2.5}$ (calculated from the estimated concentration of $\mathrm{PM}_{10}$ in Table 17.4, as described above), and $\beta_{2.5 i}$ is the slope of the concentra-tion-response function for $\mathrm{PM}_{2.5}$ (Table 17.7).

The subregion-specific relative risk for outcomes quantified in terms of $\mathrm{PM}_{10}, R R_{10 i k}$, is:

Table 17.7 Estimates of relative risk of mortality, coefficients of concentration-response functions and study types

| Health outcome | Data source |  | Relative risk per $10 \mu \mathrm{~g} / \mathrm{m}^{3}$ (95\% Cl), <br> from data source | Concentrationresponse slope ${ }^{\text {a }}$ per $\mu g / m^{3}$ (standard error) |
| :---: | :---: | :---: | :---: | :---: |
| Mortality from cardiopulmonary disease-adults | ACS study (Pope et al. 2002) | PM 2.5 | 1.059 (1.015-1.105) | Linear ${ }^{\text {b }} 79-83$ <br> 0.00575 (0.002I60) <br> Linear average ${ }^{c}$ <br> 0.008933 (0.002907) <br> Log-linear ${ }^{\text {d }} 79-83$ <br> 0.11605 (0.044790) <br> Log-linear average <br> 0.155148 (0.050460) |
| Mortality from lung cancer | ACS study (Pope et al. 2002) | PM 2.5 | 1.082 (1.011-1.158) | Linear 79-83 <br> 0.00789 (0.003447) <br> Linear average <br> 0.012673 (0.00426) <br> Log-linear 79-83 <br> 0.17114 (0.071968) <br> Log-linear average <br> 0.232179 (0.074770) |
| Mortality from acute respiratory infectionchildren aged $0-4$ years | St George's Hospital meta-analysis of five time-series studies of daily mortality | PM ${ }_{10}$ | 1.010 (0.99I-I.03I) | 0.0010 (0.0010) |
| Deaths-brought-forwardall ages | St George's Hospital meta-analysis of 165 time-series studies of daily mortality | PM ${ }_{10}$ | 1.006 (1.005-1.007) | 0.0006 (0.00005) |

[^62]\[

$$
\begin{equation*}
R R_{10 i k}=\exp \left[\beta_{10 i} \times\left(C_{k 10}-15\right)\right] \tag{8}
\end{equation*}
$$

\]

where $C_{k 10}$ is the subregion-specific population-weighted mean concentration of $\mathrm{PM}_{10}$ (Table 17.4), and $\beta_{10 i}$ is the slope of the concentra-tion-response function for $\mathrm{PM}_{10}$. The city-specific concentrations of $\mathrm{PM}_{10}$ were truncated at 15 and $100 \mu \mathrm{~g} / \mathrm{m}^{3}$ for calculation of subregionspecific population-weighted mean concentrations of $\mathrm{PM}_{10}$.

## Mortality from cardiopulmonary disease and lung cancer IN ADULTS

We used the results of the ACS study of urban air pollution and mortality (Pope et al. 2002) to estimate attributable deaths and YLL from cardiopulmonary diseases and lung cancer in adults aged $\geq 30$ years. In our base-case analyses we used the estimates of the concentrationresponse functions based on the ambient concentrations in 1979-1983, which correspond to increases of $5.9 \%$ and $8.2 \%$ in mortality from cardiopulmonary disease and lung cancer, respectively, for each $10 \mu \mathrm{~g} / \mathrm{m}^{3}$ change in the ambient concentration of $\mathrm{PM}_{2.5}$ (Table 17.7).

Deaths from cardiopulmonary disease and lung cancer in the ACS cohort were defined as persons whose underlying cause of death was coded according to the International Statistical Classification of Diseases, ninth revision, on their death certificates as ICD-9 401-440 and $460-519$, and 162 , respectively. When calculating the attributable fraction, ACS concentration-response functions for cardiopulmonary disease defined in this way were applied to baseline cause-specific rates of mortality in the GBD project. For lung cancer, this corresponded exactly to the definition used in the ACS study. For cardiopulmonary deaths, the GBD groupings of cause of death $(39,40,106-109,111)$ did not include several ICD codes (406-409, 415-417, 423-424, 426-429, 440) that were included in the ACS definition. These codes represent diverse cardiac diseases, including conduction disorders, cardiac dysrhythmias, heart failure and ill-defined cardiac causes. Together they comprise approximately $18 \%$ of all cardiopulmonary deaths in the ACS study (R. Burnett, personal communication).

The ACS study estimated concentration-response functions for $\mathrm{PM}_{2.5}$ over a range that extends from annual average concentrations of $\mathrm{PM}_{2.5}$ of about $5-30 \mu \mathrm{~g} / \mathrm{m}^{3}$ (Pope et al. 2002). The shape of the concentra-tion-response function for fine particulate air pollution outside that range is currently unknown, as noted above, and estimated annual average concentrations of $\mathrm{PM}_{2.5}$ in some subregions are outside that range (Table 17.1). In our base-case analyses we limited the risk of mortality in any city to be no greater than that attained at a concentration of $\mathrm{PM}_{2.5}$ of $50 \mu \mathrm{~g} / \mathrm{m}^{3}$ (Figure 17.7). Thus, for cities with estimated annual average concentrations of $>50 \mu \mathrm{~g} / \mathrm{m}^{3}$, we assigned a maximum concentration, or $\mathrm{C}_{\mathrm{m}}$, equal to $50 \mu \mathrm{~g} / \mathrm{m}^{3}$, regardless of their actual estimated

Figure 17.7 Alternative concentration-response curves for mortality from cardiopulmonary disease, using different scenarios

concentration. This means that the excess risk is constrained to be no greater than that associated with an annual average concentration of $50 \mu \mathrm{~g} / \mathrm{m}^{3}$, regardless of the actual estimated annual average concentration. The counterfactual or theoretical minimum concentration was set at $7.5 \mu \mathrm{~g} / \mathrm{m}^{3}$, as discussed above.

We set the maximum city-specific concentration of $\mathrm{PM}_{2.5}, \mathrm{C}_{\mathrm{m}}$, at $50 \mu \mathrm{~g} / \mathrm{m}^{3}$ to avoid producing unrealistically large estimates of mortality in the most extremely polluted subregions under a linear exposure model. With $\mathrm{C}_{\mathrm{m}}=50$, the subregion-specific attributable fraction was restricted to no more than approximately $25 \%$ of the burden of a given health outcome, while not greatly exceeding the maximum observed annual average concentration in the ACS study. We also examined alternative values for the shape of the concentration-response function for $\mathrm{PM}_{2.5}$ and mortality from cardiopulmonary disease and lung cancer in sensitivity analyses, as described below.

## MORTALITY FROM ACUTE RESPIRATORY INFECTIONS IN CHILDREN AGED 0-4 YEARS

In view of the importance of mortality from ARI among children in developing countries and the suggestion from available evidence of an association with air pollution (Romieu et al. 2002), we decided to make
a summary estimate on the basis of these studies, in spite of their heterogeneity in outcomes and age groups.

To estimate the relationship between exposure to PM and mortality from ARI among children aged $0-4$ years, we computed a summary estimate from the five published time-series studies discussed above (Table 17.6 and Figure 17.6). One study (Pereira et al. 1998) was excluded because the outcome, intrauterine mortality, was clearly unrelated to ARI. The five remaining studies were summarized as a weighted average of the estimates from individual studies (scaled to $\mathrm{PM}_{10}$, as discussed above) with the weights determined by the inverse of the reported variance in the concentration-response function. We estimate that a $10 \mu \mathrm{~g} / \mathrm{m}^{3}$ increase in ambient concentrations of $\mathrm{PM}_{10}$ results in a $1.0 \% ~(95 \% \mathrm{CI}-$ $0.9 \%-3.1 \%$ ) increase in daily mortality from ARI in children aged $0-4$ years (Table 17.7).

When calculating the attributable fraction, this concentrationresponse function was applied to GBD baseline cause-specific rates of mortality from acute respiratory infection (GBD code 38) that includes ICD-9 codes 460-466, 480-487 and 381-382.

## Numbers of deaths from all natural causes caused by short-TERM EXPOSURE TO URBAN AIR POLLUTION IN ADULTS

We also calculated an estimate of the numbers of deaths from all natural (non-injury) causes attributable to short-term exposure to urban air pollution using an estimate of concentration-response derived from international literature on air pollution and daily mortality. This estimate was not included in the total attributable deaths and disease burden because of the conceptual issue in quantifying the effects of short-term exposure discussed above.

Using the St George's Hospital Medical School database, described above, we identified 165 time-series studies of $\mathrm{PM}_{10}$ and daily mortality from all causes at all ages in all languages and countries, up to the end of July 2001. As we were concerned about the possibility of publication bias, we compared the summary estimates from 54 individual studies, a subset of the literature which would be expected to be susceptible to publication bias, with those of the two multi-city studies (Health Effects Institute 2000b; Katsouyanni et al. 2001), which selected cities from a pre-specified sampling frame, used uniform methods of analysis, and published all results. The pooled estimate for the cities of the combined APHEA and NMMAPS studies $(n=111)$ was 1.005 ( $95 \%$ CI 1.004-1.006) with no evidence of publication bias in the funnel plot or on statistical testing. For the 54 studies of individual cities, graphical analysis showed some evidence of publication bias but when formally tested, this was weak $(P=0.12)$. The pooled estimate was $1.007(95 \%$ CI 1.006-1.008) but when adjusted for publication bias using Trim and Fill analysis, it was reduced to 1.006 ( $95 \%$ CI $1.004-1.007$ ). We then examined the results for all 165 studies with results for $\mathrm{PM}_{10}$. There was
no evidence of publication bias on inspection of the funnel plot or on formal testing with Begg's or Eggar's tests (Begg and Mazumdar 1994). We calculated pooled estimates weighted according to the inverse of the variance of the individual study. Random effects models were used, as all showed significant heterogeneity. The pooled estimate was 1.006 ( $95 \%$ CI $1.005-1.007$ ) (Table 17.7). This concentration-response function was applied to all GBD baseline cause-specific rates of mortality except GBD code 148 (injuries). ${ }^{10}$

## 4. Uncertainty estimates: statistical variability and sensitivity analyses

The total uncertainty in our estimates of the burden derives from both the statistical (sampling) variability of the parameter estimates in the models we chose to quantify disease burden, and our uncertainty with regard to those choices vs plausible alternatives, i.e. the form of our models (Morgan and Henrion 1998). We therefore quantified the statistical uncertainty of our estimates in terms of a combined, or propagated, uncertainty estimate, and used sensitivity analyses to quantify model uncertainty.

### 4.1 Statistical (sampling) variability

Our estimate of statistical uncertainty combined the sampling errors from two sources to derive an uncertainty distribution:

- sampling variability in the original concentration-response estimates from the ACS and time-series studies quantified in terms of their standard errors (Table 17.5); and
- sampling variability in the estimates of subregional concentration of PM from the exposure estimation model in terms of estimates of bootstrapped standard error described above (Table 17.4).
When presenting our results we show either the complete uncertainty distribution or the intervals between the 25 th and 75 th and/or 2.5th and 97.5 th percentiles of that distribution, i.e. $50 \%$ and $95 \%$ uncertainty intervals.


### 4.2 Sensitivity analyses

We used sensitivity analyses, described below, to quantify the uncertainty in our base-case estimates, in which the burden of disease was estimated by applying the ACS concentration-response function over the range of 7.5 to $50 \mu \mathrm{~g} / \mathrm{m}^{3}$, as discussed above.

- Cases 2-4: Shape of the concentration-response function for $P M_{2.5}$ and mortality attributable to cardiopulmonary disease and lung cancer. We explored three alternatives to the base-case scenario for extrapolating the ACS concentration-response function beyond
$30 \mu \mathrm{~g} / \mathrm{m}^{3}$, the highest annual average concentration observed in the ACS study (Figure 17.7).
- Case 2: No incremental increase in excess mortality above $30 \mu \mathrm{~g} / \mathrm{m}^{3}$ $P M_{2.5}$. Under this scenario, when calculating the population-weighted subregional average concentration of $\mathrm{PM}_{2.5}$ we give the city-specific $C_{m}$ a value of $30 \mu \mathrm{~g} / \mathrm{m}^{3}$, regardless of the estimated concentration, rather than the base-case concentration of $50 \mu \mathrm{~g} / \mathrm{m}^{3}$. This means that the excess risk is constrained to be no greater than that associated with an annual average concentration of $30 \mu \mathrm{~g} / \mathrm{m}^{3}$, regardless of the actual estimated annual average concentration. The counterfactual or theoretical minimum concentration was set at $7.5 \mu \mathrm{~g} / \mathrm{m}^{3}$. We considered this the estimator that would produce the smallest (i.e. most scientifically conservative) estimate of the impact of mortality consistent with the use of the ACS concentration-response function.
- Case 3: Excess mortality increases linearly above $30 \mu \mathrm{~g} / \mathrm{m}^{3} P M_{2.5}$. Under this scenario, the city-specific concentration of $\mathrm{PM}_{2.5}$ takes its actual estimated value when calculating the population-weighted subregional averages, i.e. in contrast to the base-case and case 2 scenarios. The counterfactual or theoretical minimum concentration was set at $7.5 \mu \mathrm{~g} / \mathrm{m}^{3}$. We considered this the estimator that would produce the largest estimate of mortality impact consistent with the use of the ACS concentration-response function.
- Case 4: Excess mortality increases with the log of concentration of $P M_{2.5}$ across the entire range. Under this scenario, the city-specific concentration of $\mathrm{PM}_{2.5}$ takes the $\log$ of its actual estimated value when calculating the population-weighted subregional averages. Therefore, in contrast to case 3, the slope of the concentration-response function is constrained to decrease at higher concentrations. The counterfactual or theoretical minimum concentration was set at $7.5 \mu \mathrm{~g} / \mathrm{m}^{3}$.

We included this estimator, proposed by an external reviewer after we had made our initial estimates (R. Burnett, personal communication), because it seemed a reasonable way to characterize an excess risk that we believed may: (i) increase directly with ambient levels over the entire range of annual average concentrations that we estimated for the world's cities, but (ii) be smaller at higher ambient concentrations, as has been observed for daily mortality in time-series studies (Daniels et al. 2000; Schwartz et al. 2002).

- Cases 5 and 6: Choice of ACS concentration-response function. In the base-case analyses, we used the ACS coefficients that were based on exposure of the cohort in 1979-1983. These arguably best represented the effects of long-term past exposure that some researchers assume are responsible for the increased mortality attributable to air pollution in that cohort through to 1998. There is, however, consid-
erable uncertainty regarding the timing of exposure with regard to risk of mortality (Krewski et al. 2000), so we calculated alternative estimates using the reported ACS coefficients based on the average of past (1979-1983) and recent (1999-2000) annual average concentrations using both a linear (case 5) and log-linear (case 6) extrapolation (Table 17.7).
- Case 7: Change $P M_{2.5}: P M_{10}$ ratio. In the base-case analyses, we assumed a $\mathrm{PM}_{2.5}: \mathrm{PM}_{10}$ ratio of 0.50 , although higher and lower ratios have been observed in a number of locations, as discussed above. We examined the sensitivity of our base-case analyses by assigning cities in AMR-A, EUR-A, EUR-B, EUR-C and WPR-A a higher scaling factor of 0.65 , while assigning a lower scaling factor of 0.35 to cities in all other subregions.
- Cases 8 and 9: Choice of counterfactual or theoretical minimum concentration. We evaluated the sensitivity of the base-case estimates to two different choices of counterfactual $\mathrm{PM}_{2.5}$ concentration: $3 \mu \mathrm{~g} / \mathrm{m}^{3}$ (case 8) and $15 \mu \mathrm{~g} / \mathrm{m}^{3}$ (case 9). The former is close to the minimum background level of $\mathrm{PM}_{2.5}$ observed in the United States, and the latter is the annual concentration of $\mathrm{PM}_{2.5}$ proposed by the United States National Ambient Air Quality Standard (NAAQS) (U.S. Environmental Protection Agency 2002). Each value was substituted for the base-case concentration of $7.5 \mu \mathrm{~g} / \mathrm{m}^{3}$ in Equation 7 above, when calculating population-weighted subregional relative risks.


## 5. Results

### 5.1 Base-case estimates

We estimated that exposure to particulate air pollution caused approximately 800000 excess deaths and 6.4 million DALYs (consisting only of years of life lost to premature mortality) in the year 2000 worldwide as a result of cardiopulmonary disease, lung cancer and ARI in children aged $0-4$ years, combined (Table 17.8).

The worldwide estimates of attributable deaths and YLL from cardiopulmonary and lung cancer are subject to uncertainty contributed by the estimation of both the relative risks and the ambient concentrations of PM (Figure 17.8).

Cardiopulmonary disease in adults aged $\geq 30$ years contributed $89 \%$ ( 712000 ) and $78 \%$ ( 4.97 million) of attributable deaths and burden, respectively. Lung cancer contributed $8 \%$ and $9 \%$ and ARI in children, contributed $3 \%$ and $7 \%$ of deaths and YLL, respectively, to the total burden (Tables $17.9[\mathrm{a}]-17.9[\mathrm{c}]$ ). The number of attributable deaths from all natural causes estimated from the daily time-series studies was roughly half the total attributable deaths, 378000 vs 799000 (Table

Table 17.8 Attributable deaths and DALYs in 2000, by subregion (50\% and 95\% confidence intervals)

| Subregion | Deaths <br> (000s) | 50\% Cl | 95\% Cl | $\begin{aligned} & \text { DALYs } \\ & (000 \mathrm{~s}) \end{aligned}$ | 50\% Cl | 95\% Cl |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | 22 | I 1.1-23.7 | 4.3-34.5 | 285 | 155.1-36\|. 3 | 28.2-557.5 |
| AFR-E | 10 | 7.5-14.4 | 3.9-22.2 | 147 | 107.7-239.9 | 24.7-364.7 |
| AMR-A | 28 | 22.1-33.7 | 12.7-44.0 | 152 | 158.6-239.8 | 94.6-314.8 |
| AMR-B | 30 | 23.1-37.4 | 12.0-50.1 | 232 | 241.8-383.6 | $142.9-517.7$ |
| AMR-D | 5 | 3.2-5.3 | 1.6-7.6 | 44 | 34.2-62.6 | 14.2-87.3 |
| EMR-B | 8 | 4.0-8.4 | $2.1-13.5$ | 77 | 45.6-93.4 | 25.0-149.2 |
| EMR-D | 51 | 31.2-56.2 | 17.0-73.0 | 558 | 384.5-737.5 | 163.1-970.4 |
| EUR-A | 23 | 19.4-29.7 | 9.9-42.8 | 117 | 125.7-187.4 | 65.8-265.4 |
| EUR-B | 38 | 26.7-44.4 | 14.8-58.5 | 288 | 241.3-386.6 | 138.5-507.0 |
| EUR-C | 46 | 28.1-53.4 | $10.1-83.3$ | 320 | 229.6-432.1 | 81.9-676.0 |
| SEAR-B | 32 | 19.2-37.5 | 5.5-51.5 | 282 | 191.0-388.9 | 67.0-532.6 |
| SEAR-D | 132 | 98.3-162.1 | 54.1-212.3 | 1312 | I 185.1-\| 890.5 | 575.2-2 409.8 |
| WPR-A | 18 | 13.2-21.4 | 6.7-28.5 | 84 | 84.3-137.0 | 42.9-182.0 |
| WPR-B | 355 | 260.8-424.8 | 142.8-555.1 | 2504 | 2447.4-3848.7 | 1 431.2-5 014.1 |
| World | 799 | 574.8-942.5 | 318.2-1 196.9 | 6404 | 5955.6-9288.2 | $3199.9-11472.4$ |

17.9 [d]). The overall and cause-specific burden of disease varies across the 14 subregions, with the preponderance of the burden of air pollution contributed by cities in WPR-B, which includes China, and SEARD , which includes India. The variation in attributable deaths and YLL among the 14 subregions seen in Tables 17.9 (a)-(d) reflects a subregional variation in the attributable fraction of approximately six-fold. For example, for all mortality end-points, EUR-A and WPR-B lie at the low and high ends, respectively, of the subregional distribution of attributable fractions. This largely reflects differences in the estimated populationweighted subregional ambient concentrations of $\mathrm{PM}_{10}$ and $\mathrm{PM}_{2.5}$ ( 89 vs $26 \mu \mathrm{~g} / \mathrm{m}^{3} \mathrm{PM}_{10}$ (see Table 17.4), rather than the proportion of the population that resides in cities. The proportion of the population of WPR-B that lives in cities is, in fact, lower than that in EUR-A ( $34 \%$ vs 39\%).

### 5.2 Sensitivity analyses

Cases 2-4: Shape of the concentration-Response function for PM ${ }_{2.5}$
The estimates of attributable deaths from cardiopulmonary disease and lung cancer and YLL under the base-case and three alternative scenarios for the shape of the $\mathrm{PM}_{2.5}$ concentration-response function are presented in Table 17.10. When the city-specific estimated concentrations

Figure 17.8 Uncertainty distributions for deaths and YLL from cardiopulmonary disease and lung cancer worldwide

of $\mathrm{PM}_{2.5}$ are constrained to never exceed the concentrations observed in the most polluted city in the ACS study (annual average concentration of $\mathrm{PM}_{2.5}$ of $30 \mu \mathrm{~g} / \mathrm{m}^{3}$ ), case 2 , worldwide estimates of the number of deaths from cardiopulmonary disease and lung cancer are reduced by $29 \%$ and $27 \%$, respectively. Extrapolation of the ACS coefficients to the highest estimated city-specific concentrations of $\mathrm{PM}_{2.5}$ on the linear and logarithmic scales, cases 3 and 4, respectively, results in increases of $10 \%$ and $12 \%$ in the estimated number of attributable deaths from cardiopulmonary disease, and $8-24 \%$ increases in the estimated numbers of attributable deaths from lung cancer, relative to the base-case estimates.

These changes in worldwide estimates reflect underlying differences in the subregion-specific estimates (Table 17.11). Truncating the cityspecific annual average concentrations at a given level leaves the burden unchanged in subregions with cities with estimated concentrations of PM that are lower than the truncation point, while reducing the burden in subregions with cities with estimated concentrations of PM that are above that point. Most cities in Europe and North America have
Table 17.9(a) Attributable deaths and YLL: base-case scenario for cardiopulmonary disease (50\% and 95\% confidence intervals)

| Subregion | Relative risk | Attributable fraction (\%) | Deaths (000s) | 50\% Cl | 95\% CI | YLL (000s) | 50\% CI | 95\% CI |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | 1.148 | 2 | 18 | 8.7-20.4 | 4-31 | 162 | 80.2-188.3 | 36-283 |
| AFR-E | 1.071 | 1 | 9 | 6.1-11.7 | 3-22 | 84 | 59.0-113.8 | 25-210 |
| AMR-A | 1.029 | 2 | 23 | 17.5-29.0 | 8-38 | 116 | 87.6-144.7 | 41-190 |
| AMR-B | 1.066 | 3 | 27 | 21.3-35.5 | 10-47 | 201 | 157.4-262.7 | 77-347 |
| AMR-D | 1.108 | 3 | 4 | 2.7-5.0 | 1-8 | 31 | 20.1-36.2 | 9-60 |
| EMR-B | 1.075 | 3 | 8 | 3.7-8.3 | 1-13 | 65 | 32.0-70.9 | \|1-1|1 |
| EMR-D | 1.214 | 4 | 45 | 27.7-51.0 | 12-72 | 386 | 237.2-437.8 | 106-618 |
| EUR-A | 1.032 | 1 | 20 | 15.7-26.5 | 8-36 | 90 | 69.2-116.7 | 34-156 |
| EUR-B | 1.098 | 3 | 34 | 24.3-41.4 | \| 1-53 | 238 | 167.9-286.4 | 78-370 |
| EUR-C | 1.046 | 2 | 43 | 26.4-52.4 | 12-83 | 291 | 180.0-356.9 | 83-565 |
| SEAR-B | 1.250 | 4 | 30 | 17.1-33.8 | 5-50 | 240 | \|36.9-271.2 | 38-400 |
| SEAR-D | 1.190 | 3 | 119 | 88.0-152.7 | 44-198 | 1006 | 744.7-1293.0 | 373-1677 |
| WPR-A | 1.051 | 3 | 15 | 10.7-18.8 | 5-26 | 65 | 46.9-81.9 | 24-115 |
| WPR-B | 1.216 | 6 | 317 | 235.2-402.6 | 105-524 | 1992 | 1 477.0-2528.5 | 658-3 290 |
| World |  | 3 | 712 | 507.0-874.7 | 245-1 107 | 4966 | 3537.2-6083.2 | 1695-7700 |

Table I7.9(b)

| Subregion | Relative risk | Attributable fraction (\%) | Deaths (000s) | 50\% Cl | 95\% Cl | YLL (000s) | 50\% Cl | 95\% Cl |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | 1.210 | 5 | 0.41 | 0.2-0.5 | 0.06-0.78 | 4.2 | 1.9-5.0 | 0.6-8.0 |
| AFR-E | 1.100 | 2 | 0.27 | 0.2-0.4 | 0.05-0.69 | 2.9 | 2.1-4.2 | 0.5-7.4 |
| AMR-A | 1.040 | 3 | 4.85 | 3.8-6.7 | 0.75-8.48 | 36.9 | 28.9-50.8 | 5.7-64.7 |
| AMR-B | 1.093 | 4 | 2.10 | 1.7-2.9 | 0.39-4.25 | 19.9 | 16.0-27.5 | 3.7-40.3 |
| AMR-D | 1.153 | 6 | 0.15 | $0.1-0.2$ | 0.02-0.27 | 1.5 | $1.0-1.8$ | 0.2-2.7 |
| EMR-B | 1.105 | 4 | 0.45 | 0.2-0.5 | 0.05-0.86 | 4.7 | 2.1-5.3 | 0.5-8.8 |
| EMR-D | 1.309 | 8 | 1.55 | 1.0-1.9 | 0.18-2.70 | 16.8 | 10.4-20.6 | 2.0-29.3 |
| EUR-A | 1.044 | 2 | 3.52 | 2.8-5.0 | 0.58-6.61 | 27.4 | 22.1-38.9 | 4.5-51.5 |
| EUR-B | 1.138 | 5 | 3.00 | 2.1-3.9 | 0.41-5.43 | 30.2 | 21.5-38.9 | 4.1-54.7 |
| EUR-C | 1.065 | 3 | 2.82 | 1.6-3.6 | 0.52-5.63 | 27.4 | 15.7-34.8 | 5.1-54.8 |
| SEAR-B | 1.363 | 6 | 2.17 | 1.1-2.7 | 0.28-4.04 | 21.8 | 11.4-26.8 | 2.8-40.6 |
| SEAR-D | 1.272 | 4 | 5.63 | 4.3-7.8 | 0.90-11.60 | 55.9 | 42.4-77.2 | 8.9-115.2 |
| WPR-A | 1.071 | 4 | 2.71 | 2.0-3.7 | 0.45-4.94 | 17.6 | 13.3-23.9 | 2.9-32.1 |
| WPR-B | 1.311 | 10 | 32.37 | 24.5-44.0 | 5.15-59.01 | 308.5 | 233.2-419.1 | 49.1-562.4 |
| World |  | 5 | 62 | 47.0-83.1 | 9.95-114.32 | 576 | 436.9-767.5 | 92-1063 |

Table I7.9(c) Attributable deaths and YLL: base-case scenario for ARI in children aged 0-4 years (50\% and 95\% confidence

| Subregion | Relative risk | Attributable fraction (\%) | Deaths (000s) | 50\% Cl | 95\% CI | YLL (000s) | 50\% Cl | 95\% Cl |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | 1.050 | 0.8 | 3.5 | 0.8-4.9 | -2.9-9.0 | 119 | 26.8-165.6 | -96.0-301.2 |
| AFR-E | 1.024 | 0.3 | 1.8 | 0.7-3.4 | -2.0-5.8 | 61 | 22.6-113.5 | -65.8-193.3 |
| AMR-A | 1.010 | 0.0 | 0.0 | 0.0-0.0 | 0.0-0.0 | 0 | $0.1-0.3$ | -0.2-0.5 |
| AMR-B | 1.023 | 0.3 | 0.3 | 0.1-0.6 | -0.3-0.8 | 11 | 3.9-19.1 | -11.3-27.4 |
| AMR-D | 1.037 | 0.9 | 0.3 | 0.1-0.5 | -0.3-0.9 | 11 | 3.2-18.2 | -10.9-30.3 |
| EMR-B | 1.026 | 0.6 | 0.2 | 0.1-0.3 | -0.2-0.6 | 7 | 1.8-9.6 | -6.3-20.7 |
| EMR-D | 1.070 | 1.4 | 4.6 | 1.5-7.0 | $-4.1-11.5$ | 155 | 51.5-237.9 | -137.6-389.2 |
| EUR-A | 1.011 | 0.0 | 0.0 | 0.0-0.0 | 0.0-0.0 | 0 | 0.0-0.1 | -0.1-0.2 |
| EUR-B | 1.033 | 0.7 | 0.6 | 0.2-1.0 | -0.6-1.5 | 20 | 7.0-33.1 | -20.1-51.1 |
| EUR-C | 1.016 | 0.1 | 0.1 | 0.0-0.1 | -0.1-0.2 | 2 | 0.6-3.5 | -1.9-5.9 |
| SEAR-B | 1.082 | 0.5 | 0.6 | 0.2-1.0 | -0.6-1.7 | 21 | 5.4-35.3 | -19.7-58.7 |
| SEAR-D | 1.063 | 0.6 | 7.4 | 2.6-13.3 | -7.2-20.1 | 250 | 86.6-448.4 | -243.8-678.5 |
| WPR-A | 1.018 | 0.0 | 0.0 | 0.0-0.0 | 0.0-0.0 | 0 | 0.1-0.2 | -0.1-0.4 |
| WPR-B | 1.071 | 1.2 | 6.1 | 2.0-10.7 | -6.1-16.5 | 204 | 68.8-358.8 | -203.3-555.0 |
| World |  | 0.7 | 25.6 | 8.2-43.9 | -23.7-66.1 | 862 | 277.7-1480.7 | -798.6-2228.0 |

Table 17.9(d) Attributable deaths: base-case scenario for mortality from all natural causes

| Subregion | Relative risk | Attributable fraction (\%) | Deaths (000s) |
| :--- | :---: | :---: | :---: |
| AFR-D | 1.029 | 0.67 | 26 |
| AFR-E | 1.015 | 0.29 | 16 |
| AMR-A | 1.006 | 0.42 | 11 |
| AMR-B | 1.014 | 0.68 | 15 |
| AMR-D | 1.022 | 0.90 | 4 |
| EMR-B | 1.015 | 0.62 | 4 |
| EMR-D | 1.042 | 1.20 | 37 |
| EUR-A | 1.007 | 0.26 | 10 |
| EUR-B | 1.019 | 0.74 | 14 |
| EUR-C | 1.009 | 0.43 | 13 |
| SEAR-B | 1.048 | 0.87 | 17 |
| SEAR-D | 1.037 | 0.64 | 70 |
| WPR-A | 1.011 | 0.68 | 7 |
| WPR-B | 1.042 | 1.43 | 133 |
| World |  | 0.75 | 378 |

estimated annual average concentrations of $\mathrm{PM}_{2.5}$ of $<30 \mu \mathrm{~g} / \mathrm{m}^{3}$, while estimated concentrations in the cities of developing countries are frequently much greater. More than $95 \%$ of the decrease in the worldwide burden in case 2 and the increase in case 3 occurs in four subregions: WPR-B, EMR-D, SEAR-B and SEAR-D.

Log-linear specification of the concentration-response function, as in case 4 , allows for a more gradual increase in the relative risk at concentrations of PM of $>30 \mu \mathrm{~g} / \mathrm{m}^{3}$ than does the linear extrapolation model of case 3. This specification also means that the relative risk increases more steeply at concentrations of $<30 \mu \mathrm{~g} / \mathrm{m}^{3}$. Since the estimates for burden in both the log-linear case and the base case are measured with reference to a counterfactual annual average concentration of $\mathrm{PM}_{2.5}$ of $7.5 \mu \mathrm{~g} / \mathrm{m}^{3}$, the burden under the log-linear specification is higher than that under the base case at low levels of exposure, but lower than the base case at high levels of exposure. Differences in the subregion-specific estimates for burden under the log-linear specification relative to the base case depend on the subregion-specific distributions of the city-specific concentrations of PM. The burden of disease in subregions where exposure is relatively low (AMR-A, EUR-A, EUR-C and WPR-A) increases by $63 \%$, relative to the base case, while the burden in subregions where exposure is high remains unchanged or is slightly reduced.
Table 17.10 Sensitivity analyses of base-case estimates of attributable deaths and YLL, by cause

| Case | Conditions | Cardiopulmonary disease |  | Lung cancer |  | Acute respiratory infections ${ }^{\text {a }}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Attributable deaths (000s) (\% change) | $\begin{aligned} & \text { YLL (000s) } \\ & \text { (\% change) } \end{aligned}$ | Attributable deaths (000s) (\% change) | $\begin{aligned} & \text { YLL (000s) } \\ & \text { (\% change) } \end{aligned}$ | Attributable deaths (000s) (\% change) | $\begin{aligned} & \text { YLL (000s) } \\ & \text { (\% change) } \\ & \hline \end{aligned}$ |
| Base-case | Maximum concentration of $\mathrm{PM}_{2.5}=50 \mu \mathrm{~g} / \mathrm{m}^{3}$ | 712 | 4966 | 62 | 576 | 26 | 862 |
| Case 2 | Maximum concentration of $\mathrm{PM}_{2.5}=30 \mu \mathrm{~g} / \mathrm{m}^{3}$ | 506 (-29) | 3498 (-30) | 45 (-27) | 414 (-28) | NA | NA |
| Case 3 | Linear extrapolation | 783 (10) | 5507 (11) | 67 (8) | 623 (8) | NA | NA |
| Case 4 | Log-linear extrapolation | 794 (12) | 5476 (10) | 77 (24) | 698 (21) | NA | NA |
| Case 5 | Change ACS coefficient/linear extrapolation | 1132 (59) | 7908 (59) | 101 (63) | 939 (63) | NA | NA |
| Case 6 | Change ACS coefficient/log-linear extrapolation | 1069 (50) | 7385 (49) | 105 (69) | 955 (66) | NA | NA |
| Case 7 | Change $\mathrm{PM}_{2.5}$ : $\mathrm{PM}_{10}$ ratio | 609 (-15) | 4109 (-17) | 58 (-7) | 521 (-10) | NA | NA |
| Case 8 | Theoretical minimum concentration of PM $=3 \mu \mathrm{~g} / \mathrm{m}^{3}$ | 882 (24) | 6081 (23) | 80 (29) | 731 (27) | 30 (-15) | 1012 (-17) |
| Case 9 | Theoretical minimum concentration of $P M=15 \mu \mathrm{~g} / \mathrm{m}^{3}$ | 474 (-33) | 3365 (-32) | 39 (-37) | 369 (-36) | 19 (27) | 627 (27) |
| NA Not applicable. <br> In children aged-4 years. |  |  |  |  |  |  |  |

Table I7.II Subregion-specific estimates for number of deaths and YLL from cardiopulmonary disease, lung cancer and ARI ${ }^{2}$ for

| Subregion | Cardiopulmonary deaths (000s) and YLL (000s) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Base-case |  | Case 2 |  | Case 3 |  | Case 4 |  | Case 5 |  | Case 6 |  | Case 7 |  | Case 8 |  | Case 9 |  |
|  | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL |
| AFR-D | 18 | 162 | 13 | 123 | 19 | 179 | 19 | 180 | 28 | 259 | 26 | 243 | 12 | 106 | 21 | 194 | 12 | 112 |
| AFR-E | 9 | 84 | 8 | 79 | 9 | 84 | 12 | 113 | 14 | 132 | 16 | 152 | 5 | 44 | 12 | 115 | 4 | 41 |
| AMR-A | 23 | 116 | 23 | 116 | 23 | 116 | 41 | 205 | 36 | 181 | 55 | 273 | 40 | 202 | 44 | 220 | 3 | 13 |
| AMR-B | 27 | 201 | 25 | 187 | 27 | 203 | 35 | 261 | 43 | 315 | 47 | 349 | 15 | 109 | 37 | 277 | 14 | 106 |
| AMR-D | 4 | 31 | 4 | 28 | 4 | 31 | 5 | 38 | 7 | 49 | 7 | 51 | 2 | 18 | 5 | 39 | 3 | 18 |
| EMR-B | 8 | 65 | 7 | 62 | 8 | 65 | 10 | 89 | 12 | 102 | 14 | 120 | 4 | 34 | 10 | 88 | 4 | 30 |
| EMR-D | 45 | 386 | 26 | 219 | 64 | 545 | 46 | 397 | 72 | 617 | 63 | 537 | 37 | 317 | 51 | 440 | 36 | 306 |
| EUR-A | 20 | 90 | 20 | 90 | 20 | 90 | 34 | 151 | 32 | 141 | 46 | 202 | 35 | 154 | 37 | 164 | 4 | 18 |
| EUR-B | 34 | 238 | 33 | 229 | 34 | 238 | 44 | 304 | 54 | 375 | 59 | 408 | 49 | 339 | 44 | 304 | 20 | 136 |
| EUR-C | 43 | 291 | 43 | 291 | 43 | 291 | 68 | 460 | 67 | 457 | 91 | 616 | 67 | 458 | 66 | 449 | 11 | 74 |
| SEAR-B | 30 | 240 | 17 | 132 | 36 | 287 | 28 | 226 | 49 | 390 | 39 | 309 | 22 | 174 | 34 | 269 | 24 | 191 |
| SEAR-D | 119 | 1006 | 79 | 667 | 136 | 1150 | 123 | 1037 | 192 | 1625 | 167 | 1413 | 82 | 697 | 137 | 1164 | 88 | 749 |
| WPR-A | 15 | 65 | 15 | 65 | 15 | 65 | 23 | 99 | 23 | 102 | 30 | 132 | 23 | 101 | 23 | 98 | 5 | 21 |
| WPR-B | 317 | 1992 | 192 | 1209 | 344 | 2163 | 305 | 1915 | 504 | 3163 | 411 | 2580 | 216 | 1357 | 360 | 2261 | 247 | 1550 |
| World | 712 | 4966 | 506 | 3498 | 783 | 5507 | 794 | 5476 | 1132 | 7908 | 1069 | 7385 | 609 | 4109 | 882 | 6081 | 474 | 3365 |

Table I7.II Subregion-specific estimates for number of deaths and YLL from cardiopulmonary disease, lung cancer and ARI ${ }^{2}$ for

| Subregion | Lung cancer deaths (000s) and YLL (000s) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Base-case |  | Case 2 |  | Case 3 |  | Case 4 |  | Case 5 |  | Case 6 |  | Case 7 |  | Case 8 |  | Case 9 |  |
|  | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL | Deaths | YLL |
| AFR-D | 0.4 | 4.2 | 0.3 | 3.2 | 0.5 | 4.7 | 0.5 | 5.0 | 1 | 7 | 1 | 7 | 0.3 | 2.8 | 0 | 5 | 0 | 3 |
| AFR-E | 0.3 | 2.9 | 0.3 | 2.7 | 0.3 | 2.9 | 0.4 | 4.1 | 0 | 5 | 1 | 6 | 0.1 | 1.5 | 0 | 4 | 0 | 1 |
| AMR-A | 4.8 | 36.9 | 4.8 | 36.9 | 4.8 | 36.9 | 9.1 | 69.5 | 8 | 59 | 12 | 94 | 8.4 | 64.4 | 9 | 70 | 1 | 4 |
| AMR-B | 2.1 | 19.9 | 2.0 | 18.6 | 2.1 | 20.1 | 2.9 | 27.5 | 3 | 32 | 4 | 37 | 1.1 | 10.8 | 3 | 27 | 1 | 11 |
| AMR-D | 0.1 | 1.5 | 0.1 | 1.4 | 0.1 | 1.5 | 0.2 | 2.0 | 0 | 2 | 0 | 3 | 0.1 | 0.9 | 0 | 2 | 0 | 1 |
| EMR-B | 0.5 | 4.7 | 0.4 | 4.5 | 0.5 | 4.7 | 0.7 | 6.8 | 1 | 7 | 1 | 9 | 0.2 | 2.4 | 1 | 6 | 0 | 2 |
| EMR-D | 1.6 | 16.8 | 0.9 | 9.5 | 2.2 | 23.8 | 1.7 | 18.4 | 3 | 28 | 2 | 25 | 1.3 | 13.8 | 2 | 19 | 1 | 13 |
| EUR-A | 3.5 | 27.4 | 3.5 | 27.4 | 3.5 | 27.4 | 6.3 | 49.2 | 6 | 44 | 9 | 67 | 6.0 | 46.9 | 6 | 50 | 1 | 6 |
| EUR-B | 3.0 | 30.2 | 2.9 | 29.0 | 3.0 | 30.2 | 4.1 | 41.0 | 5 | 49 | 6 | 56 | 4.3 | 43.0 | 4 | 39 | 2 | 17 |
| EUR-C | 2.8 | 27.4 | 2.8 | 27.4 | 2.8 | 27.4 | 4.8 | 46.2 | 5 | 44 | 6 | 63 | 4.4 | 43.1 | 4 | 42 | 1 | 7 |
| SEAR-B | 2.2 | 21.8 | 1.2 | 11.9 | 2.6 | 26.2 | 2.2 | 22.0 | 4 | 37 | 3 | 31 | 1.6 | 15.7 | 2 | 25 | 2 | 17 |
| SEAR-D | 5.6 | 55.9 | 3.7 | 36.8 | 6.4 | 64.0 | 6.2 | 61.6 | 9 | 94 | 9 | 86 | 3.9 | 38.5 | 7 | 65 | 4 | 41 |
| WPR-A | 2.7 | 17.6 | 2.7 | 17.6 | 2.7 | 17.6 | 4.4 | 28.3 | 4 | 28 | 6 | 38 | 4.2 | 27.1 | 4 | 26 | 1 | 6 |
| WPR-B | 32.4 | 308.5 | 19.6 | 186.7 | 35.2 | 335.2 | 33.2 | 316.1 | 53 | 503 | 45 | 433 | 22.0 | 209.6 | 37 | 350 | 25 | 240 |
| World | 62.0 | 576 | 45 | 414 | 67 | 623 | 77 | 698 | 101 | 939 | 105 | 955 | 58 | 521 | 80 | 731 | 39 | 369 |

Deaths (000s) and YLL (000s) from acute respiratory infections

| Subregion | Deaths (000s) and YLL (000s) from acute respiratory infections |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Base-case |  | Case 8 |  | Case 9 |  |
|  | Attributable deaths | YLL | Attributable deaths | YLL | Attributable deaths | YLL |
| AFR-D | 3.5 | 118.8 | 4 | 141 | 2 | 82 |
| AFR-E | 1.8 | 61.2 | 3 | 84 | 1 | 30 |
| AMR-A | 0.0 | 0.2 | 0 | 0 | 0 | 0 |
| AMR-B | 0.3 | 10.6 | 0 | 15 | 0 | 6 |
| AMR-D | 0.3 | 11.2 | 0 | 14 | 0 | 7 |
| EMR-B | 0.2 | 7.3 | 0 | 10 | 0 | 3 |
| EMR-D | 4.6 | 154.8 | 5 | 176 | 4 | 123 |
| EUR-A | 0.0 | 0.1 | 0 | 0 | 0 | 0 |
| EUR-B | 0.6 | 20.2 | 1 | 26 | 0 | 12 |
| EUR-C | 0.1 | 2.0 | 0 | 3 | 0 | 1 |
| SEAR-B | 0.6 | 21.3 | 1 | 24 | 1 | 17 |
| SEAR-D | 7.4 | 249.9 | 9 | 288 | 6 | 188 |
| WPR-A | 0.0 | 0.1 | 0 | 0 | 0 | 0 |
| WPR-B | 6.1 | 204.3 | 7 | 231 | 5 | 159 |
| World | 25.6 | 862.1 | 30 | 1012 | 19 | 627 |
| ${ }^{\text {a }}$ In children aged 0-4 years. |  |  |  |  |  |  |

Linear extrapolation beyond concentrations of PM of $30 \mu \mathrm{~g} / \mathrm{m}^{3}$ of larger alternative coefficients from the ACS study on the basis of the average of ambient concentrations measured in 1979-1983 and 1999-2000 resulted in increases of $59 \%$ and $63 \%$ in deaths attributable to cardiopulmonary disease and lung cancer, respectively, relative to the base-case estimates. Log-linear extrapolation of the larger coefficients produced increases of $50 \%$ and $69 \%$ in the number of deaths attributable to cardiopulmonary disease and lung cancer, respectively (Table 17.10).

Attributable burdens increased in all subregions (Table 17.11). The differences between the linear and log-linear estimates followed the same subregional patterns as in cases 3 and 4, discussed above.

## Case 7: Choice of PM $_{2.5}:$ PM $_{10}$ Ratio

Allowing limited subregional variation in the ratio of $\mathrm{PM}_{2.5}$ to $\mathrm{PM}_{10}$ produced reductions of $15 \%$ and $7 \%$ in the worldwide estimates of numbers of deaths attributable to cardiopulmonary disease and lung cancer, respectively, relative to the base-case scenario in which this ratio was fixed at 0.50 (Table 17.10).

As one might expect, the burden of disease increases by $57 \%$ in those subregions assigned a ratio of 0.65 , that is, AMR-A, all of Europe, and WPR-B. This increase is more than offset by the rest of the world, assigned a ratio of 0.35 , where the burden of disease falls by $31 \%$ (Table 17.11).

## Cases 8 and 9: Choice of theoretical minimum level of exposure

Halving and doubling the base-case theoretical minimum concentration of $\mathrm{PM}_{2.5}$ of $7.5 \mu \mathrm{~g} / \mathrm{m}^{3}$ resulted in a $24 \%$ increase and a $33 \%$ decrease in the number of deaths attributable to cardiopulmonary disease, and a $29 \%$ increase and a $37 \%$ decrease in deaths attributable to lung cancer, but only minor variations in mortality from ARI (Table 17.10).

All subregions experienced increases in attributable burden when the theoretical minimum concentration was halved, with the largest proportional increases in less polluted subregions (AMR-A, EUR-A and WPR-A). These subregions also experienced the largest reductions in burden when the theoretical minimum concentration was doubled. Highly polluted subregions (WPR-B and SEAR-D) also experienced marked reductions in the estimated burden when the theoretical minimum concentration was doubled (Table 17.11).

## 6. Discussion

Previously, most large-scale estimates of the health impacts of urban air pollution were conducted for countries or regions where data on expo-
sure and estimates of effect required for impact estimation were available (e.g. Brunekreef 1997; COMEAP 1998; Künzli et al. 2000; Ostro and Chestnut 1998). In the few previous global estimates, systematic methods were not applied to extrapolate exposures and exposureresponse functions to other parts of the world (Hong 1995; WHO 1997; Working Group on Public Health and Fossil Fuel Combustion 1997). Although our estimates exceed those reported earlier, the differences are not large, given the variation in the approaches that were taken (Smith and Mehta 2003). The global scope of the present analysis required new approaches for estimating exposure, absent measurements of air pollution in many developing countries, extrapolating the results of epidemiological studies more widely than had previously been attempted, and describing and attempting to quantify, the many uncertainties this entailed. The results indicate that the impact of urban air pollution on the burden of disease in the cities of the world is large and, for a variety of reasons discussed below, have probably underestimated the burden. There is also considerable variation in our estimates among the 14 subregions, with the greatest burden occurring in the more polluted and rapidly growing cities of developing countries.

The availability of actual measurements of outdoor concentrations of PM varied widely across the globe. In order to have estimates for all 14 subregions, models developed by the World Bank were used to estimate concentrations of inhalable particles ( $\mathrm{PM}_{10}$ ) using economic, meteorological and demographic data and the available measurements of PM for 3211 cities with populations of $>100000$, and also capital cities. To allow the most appropriate epidemiological studies to be used for the estimation of the burden of disease, the $\mathrm{PM}_{10}$ estimates were converted to estimates of fine particles $\left(\mathrm{PM}_{2.5}\right)$ using information on geographic variation in the ratio of $\mathrm{PM}_{2.5}$ to $\mathrm{PM}_{10}$. Population-weighted subregional annual average exposure estimates for $\mathrm{PM}_{2.5}$ and for $\mathrm{PM}_{10}$ were obtained using the population of the city in the year 2000.

The estimates of the burden of disease were based on three health outcomes: mortality from cardiopulmonary causes in adults, mortality from lung cancer and mortality from ARI in children aged $0-4$ years. Attributable numbers of deaths and YLL for adults and children (aged $0-4$ years) were estimated using risk coefficients from a large cohort study of adults in the United States (Pope et al. 2002) and a metaanalytic summary of five time-series studies of mortality in children, respectively. Base-case estimates were calculated assuming that the risk of death increases linearly over a range of annual average concentrations of $\mathrm{PM}_{2.5}$ between a counterfactual (or referent) concentration of 7.5 and a maximum of $50 \mu \mathrm{~g} / \mathrm{m}^{3}$. For comparison, an additional estimate of attributable deaths was calculated from time-series studies of daily mortality, on the basis of results of a meta-analysis of the world literature, but was not used in the final calculations. Worldwide and subregion-
specific estimates of attributable deaths and burden of disease in terms of YLL were calculated based on the standard methodology developed for this project (see chapters 1 and 25).

We estimated that urban air pollution, as measured by PM, is responsible for about $3 \%$ of mortality caused by cardiopulmonary disease in adults, about $5 \%$ of mortality caused by cancers of the trachea, bronchus and lung, and about $1 \%$ of mortality caused by ARI in children worldwide in the year 2000. The total burden was about 0.80 million ( $1.2 \%$ of total) premature deaths and 6.4 million ( $0.5 \%$ of total) DALYs. This burden occurred predominantly in developing countries, with $30 \%$ of attributable disease burden occurring in WPR-B and 19\% in SEAR-D. The greatest contributions to the total burden of disease occurred in WPR-B, EUR-B and EUR-C, where urban air pollution caused 0.7-0.9\% of the total burden of disease.

### 6.1 IDENTIFYING AND QUANTIFYING UNCERTAINTY IN THE ESTIMATES

These estimates are subject to considerable uncertainty given the need to estimate exposures and to extrapolate concentration-response relationships. This is almost invariably the case in quantitative risk assessment of complex environmental exposures; and is certainly to be expected in this particular exercise, for reasons discussed above.

We quantified the statistical uncertainty of our base-case estimates by estimating the joint uncertainty in the estimates of annual average concentration and the estimates of the relative risks. Worldwide and most subregional estimates vary by less than two-fold ( $50 \%$ uncertainty intervals (Tables 17.8 and $17.9[\mathrm{a}]-[\mathrm{c}])$. Uncertainty of the model owing to assumptions about the shape of the concentration-response function, the magnitude of the relative risk of disease attributable to urban air pollution, the choice of counterfactual level for PM , and the ratio of $\mathrm{PM}_{2.5}$ to $\mathrm{PM}_{10}$ was assessed in sensitivity analyses. For the most part, the estimated worldwide burdens in the various sensitivity analyses are within the $50 \%$ uncertainty interval for the base-case estimate of worldwide burden. The sensitivity analyses indicate that base-case estimates were most sensitive to choice of coefficient from the ACS study and theoretical minimum concentration.

Although some sources of uncertainty could be quantified, others that were no less important or were perhaps more important, could not. These additional sources of uncertainty arise from the methods we used to estimate annual average exposure of the population and our choice of health end-points and concentration-response functions.

Estimates of exposure to urban air pollution
There are four key uncertainties related to exposure that have not been quantified, and that could affect the estimates of burden of disease.

First, we used PM as the sole indicator of exposure to urban air pollution, although urban air pollution is a complex mixture, as noted above. Other frequently measured pollutants, notably ozone, carbon monoxide, oxides of sulfur and nitrogen, and lead are associated with mortality and morbidity, albeit not as consistently as PM, although the effects of a number of these pollutants may be at least partially captured via the use of a PM metric (Sarnat et al. 2001). Estimating the health impacts of specific components poses challenges for both scientific research and risk assessment, including how to avoid attributing the same burdens to multiple pollutants (i.e. double counting), and how to quantify the effects of possible interactions (i.e. synergistic effects) among pollutants. Nonetheless, there is evidence for an effect of ozone on daily mortality that is independent of PM (e.g. Health Effects Institute 2000b). Future estimates of the burden of disease should include the health impacts of ozone. Unfortunately, lead in petrol remains an important toxic component of air pollution in some cities of the developing world and its contribution to the burden of disease has been estimated elsewhere in this book (see chapter 19). Some combination of the GBD estimates for urban air pollution and lead probably provide the best overall estimate of the burden of disease attributable to urban air pollution.

Second, use of estimated levels of exposure introduces some uncertainties and biases in the predicted levels of exposure that could not be addressed, owing to lack of data. The most important of these is the lack of city-specific data on the structure of economic activity and on fuel consumption. The exposure model uses national average data for these variables as a reasonable proxy, which can lead to bias in unknown directions with regard to city-specific estimates. The net bias in estimates of the aggregate burden at the subregional level is unclear. The use of long-run average climatic conditions instead of time-varying local data may result in biased estimates for specific years, but may not pose a serious problem as we are interested mainly in the long-term average health effects of air pollution. We have also explicitly examined our uncertainty regarding spatial variations in the size composition of PM through sensitivity analysis. The model for the estimation of exposure clearly suggests that coarser particles account for a larger fraction of the TSP in developing countries than in developed countries, all other things being equal. The limited data from monitoring available on $\mathrm{PM}_{2.5}$ also indicate that spatial variations may also exist in the sizes of finer particles. Consequently, we have used conservative estimates for the fraction of $\mathrm{PM}_{10}$ accounted for by finer particles in our overall estimates and further tested the implications of using an even more conservative estimate. The burden estimates should be relatively insensitive to PM size fraction.

Third, misclassification of exposure may have led to underestimation of the burden of disease. Like the epidemiological studies used to quantify the estimates of health impact, we used the annual average ambient
concentration measured from a few stationary sources in each city to estimate average personal levels of exposure. Differences between personal levels of exposure and concentrations measured at fixed points depend on how well the pollutant mixes in the environment and the efficiency with which the pollutant penetrates indoors. The exposure estimates are based on a model developed from population-oriented monitors. Measurements of PM from these sites in well-designed monitoring networks would provide representative city-wide levels of exposure for a pollutant that mixes uniformly in the environment. They would underestimate the actual level of exposure of people living near pollution hotspots, such as busy roads or local sources of pollutant, which can contribute to spatial heterogeneity of exposure within cities (Hoek et al. 2002; Jerrett et al. 2001). This underestimate would probably be more pronounced for cities in developing countries, where nearly one third of the population resides in slums, which are often in heavilypolluted parts of cities, and even larger populations work near pollution hotspots.

Exposure misclassification from using outdoor concentrations to represent personal exposure to urban air pollution also results from differences in the efficiency with which PM penetrates indoors. Use of ambient concentrations as surrogates for exposure tends to underestimate the risk per unit exposure because the penetration of particles indoors, where most exposure occurs, is less than $100 \%$. If average penetration is $66 \%$, for example, actual exposure-response per $1 \mu \mathrm{~g} / \mathrm{m}^{3}$ would be 1.5 -fold that indicated by outdoor concentrations of pollution. However, because of climate and housing, the rates of penetration of pollution in most, but not all, cities in developing countries can be expected to be somewhat greater than those in the average city in the United States where the epidemiology used here has been undertaken. Not being able to consider this factor because of lack of data on penetration of the pollutant would bias estimates of burden downward if actual changes in exposures in developing countries are better indicated by changes in outdoor concentrations than in developed countries.

An additional source of misclassification concerns the time referent of our exposure estimate. The current burden is related to past exposure, but our model estimates current (i.e. 1999) levels only. However, even if we had been able to retrospectively estimate a time series of annual average concentrations for each subregion, the ACS study provides little information as to how the concentration-response function varies over time (Krewski et al. 2000). It is not clear how this source of misclassification would affect our estimates.

Fourth, our estimates do not include the attributable burden of disease among the 800 million additional urban residents living either in suburban areas of some of the cities or in cities with populations of $<100000$ and in the $>3$ billion residents of rural areas. Although lower levels of emissions per area combined with the differences in the built-up envi-
ronment in rural areas probably results in a small average exposure to ambient pollution and a modest increase in the global burden of disease from such pollution in rural areas, the same is not true for the urban residents that were not included in the target population. The magnitude of the missing burden depends on the actual exposures of those living in smaller cities. The target population was identified for the study on the basis of data available from the United Nations, which compiles the data reported by Member States from national censuses and makes projections from them on the basis of expected changes in demographics. In compiling the population statistics, Member States, hence the United Nations, do not use uniform definitions either for city area (the characteristic such as city size that defines an urban area) or the population included for each identified city (whether political boundaries are used or agglomerations of contiguous urban areas are used). For the target population, we have used the population of the city agglomeration when this choice was available. If all of the remaining 800 million residents lived in suburban areas next to a targeted city, exposures and hence, estimates of burden, for these residents could be expected to be similar to those in the identified city resulting in an aggregate underestimate of the attributable fraction of the population of about $28 \%$. The exposure model suggests, however, that concentrations gradually decrease as the local population density decreases, suggesting that levels of exposure and hence estimates of burden are lower for these residents compared to those living in larger cities. The net result is that our focus on residents in cities with $>100000$ inhabitants may underestimate the aggregate burden by between $0 \%$ and $28 \%$.

## Choice of health end-Points and ConcentrationRESPONSE FUNCTIONS

Our base-case estimates of burden in terms of disease burden considered only the impact of air pollution on mortality. This approach is likely to have underestimated the true attributable burden, since there is evidence from studies of both epidemiology and toxicology, to suggest that air pollution may play a role in the incidence of cardiopulmonary disease, and thus contribute to years lived with disability (YLD). Lacking estimates of the concentration-response function for air pollution and the incidence of cardiopulmonary disease, lung cancer, and ARI in children, we calculated disease burden under the assumption that air pollution multiplies both incidence and mortality to the same extent, i.e. the relative risk of unobserved morbidity equals the observed relative risk of mortality (Table 17.12), an approach taken to estimating the attributable burden caused by other factors other than urban air pollution. The total disease burden, including YLD for cardiopulmonary disease, exceeds the YLL by $23 \%$ worldwide in the base-case analyses. The effect on the estimated burden for lung cancer and ARI in children is negligible.

Table 17.12 Attributable YLL and DALYs for cardiopulmonary disease, lung cancer, $\mathrm{ARI}^{1}$ and total mortality

| Subregion | Cardiopulmonary disease |  | Lung cancer |  | Acute respiratory infections |  | Total |  | \% change |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | YLL | DALYs | YLL | DALYs | YLL | DALYs | YLL | DALYs |  |
|  | (000s) |  | (000s) |  | (000s) |  | (000s) |  |  |
| AFR-D | 162 | 193 | 4 | 4 | 119 | 121 | 285 | 319 | 12 |
| AFR-E | 84 | 100 | 3 | 3 | 61 | 62 | 147 | 166 | 13 |
| AMR-A | 116 | 161 | 37 | 38 | 0 | 0 | 152 | 200 | 32 |
| AMR-B | 201 | 273 | 20 | 20 | 11 | 14 | 232 | 307 | 32 |
| AMR-D | 31 | 39 | I | 2 | 11 | 12 | 44 | 53 | 21 |
| EMR-B | 65 | 77 | 5 | 5 | 7 | 9 | 77 | 91 | 18 |
| EMR-D | 386 | 457 | 17 | 17 | 155 | 162 | 558 | 636 | 14 |
| EUR-A | 90 | 122 | 27 | 28 | 0 | 0 | 117 | 151 | 29 |
| EUR-B | 238 | 286 | 30 | 31 | 20 | 21 | 288 | 338 | 17 |
| EUR-C | 291 | 340 | 27 | 28 | 2 | 2 | 320 | 360 | 13 |
| SEAR-B | 240 | 291 | 22 | 22 | 21 | 25 | 282 | 339 | 20 |
| SEAR-D | 1006 | 1195 | 56 | 57 | 250 | 261 | 1312 | 1513 | 15 |
| WPR-A | 65 | 95 | 18 | 18 | 0 | 0 | 84 | 114 | 36 |
| WPR-B | 1992 | 2732 | 304 | 317 | 204 | 224 | 2504 | 3272 | 31 |
| World | 4966 | 6360 | 572 | 591 | 862 | 913 | 6404 | 7865 | 23 |

a In children aged 0-4 years.

The estimates of the attributable burden caused by cardiopulmonary disease and lung cancer were derived from a single cohort study in the United States (the largest and most extensively reviewed study suitable for the estimation of the burden of disease). This raises questions concerning whether these results can be generalized to other populations, especially those in developing countries, owing to differences in susceptibility to the effects of air pollution and the nature of the mixture of air pollutants. The apparent qualitative and quantitative similarity of the relative risks of daily mortality in developed and developing countries, discussed above, provides some evidence that these results are generally applicable. In addition, trends in known risk factors for chronic cardiovascular and respiratory disease, such as diet and cigarette smoking, suggest that the populations of cities in developing countries may now be more comparable to populations of cities in Europe and North America with regard to susceptibility to air pollution conferred by preexisting cardiovascular and respiratory morbidity (Reddy and Yusuf 1998). The increasing contribution of mobile sources to urban air pollution in developing countries also increases the similarity with cities in North America and Europe.

Other sources of uncertainty in our estimates cannot be readily quantified for the following reasons:

- Lack of knowledge concerning differences between developed and developing countries in the physicochemical nature of PM produced by different sources. The relative toxicity of PM may well vary according to the type of fuel burned and the type technology used to burn it. Increased burning of refuse outdoors and the prevalence of motor vehicles without emissions controls (e.g. vehicles powered by twostroke engines) are two examples.

Inhalable particles that are not the direct or indirect product of combustion sources may also be important. These particles are mainly of crustal origin and may be important, for example, in desert areas, or where there is disturbance of surface material owing to construction, use of unsurfaced roads, etc. They are largely found in the coarse fraction of inhalable particles, whereas combustion-derived particles tend to be found in the fine and ultra-fine fractions. The evidence concerning the toxicity of this fraction is mixed (Anderson 2000). Data on worldwide variation in the ratio of fine to coarse particles is limited, as discussed above, and our sensitivity analyses, which suggest relatively minor differences with our base-case estimates, may understate the uncertainty.

- Lack of knowledge concerning differences in the susceptibility of the population. Despite the trends discussed above, differences in demography and in the patterns of the incidence and prevalence of disease may be associated with differences in short-term and long-term vulnerability to air pollution. There exists the possibility of effectmodification factors related to health status, and behavioural factors, such as smoking and diet (Katsouyanni et al. 2001). The effects of previous or concurrent exposure to high levels of indoor air pollution may also play a role in determining susceptibility to urban air pollution. Poverty, which is a determinant of the factors just discussed, may also determine susceptibility in other ways. If the effects of air pollution are more severe among the poor, who comprise a large part of the world's population, then the magnitude of the burden would likely be greater than that which we estimated (Krewski et al. 2000; O’Neill et al. 2003).
- The shape of the exposure-response relationship may differ between developing and developed countries in ways that were not captured in the sensitivity analyses. For example, a recent time-series study of daily mortality in Mexico City did not observe a flattening of the $\mathrm{PM}_{10}$ concentration-response curve until $175 \mu \mathrm{~g} / \mathrm{m}^{3}$ (the daily mean) (Tellez-Rojo et al. 2000). These concentrations are measured in many mega-cities in developing countries.

We did not know which form of the concentration-response relationship should be used in extrapolating the results of the ACS study to the much higher concentrations observed in cities in India and China, for example. For this reason, we examined the sensitivity of the estimates to a range of scenarios, presenting a "base case", which we thought was a reasonable compromise between the conditions of the ACS and those of the rest of the world. Cohort evidence has recently been reported from Europe, although unfortunately it was unable to estimate concentrations of PM (Hoek et al. 2002). This study provides evidence that long-term exposure to urban pollution is associated with health effects elsewhere in developed countries, but we still lack cohort evidence from developing countries.

## Mortality from acute respiratory infections in children aged 0-4 YEARS

Despite limited evidence, discussed above, linking mortality from ARI to exposure to urban air pollution, we used the results from the small number of time-series studies in developing countries to estimate attributable deaths and YLL in children aged 0-4 years. In our view, most of these deaths are likely to be among children with temporary vulnerability owing to chest infections which would resolve eventually, and therefore represent, on average, the loss of many potential years of life, but this view is largely speculative.

Several studies that we used to derive the concentration-response function for ARI mortality actually reported results for all causes mortality in the $0-4$ years age group (Ostro et al. 1999a), or total mortality in the first year of life (Loomis et al. 1999). We have assumed that the relationship between $\mathrm{PM}_{10}$ and mortality from ARI in children aged $0-4$ years is similar to that for all-cause mortality. To some extent this is justified by the knowledge that mortality from ARI is an important component of all-cause mortality in developing countries.

### 6.2 Generalizability of our results

As a consequence of the uncertainties in this global assessment, its quantitative results cannot be confidently extrapolated to smaller geographic areas, such as specific countries or cities. The methods for estimation of exposure and extrapolation of concentration-response functions were developed specifically for estimating burdens for large geographic regions, often in the absence of essential data on exposure and response. Where better data exist, as they currently do in some parts of the world, they can, of course, be used.

Differences between our estimates and those of other groups may reflect other differences in methodology. For example, a tri-national European assessment recently estimated that some 40000 deaths per year were attributable to exposure to ambient air pollution in a population of approximately 72 million, whereas the burden in EUR-A, in an urban
population of 80 million, was estimated to be 23000 deaths per year, despite similar estimates of the concentration of ambient pollution (Künzli et al. 2000). The difference is largely owing to the different assumptions regarding the exposure reference level of $15 \mu \mathrm{~g} / \mathrm{m}^{3} \mathrm{PM}_{10}$ in this work vs $7.5 \mu \mathrm{~g} / \mathrm{m}^{3}$ in the European study. In addition, the concentration-response functions were slightly higher in the tri-national project, which used estimates of total mortality from both the first ACS publication and the Harvard Six City estimates (Dockery et al. 1993; Pope et al. 1995).

### 6.3 Avoidable disease burden

We did not attempt to estimate the avoidable burden of disease, despite this being a specific objective of the project. Estimating the avoidable burden would have required making projections of concentrations of ambient air pollution and providing a model for the exposure time-response function for PM and mortality. Time constraints did not allow us to undertake the former task, although it is feasible. The latter information is not currently available from the existing cohort studies, although there is limited evidence that the induction time for mortality from lung cancer attributable to exposure to urban air pollution is in the order of decades (Nyberg et al. 2000), and that it is perhaps in the order of years for mortality from cardiovascular disease (Krewski et al. 2000). Evaluations of both "natural experiments" (Heinrich et al. 2000; Pope 1989), and regulatory interventions (Clancy et al. 2002; Hedley et al. 2002) provide further support for relatively rapid improvements in cardiovascular and respiratory outcomes. The latter studies also suggest that although rates of mortality may decrease after the successful implementation of air pollution reductions, the long-term benefits may extend well beyond that observed during the first years after the intervention is implemented.

### 6.4 How could a future risk assessment exercise PROVIDE BETTER ESTIMATES?

There is a critical need for more information on the health effects of air pollution in developing counties. Research on exposure should aim to provide better estimates not only of ambient concentrations of pollutants, but also the characteristics of urban air pollution, including the contribution of various sources and the size distribution of PM. Epidemiological studies of mortality should be designed to provide ageand disease-specific estimates of the effects of air pollution, as well as identifying factors that confer susceptibility to air pollution. There is an obvious need for epidemiological studies of the effect of air pollution on the incidence of chronic cardiovascular and respiratory disease, and on the growth and development of children. Future estimates of the burden of disease attributable to urban air pollution should include morbidity outcomes, such as asthma exacerbation, which most certainly contribute to morbidity.

Estimates of uncertainty distributions should more fully incorporate model uncertainties, such as those related to the choice of concentra-tion-response function. This could be accomplished via the elicitation and weighting of expert opinions in the context of a Bayesian approach to quantifying model uncertainty (Morgan and Henrion 1998; National Research Council 2002).

## Acknowledgements

The authors wish to acknowledge the contribution of Richard Burnett, Uwe Deichmann, Majid Ezzati, Stephen Vander Hoorn and David Wheeler.

## Disclaimer

The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Health Effects Institute (HEI) or its sponsors nor those of the World Bank.

## Notes

1 Throughout the chapter we refer to urban air pollution using the terms "ambient air pollution" or "urban air pollution". For our current purposes, these terms are fully interchangeable.
2 See Preface for an explanation of this term.
3 Ambient particles fall into a trimodal size distribution, according to their aerodynamic diameter: coarse particles ( $>1 \mu \mathrm{~m}$ ), fine particles $(0.1-1 \mu \mathrm{~m})$, and ultrafine particles $(<0.1 \mu \mathrm{~m})$. Ultrafine particles constitute a small percentage of the total mass of PM, but are present in very high numbers. Because of health concerns, the ambient concentrations (mass) of both coarse and fine PM are regulated by the United States Environmental Protection Agency (EPA) through the National Ambient Air Quality Standards for $\mathrm{PM}_{10}$ (PM $<10 \mu \mathrm{~m}$ ) and $\mathrm{PM}_{2.5}(\mathrm{PM}<2.5 \mu \mathrm{~m})$ (USEPA 1997), and by the European Union through limit values for $\mathrm{PM}_{10}$. $\mathrm{PM}_{2.5}$, which includes only fine and ultrafine particles, is dominated by emissions from combustion processes; $\mathrm{PM}_{10}$, which includes coarse as well as fine and ultrafine particles, has a much higher proportion of particles generated by mechanical processes from a variety of noncombustion sources. It is currently not clear how much particles of different sizes and composition differ in the effects on health that they cause.
4 Cities in the United States account for about $40 \%$ of the observations in the estimation model, even after this exclusion.
5 Data from monitoring were available for one additional city/country, Skopje in The former Yugoslav Republic of Macedonia, but were not used in the estimation model because of missing explanatory variables. In addition, 150 observations primarily from Germany (94), Lithuania (30) and other eastern European states (26) made during the early 1990s were excluded because of uncertainties in defining appropriate explanatory variables.

6 The model presented here is one of several versions of the GMAPS model developed at the World Bank. An alternative model jointly estimates concentrations of $\mathrm{PM}_{10}$ and TSP at residential and non-residential sites.

7 The climatic variables have been constructed from a global mean monthly climatology map with a resolution of $0.5^{\circ}$ latitude $\times 0.5^{\circ}$ longitude developed by researchers at the Climate Research Unit of the University of East Anglia. These data are available at http://ipcc-ddc.cru.uea.ac.uk/cru_ data/examine/have_index.html. All of the climate variables are based on the conditions for the city centre.
8 Residential monitoring sites are located in residential areas but do not include pollution hotspots, such as locations that are immediately adjacent to industrial and commercial pollution sources or high traffic corridors. In contrast, mixed residential sites are characterized by both high population densities and the presence of some pollution sources that may result in elevated concentrations of PM in the immediate vicinity of the pollution source. Neither site includes areas of high pollution activity located in sparsely populated areas.
9 Had we instead included data from monitoring for cities with measured data for 1999 , there would be an insignificant difference in the subregional average concentration, because a small fraction of the population in each subregion lives in monitored cities and most of the monitored cities are located in North America and western Europe, where the estimates of PM are more precise.
10 After these estimates had been made, investigators in the United States and Canada discovered several problems with the statistical software that had been used to estimate the relative risks associated with air pollution in the time-series studies (Health Effects Institute 2003). Correcting these problems reduced the magnitude of estimated relative risks and increased their standard errors.

## References

Abbey DE, Nishino N, McDonnell WF et al. (1999) Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. American Journal of Respiratory and Critical Care Medicine, 159:373-382.
Aerometric Information Retrieval System (2001) U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, AIRS. Available at http://www.epa.gov/oar/oaqps/airpdata.html\#airs.
American Thoracic Society (ATS) (2000) What constitutes an adverse health effect of air pollution? American Journal of Respiratory and Critical Care Medicine, 161:665-673.
Anonymous (1996a) Health effects of outdoor air pollution. Part 1. Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society. American Journal of Respiratory and Critical Care Medicine, 153:3-50.

Anonymous (1996b) Health effects of outdoor air pollution. Part 2. Committee of the Environmental and Occupational Health Assembly of the American

Thoracic Society. American Journal of Respiratory and Critical Care Medicine, 153:477-498.

Anonymous (1998) Worldwide variations in the prevalence of asthma symptoms: the International Study of Asthma and Allergies in Childhood (ISAAC). European Respiratory Journal. 12:315-335.
Anderson HR (2000) Differential epidemiology of ambient aerosols. Philosophical Transactions of the Royal Society of London, Series A, 358: 2771-2785.
Atkinson RW, Anderson HR, Sunyer J et al. (2001) Acute effects of particulate air pollution on respiratory admissions: results from APHEA 2 project. Air pollution and health: a European approach. American Journal of Respiratory and Critical Care Medicine, 164:1860-1866.

Awasthi S, Glick HA, Fletcher RH, Ahmed N (1996) Ambient air pollution and respiratory symptoms complex in preschool children. Indian Journal of Medical Research, 104:257-262.

Bacharova L, Fandakova K, Bratinka J, Budinska M, Bachar J, Gud-aba M (1996) The association between air pollution and the daily number of deaths: findings from the Slovak Republic contribution to the APHEA project. Journal of Epidemiology and Community Health, 50:S19-21.
Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. Biometrics, 50:1088-1101.
Bobak M, Leon DA (1999) The effect of air pollution on infant mortality appears specific for respiratory causes in the post neonatal period. Epidemiology, 10:666-670.

Borja-Aburto VH, Castillejos M, Gold DR, Bierzwinski S, Loomis D (1998) Mortality and ambient fine particles in southwest Mexico City, 1993-1995. Environmental Health Perspectives, 106:849-855.

Borja-Aburto VH, Loomis DP, Bangidiwala SI, Shy CM, Rascon-Pacheco RA (1997) Ozone, suspended particulates, and daily mortality in Mexico City. American Journal of Epidemiology, 145:258-268.
Braga AL, Zanobetti A, Schwartz J (2001) The lag structure between particulate air pollution and respiratory and cardiovascular deaths in 10 US cities. Journal of Occupational and Environmental Medicine, 43:927-933.
Brunekreef B (1997) Air pollution and life expectancy: is there a relation? Occupational and Environmental Medicine, 54:781-784.
Burnett RT, Smith-Doiron M, Stieb D, Cakmak S, Brook JR (1999) Effects of particulate and gaseous air pollution on cardiorespiratory hospitalizations. Archives of Environmental Health, 54:130-139.
Burney PG, Luczynska C, Chinn S, Jarvis D (1994) The European Community Respiratory Health Survey. European Respiratory Journal, 7:954-960.
California Air Resources Board (2002) Staff report: Public hearing to consider amendments to the ambient air quality standards for particulate matter and sulfates. California Air Resources Board and Office of Environmental Health Hazard Assessment, Sacramento, CA.

Calle EE, Rodriguez C, Jacobs EJ et al. (2002) The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. Cancer, 94:2490-2501.

Castillejos M, BorjaAburto VH, Dockery DW, Gold DR, Loomis D (2000) Airborne coarse particles and mortality. Inhalation Toxicology, 12:61-72.
CIESIN (2000) Gridded population of the world (GPW), version 2. Center for International Earth Science Information Network (CIESIN), Columbia University; International Food Policy Research Institute (IFPRI), World Resources Institute (WRI) Columbia University, Palisades, NY. Available at http://sedac.ciesin.columbia.edu/plue/gpw.
Chay KY, Greenstone M (1999) The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. Unpublished document. (Working Paper No. 17.) Available from: Center for Labor Economics, University of California, Berkeley, CA.
Chay KY, Greenstone M (2001) Air quality, infant mortality, and the Clean Air Act of 1970. Unpublished document. (Working Paper No. 42.) Available from: Center for Labor Economics, University of California, Berkeley, CA.
Chhabra SK, Chhabra P, Rajpal S, Gupta RK (2001) Ambient air pollution and chronic respiratory morbidity in Delhi. Archives of Environmental Health, 56:58-64.
Cifuentes L, Borja-Aburto VH, Gouveia N, Thurston G, Davis DL (2001) Assessing the health benefits of urban air pollution reductions associated with climate change mitigation (2000-2020): Santiago, São Paulo, Mexico City, and New York City. Environmental Health Perspectives, 109: S419-425.
Cifuentes L, Vega J, Kopfer K, Lava LB (2000) Effect of the fine fraction of particulate matter versus the coarse mass and other pollutants on daily mortality in Santiago, Chile. Journal of the Air Waste Management Association, 50:1287-1298.

Clancy L, Goodman P, Sinclair H, Dockery DW (2002) Effect of air pollution control on death rates in Dublin, Ireland: an intervention study. The Lancet, 360:1210-1214.

Cohen AJ, Nikula K (1999) Health effects of diesel exhaust: laboratory and epidemiologic studies. In: Air pollution and health. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.
COMEAP (1998) Quantification of the effects of air pollution on health in the United Kingdom. Department of Health Committee on the Medical Effects of Air Pollutants. Stationery Office, London.
COMEAP (2001) Statement on long term effects of particles on mortality. Available at http://www.doh.gov.uk/comeap/state.htm, or as a printable document at http://www.doh.gov.uk/comeap/statementsreports/longtermeffects.pdf.

Conceiao GMS, Miraglia SGEK, Kishi HS, Saldiva PNH, Singer JM (2001) Air pollution and child mortality: a time series study in São Paulo, Brazil. Environmental Health Perspectives, 109:347-350.

Daniels MJ, Dominici F, Samet JM, Zeger SL (2000) Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. American Journal of Epidemiology, 152:397-406.
Derwent RG (1999) Atmospheric chemistry. In: Air pollution and health. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.
Diesel Working Group (1995) Diesel exhaust: a critical analysis of emissions, exposure, and health effects. Special report. Health Effects Institute, Cambridge, MA.
Dockery DW, Pope AC, Xu X et al. (1993) An association between air pollution and mortality in six US cities. New England Journal of Medicine, 329: 1753-1759.

Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJL, Comparative Risk Assessment Collaborating Group (2002). Selected major risk factors and global and regional burden of disease. The Lancet, 360: 1347-1360.

Gao J (1993) [Relationship between air pollution and mortality in Dongcheng and Xicheng Districts, Beijing]. Chung-Hua Yu Fang i Hsueh Tsa Chih [Chinese Journal of Preventive Medicine], 27:340-343.
Gouveia N, Fletcher T (2000) Time series analysis of air pollution and mortality: effects by cause, age and socioeconomic status. Journal of Epidemiology and Community Health, 54:750-755.

Grant LD, Shoaf CR, Davis JM (1999) Sources of air pollution. In: Air pollution and health. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.

Grønskei KE, Gram F, Hagen LO et al. (1997a) URBAIR Urban air quality management strategy in Asia: Jakarta report. Shah J, Nagpal T, eds. World Bank, Washington, DC.
Grønskei KE, Gram F, Hagen LO et al. (1997b) URBAIR Urban air quality management strategy in Asia: Kathmandu Valley report. Shah J, Nagpal T, eds. World Bank, Washington, DC.
Hajat S, Anderson HR, Atkinson RW, Haines A (2002) Effects of air pollution on general practitioner consultations for upper respiratory diseases in London. Occupational and Environmental Medicine, 59:294-299.
Hajat S, Haines A, Atkinson RW, Bremner SA, Anderson HR, Emberlin J (2001) Association between air pollution and daily consultations with general practitioners for allergic rhinitis in London, United Kingdom. American Journal of Epidemiology, 153:704-714.
Hales S, Salmond C, Town GI, Kjellstrom T, Woodward A (2000) Daily mortality in relation to weather and air pollution in Christchurch, New Zealand. Australia and New Zealand Journal of Public Health, 24: 89-91.

Health Effects Institute (2000a) National morbidity, mortality and air pollution study. (HEI Report No. 94, Part 1. Methods and methodologic issues) Health Effects Institute. Cambridge, MA.

Health Effects Institute (2000b) National morbidity, mortality and air pollution study. (HEI Report No. 94. Part 2.) Health Effects Institute. Cambridge, MA.
Health Effects Institute (2001) Airborne particles and health: HEI epidemiologic evidence. HEI perspectives. Health Effects Institute, Cambridge, MA.

Health Effects Institute (2002) Understanding the health effects of components of the particulate matter mix: progress and next steps. HEI Perspectives. Health Effects Institute, Boston, MA.
Health Effects Institute (2003) Revised analyses of time-series studies of air pollution and health. Health Effects Institute. Boston, MA.
Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR (2002) Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: an intervention study. The Lancet, 360: 1646-1652.
Heinrich J, Hoelscher B, Wichmann HE (2000) Decline of ambient air pollution and respiratory symptoms in children. American Journal of Respiratory and Critical Care Medicine, 161:1930-1936.

Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt PA (2002) Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. The Lancet, 360:1203-1209.
Holman C (1999) Sources of air pollution. In: Air pollution and health. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.
Hong C (1995) Global burden of disease from air pollution. World Health Organization, Geneva. Available at http://www.cru.uea.ac.uk/~markn/cru05/cru05_intro.html.
Hong YC, Leem JH, Ha EH (1999) Air pollution and daily mortality in Inchon, Korea. Journal of Korean Medical Science, 14:239-244.
Hong YC, Leem JH, Ha EH, Christiani DC (1999) PM10 exposure, gaseous pollutants, and daily mortality in Inchon, South Korea. Environmental Health Perspectives, 107:873-888.

Hurley F, Holland MR, Markandya A et al. (2000) Towards assessing and costing the health impacts of ambient particulate air pollution in the UK. (Report TM/00/07.) Institute of Occupational Medicine, Edinburgh.
Instituto Nacional de Ecología (2000) Almanaque de datos y tendencias de la calidad del aire en ciudades Mexicanas [Temperature data and trends in Mexican cities], Dirección General de Gestión e Información Ambiental del Instituto Nacional de Ecología, Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAP) [The Secretariat of Environment and Natural Resources], Mexico. Unpublished document, available at http://www.ine.gob.mx.
International Energy Agency (2001a) Energy balance of non-OECD countries 1995-1999. International Energy Agency, Vienna.
International Energy Agency (2001b) Energy Balance of OECD Countries 1995-1999. International Energy Agency, Vienna.

Jerrett M, Burnett RT, Kanaroglou P et al. (2001) A GIS-environmental justice analysis of particulate air pollution in Hamilton, Canada. Environment and Planning A, 33:955-973.

Katsouyanni K, Schwartz J, Spix C et al. (1996) Short term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol. Journal of Epidemiology and Community Health, 50:S12-18.
Katsouyanni K, Touloumi G, Samoli E et al. (2001) Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. Epidemiology, 12:521-531.

Krewski D, Burnett RT, Goldberg MS et al. (2000) Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of particulate air pollution and mortality. A special report of the institute's particle epidemiology reanalysis project. Health Effects Institute, Cambridge, MA.
Krzyzanowski M, Schwela D (1999) Patterns of air pollution in developing countries. In: Air pollution and health. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.
Krzyzanowski M, Wojtyniak B (1991) Air pollution and daily mortality in Cracow. Public Health Reviews, 19:73-81.

Künzli N, Kaiser R, Medina S et al. (2000) Public-health impact of outdoor and traffic-related air pollution: a European assessment. The Lancet, 356: 795-801.

Künzli N, Medina S, Kaiser R et al. (2001) Assessment of deaths attributable to air pollution: should we use risk estimates based on time series or cohort studies? American Journal of Epidemiology, 153:1050-1055.
Lee J-T, Kim H, Hong YC, Kwon HJ, Schwartz J, Christiani DC (2000) Air pollution and daily mortality in seven major cities of Korea, 1991-1997. Environmental Research, 84:247-254.
Lee J-T, Shin D, Chung Y (1999) Air pollution and daily mortality in Seoul and Ulsan, Korea. Environmental Health Perspectives, 107:149-154.
Leksell I, Rabl A (2001) Air pollution and mortality: quantification and valuation of years of life lost. Risk Analysis, 21:843-857.
Levy JI, Hammitt JK, Spengler JD (2000) Estimating the mortality impacts of particulate matter: what can be learned from between-study variability? Environmental Health Perspectives, 108:109-117.

Lipfert FW, Perry HM Jr, Miller JP, Baty JD, Wyzga RE, Carmody SE (2003) Air pollution, blood pressure, and their long-term associations with mortality. Inhalation Toxicology, 15:493-512.
Loomis D, Castillejos M, Gold DR et al. (1999) Air pollution and infant mortality in Mexico City. Epidemiology, 10:118-123.
McDonnell WF, Nishino-Ishikawa N, Petersen FF, Chen LH, Abbey DE (2000) Relationships of mortality with the fine and coarse fractions of long-term ambient PM10 concentrations in nonsmokers. Journal of Exposure Analysis and Environmental Epidemiology, 10:427-436.

McMichael AJ, Anderson HR, Brunekreef B, Cohen AJ (1998) Inappropriate use of daily mortality analyses to estimate longer-term mortality effects of air pollution. International Journal of Epidemiology, 27:450-453.
Morgan MG, Henrion M (1998) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, Cambridge.
National EPA of China (2000) Environmental statistical yearbook. China Environmental Publishing House, Beijing.
National Research Council (2002) Estimating the public bealth benefits of proposed air pollution regulations. National Academies Press, Washington, DC.

Navidi W, Lurmann F (1995) Measurement error in air pollution exposure assessment. Journal of Exposure Analysis and Environmental Epidemiology, 5:111-124.

New M, Hulme M, Jones P (1999) Representing twentieth-century space-time climate variability. Part I. Development of a 1961-90 mean monthly terrestrial climatology, Journal of Climate, 12:829-856.
Nyberg F, Gustavsson P, Jarup L et al. (2000) Urban air pollution and lung cancer in Stockholm. Epidemiology, 11:487-495.
O’Neill M, Jerrett M, Kawachi I et al. (2003) Health, wealth and air pollution. Environmental Health Perspectives, 111:1861-1870.
Ostro B, Chestnut L (1998) Assessing the health benefits of reducing particulate matter air pollution in the United States. Environmental Research, 76:94-106.
Ostro B, Chestnut L, Vichit-Vadakan N, Laixuthai A (1999a). The impact of particulate matter on daily mortality in Bangkok, Thailand. Journal of the Air and Waste Management Association 49:100-107.
Ostro B, Hurley S, Lipsett MJ (1999b) Air pollution and daily mortality in the Coachella Valley, California: a study of PM10 in an area dominated by coarse particles. Environmental Research, 81:231-238.
Ostro BD, Sanchez JM, Aranda C, Eskeland GS (1996) Air pollution and mortality: results from a study of Santiago, Chile. Journal of Exposure Analysis and Environmental Epidemiology, 6:97-114.
Ozkaynak H (1999) Exposure assessment. In: Air pollution and health. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.
Pandey KD, Wheeler D, Ostro B, Deichmann U, Hamilton K, Bolt K (forthcoming) Ambient particulate matter concentrations in residential areas of world cities: new estimates based on global model of ambient particulates (GMAPS). The Development Research Group and the Environment Department, World Bank, Washington, DC.
Pereira LA, Loomis D, Conceicao GM et al. (1998) Association between air pollution and intrauterine mortality in São Paulo, Brazil. Environmental Health Perspectives, 106:325-329.
Peters A, Dockery DW, Muller JE, Mittleman MA (2001) Increased particulate air pollution and the triggering of myocardial infarction. Circulation, 103: 2810-2815.

Pope CA III (1989) Respiratory disease associated with community air pollution and a steel mill, Utah Valley. American Journal of Public Health, 79:623-628.

Pope CA III, Burnett RT, Thun MJ et al. (2002) Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Journal of the American Medical Association, 287:1132-1141.
Pope CA III, Dockery DW (1999) Epidemiology of particle effects. In: Air pollution and health. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.
Pope CA III, Thun MJ, Namboodiri MM et al. (1995) Particulate air pollution as a predictor of mortality in a prospective study of US adults. American Journal of Respiratory and Critical Care Medicine, 151:669-674.

Qian Z, Chapman RS, Tian Q, Chen Y, Lioy PJ, Zhang J (2000) Effects of air pollution on children's respiratory health in three Chinese cities. Archives of Environmental Health, 55:126-133.
Qian Z, Zhang J, Wei F, Wilson WE, Chapman RS (2001) Long term ambient air pollution levels in four Chinese cities: inter-city and intra-city concentration gradients for epidemiological studies. Journal of Exposure Analysis and Environmental Epidemiology, 11:341-351.
Reddy KS, Yusuf S (1998) Emerging epidemic of cardiovascular disease in developing countries. Circulation, 97:596-601.
Roemer W, Clench-Aas J, Englert N et al. (1999) Inhomogeneity in response to air pollution in European children (PEACE project). Occupational and Environmental Medicine, 56:86-92.

Romieu I, Samet JM, Smith KR, Bruce N (2002) Outdoor air pollution and acute respiratory infections among children in developing countries. Journal of Occupational and Environmental Medicine, 44:640-649.

Saldiva PHN, Lichtenfels AJFC, Paiva PSO et al. (1994) Association between air pollution and mortality due to respiratory diseases in children in São Paulo, Brazil: a preliminary report. Environmental Research, 65:218-225.
Saldiva PH, Pope CA, III, Schwartz J et al. (1995) Air pollution and mortality in elderly people: a time-series study in São Paulo, Brazil. Archives of Environmental Health, 50:159-163.
Samet JM (1999) Air pollution and risk assessment. In: Air pollution and health. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.
Samet JM, Cohen AJ (1999) Air pollution and lung cancer. In: Air pollution and health. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.

Sanhueza P, Vargas C, Jimenez J (1999) Daily mortality in Santiago and its relationship with air pollution. Revista Medica de Chile, 127:235-242.
Sarnat JA, Schwartz J, Catalano PJ, Suh HH (2001) Gaseous pollutants in particulate matter epidemiology: confounders or surrogates? Environmental Health Perspectives, 109:1053-1061.
Schwartz J (2000) Harvesting and long-term exposure effects in the relation between air pollution and mortality. American Journal of Epidemiology, 151:440-448.

Schwartz J, Laden F, Zanobetti A (2002) The concentration-response relation between PM(2.5) and daily deaths. Environmental Health Perspectives, 110:1025-1029.

Simpson R, Denison L, Petroeschevsky A, Thalib L, Williams G (2000) Effects of ambient particle pollution on daily mortality in Melbourne, 1991-1996. Journal of Exposure Analysis and Environmental Epidemiology, 10:488-496.
Smith KR, Mehta S (2003) The burden of disease from indoor air pollution in developing countries: comparison of estimates. International Journal of Hygiene and Environmental Health, 206:279-89.
Sommer H, Künzli N, Seethaler R et al. (2000) An impact assessment project of Austria, France and Switzerland. Ancillary benefits and costs of greenhouse gas mitigation (Proceedings from an OECD Expert Workshop). OECD, New York.

Stanfield JP (1993) Some aspects of the long-term effects of malnutrition on the behaviour of children in the Third World. Proceedings of the Nutritional Society, 52:201-210.
Stieb DM, Judek S, Burnett RT (2003) Meta-analysis of time-series studies of air pollution and mortality: update in relation to the use of generalized additive models. Journal of the Air and Waste Management Association, 53:258261.

Szafraniec K, Tecza W (1999) The effect of short-term changes in levels of air pollution on mortality from cardiovascular diseases among inhabitants of Krakow [in Polish]. Przeglad Lekarski, 56:698-703.

Tager IB, Künzli N, Lurmann F, Ngo L, Segal M, Balmes J (1998) Methods development for epidemiologic investigations of the health effects of prolonged ozone exposure. Part II. An approach to retrospective estimation of lifetime ozone exposure using a questionnaire and ambient monitoring data (California sites). (Research Report No. 81.) Health Effects Institute, Cambridge, MA.

Tellez-Rojo MM, Romieu I, Ruiz-Velasco S, Lezana MA, Hernandez-Avila MM (2000) Daily respiratory mortality and PM10 pollution in Mexico City: importance of considering place of death. European Respiratory Journal, 16:391-396.

UN (2001) Demographic yearbook 1998. (United Nations Publication, Sales No. E/F.97.XIII.1.) Data also available at http://unstats.un.org/unsd/citydata/.
U.S. Environmental Protection Agency (1996) Review of the national ambient air quality standards for particulate matter: policy assessment of scientific and technical information: OAQPS staff paper (EPA-452\R-96-013.) Office of Air Quality Planning and Standards, Research Triangle Park, NC.
U.S. Environmental Protection Agency (1997) National ambient air quality standards for particulate matter: final rule. (CFR Part 50, Federal register 62:1-54.) U.S. Environmental Protection Agency, Washington, DC.
U.S. Environmental Protection Agency (2002) Third external review draft of air quality criteria for particulate matter, April 2002. (EPA/600/p-99/002aC.) Office of Research and Development, Washington, DC.

USGS (1996) GTOPO30—Global digital elevation model, US geological survey. Eros Data Center, Sioux Falls, SD. Available at http://edcdaac.usgs.gov/gtopo30/gtopo30.asp.

WHO (1997) Health and environment in sustainable development: five years after the earth summit. World Health Organization, Geneva.
WHO (2000a) Air quality guidelines for Europe. Second edition. (WHO Regional Publication, European Series, No. 91.) WHO Regional Office for Europe, Copenhagen.
WHO (2000b) Evaluation and use of epidemiological evidence for environmental health risk assessment. (EUR/00/5020369.) WHO Regional Office for Europe, Copenhagen.

WHO (2001a) Health impact assessment of air pollution in the WHO European Region. (Technical Report from WHO/ECEH Project.) WHO European Centre for Environment and Health, Bonn.

WHO (2001b) Quantification of health effects of exposure to air pollution. WHO Regional Office for Europe. Available at http://www.euro.who.int/document/e74256.pdf.
WHO (2001c) Air quality and bealth. Air Management Information System AMIS3.0. World Health Organization, Geneva.
WHO (2002) World health report 2002. Reducing risks, promoting healthy life. World Health Organization, Geneva.
WHO (2003) Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide. WHO Regional Office for Europe, Copenhagen. Available at http://www.euro.who.int/document/e79097.pdf.
Wilhelm M, Ritz B (2003) Residential proximity to traffic and adverse birth outcomes in Los Angeles county, California, 1994-1996. Environmental Health Perspectives, 111:207-216.
Wojtyniak B, Piekarska T (1996) Short term effect of air pollution on mortality in Polish urban populations-what is different? Journal of Epidemiology and Community Health, 50:S36-41.
Wong CM, Atkinson RW, Anderson HR et al. (2002) A tale of two cities: effects of air pollution on hospital admissions in Hong Kong and London compared. Environmental Health Perspectives, 110:67-77.
Wong CM, Ma S, Hedley AJ, Lam TH (2001) Effect of air pollution on daily mortality in Hong Kong. Environmental Health Perspectives, 109:335340.

Woodruff TJ, Grillo J, Schoendorf KC (1997) The relationship between selected causes of post neonatal mortality and particulate air pollution in the United States. Environmental Health Perspectives, 105:608-612.

Working Group on Public Health and Fossil Fuel Combustion (1997) Short-term improvements in public health from global-climate policies on fossil fuel combustion: an interim report. The Lancet, 350:1341-1349.
World Bank (2002) World development indicators 2002. World Bank, Washington.

Xu X, Gao J, Dockery DW, Chen Y (1994) Air pollution and daily mortality in residential areas of Beijing, China. Archives of Environmental Health, 49:216-222.

Xu ZY, Yu DG, Jing LB, Xu XP (2000) Air pollution and daily mortality in Shenyang, China. Archives of Environmental Health, 55:115-120.

Zeger SL, Dominici F, Samet J (1999) Harvesting-resistant estimates of air pollution effects on mortality. Epidemiology, 10:171-175.
Zeger SL, Thomas D, Dominici F et al. (2000) Exposure measurement error in time-series studies of air pollution: concepts and consequences. Environmental Health Perspectives, 108:419-426.
Zhang J, Qian Z, Kong L, Zhou L, Yan L, Chapman RS (1999) Effects of air pollution on respiratory health of adults in three Chinese cities. Archives of Environmental Health, 54:373-381.

Zhang JJ, Hu W, Wei F, Wu G, Korn LR, Chapman RS (2002) Children's respiratory morbidity prevalence in relation to air pollution in four Chinese cities. Environmental Health Perspectives, 110:961-967.

## Chapter I 8

# INDOOR AIR POLLUTION FROM HOUSEHOLD USE OF SOLID FUELS 

Kirk R. Smith, Sumi Mehta and<br>Mirjam Maeusezahl-Feuz

## Summary

This chapter summarizes the methodology used to assess the burden of disease caused by indoor air pollution from household use of solid fuels. Most research into and control of indoor air pollution worldwide has focused on sources of particular concern in developed countries, such as environmental tobacco smoke (ETS), volatile organic compounds from furnishings and radon from soil. Although these pollutants have impacts on health, little is known about their global distribution. Thus, we focus solely on indoor smoke from household use of solid fuels, the most widespread traditional source of indoor air pollution on a global scale.

In order to be consistent with the epidemiological literature, binary classifications of household use of solid fuels (biomass and coal) were used as a practical surrogate for actual exposure to indoor air pollution. Specifically, household solid fuel use was estimated at the national level using binary classifications of exposure to household fuel use, i.e. solid fuel and non-solid fuel (gas, kerosene, electricity). We estimated exposure to smoke from solid fuel by combining a number of national surveys of household fuel use into a regression model that predicts use according to independent, development-related variables, such as income and urbanization. Although this method was necessary owing to the current paucity of quantitative data on exposure, we acknowledge that it overlooks the large variability of exposure within households using solid fuels. As pollution emissions from the use of solid fuel may not always indicate high exposures, we have adjusted exposure estimates by a second term, the ventilation factor, which is based on qualitative measures of ventilation.

Estimates of relative risk obtained from epidemiological studies were combined in meta-analyses for three disease end-points for which there is strong evidence of an association with use of solid fuels: acute lower respiratory infections (ALRI) in children aged $<5$ years, chronic
obstructive pulmonary disease (COPD) and lung cancer (estimates for lung cancer are only for use of coal).

More than 1.6 million deaths and over 38.5 million disability-adjusted life years (DALYs) were attributable to indoor smoke from solid fuels in 2000. Cooking with solid fuels is thus responsible for a significant proportion, about $3 \%$, of the global burden of disease. Although trends are highly uncertain, attributable risks are likely to be greater than avoidable risks.

Several potentially important health outcomes, including tuberculosis, cardiovascular disease, and adverse pregnancy outcomes, were not included, owing to insufficient epidemiological evidence. In addition, there was insufficient evidence to assess the associated health effects for children aged $5-14$ years. The burden of disease caused by use of solid fuel is thus likely to be underestimated.

## 1. Introduction

The use of solid fuels for cooking and heating is likely to be the largest source of indoor air pollution on a global scale. Nearly half the world continues to cook with solid fuels such as dung, wood, agricultural residues and coal. When used in simple cooking stoves, these fuels emit substantial amounts of toxic pollutants. These pollutants, called solid-fuel "smoke" in this chapter, include respirable particles, carbon monoxide, oxides of nitrogen and sulfur, benzene, formaldehyde, 1,3butadiene, and polyaromatic compounds, such as benzo $(\alpha)$ pyrene (Smith 1987). In households with limited ventilation (as is common in many developing countries), exposures experienced by household members, particularly women and young children who spend a large proportion of their time indoors, have been measured to be many times higher than World Health Organization (WHO) guidelines and national standards (Bruce et al. 2000; Smith 1987).

Most research into and control of indoor air pollution worldwide has focused on sources of particular concern in developed countries, such as ETS, volatile organic compounds from furnishings and radon from soil (Table 18.1) (Spengler et al. 2001). Although these pollutants have impacts upon health, little is known about their global distribution.

In an initial attempt to estimate the burden of disease and death caused by indoor sources of air pollution, this chapter focuses solely on the burning of solid fuels. Studies of the health effects of exposure to emissions from the two major sources of energy used for cooking in developed countries, gas and electricity, have been inconsistent, although small but statistically significant increased risks of childhood respiratory disease and other effects associated with use of gas have emerged from meta-analyses (Basu and Samet 1999). This is in contrast to the growing quantity of literature reporting reasonably consistent and strong relationships for a number of health end-points in households burning solid

Table 18.I Major toxic pollutants of indoor air

| Pollutant | Major indoor sources |
| :---: | :---: |
| Fine particles | Fuel/tobacco combustion, cleaning, fumes from food being cooked, e.g. from cooking oil |
| Carbon monoxide | Fuel/tobacco combustion |
| Polycyclic aromatic hydrocarbons | Fuel/tobacco combustion, fumes from food being cooked, e.g. from cooking oil |
| Nitrogen oxides | Fuel combustion |
| Sulfur oxides | Coal combustion |
| Arsenic and fluorine | Coal combustion |
| Volatile and semi-volatile organic compounds | Fuel/tobacco combustion, consumer products, furnishings, construction materials, fumes from food being cooked, e.g. from cooking oil |
| Aldehydes | Furnishing, construction materials, cooking |
| Pesticides | Consumer products, dust from outside |
| Asbestos | Remodelling/demolition of construction materials |
| Lead ${ }^{\text {a }}$ | Remodelling/demolition of painted surfaces |
| Biological pollutants | Moist areas, ventilation systems, furnishings |
| Free radicals and other short-lived, highly reactive compounds | Indoor chemistry |
| Radon | Soil under building, construction materials |
| Lead-containing dust from deteriorating paint is an important indoor pollutant for occup many households, but the most critical exposure pathways are not usually through air. S chapter 19. |  |
| Source: Zhang and Smith (2003). |  |

fuels (biomass or coal), particularly those with poorly-vented stoves and homes, which are common throughout developing countries. In many circumstances, it is difficult to distinguish use of solid fuels for cooking from use for heating the home. There may also be effects associated with the use of kerosene, a common cooking fuel in many parts of the world, for which emissions and exposures are intermediate between those for solid and for gaseous fuels (Smith 1987), but on which few studies of health effects seem to have been conducted.

## 2. Estimating risk factor levels

### 2.1 Exposure variables

One way to determine the health effects of indoor smoke from solid fuels would be to apply the well-established exposure-response relationships from epidemiological studies of outdoor, or ambient, concentrations of
the same pollutants (see chapter 17) to the household exposures, called here the "pollutant-based approach" (Smith and Mehta 2003).

There are a number of potential problems with such an approach, however, including:

- Differences in pollutant mixtures: Although particles are often used as an indicator pollutant, the composition of particles (size, chemical composition, etc.) as well as that of other pollutants varies from source to source, and also changes with dispersion (Rossi et al. 1999).
- Differences in exposure patterns: The daily pattern of indoor air pollution sources varies from that of ambient sources, with large peaks corresponding to cooking and heating schedules (Naeher et al. 2000b).
- Differences in exposure levels: Concentrations of particulates from the indoor combustion of biomass have been measured at levels that are 10-50 times greater than in urban areas of developed countries, where the main epidemiology of pollutants has been performed. Extrapolating exposure-response relationships by such a large factor is problematic, particularly as there are indications that the relationship becomes more shallow at higher exposures (Bruce et al. 2000).
- Relevance of health outcomes addressed: Most studies of outdoor air pollution have attempted to associate short-term changes in exposure with acute health outcomes. This does not address the long-term impact on chronic health outcomes, nor does it necessarily focus on the health outcomes that are responsible for the bulk of the burden of disease. In particular, ALRI, mostly in the form of pneumonia, are likely to be responsible for the largest burden of disease caused by exposure to indoor air pollution.
- Data on concentrations of particulate matter (PM) in indoor air ${ }^{1}$ are sparse. In addition, most measurements have been made for concentrations of total particulates, which are less reliable indicators of risk than smaller particles $\left(\mathrm{PM}_{10}\right.$ or $\left.\mathrm{PM}_{2.5}\right)$.

An alternative approach, consistent with that used in most epidemiological studies in developing countries, is to divide the population into categories of people that are exposed or not exposed to smoke from solid fuel, on the basis of fuel use and ventilation. Although necessary here, owing to the current lack of exposure data, this method overlooks the large variability of exposure within each of these groups (Naeher et al. 2000a). Furthermore, the method based on use of fuel is affected by the first of the shortcomings listed above, as the same broad category of fuels may produce different mixtures of pollutants in different settings. We also recognize that exposures from cooking and heating
can differ considerably because of different conversion technologies. It was not possible to distinguish between the two end-uses in most cases, however.

To account for differences in other factors (e.g. housing) that would affect levels of pollution (Mehta and Smith 2002), we included a second component in the exposure variable, which we refer to as the "ventilation factor". The final exposure variable in the population was defined as:

## Household-equivalent solid-fuel exposed population = (Population using solid fuel $) \times($ Ventilation factor $)$

We compiled a database of household use of solid fuel, from which the prevalence of household use of solid fuel was estimated for each subregion. ${ }^{2}$ Using known values from this database, a statistical model was developed to predict national use of solid fuel for countries without data. Ventilation factors were assigned on the basis of qualitative evidence, to account for differences in types of cooking and heating appliances and housing.

### 2.2 ThEORETICAL-MINIMUM-RISK EXPOSURE DISTRIBUTION

The theoretical minimum for this risk factor is clearly no use of solid fuels for the production of household energy; this has already been achieved in many populations. In reality, of course, there would still be exposure to pollution from liquid and gaseous fuels, which might be further reduced through a switch to use of electricity or of very wellventilated cooking conditions.

### 2.3 A database of household use of solid fuel

A database of households using solid fuel, expressed as a percentage of all households, was compiled for 52 countries in 10 subregions, in order to estimate global household use of solid fuel (see Table 18.2). Although the data were acquired from studies conducted at different times in the past decade, fuel-use patterns are unlikely to have changed drastically within this time frame (International Energy Agency 2002; World Resources Institute 2000). Out of necessity, the data were gathered from various sources using different and, at times, non-validated methodology. We thus had to make many assumptions in order to facilitate subregional comparison and data manipulation associated with solid fuel use. No households were reported to be using solid fuels for cooking in AMR-A, EUR-A, EUR-C and WPR-A, presumably because countries in these subregions have already shifted to cleaner fuels.

In many countries where large proportions of the population cook with solid fuels, data on household energy are widely, although not
Table I8.2 Estimates of data for the database of households using solid fuels

| Subregion | Country | Households using solid fuel (\%) | Type of data source | Year | Reference |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | Algeria | 4 | National energy statistics | 1999 | World Resources Institute (2003) |
|  | Angola | 100 | National energy statistics | 1999 | International Energy Agency (1999) |
|  | Burkina Faso | 97 | Household survey | 1994/1995 | World Bank (2000) |
|  | Chad | 95 | Household survey | 1991 | World Bank (1998) |
|  | Gambia | 98 | Household survey | 1992 | World Bank (2000) |
|  | Ghana | 95 | Household survey | 1997 | World Bank (2000) |
|  | Guinea | 99 | Household survey | 1994/1995 | World Bank (2000) |
|  | Guinea-Bissau | 95 | Household survey | 1992 | World Bank (2000) |
|  | Madagascar | 99 | Household survey | 1993/1994 | World Bank (2000) |
|  | Mali | 100 | Household survey | 1994 | World Bank (2000) |
|  | Mauritania | 69 | Household survey | 1995 | World Bank (2000) |
|  | Niger | 98 | Household survey | 1995 | World Bank (2000) |
|  | Nigeria | 67 | Household survey and census data | 1992 | World Bank (2000) |
|  | Senegal | 79 | Household survey | 1994/1995 | World Bank (2000) |
|  | Sierra Leone | 92 | Household survey | 1989/1990 | World Bank (2000) |
| AFR-E | Botswana | 65 | National census | 1991 | Government of Botswana (1991) |
|  | Central African Republic | 99 | Household survey | 1993 | World Bank (2000) |
|  | Congo | 67 | Household survey | 1988 | World Bank (1988) |
|  | Côte d'lvoire | 93 | Household survey | 1995 | World Bank (2000) |
|  | Democratic Republic of the Congo | 100 | National energy statistics | 1999 | World Resources Institute (2003) |
|  | Ethiopia and Eritrea | 97 | Household survey and census data | 1994 | Government of Ethiopia (1998) |
|  | Kenya | 85 | Household survey | 1994 | World Bank (2000) |
|  | South Africa | 28 | Household survey | 1993 | World Bank (2000) |
|  | Swaziland | 88 | Household survey | 1994 | World Bank (2000) |
|  | United Republic of Tanzania | 96 | Household survey | 1993 | World Bank (2000) |
|  | Uganda | 97 | Household survey | 1992/1993 | World Bank (2000) |
|  | Zambia | 87 | Household survey | 1996 | World Bank (2000) |
|  | Zimbabwe | 67 | National census | 1992 | Government of Zimbabwe (1992) |


| AMR-A | - |
| :--- | :--- |
| AMR-B | Brazil <br> Mexico |
| AMR-D | Ecuador |
| EMR-B | Iran (Islamic Republic of) <br> Lebanon <br>  <br>  <br>  <br>  <br> Libyan Arab Jamahiriya <br> Tunisia |
| EMR-D | Afghanistan <br> Djibouti |
|  | Egypt <br> Iraq |
|  | Morocco <br> Pakistan |
|  | Sudan |
| EUR-A | - |
| EUR-B | Turkey |
| EUR-C | - |
| SEAR-B | Indonesia <br> Thailand |
| SEAR-D | Bangladesh |
|  | India |
|  | Myanmar |
| Nepal |  |

universally, available. In some cases, the data come directly from national census information or energy use statistics, which state explicitly the number or fraction of households that rely predominantly on solid fuels for their energy needs (Government of Botswana 1991; Government of Brazil 1991; Government of Ecuador 1990b; Government of Ethiopia 1998; Government of India 1991b; Government of Mexico 1990b; Government of Nigeria 1990; Government of Zimbabwe 1992). For example, information on the main fuel used for cooking is collected during the house listing of the census of India each decade (Government of India 1991b). These data, disaggregated into urban and rural sectors, are available at the district level (in India, a district contains about 2 million people).

In some countries, where censuses are infrequent and/or data on residential energy use are not collected, household surveys are an important source of information. Some of these household surveys, such as the widely conducted Demographic Health Surveys are repeated, while others may be conducted only once. For example, primary household energy estimates for 22 countries in Africa, based on household surveys with sample sizes ranging from 1000 to $>14000$ households, are included in a database of development indicators for Africa, compiled by the World Bank (2000). In China, data are available in the form of aggregate annual residential fuel consumption at the provincial level, disaggregated by urban and rural areas (Government of China 1996). Cooking and heating energies were distinguished using a simple model that accounted for the average number of "heating days" in each province, based on a 30 -year average from 1951-1980 (Lin 1995). A small amount of energy ( 2 kg -coal equivalent per household per heating day) was considered to be heating fuel and subtracted from the mix of solid fuels in each province. The remaining heating-adjusted cooking fuel was then normalized to "useful energy" using typical conversion efficiencies for each fuel-stove combination reported (Zhang et al. 2000). The proportion of useful cooking energy attributed to each fuel type per household in each province was taken to represent the number of households using that fuel. This analysis was repeated for each of the provinces in China ${ }^{3}$ and aggregated to give a national total. It was estimated that in 1996 nearly $80 \%$ of the households in China used solid fuels.

Many countries produce national estimates of solid-fuel use, but only a minority collect specific information on fuel use at the household level. Evidence from 10 countries (Bangladesh, Ecuador, Indonesia, Mexico, Myanmar, Nepal, Pakistan, the Philippines, Thailand and Viet Nam) indicates that national and household levels of solid-fuel use are highly correlated ( $\mathrm{R}^{2}=0.75$ ). It should be noted, however, that this relationship holds true when solid fuels are not heavily used in industry. This correlation was used to estimate use of solid fuel by households in nine coun-
tries (Afghanistan, Algeria, Egypt, the Islamic Republic of Iran, Lebanon, the Libyan Arab Jamahiriya, Morocco, Tunisia and Turkey) where only information on national use of solid fuel was available. For three countries (Angola, the Democratic Republic of the Congo and the Sudan), in which a large fraction of the total national energy consumed ( $>70 \%$ ) comprised biomass fuels (World Resources Institute 2003), household use of solid fuel was assumed to be $100 \%$. In other countries, including Bangladesh, Indonesia, Myanmar, Nepal, the Philippines, Thailand, Viet Nam and Pakistan (FAO 1997a, 1997b; Government of Indonesia 1995, 1996a; Government of Pakistan 1997), aggregate data on annual residential fuel consumption are available. In these cases, the percentage of households using solid fuels was estimated according to the quantity of fuel consumed.

The fraction of the population of each subregion covered by the countries for which some data were available, and the prevalence of solid-fuel use according to these data are given in Table 18.3. Data on specific types of solid fuel (i.e. use of coal vs biomass) are limited to India and China, but this factor is also likely to be important in other countries in which no estimates were made, including South Africa and Pakistan.

Table 18.3 Estimates of the prevalence of households using solid fuel, by subregion, using the household fuels database

|  | Population covered <br> by available <br> data (000s) | Population covered by <br> available data (\% of total <br> population of subregion) | Households using solid <br> fuel in population <br> covered (\%) |
| :--- | :---: | :---: | :---: |
| Subregion | 260515 | 88.8 | 72.5 |
| AFR-D | 284784 | 84.4 | 84.5 |
| AFR-E | - | - | - |
| AMR-A | 268997 | 62.5 | 24.9 |
| AMR-B | 12646 | 17.7 | 28.1 |
| AMR-D | 86174 | 61.8 | 5.6 |
| EMR-B | 260797 | 73.0 | 66.8 |
| EMR-D | - | - | - |
| EUR-A | 66591 | 30.7 | 10.8 |
| EUR-B | - | - | - |
| EUR-C | 273507 | 93.6 | 64.9 |
| SEAR-B | 1212359 | 97.9 | 83.8 |
| SEAR-D | - | - | - |
| WPR-A | 1433356 |  | 83.8 |
| WPR-B |  |  | 81.1 |
| - No data. |  |  |  |

### 2.4 A model to predict national use of solid fuel

Using known values from the database of households using solid fuel, a statistical model was built to predict national use of solid fuel according to a number of development parameters. The model was then applied to countries where no data on household fuel use existed. This method also allowed for the estimation of statistical uncertainty (i.e. excluding uncertainty in available data and the validity of model) surrounding each prediction. ${ }^{4}$

As a country develops, households gradually switch from using solid fuels to using cleaner liquid and/or gaseous fuels. Although the picture is often more complex at local and household levels, it is assumed here that this generally holds true over the long term on a subregional scale, a trend well-established by current, albeit cross-sectional, international comparisons. After a certain level of economic growth has been achieved, it is assumed that countries will shift away from cooking entirely with solid fuels. The use of solid fuel for heating may continue, however, especially in areas that are rich in coal and wood.

For countries for which data were not available, a model based on the parameters described in Table 18.4 was used with stepwise linear regression. With a gross national product (GNP) of US\$ 4420 per capita

Table 18.4 Parameters in the fuel use prediction model ${ }^{2}$

| Indicator | Source |
| :--- | :--- |
| Solid-fuel use (dependent variable) | Table I8.3 |
| Adult female illiteracy, I998 | World Bank (200I) |
| Average annual growth rate, I998-I999 | World Bank (200I) |
| Dummy variables for all subregions | NA |
| Electricity consumption, per capita, I997 (kilowatt-hours) | World Bank (200I) |
| Fuel-wood production | UN (I993) |
| Population in 2000 | UN (I998) |
| Fuel-wood production per capita (kg) | Author calculation |
| Gini coefficient | World Bank (200I) |
| GNP per capita, I999 | World Bank (200I) |
| In (GNP per capita, I999) | Author calculation |
| Petroleum use per capita | UN (I993) |
| In (petroleum use per capita) | Author calculation |
| Rural population, I999 | World Bank (200I) |
| Traditional fuel use (national), I993 | UN (I993) |
| NA Not applicable. |  |
| Variables already entered were tested for removal at each step, so that variables in the model that |  |
| became insignificant with inclusion of additional variables were removed. Missing values were replaced |  |
| with mean values for each variable. |  |

## Table 18.5 Models to predict fuel use: GNP per capita vs use of traditional fuel as a predictor variable

| Model | Predictors $^{\mathrm{a}}$ | $R$ | $R^{2}$ | ${\text { Adjusted } R^{2}}^{\prime 2}$ |
| :--- | :--- | :---: | :---: | :---: |
| I | Use of traditional fuel, EMR, <br> bapita, rural population, constant |  |  |  |
| 2 | GNP per capita, EMR, petroleum use per capita, <br> rural population, constant | 0.869 | 0.756 | 0.735 |
|  |  | 0.746 | 0.724 |  |

${ }^{\text {a }}$ Dependent variable in both models is the percentage of households using solid fuels.
b Each subregional dummy variable was entered separately into the model. EMR was the only subregional dummy variable that was significant in the final model, perhaps because of a combination of low biomass resources and high access to petroleum fuels in some countries in these subregions.
in 1999, Brazil was the richest country in the database to have significant levels of cooking with solid fuels ( $27 \%$ of households). To avoid extrapolating the model to areas where it may be inappropriate, estimates were made only for countries with a GNP of $<$ US $\$ 5000$ per capita in 1999. All countries with a GNP of $>$ US $\$ 5000$ per capita in 1999 were assumed to have made a complete transition to clean household-cooking systems, either with cleaner liquid or gaseous fuels, or electricity or, where solid fuel was still used for cooking or heating, to fully ventilated appliances.

As use of traditional fuel (as a percentage of national energy use) is highly correlated with GNP per capita, stepwise linear regression eliminates GNP per capita when both variables are entered together. If use of traditional fuel is not entered, it is essentially replaced by GNP per capita in the model, with little impact on model fit or standard error (Table 18.5). Two models to predict fuel use were assessed, one employing GNP per capita and the other use of traditional fuel (as a percentage of national energy use) as predictor variables. Use of traditional fuel, which includes use of fuel-wood, bagasse (biomass remaining after processing sugar-cane), charcoal, animal wastes, agricultural residues, and other vegetable biomass wastes, is expressed as a percentage of total fuel use at the national (as opposed to the household) level, on an energy-equivalent basis. Like household use of solid fuel, use of traditional fuel at the national level is highly correlated with GNP per capita (Figures 18.1 and 18.2).

Information on GNP per capita is more reliable, is updated more routinely, and is available at the national level for nearly all countries. Therefore, we used the model including GNP per capita as a predictor, rather than the model using use of traditional fuel. The final model is shown in Table 18.6 and includes percentage of the rural population, GNP per capita (log-transformed), petroleum use per capita, and location within the EMR subregions (entered as a dummy variable). Other

Figure 18.I The relationship between use of traditional fuel at the national level (as a fraction of national energy use) and GNP per capita


Table 18.6 Final model used to predict household use of solid fuel at the national level ${ }^{2}$

|  | $\begin{array}{c}\text { Unstandardized } \\ \text { coefficients }\end{array}$ |  | Standardized coefficients |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |$)$

[^63]Figure 18.2 Relationship between use of traditional fuel at the household level and GNP per capita

potential variables were dropped from the model in stepwise linear regression.

This model was used to predict percentages of households using solid fuel in all countries where these values were unknown (see Figure 18.3). In order to force the percentage of households using solid fuel to lie between $0 \%$ and $100 \%$, estimates for the 23 countries with predicted values of $<0$ or $>100$ were converted to 0 and 100 , respectively.

Known (for all countries in the household fuel-use database) and predicted estimates of use of solid fuel at the country level were aggregated into subregional estimates of household solid-fuel use (Table 18.7). The subregions with the least coverage are those that have the highest levels of economic development, i.e. those subregions that are least likely to have high proportions of household solid-fuel use because people have, for the most part, already shifted to cleaner fuels and cooking technologies.

We assumed that the fraction of the population exposed is the same as the fraction of households using solid fuel. This assumption is likely

Figure 18.3 Household use of solid fuel, at the national level, 2000


Note: Household solid fuel use estimates are predictions in areas without striations.

Table 18.7 Estimated household use of solid fuel, by subregion

| Subregion | Subregional population (000s) | Total population covered by fuel use prediction and by available data |  | Household use of solidfuel (\% of population) |
| :---: | :---: | :---: | :---: | :---: |
|  |  | n (000s) | \% | Point estimate |
| AFR-D | 293440 | 292317 | 99.6 | 73.4 (68.1-77.7) |
| AFR-E | 337547 | 333697 | 98.9 | 85.8 (80.5-89.2) |
| AMR-A | 320704 | 11201 | 3.5 | 1.5 (0.9-2.0) |
| AMR-B | 430674 | 388897 | 90.3 | 24.6 (18.8-30.8) |
| AMR-D | 71318 | 71318 | 100.0 | 52.9 (42.6-63.2) |
| EMR-B | 139532 | 145137 | 100.0 | 6.1 (2.0-12.1) |
| EMR-D | 357476 | 278909 | 78.0 | 55.2 (49.8-60.1) |
| EUR-A | 410714 | 10689 | 2.6 | 0.2 (0.0-0.5) |
| EUR-B | 216930 | 216930 | 100.0 | 41.5 (32.0-50.7) |
| EUR-C | 245688 | 245688 | 100.0 | 22.8 (13.9-41.0) |
| SEAR-B | 292334 | 292334 | 100.0 | 66.5 (6I.1-71.8) |
| SEAR-D | 1238808 | 1 236398 | 99.8 | 83.5 (78.3-88.3) |
| WPR-A | 153357 | 328 | 0.2 | 0.2 (0.1-0.2) |
| WPR-B | 1528144 | 1479669 | 96.8 | 78.1 (73.0-82.8) |
| World | 6036664 | 5003510 | 82.9 | 56.5 (51.7-61.5) |

to underestimate exposure since solid-fuel-using households are more likely to be rural and of low socioeconomic status, and are thus likely to have more members than the subregional average.

### 2.5 Assigning ventilation factors

Since people in different parts of the world use different cooking and heating appliances and have different types of housing, ventilation must also be taken into account when estimating exposure. Here, the term "ventilation" encompasses both ventilation-related characteristics of the stove (such as the presence of a chimney that vents to the outside of the house) and characteristics of the kitchen (building material, architectural features that influence indoor air quality, location of the kitchen with relation to living area, etc.).

Although we had no data on ventilation conditions according to subregion, we hypothesized that ventilation was a function of climate and development (UNCHS 1996). As described above, countries with a GNP per capita of >US $\$ 5000$ were essentially assigned an estimated exposure of 0 , that is, any use of solid fuel in the household was assumed to be undertaken in fully-vented appliances, with no re-entry of the pollution into the household. In the absence of further information (as described below), all other countries were assigned a ventilation factor of 1.0.

In countries of eastern Europe and the former Soviet Union, a long history of household use of solid fuel under cold climatic conditions and relatively high standards of living, before the recent economic decline, led to the development of energy technologies with far fewer indoor emissions and, consequently, less exposure per unit of solid fuel burned. Therefore, we set the ventilation factor at 0.2 for EUR-B and EUR-C.

In China, the widespread national improved-stove programme has disseminated cooking stoves with chimneys to three-quarters of rural households using solid fuel since 1981 (Goldemberg et al. 2000; Smith et al. 1993), resulting in decreased effective exposure. The ventilation factor for China was set at 0.25 for child health outcomes, because even welloperating, improved biomass stoves with chimneys are still responsible for some exposure (Sinton et al. 1995). We set China's ventilation factor at 0.5 for adult health outcomes, as current disease patterns for adults partly reflect exposure before the introduction of improved stoves. India, the only other country with a long-term national stove-improvement programme, has had only mixed success, with relatively low stove lifetimes and national coverage (NCAER 2002). The ventilation factor was therefore maintained at 1.0 for India.

Tables 18.8 and 18.9 detail estimated exposures as defined above for children aged $<5$ years and for adults. Separate estimates of exposure resulting from use of coal are presented in Table 18.10 for adults only, as adults are affected by chronic health outcomes (see section 3 ).

Table 18.8 Exposure of children (aged <5 years) to indoor smoke from solid fuels

| Subregion | Household solid-fuel <br> use (\%) | Ventilation factor | Exposure (\% population) <br> Point estimate (95\% Cl) |
| :--- | :---: | :---: | :---: |
| AFR-D | 73.4 | 1.00 | $73.4(68.1-77.7)$ |
| AFR-E | 85.8 | 1.00 | $85.8(80.5-89.2)$ |
| AMR-A | 1.5 | 1.00 | $1.5(0.9-2.0)$ |
| AMR-B | 24.6 | 1.00 | $24.6(18.8-30.8)$ |
| AMR-D | 52.9 | 1.00 | $52.9(42.6-63.2)$ |
| EMR-B | 6.1 | 1.00 | $6.1(2.0-12.1)$ |
| EMR-D | 55.2 | 1.00 | $55.2(49.8-60.1)$ |
| EUR-A | 0.2 | 0.97 | $0.0(0.0-0.5)$ |
| EUR-B | 41.5 | 0.65 | $26.0(20.6-31.1)$ |
| EUR-C | 22.8 | 0.25 | $7.2(5.0-11.3)$ |
| SEAR-B | 66.5 | 1.00 | $66.5(61.1-71.8)$ |
| SEAR-D | 83.5 | 1.00 | $83.5(78.3-88.3)$ |
| WPR-A | 0.2 | 1.00 | $0.2(0.1-0.2)$ |
| WPR-B | 78.1 | 0.37 | $28.0(26.1-29.6)$ |

Table I8.9 Exposure of adults (aged $\geq 15$ years) to indoor smoke from solid fuels

| Subregion | Household solid-fuel <br> use (\%) | Ventilation factor | Exposure (\%) <br> Point estimate (95\% CI) |
| :--- | :---: | :---: | :---: |
| AFR-D | 73.4 | 1.00 | $73.4(68.1-77.7)$ |
| AFR-E | 85.8 | 1.00 | $85.8(80.5-89.2)$ |
| AMR-A | 1.5 | 1.00 | $1.5(0.9-2.0)$ |
| AMR-B | 24.6 | 1.00 | $24.6(18.8-30.8)$ |
| AMR-D | 52.9 | 1.00 | $52.9(42.6-63.2)$ |
| EMR-B | 6.1 | 1.00 | $6.1(2.0-12.1)$ |
| EMR-D | 55.2 | 1.00 | $41.4(37.4-45.1)$ |
| EUR-A | 0.2 | 0.97 | $0.0(0.0-0.5)$ |
| EUR-B | 41.5 | 0.65 | $26.0(20.6-31.1)$ |
| EUR-C | 22.8 | 0.25 | $7.2(5.0-11.3)$ |
| SEAR-B | 66.5 | 1.00 | $66.5(61.1-71.8)$ |
| SEAR-D | 83.5 | 1.00 | $83.5(78.3-88.3)$ |
| WPR-A | 0.2 | 1.00 | $0.2(0.1-0.2)$ |
| WPR-B | 78.1 | 0.58 | $44.7(41.7-47.4)$ |

Table 18.10 Exposure of adults (aged $\geq 15$ years) to coal smoke ${ }^{\text {a }}$
Exposure (\%)

| Subregion | Point estimate (95\% CI) |
| :--- | :---: |
| SEAR-D | $2.1(0.0-7.1)$ |
| WPR-B | $12.9(7.9-17.9)$ |

a Assumed to be zero in all other subregions owing to lack of disaggregated data.

### 2.6 Quantitative and Qualitative sources of uncertainty

Estimates of use of solid fuel for countries in the household fuel-use database were arbitrarily assigned an uncertainty range of $5 \%$. The exposure classification system used here is binary (exposed to solid fuels or not exposed), which is consistent with the available epidemiological literature. In reality, exposure to indoor air pollution from use of solid fuel results in a wide range of exposures, which vary according to different types and quality of fuel and stove housing characteristics (e.g. ventilation and size), cooking and heating methods, differences in timeactivity patterns (time spent within the household and in close proximity to the pollution source) and season (Saksena et al. 1992). Since the distribution of exposures is continuous, exposures would best be characterized as a continuous outcome, or at least better characterized by multiple categories. As a result, the above binary categorization and uncertainty values significantly underestimate the true uncertainty in levels of exposure. In addition, the need to use the fuel-prediction model for countries without data obviously introduces uncertainty, only part of which may be reflected in the variance of the results obtained from the model.

## 3. Estimating Risk Factor-Disease RELATIONSHIPS

### 3.1 Health outcomes: EVIDENCE FOR CAUSALITY and INCLUSION CRITERIA

Health outcomes caused by indoor exposure to smoke from use of solid fuel were chosen after a review of the epidemiological evidence available for each end-point, using electronic databases, including Medline and TCMLARS (Traditional Chinese Medical Literature Analysis and Retrieval System, an electronic database of Chinese journals). In addition, given that a large body of evidence comes from developing countries, literature was also obtained from other researchers and reputable developing-country journals not currently indexed in international databases. Only articles written or abstracted in English were used, except for articles on lung cancer, for which both the Chinese and the English

Table 18.1I Diseases associated with use of solid fuels and populations affected that were included in the analysis

| Disease | Population affected |
| :--- | :--- |
| Acute lower respiratory infections (ALRI) | Children aged $<5$ years |
| Chronic obstructive pulmonary disease (COPD) | Females and males aged $\geq 30$ years |
| Lung cancer (coal use only) | Females and males aged $\geq 30$ years |

literature were accessed, since, to our knowledge, only in China has there been significant use of coal in unvented household devices in recent decades.

## GENERAL ASSESSMENT OF CAUSALITY

The strength of the evidence for each end-point was determined on the basis of a structured assessment of causality, using Bradford Hill's criteria for causality, including temporal relationship, strength of association, specificity, the presence of a dose-response relationship, biological plausibility, coherence, the existence of experimental evidence and consistency of association.

As specificity, dose-response relationships, and experimental evidence are often difficult to assess for environmental exposures and health outcomes with multiple causes or long latency periods, we used the epidemiological evidence in conjunction with available information on emissions, exposures and mechanisms for indoor air pollution (Smith et al. 2000; Zelikoff et al. 2003). Three health outcomes were determined to have strong enough evidence to be included: ALRI, COPD and lung cancer (Table 18.11). Information on assessing causality for these outcomes is given in section 3.3 and excluded outcomes are discussed in section 3.2.

Children aged $>5$ years (of school-age) were excluded as they spend less time in the house than women and children aged $<5$ years; this is a conservative assumption as there is some exposure of this group, although levels are unknown on a global scale (Ezzati and Kammen 2001; Saksena et al. 1992). Because of the limitations of the available epidemiological studies, only risks in young children (aged $<5$ years) and adults were included. Available data indicate that men are also affected by those outcomes considered for women, but presumably at lower risks than women because of lower exposures. Adults aged 15-30 years were excluded because the chronic diseases of concern (COPD and lung cancer) have not yet become manifest in this group. Obviously, however, development of these diseases in later years is partly caused by exposures at these and younger ages.

### 3.2 Excluded health outcomes

## OUTCOMES WITH INSUFFICIENT EVIDENCE

A number of important diseases that are potentially associated with use of solid fuels have not been included in this analysis owing to insufficient or lack of direct evidence on causality. Lack of inclusion does not necessarily imply inconclusive findings. Rather, it refers to a relatively small set of findings, suggesting that additional, carefully conducted studies are needed to strengthen the evidence base.

## Asthma

On the basis of the usual measures (concentrations of small particles, $\mathrm{PM}_{2.5}$ ), typical exposures to indoor smoke from use of solid fuels are much higher than those for urban outdoor pollution (García-Marcos et al. 1999) and ETS (Strachan and Cook 1998), with which asthma has been frequently associated. In addition, a study of children aged $<5$ years in Malaysia found increased risk associated with the burning of mosquito coils, another important indoor source of $\mathrm{PM}_{2.5}$ (Azizi et al. 1995). Studies in China (Xu et al. 1996a) and Kenya (Mohamed et al. 1995) have quantitatively associated asthma in children of school age and in adults with various measures of indoor pollution from solid-fuel use. As the reported background rate is low in most developing countries, however, asthma contributes relatively little to the total burden of deaths or DALYs from indoor air pollution.

## Cataracts and other visual impairments

Two case-control studies in India have found an increased risk of cataracts among people using biomass fuel; Mohan et al. (1989) determined an odds ratio of 1.6; Zodpey and Ughade (1999) found an adjusted odds ratio of 2.4. Evaluation of the National Family Health Survey of India (NFHS 1995) found a somewhat lower rate for partial blindness (odds ratio of 1.3; Mishra et al. 1999a), but no significant difference for total blindness. There is also evidence that exposure to ETS is associated with cataracts (West 1992) and animal studies show that cataracts can be caused by wood smoke (Rao et al. 1995; Shalini et al. 1994).

Indoor air pollution may also be linked to blindness through trachoma (Prüss and Mariotti 2000). Two unadjusted studies in the United Republic of Tanzania found such a link (Taylor and West 1989; West and Lynch 1989) although another in Ethiopia found cooking in a central room to be protective, perhaps through reduction of flies (Sahlu and Larson 1992).

## Perinatal effects

One study in India found an adjusted excess risk of stillbirth of $50 \%$ among women using biomass fuels during pregnancy (Mavalankar et al. 1991), and two Chinese studies of urban ambient pollution, from the
same group of researchers, also found a strong relationship between concentrations of particulates and pre-term delivery ( Xu et al. 1995) and low birth weight (Wang et al. 1997). Low birth weight was also found to be associated with household exposure to biomass smoke in Guatemala (Boy et al. 2002). Intrauterine mortality, low birth weight, prematurity, and early infant death have been significantly associated with urban outdoor pollution at much lower concentrations than those typically found in households that use biomass (Bobak 2000; Loomis et al. 1999; Pereira et al. 1998; Ritz and Yu 1999; Scram 1999; Woodruff et al. 1997). Exposure of non-smoking pregnant women to ETS has been associated with low birth weight in a meta-analysis of 17 studies (Windham et al. 1999a), with low cognitive development (Johnson et al. 1999), but not with spontaneous abortion (Windham et al. 1999b).

Low birth weight is a risk factor for a number of childhood (Walsh 1993) and, probably, adult (Barker 1997) diseases, not just those of the respiratory system. The potential pathways by which indoor cooking smoke may cause low birth weight are given in Figure 18.4. Although this mechanism seems likely to be important in some parts of the world,

Figure 18.4 Possible mechanisms for indoor air pollution and low birth weight


[^64]at present it is difficult to provide a quantitative estimate of the potential burden, and it is not attempted here.

## Tuberculosis

Recent studies in India and Mexico have suggested that indoor air pollution from use of solid fuel may be a risk factor for active tuberculosis. A statistically significant relationship has been found between reported use of biomass fuel and incidence of tuberculosis in 260000 adults aged $>20$ years. Indeed, women in households using biomass fuels were found to be 2.7 ( $95 \%$ CI 1.9-4.0) times more likely to have tuberculosis than women in households using cleaner fuels, even after correction for a range of socioeconomic factors (Mishra et al. 1999b). In addition, an unadjusted but significant odds ratio of 2.5 has been reported for clinically-confirmed tuberculosis in adult male and female householders aged 16-60 years using wood or dung cakes as fuel (Gupta and Mathur 1997). Although these studies were not able to address smoking as a possible confounder, two studies in Mexico City have found an association between exposure to wood smoke and incidence of tuberculosis, after taking smoking into account (Perez-Padilla et al. 1996, 2001). A study in China also found exposure to outdoor air pollution to be associated with tuberculosis (Xu et al. 1995). Animal studies have shown that wood smoke causes immune suppression in the respiratory system (Thomas and Zelikoff 1999; Zelikoff 1994).

## Other health effects not included

- Interstitial lung disease has been associated with long-term exposures in several studies (Dhar and Pathania 1991; Gold 2000; Ramage et al. 1988; Sandoval et al. 1993).
- Early studies in Africa seemed to implicate wood smoke as a cause of nasopharyngeal cancer, but this association was not borne out by later studies in Asia (Smith 1987; Smith and Liu 1994).
- Two studies in Brazil have shown increased risk of upper aerodigestive tract cancers, with adjusted odds ratios of 2.7 (Pintos et al. 1998) and 2.5 (Franco et al. 1989).
- An association has been shown with cervical neoplasia among HPVinfected women in Honduras, with an adjusted odds ratio of 5.7 after 35 years or more of cooking over an open fire (Velema et al. 2002).
- Ischaemic heart disease has been associated with exposure to outdoor particulate air pollution (Ponka and Virtanen 1996; Pope et al. 1992; Schwartz 1993; Schwartz and Dockery 1992; Schwartz and Morris 1995) and ETS (Steenland et al. 1998) in some studies, both at much lower levels of exposure than for indoor air pollution (see chapter 17).


## EXCLUDED OUTCOMES ASSOCIATED WITH USE OF SOLID FUEL, BUT NOT CAUSED BY EXPOSURE TO AIR POLLUTION

The use of solid fuels for household cooking and heating involves a range of activities with potential health implications that are separate from those involving the pollution created. The most important involve the harvesting of the two major types of fuel.

- The harvesting of biomass in rural settings in developing countries may involve regular carriage of heavy loads for long distances, with consequent physical strain and food energy demands, along with exposure to such hazards as snake-bite, leeches and assault (crime). Women and children typically bear the greatest burden of such harvesting, although there is much variation across the world.
- Coal mining is one of the most hazardous occupations in the world, particularly in developing countries in small mines from which much household fuel is obtained.

In addition, the extra time taken to harvest, store, and prepare solid fuels is time that is potentially deducted from other pursuits that are associated with health benefits, such as child care or the generation of the household income.

Considering that the counterfactual distribution is cooking with nonsolid fuels (rather than no cooking at all), there are also categories of health risk that are avoided by the use of solid fuels:

- fires and explosions related to household use of liquid and gaseous fuels;
- poisoning caused by ingestion of household kerosene;
- risk inherent in the operation of the national and international petroleum fuel cycles required to provide liquid and gaseous fuels;
- risks involved in providing electricity for household cooking, such as coal mining, air pollution from power plants, accidents involving nuclear and hydroelectric dams, etc.; and
- additional risk of mosquito-borne diseases owing to absence of repellence from household smoke produced by solid fuel.

In its current form, the system limits of this comparative risk assessment (CRA) do not encompass any of these health effects, positive or negative, that are not directly caused by exposure of humans to pollution in the household.

EXCLUDED OUTCOMES ASSOCIATED WITH SPECIALIZED AIRBORNE PRODUCTS OF INDOOR COMBUSTION

There are several related sources of indoor pollution not covered by this analysis that may be locally important in some countries. However, too
few data are available regarding exposures to extrapolate these risks to global burdens, although we suggest that these sources represent potential research topics, as well as priorities for determining exposure distributions, in order to improve the estimated burden of indoor air pollution.

- Smoke from cooking oil: studies in China (including the Province of Taiwan) show relative risks for lung cancer of 3-5 for Chinese-style cooking in a wok with certain cooking oils (Ko 2000; Zhong et al. 1999b).
- Exposures to trace quantities of toxic elements resulting from indoor use of coal in China and elsewhere: significant and widespread impacts from exposures to fluorine and arsenic have been reported in China (Finkelman et al. 1999) and can be expected to occur wherever coal fuels are contaminated with such toxic elements.
- Smoke from incense and mosquito coils, which have been associated with ill-health in some Asian studies (Azizi et al. 1995).


### 3.3 Evidence and exposure-Risk relationships

The estimates of relative risk ${ }^{5}$ and confidence intervals used for ALRI, COPD and lung cancer were derived through formal meta-analyses of the available literature.

Searches of the scientific literature were conducted using the Medline computerized bibliographic database, review of bibliographies from previously-retrieved articles and personal communications. In some cases, the authors of articles that were lacking data that were of interest for this analysis were contacted and asked for clarification, and specific requests for information were sent to researchers in this field.

Medline searches were conducted using the following key words:

- For ALRI: indoor air pollution, household fuel, smoke, acute respiratory infections (ARI), pneumonia and ALRI
- For COPD: indoor air, fuel, COPD, chronic obstructive lung disease (COLD) and chronic bronchitis
- For lung cancer: indoor, air, fuel and lung cancer

To be eligible for inclusion in the meta-analysis, studies had to fulfil the following criteria:

- to be a primary study, not a re-analysis or review;
- to examine some proxy for exposure to indoor smoke from the use of solid fuels for cooking and/or heating purposes;
- to report an odds ratio and its variance, or sufficient data with which to estimate them; and
- to be written or abstracted in English. Additionally, for lung cancer only, a Chinese colleague assisted in a comprehensive search of the Chinese literature, extraction of the relevant data and translation.

We considered both fixed- and random-effects models for the metaanalysis. As the results from both were similar, we used those from the fixed-effects model only. Owing to heterogeneity within studies, we performed sensitivity analyses by stratifying the studies by potential sources of heterogeneity, including assessment of exposure and adjustment for confounders. We did not use a random-effects model, even when statistical significance for heterogeneity was present, for the following reasons.

- Random-effects models assume that studies are selected from a population with a single underlying variance. This would be violated given the heterogeneity among the studies in measuring exposure.
- Random-effects models assign the same weight to small and large studies. This would be problematic for the studies of this analysis because the number of cases ranged from 45 to 500 .

Smoking is an important risk factor for the diseases associated with indoor smoke from use of solid fuel, especially lung cancer and COPD. At present, information on the combined effects of smoking and use of solid fuel is rare. To avoid possible overestimation of the burden of disease, therefore, attributable fractions for lung cancer and COPD caused by use of solid fuel were applied to disease burdens remaining after removal of the burden attributable to smoking. This is conservative in that some of the effect attributable to smoking could also be attributed to use of solid fuel. To ensure internal consistency within the CRA project, burdens attributable to smoking were obtained from chapter 11. Globally, about $51 \%$ and $62 \%$, for men and women respectively, of the total burden of COPD is not attributable to tobacco.

## ACUTE LOWER RESPIRATORY INFECTIONS

A number of studies in developing countries (Argentina, Brazil, the Gambia, India, Kenya, Nepal, Nigeria, South Africa, the United Republic of Tanzania and Zimbabwe) have quantified the relative risk of ALRI for children in households that burn biomass (Armstrong and Campbell 1991; Campbell 1997; Cerqueiro et al. 1990; Collings et al. 1990; de Francisco et al. 1993; Ezzati and Kammen 2001; Johnson and Aderele 1992; Kossove 1982; Mtango et al. 1992; O’Dempsey et al. 1996; Pandey et al. 1989b; Shah et al. 1994; Victora et al. 1994). Some work has also been done to identify possible mechanisms in the developing countries (Verma and Thakur 1995).

Studies among native Americans (Navajos in the south-western United States of America) show a strong and significant association between ALRI and use of wood stoves, at much lower levels of indoor pollution than found in developing countries (Morris et al. 1990; Robin et al.
1996). There is a larger group of studies that show various childhood respiratory symptoms (e.g. cough, wheezing) to be associated with exposure to smoke from solid fuel, but do not provide sufficient evidence to calculate odds ratios of ALRI itself.

As all studies included here used either ARI or ALRI, or death caused by ARI or ALRI, in children aged $<5$ years as a health outcome, we only estimated the burden of disease for children in this age group. A recent study in Kenya (Ezzati and Kammen 2001) found associations between use of solid fuels and ARI in adults (both men and women), suggesting that, once time-activity patterns and spatial dispersion of smoke have been taken into account, men and women may have similar patterns of exposure-response.

A single statistical analysis of all 15 studies identified (Table 18.12) was not appropriate because of the heterogeneous exposure variables and the diverse analytical strategies used by the investigators, especially with respect to potential confounding factors. To address this diversity, different subgroups of these studies were used to conduct several metaanalyses, the results of which were remarkably consistent; pooled relative risk estimates increased with improved precision of exposure measure. ${ }^{6}$

## Characteristics of excluded studies

Of the 15 studies identified (Table 18.12), we excluded the study by Kossove (1982), which had an inappropriately-small comparison group. Two studies in South America focused on use of solid fuels in urban populations (Cerqueiro et al. 1990; Victora et al. 1994). The study in Buenos Aires, Argentina, was excluded owing to a very low prevalence of households using solid fuels and, in one of the case groups, missing data on exposure to heating fuelled by charcoal (Cerqueiro et al. 1990). In the study in Brazil (Victora et al. 1994), only a small proportion of the study population was exposed $(6 \%)$ and exposure was defined loosely, encompassing a wide range of sources of pollution, from open fires to enclosed metal heating stoves and vented fireplaces. The study by Shah et al. (1994) was excluded because its definition of nonexposure (use of stove with chimney provided by the government improved-stove programme) has been shown to produce concentrations of indoor pollutants that were not statistically different from those produced by open fires at that time in India (Ramakrishna et al. 1989) and no observations of direct pollution were made. The study by Mtango et al. (1992) was excluded because, as the study focused on mortality from all causes, no information was given on exposure status for the proportion of deaths caused by ALRI (in this case, pneumonia). Two studies reported on the same study population (Armstrong and Campbell 1991; Campbell et al. 1989). We chose to include the older report by Campbell, which included the odds ratio for girls and boys combined. A recently-published longitudinal study examining rates of episodes of
Table 18.12 Studies on the risk of acute lower respiratory infection associated with use of solid fuels, in children aged $<5$ years

| Study location | Reference | Study design (n) Study population | Exposure assessment | Outcome assessment | Covariates adjusted for | Odds ratio (95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Argentina | Cerqueiro et al. (1990) Excluded | Case-control (616-669) Children aged $<5$ years | Questionnaire: type of cooking fuel used (wood, kerosene, gas) | ALRI within the last 12 days, at a wellbaby clinic | None | 9.9 (1.8-31.4) |
| Brazil (urban) | Victora et al. (1994) Excluded | Case-control (510-5IO) Children aged $<2$ years | Questionnaire: presence of indoor smoke | ALRI hospital cases, clinical signs and X-ray | Smoking, housing, no. of siblings, income, education, history of respiratory illness | I.I (0.6-2.0) |
| Gambia | Armstrong and Campbell (1991) Excluded | Cohort (500) <br> Children aged <5 years | Questionnaire: mother carries child on her back while cooking | ALRI, by weekly home visits | Birth interval, ETS, crowding, socioeconomic status, nutrition, vaccination, education | Males: 0.5 (0.2-I.2) <br> Females: <br> 1.9 (1.0-3.9) |
| Gambia | Campbell et al. (1989) | Cohort (27I) <br> Children aged <l year | Questionnaire: mother carries child on her back while cooking | ALRI, by weekly home visits | Birth interval, ETS, crowding, socioeconomic status, nutrition, vaccination, education | 2.8 (1.3-6.1) |
| Gambia |  |  |  |  |  |  |
| Upper <br> River <br> Division | de Francisco <br> et al. (1993) | Case-control (129-270) Children aged <2 years | Questionnaire: mother carries child on her back while cooking | Death from ALRI by verbal autopsy confirmed by three independent physicians | Socioeconomic status, ETS, maternal education, crowding, nutrition | 5.2 (1.7-15.9) |
| Gambia |  |  |  |  |  |  |
| Upper River Division | O'Dempsey et al. (1996) | Case-control (80-159) Children aged <5 years | Questionnaire: mother carries child on her back while cooking | ALRI hospital cases, clinical signs, X-ray and laboratory | ETS, mother's income, weight slope, recent illness, nutrition | 2.5 (1.0-6.6) |
| India | Shah et al. (1994) <br> Excluded | Case-control (400) Children aged $\leq 5$ years | Household has a smokeproducing stove | Severe ARI hospital cases, clinical symptoms | Smoking, housing, no. of siblings, income, education, birth weight | 1.2 (0.7-2.3) |


| Kenya | Ezzati and Kammen (200I) Excluded | Cohort (93) <br> Children aged <5 years | Mean daily personal $\mathrm{PM}_{10}$ exposure from pollution and time-location data | Rate of ALRI during study period by Integrated Management of Childhood Illness (IMCI) diagnosis criteria | Age, sex, crowding, smoking, village type | 2.93 (1.34-6.39) <br> Highest vs lowest exposure category plus exposureresponse trend |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Nepal | Pandey et al. (1989b) | Cohort (280) <br> Children aged <2 years | Questionnaire: Average time spent near the fireplace | ARI, by bi-weekly home visits | None | 2.3 (1.8-2.9) |
| Nigeria | Johnson and Aderele <br> (1992) | Case-control (103-103) <br> Children aged <5 years | Questionnaire: type of cooking fuel used (wood, kerosene, gas) | ALRI hospital cases, clinical signs, X -ray and laboratory | None | 0.8 (0.4-I.7) |
| South Africa | Kossove (1982) Excluded | Case-control (132-18) Children aged $\leq 1$ year | Questionnaire: does the child stay in the smoke? | ALRI hospital cases, clinical signs and X-ray | None | 4.8 (1.7-13.6) |
| United <br> Republic of Tanzania | Mtango et al. (1992) <br> Excluded | Case-control (456-II60) Children aged <5 years | Questionnaire: child sleeps in room where cooking is done | Death from all causes, by verbal autopsy and physician | Village, age, questionnaire respondent, maternal education, parity, water source, child eating habits | 2.8 (1.8-4.3) |
| USA |  |  |  |  |  |  |
| Arizona | Morris et al. (1990) | Case-control (58-58) Children aged <2 years | Questionnaire: primary source for heating and cooking | ALRI hospital cases, clinical signs and X-ray | Family history of asthma, recent respiratory illness, dirt floor, running water | 4.9 (1.7-12.9) |
| USA |  |  |  |  |  |  |
| Arizona | Robin et al. (1996) | Case-control (45-45) Children aged <2 years | Questionnaire: household uses wood for cooking | ALRI hospital cases | No. of siblings, electricity, running water, difficulty in transport to clinic, ETS, housing | 5.0 (0.6-42.8) |
| Zimbabwe | Collings et al. (1990) | Case-control (244-500) <br> Children aged <3 years | Questionnaire: household uses open wood-fire for cooking | ALRI hospital cases, clinical signs and X-ray | ETS, crowding, housing, number of siblings | 2.2 (1.4-3.3) |

ALRI in a range of age groups across several categories of exposure to smoke from combustion of biomass in Kenya (Ezzati and Kammen 2001) was excluded from the formal meta-analysis because the outcome, expressed as "fraction of weeks with illness", could not be translated into an odds ratio in a manner consistent with the other epidemiological studies. This study did provide strong collaborative evidence, nevertheless, for it showed effects in older children and women as well as in young children and demonstrated a statistically significant trend in the exposure-response relationship. In a subsequent analysis, the authors reported an odds ratio of 2.14 for children exposed to $\mathrm{PM}_{10}$ concentrations of $>1000 \mu \mathrm{~g} / \mathrm{m}^{3}$ (Ezzati 2002).

## Estimating risk factor-disease relationships

After the exclusions noted above, there remained eight studies that reported relative risks of acute respiratory illness for young children exposed to indoor smoke from use of solid fuel (Campbell et al. 1989; Collings et al. 1990; de Francisco et al. 1993; Johnson and Aderele 1992; Morris et al. 1990; O’Dempsey et al. 1996; Pandey et al. 1989b; Robin et al. 1996). Of these, the majority were case-control studies. One study used the outcome "pneumococcal infection", which includes meningitis and septicaemia (O’Dempsey et al. 1996). However, $80 \%$ of patients in this study were diagnosed with pneumonia. Although most of the studies were conducted in developing countries, two were carried out in populations of Navajo and Hopi Indians in the United States (Morris et al. 1990; Robin et al. 1996). The populations in the United States are likely to differ in socioeconomic characteristics from the rest of the studies, thus potentially influencing the rates of incidence of ALRI. As the overall odds ratio did not change substantially with the exclusion or inclusion of these studies, all subsequent analyses included these two studies.

## Exposure assessment used in the studies

The studies provide relatively little information on the indoor concentrations of or exposures to specific pollutants produced by use of solid fuels, or on the baseline concentrations within similarly-constructed households that do not use solid fuels. All but one study used binary classifications of exposure (Table 18.12). On the basis of evidence for an exposure-response relationship between ARI and exposure to smoke from solid fuels (Ezzati and Kammen 2001; Pandey et al. 1989a), we attempted to analyse the studies according to the precision of the exposure measure used and the likely intensity of exposure. Exposure measures used were grouped in three major categories, in what was assumed to be an increasing order of precision: type of fuel used, duration of exposure to smoke from solid fuels, and using solid fuel and carrying the child on the mother's back (Table 18.13). Although it is generally true that concentrations of pollutants are likely to be lower in households using cleaner fuels, such as kerosene or gas, there is a wide variation in con-

Table 18.13 The risk of ALRI associated with use of solid fuels, in children aged $<5$ years: subgroup analyses

| Subgroup analyses | Studies included | Odds ratio (95\% Cl) |
| :---: | :---: | :---: |
| All studies | Campbell et al. (1989); Collings et al. (1990); de Francisco et al. (1993); Johnson and Aderele (1992); Morris et al. (1990); O’Dempsey et al. (1996); Pandey et al. (I989b); Robin et al. (1996) | 2.3 (1.9-2.7) |
| Use of solid fuel | Johnson and Aderele (I992); Collings et al. (1990); <br> Morris et al. (1990); Robin et al. (1996) | 2.0 (1.4-2.8) |
| Duration of time child spent near the cooking fire | Pandey et al. (1989b) | 2.3 (1.8-2.9) |
| Child is carried on the mother's back | Campbell et al. (I989); de Francisco et al. (I993); O'Dempsey et al. (1996) | 3.1 (1.8-5.3) |
| Studies adjusting for nutritional status | Campbell et al. (I989); de Francisco et al. (1993); O'Dempsey et al. (1996) | 3.1 (1.8-5.3) |
| Studies not adjusting for nutritional status | Collings et al. (1990); Johnson and Aderele (1992); Morris et al. (1990); Pandey et al. (1989b); Robin et al. (1996) | 2.2 (1.8-2.6) |
| Children aged $<2$ years old | Campbell et al. (1989); de Francisco et al. (1993); Morris et al. (I990); Pandey et al. (1989b); Robin et al. (1996) | 2.5 (2.0-3.0) |
| Children aged <5 years old | Collings et al. (1990); Johnson and Aderele (1992); O'Dempsey et al. (1996) | 1.8 (1.3-2.5) |

centrations reported in households using solid fuels (Mehta and Smith 2002). Some studies report whether or not children remained indoors when the mother was cooking, but, for reasons noted above, all of these studies were excluded (Awasthi et al. 1996; Kossove 1982; Mtango et al. 1992). Only one study reported the average time that the child spent near the cooking fire (Pandey et al. 1989b). We assumed that carrying the child on the mother's back during cooking represented the most precise measure of exposure, as this suggests that the child was in close proximity to the fire, where exposures are generally higher (although the type of fuel used in control households in these studies was not specified).

We performed separate analyses for each category of exposure, as summarized in Table 18.13. Cooking with wood or other biomass was associated with an odds ratio of $2.0,95 \%$ CI 1.4-2.8. The Pandey study reported an intermediate estimate of relative risk of $2.3,95 \%$ CI 1.8-2.9, for children spending more than two hours near the cooking fire each day. The highest odds ratio was found to be associated with the child being carried on the mother's back during cooking (odds ratio of 3.1, 95\% CI 1.8-5.3).

In only three of the studies were the results adjusted for nutritional status in multivariate analyses, an important confounding variable for

ARI in young children (Victora et al. 1999). The odds ratio found by those studies that did adjust was $3.1,95 \%$ CI 1.8-5.3, whereas the effect was slightly smaller in the studies that did not adjust, with an odds ratio of $2.2,95 \%$ CI $1.8-2.6$. This may be explained, however, by the fact that the studies that controlled for nutrition also used a different exposure proxy (child was carried on mother's back during cooking).

Age is another potential confounding variable because younger children are more likely to remain close to their mothers and are therefore also more likely to be exposed to indoor smoke from cooking or heating, and because age is independently associated with ALRI, with younger children being more susceptible than older children. Most case-control studies adjusted for age by matching controls to cases. When the analysis was restricted to include only studies in children aged $<2$ years, the risk of ALRI was found to be slightly higher (odds ratio of $2.3,95 \%$ CI 1.9-2.7) than that obtained from studies in children aged $\leq 5$ years (odds ratio of $1.6,95 \%$ CI 1.2-2.2). Armstrong and Campbell (1991) noted that, in their study population, girls were more likely to be carried on their mothers' backs than boys and were thus exposed to higher concentrations of pollutants for a longer duration of time. This study found that girls who were carried on the mother's back during cooking had an increased risk of ALRI; no association was observed for boys. The risk in girls was much higher (odds ratio of 6.0 vs odds ratio of 1.9 ) when only the first episode of ALRI (rather than all episodes) was included in the analysis, although the confidence interval was also much wider, owing to the smaller sample. Data were not disaggregated by sex in any of the other studies (although several did control for sex in the multivariate analyses).

As we could not separate the effects of measures of exposure from adjustment for nutritional status, we used the combined odds ratios for all eight studies remaining after exclusions. The results of this approach are similar to those that would be produced if the difference between the most and least precise exposure measures were to be used as the range, i.e. 2.0-3.1 (geometric mean, $\mathrm{GM}=2.4$ ). This is also consistent with the differences in the odds ratios for the two age groups, that is, 1.8 for children aged $<5$ years and 2.5 for children aged $<2$ years. The overall estimate, from all eight studies, of the risk of ALRI in young children exposed to indoor air pollution caused by use of solid fuels was 2.3 , CI 95\% 1.9-2.7.

## Chronic obstructive pulmonary disease

Globally, the most important risk factor for COPD is thought to be smoking of tobacco (NHLBI/WHO 2001) (see also chapter 11). A number of studies have examined various symptoms of chronic respiratory ill-health in women who cook with open stoves burning biomass (Smith 2000). Eight studies in six countries-Bolivia (Albalak et al. 1999), Colombia (Dennis et al. 1996), India (Gupta and Mathur 1997;

Malik 1985), Mexico (Perez-Padilla et al. 1996), Nepal (Pandey 1984b; Pandey et al. 1988) and Saudi Arabia (Døssing et al. 1994)—have quantified the association between indoor air pollution and COPD. Although there are no comparable studies reporting odds ratios in China, the high rates of COPD in non-smoking Chinese women argue that this risk can be related to exposure to coal smoke (Liu et al. 1998).

Cor pulmonale, a heart condition that is secondary to COPD and that is also found among non-smoking rural women in south Asia (Smith 1987), has long been attributed to long-term exposure to smoke from biomass (Padmavati and Pathak 1959). Other studies have attributed silicosis (Norboo et al. 1991; Saiyed et al. 1991), reductions in lung function, cough and various other respiratory conditions to exposure to smoke from biomass, in women, ${ }^{7}$ but were not however included here, owing to limited evidence and the relatively small burden of disease associated with these conditions.

Studies that were included in the meta-analysis used a specific definition of COPD or chronic bronchitis, such as cough and sputum on every day for at least three consecutive months for two successive years, and/or a forced expiratory volume in first second/forced vital capacity (FEV1/FVC) ratio of $<70 \%$ or a FEV1 of $<70 \%$ of the predicted value. We identified 11 studies reporting the relative risks of chronic airway disease in adults exposed to indoor smoke (Albalak et al. 1999; Behera et al. 1991; Dennis et al. 1996; Døssing et al. 1994; Dutt et al. 1996; Gupta and Mathur 1997; Malik 1985; Menezes et al. 1994; Pandey 1984a; Perez-Padilla et al. 1996; Qureshi 1994). Of these, one was a cohort study (Dutt et al. 1996) and three were case-control studies (Dennis et al. 1996; Døssing et al. 1994; Perez-Padilla et al. 1996). The remaining six studies were cross-sectional (Table 18.14).

Where studies reported exposure as a continuous variable, categories were constructed post hoc to be consistent with studies that presented the same exposure or a similar exposure as a categorical variable (e.g. average time spent daily near the stove, $<2$ hours and $>2$ hours). More than half of the study populations in Table 18.14 originated from rural areas where cooking on an open fire in ill-ventilated huts was common. Five study sites, however, were in urban or peri-urban settings where a mixture of fuels might be used (see Table 18.14) and where exposure to indoor smoke is likely to be lower than for women living in rural areas.

## Estimating the relationship between risk factor and disease

Smoking is an important potential confounding variable for COPD and particularly so if men are included in the analysis, given the higher prevalence of smoking in men than in women in developing countries. Only two studies adjusted for smoking (Dennis et al. 1996; Menezes et al. 1994). Of the studies that did not adjust for smoking, two included nonsmokers only (Behera et al. 1991; Dutt et al. 1996), another reported an overall prevalence of smoking of $<1 \%$ in the entire study population
Table 18.14 Studies on the risk of chronic obstructive pulmonary disease associated with use of solid fuels

| Study location | Author (year of publication) | Study design ( n ) <br> Study population | Exposure assessment | Outcome assessment | Covariates adjusted for | Odds ratio (95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Bolivia | Albalak et al. (1999) | Cross-sectional (24I) <br> Females+males aged $>20$ years | Cooking inside or outside | CB | Age, sex | 2.5 (1.25-5) |
| Brazil (urban) | Menezes et al. (1994) | Cross-sectional (1053) <br> Females + males aged $>40$ years | Presence of at least two of the following: open fire, charcoal stove, paraffin lamp or coal heater | CB | Age, sex, race, income, schooling, smoking, childhood respiratory illnesses, occupational exposures | 1.3 (0.75-2.27) |
| Colombia (urban) | Dennis et al. (1996) | Case-control (104-104) <br> Females aged $>35$ years | Use of solid bio-fuel for cooking (wood) | COPD, ${ }^{\text {a }}$ COPD + CB | Age, smoking, hospital | 3.92 (1.16-9.1) |
| India (rural) | Gupta and Mathur (1997) | Cross-sectional (707) Females + males aged $>15$ years | Use of solid bio-fuel for cooking (wood+dung) | $C B+$ bronchial asthma | Age | 7.9 (2.84-21.8) |
| Northern India | Behera et al. (1991) | Cross-sectional (3718) Females involved in cooking | Use of solid bio-fuel for cooking (wood+dung) | CB | None | 1.97 (1.16-3.22) |
| Northern India | Malik (1985) | Cross-sectional (2 180) <br> Females aged $>20$ years | Use of solid bio-fuel for cooking (wood) | $C B, C O P D+C B$ | None | 3.0 (1.77-4.93) |
| Southern India (urban) | Dutt et al. (1996) | Cohort (315) <br> Females aged I5-60 years | Use of solid bio-fuel (wood) for cooking | CB | None, age-stratified sampling | 2.8 (0.7-11.4) |

India

| NA | NA |
| :--- | :--- |
| None | $3.5(1.4-8.77)$ |
| Age, place of residence, <br> education, income, <br> smoking | $4.1(2.3-9.4)$ |
| Economic and <br> environmental variables | NA |
| Age | $5.4(2.96-9.78)$ |
| None | NA |
| None, matched for age <br> and sex | I4.4 (5.5-37.5) |

(Albalak et al. 1999). Pandey (1984a) reported the data stratified by smoking status and finally, the study by Perez-Padilla et al. (1996) reported that $70-80 \%$ of the subjects indicated that they had never smoked.

Two studies (Døssing et al. 1994; Gupta and Mathur 1997), which included men and women and reported a relatively high prevalence of smoking in their study populations (not equally distributed between COPD cases and controls), did not adjust for smoking. The combined estimate of risk from the group of studies that accounted for smoking, and excluding the Døssing et al. and Gupta and Mathur studies, was $2.5,95 \%$ CI 1.9-3.3. The combined estimate of relative risk for the studies by Døssing et al. and Gupta and Mathur that did not adjust for smoking, and which is thus likely to be an overestimation, was substantially higher at $10.8,95 \%$ CI 5.4-21.8. Another major confounding variable in the association between risk of COPD and exposure to indoor smoke is age, with absolute risk increasing with age. Most studies adjusted for age by matching, stratified sampling (Dutt et al. 1996), or by adjustment in the analysis; two studies (Malik 1985; Qureshi 1994) reported the mean age to be similar in the exposed vs unexposed subjects. A potential problem of confounding by age remains with the studies by Pandey (1984a) and Behera et al. (1991), which showed no data on the age distribution. The combined estimate of the relative risk excluding these two studies was $2.9,95 \%$ CI 2.2-3.6.

This analysis primarily included women as they comprise the population that is most frequently exposed to smoke from wood during cooking and which is thus at greatest risk of developing chronic airway disease. Therefore, we included estimates for women or the combined estimate adjusted for sex, if available. With two exceptions (Døssing et al. 1994; Gupta and Mathur 1997), all studies reported the data separately for men and women, or combined the data while adjusting for sex. The overall estimate of relative risk for all studies included was 2.9 , $95 \%$ CI $2.2-3.8$. For men, it was $2.8,95 \%$ CI $1.4-5.7$, but this was based on only two studies, one of which did not correct for age (Døssing et al. 1994; Qureshi 1994). See Table 18.15 for details.

All three case-control studies were hospital-based; control groups consisted of visitors to patients other than the study subjects (Døssing et al. 1994), patients with illnesses other than those of the respiratory tract (Dennis et al. 1996) and a mixture of visitors, patients diagnosed with tuberculosis or interstitial lung disease and patients with otolaryngological problems (Perez-Padilla et al. 1996). Bias could have been introduced by the choice of visitor controls if exposure to indoor smoke was related to the likelihood to come to the hospital to visit a patient, or by the selection of inpatient controls, if exposure to indoor smoke made the patients with the control diseases less or more likely to be referred to the hospital (e.g. tuberculosis).

Table 18.15 The risk of chronic obstructive pulmonary disease associated with use of solid fuels: subgroup analyses

|  | Subgroup analyses | Studies included | Odds ratio (95\% Cl) |
| :---: | :---: | :---: | :---: |
| Males and females | Rural population | Too few studies available to allow odds ratio to be calculated | NA |
|  | Urban population | Too few studies available to allow odds ratio to be calculated | NA |
|  | Adjusted for smoking | Albalak et al. (1999); <br> Menezes et al. (1994); <br> Pandey (1984a) | 2.51 (1.76-3.56) |
|  | Not adjusted for smoking | Qureshi (1994); Døssing et al. (1994); Gupta and Mathur (1997) | 5.8 (3.74-8.99) |
|  | Adjusted for age | Albalak et al. (1999); <br> Døssing et al. (I994); <br> Gupta and Mathur (1997); <br> Menezes et al. (1994) | 3.3 (2.32-4.69) |
| Females only | Adjusted for smoking but not for age | Behera et al. (1991); Pandey (1984a); Qureshi (1994) | 2.56 (1.75-3.75) |
|  | Adjusted for smoking and age | Dutt et al. (1996); Perez-Padilla et al. (1996); Dennis et al. (I996); Malik (I985) | 2.83 (2.0-3.97) |
| Males only | Not adjusted | NA, too few studiesDøssing et al. (1994) adjusted for age; Qureshi (1994) adjusted for none | NA, see also text |
|  | Adjusted for smoking and age | None of the studies in males adjusted for both age and smoking | NA |

[^65]The final model for women excluded the three studies that did not adjust for age and/or smoking status. The overall risk of COPD in women exposed to indoor air pollution from use of solid fuels was estimated as $3.2,95 \%$ CI $2.3-4.8$. There is much less evidence available about the impact on men, but the risk seems to be lower, $1.8,95 \%$ CI $1.0-3.2,{ }^{8}$ presumably because of lower exposure.

## LUNG CANCER

Lung cancer in women has been associated with cooking with open coal stoves in China on the basis of a number of studies. In China, there is also evidence that lung cancer is caused by use of certain cooking oils
(Zhong et al. 1999a, 1999b) as well as by exposures to known carcinogens contained in coal smoke, such as arsenic (Finkelman et al. 1999). There is limited evidence available for an association between lung cancer and use of biomass fuels in women, but not in men (Gao et al. 1987; Liu et al. 1993; Sobue 1990), although several pollutants in biomass smoke are known or suspected human carcinogens (Smith 1987).

The majority of the internationally published studies on lung cancer and indoor air pollution that we were able to locate were conducted in China. One took place in Japan (Sobue 1990) and one in the United States (Wu et al. 1985). Two eligible studies were published in Chinese only (Huang 1999; Wu et al. 1999). All 19 studies identified were case-control studies, including either newly-diagnosed cases of lung cancer at a hospital or using death registries, and of these, 14 studies were hospital-based. Inherent in the choice of this design is Berkson's bias, referring to the possibility that controls (men and women hospitalized with other diseases) are not selected independently of exposure in the source population. With two exceptions (Ko et al. 1997; Sobue 1990), all studies used population controls, which minimizes such bias (Table 18.16).

## Characteristics of excluded studies

Of the 19 studies identified, we excluded three (Du et al. 1988; Xu et al. 1996b; Yang et al. 1990). The ecologic study by Yang et al. (1990) neither adjusted for smoking or other risk factors nor provided sufficient information to calculate odds ratios. Of two articles which reported on the same study population (Du et al. 1988, 1996), we included the more recent, which provided $95 \%$ CIs for the relative risk. More than one article reported on a collaborative study that included men and women of two major cities in the Province of Liaoning (Wu-Williams et al. 1990; Xu et al. 1996b); we included only the study by Wu-Williams et al. (1990), which combined all female lung cancer cases from the death registries of the two cities. The study by Xu et al. (1996b) considered cases in males and females from one city only.

## Estimating risk factor-disease relationship

Although the 16 studies included in this analysis were all case-control designs, measurement of exposure to indoor air pollution was carried out by a multitude of methods. Seven studies assessed exposure to indoor air pollution in terms of years of exposure (Dai et al. 1996; Ko et al. 1997; Lei et al. 1996; Liu et al. 1991; Sobue 1990; Wu et al. 1999; Wu-Williams et al. 1990). The remaining eight studies merely determined whether coal and/or bio-fuel were generally used for cooking or heating (Du et al. 1996; Gao et al. 1987; Huang 1999; J. Liu and H. Hu, unpublished data, 1996; Liu et al. 1993; Shen et al. 1996; Wang et al. 1996; Wu et al. 1999). In order to explore the characteristics responsible for
Table 18.16 Studies on the risk of lung cancer associated with use of solid fuels

| Study location | Reference | Study design ( n ) Study population | Exposure assessment | Outcome assessment ${ }^{\text {a }}$ | Covariates adjusted for | Odds ratio (95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| China Fujian Province Fuzhou | Luo et al. (1996) | Case-control (102-306) <br> Females+males | Indoor combustion of coal | Newly-diagnosed lung cancer | Smoking, passive smoking, chronic bronchitis and matched for age and sex | $\begin{aligned} & \text { ADC: } 6.0 \\ & (1.36-23.49) \\ & \text { SCC: } 14.1 \\ & (1.67-119.4) \end{aligned}$ |
| China Gaunxi Province Nanning | Huang (1999) | Case-control (122-244) <br> Females+males | Use of coal | Newly-diagnosed lung cancer | Smoking, chronic lung disease, meat consumption, depression, SES, BMI, exercise | 1.76 (1.3-2.38) |
| China Guangzhou | Du and Ou (1990) <br> Excluded | Case-control (662-662) <br> Females+males | Exposed to coal fumes yes/no | Deaths from lung cancer over 5 years | Matched for age, sex, residence | 14.52 (-) |
| China Guangzhou | Du et al. (1996) | Case-control (120-240) Non-smoking females + males | Exposed to coal fumes yes/no | Death from lung cancer | Smoking and chronic respiratory disease | $\begin{aligned} & \text { Females: I. } 56 \\ & (0.57-4.25) \\ & \text { Males: } 1.5 \\ & (0.69-3.27) \end{aligned}$ |
| China Guangzhou | Lei et al. (1996) | Case-control (792-792) <br> Females+males | Cooking for $>40$ years | Death from lung cancer | Matched for age and sex | 0.93 (0.67-1.21) |
| China Guangzhou | Liu et al. (1993) | Case-control (3\|6-3|6) Females+males | Use of coal and wood for cooking | Newly-diagnosed lung cancer | Smoking, passive smoking, education, SES, history of cancer | Coal: 1.46 (0.83-2.56) <br> Bio-fuel: I.19 <br> (0.46-3.1I) |
| China Guangzhou | Wu et al. (1999) | Case-control (258-258) Females | Use of coal as residential fuel | Newly-diagnosed lung cancer | Smoking, history of tuberculosis, fruit consumption, ventilation of kitchen | 1.57 (0.89-2.82) |

Table 18.16 Studies on the risk of lung cancer associated with use of solid fuels (continued)

| Study location | Reference | Study design (n) Study population | Exposure assessment | Outcome assessment ${ }^{\text {a }}$ | Covariates adjusted for | Odds ratio (95\% CI) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| China Liaoning Province Harbin | Dai et al. (1996) | Case-control (120-120) <br> Non-smokers, females | Use of coal heater for 25-34 years | Newly-diagnosed lung cancer | History of family cancer, income, carrot consumption, deep fried cooking | 4.7 (1.28-17.18) |
| China <br> Liaoning Province Harbin and Shenyang | Wu-Williams et al. (1990) | Case-control (956-952) Females | Use of coal stove for $>40$ years | Newly-diagnosed lung cancer | Age, education, smoking | 1.3 (1-1.7) |
| China Liaoning Province Shenyang | Wang et al. (1996) | Case-control (135-135) Females | Use of coal for cooking | Newly-diagnosed lung cancer | Family history of cancer, ETS | 0.75 (0.43-1.31) |
| China Liaoning Province Shenyang | $X u$ et al. <br> (I996b) <br> Excluded | Case-control (1249-1345) Females+males | Use of coal stove for cooking | Newly-diagnosed lung cancer from cancer registry | None | Females: I. 5 (-) <br> Males: 2.3 (一) |
| China Shanghai | Gao et al. (1987) | Case-control (672-735) Females | Cooking with coal or bio-fuel | Newly-diagnosed lung cancer | Smoking, education, age | Coal: 0.9 (0.7-I.3) <br> Bio-fuel: 1.0 <br> (0.6-1.8) |
| China <br> Beijing and Shunyi | J. Liu and H. Hu, unpublished data, 1996 | Case-control (220-440) Females+males, farmers | Combustion of coal cakes | Death from lung cancer | Smoking, chronic respiratory disease and matched for age | 1.9 (1.16-3.43) |


| China <br> Hubei Province <br> Wuhan | Yang et al. <br> $(1990)$ <br> Excluded | Cross-sectional; ecologic <br> design (50-200) <br> Females+males, two <br> parts of city | Use of coal for <br> cooking | Death from lung <br> cancer | None |
| :--- | :--- | :--- | :--- | :--- | :--- |

heterogeneity found in the results of a meta-analysis of all studies, several subgroup analyses were conducted, in which stratification by type of fuel used (mostly coal and some wood) and sex was used. The variability of exposure categories was too great and the number of studies too small to be grouped for duration of exposure. If a study reported an estimate of relative risk for several exposure categories, the odds ratio for the category representing the longest period of exposure was used (Lei et al. 1996; Wu-Williams et al. 1990). Two studies (Luo et al. 1996; Wu et al. 1985) reported separate estimates for adenocarcinoma and squamous cell carcinoma; these were entered as separate studies as we were unable to achieve a combined estimate. Whenever possible, separate estimates for men and women were extracted and entered as individual studies (Du et al. 1996; Liu et al. 1991).

In a recent review of the literature on indoor air pollution and several health outcomes (Bruce et al. 2000), the most prominent concern voiced was regarding the lack of control for confounders. Therefore, we conducted stratified analyses based on studies that accounted for the most common potential confounders, such as smoking and the presence of a chronic respiratory disease. All studies included in the meta-analysis either adjusted for smoking or included only non-smokers. It has been suggested that chronic respiratory diseases such as chronic bronchitis, tuberculosis, asthma and emphysema that originate from infections or other predispositions may increase the probability of developing lung cancer later in life (Luo et al. 1996). We examined the effect of indoor air pollution from coal smoke on men and women separately. Nine studies either only included women or presented risk estimates for men and women separately (Dai et al. 1996; Du et al. 1996; Gao et al. 1987; Ko et al. 1997; Liu et al. 1991; Sobue 1990; Wang et al. 1996; Wu et al. 1985, 1999). The overall estimate for females was 1.17, $95 \%$ CI 1.02-1.35. The analysis restricted to studies that adjusted for smoking and chronic respiratory disease indicated a substantial increase in risk for women of almost two-fold (odds ratio of 1.94, $95 \%$ CI 1.09-3.47).

Five studies presented a combined risk estimate for men and women (Huang 1999; Lei et al. 1996; Liu et al. 1993; Luo et al. 1996; Shen et al. 1996), producing a summary odds ratio of 1.86 ( $95 \%$ CI 1.48-2.35). Restricting the analysis to the three studies that controlled for smoking and chronic respiratory disease showed a substantial increase in risk (odds ratio of $2.55,95 \%$ CI 1.58-4.10).

Only three studies either included males only (Wu et al. 1999) or presented sufficient data to extract a separate estimate for males (Du et al. 1996; Liu et al. 1991). The risk associated with coal use for the male population was $1.79,95 \%$ CI 1.18-2.72, and slightly lower when taking into account confounding by smoking and chronic airway disease (odds ratio of $1.51,95 \% \mathrm{CI} 0.97-2.46)$. Although the results of the two studies

Table 18.17 Summary of results of subgroup meta-analyses

|  | Odds ratio (95\% CI) |  |
| :--- | :---: | :---: |
| Subgroup analyses | Not adjusted | Adjusted for smoking and chronic <br> airway disease |
| Males and females—coal use | $1.86(1.48-2.35)$ | $2.55(1.58-4.10)$ |
| Males only—coal use | $1.79(1.18-2.72)$ | $1.51(0.97-2.46$ |
| Females only-coal use | $1.17(1.02-1.35)$ | $1.94(1.09-3.47)$ |

comprising this model were not quite statistically significant (lower confidence limit was 0.97 ), the pattern of significance of the five studies assessing risks for men and women combined, give confidence that there is likely to be a real effect on men. Odds ratios are shown in Table 18.17.

### 3.4 Sources of uncertainty

Uncertainty estimates were generated through the use of meta-analyses for all the disease end-points included. A critical problem with extrapolating the results of epidemiological studies from one subregion to another, particularly between developed and developing regions, is the difference in other potentially interactive risk factors, such as malnutrition, which are not addressed by the methodology. That all the studies used for the calculations of solid-fuel use were done in developing countries, however, does provide some confidence that differences in competing risks were not excessive. Meta-analytical confidence intervals probably underestimate true uncertainty because of variations in the way different studies dealt with measures of exposure, adjustment for confounding, and outcome definitions, as well as the need to extrapolate results across populations.

### 3.5 Risk Reversibility

There are few studies on the reversibility of the health effects of smoke from solid fuel. For acute outcomes (ALRI), evidence from risk factors for other childhood infectious diseases may provide some guidance (Jones et al. 2003). For the chronic conditions, COPD and lung cancer, the timing is less clear, however, since the increased risk presumably results from many years of exposure. A retrospective cohort study in China, however, did find a statistically significant drop in lung cancer rates associated with introduction of improved stoves with flues in around 1980 (Lan et al. 2002). The delay between intervention and a discernible reduction in lung cancer incidence was about 10 years, consistent with that observed after smoking cessation (see chapter 11).

## 4. Results

### 4.1 Attributable burden of disease

As shown in Table 18.18, the burden of disease attributed to use of household solid fuels is dominated by that caused by ALRI in young children, which accounts for $59 \%$ of all attributed premature deaths and $78 \%$ of DALYs. COPD accounts for nearly all the remainder, with the burden from lung cancer a relatively minor contributor, owing to the concentration of estimated use of coal in two subregions only. Because ALRI in children does not cause many years lost due to disability, however, COPD is responsible for a much larger portion of the total disability.

As shown in Table 18.18, five subregions account for nearly all deaths ( $94 \%$ ) and DALYs ( $93 \%$ ) attributable to indoor air pollution from solid fuel. The subregions with the largest numbers of DALYs, in descending order, are SEAR-D, WPR-B, AFR-E, AFR-D and EMR-D. When the subregions are ranked according to numbers of deaths, the relative positions of SEAR-D and WPR-B shift, because there are more deaths in SEAR-D in a younger age group (ALRI-related deaths in children) compared to WPR-B (mortality is dominated by COPD in adults).

As shown in Figure 18.5, because of differences in baseline rates of disease, not exposure or risk from use of solid fuel, effects on mortality

Table 18.18 Burden of disease from use of solid fuel, 2000

| Deaths (000s) |  |  |  |  | DALYs (000s) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Subregion | ALRI | COPD | Lung cancer | All causes | ALRI | COPD | Lung cancer | All causes |
| AFR-D | 153 | 20 | NA | 173 | 5221 | 173 | NA | 5394 |
| AFR-E | 198 | 21 | NA | 219 | 6746 | 178 | NA | 6924 |
| AMR-A | 0 | 0 | NA | 1 | 1 | 6 | NA | 6 |
| AMR-B | 6 | 9 | NA | 16 | 291 | 153 | NA | 444 |
| AMR-D | 9 | 2 | NA | 10 | 314 | 16 | NA | 330 |
| EMR-B | 2 | 0 | NA | 2 | 59 | 5 | NA | 64 |
| EMR-D | 94 | 22 | NA | 116 | 3306 | 203 | NA | 3508 |
| EUR-A | 0 | 0 | NA | 0 | 0 | 0 | NA | 0 |
| EUR-B | 12 | 5 | NA | 17 | 417 | 60 | NA | 477 |
| EUR-C | 1 | 4 | NA | 4 | 22 | 44 | NA | 67 |
| SEAR-B | 19 | 17 | NA | 37 | 761 | 229 | NA | 990 |
| SEAR-D | 355 | 167 | 1 | 522 | 12506 | 1724 | 8 | 14237 |
| WPR-A | 0 | 0 | NA | 0 | 0 | 0 | NA | 0 |
| WPR-B | 62 | 426 | 15 | 503 | 2275 | 3662 | 160 | 6097 |
| World | 910 | 693 | 16 | 1619 | 31919 | 6453 | 168 | 38539 |

NA Not applicable.

Figure 18.5 Deaths from acute lower respiratory infection attributable to indoor smoke from use of solid fuels, 2000

attributable to ALRI are larger for males than females in AFR-D and AFR-E, similar in EMR-D and WPR-B, and greater for females in SEAR-D.

As shown in figure 18.6, the vast majority of attributable deaths from COPD and lung cancer appear to be experienced by the women of SEARD and WPR-B. This is partially because lung cancer deaths associated with solid fuel use were only estimated in these two subregions, due to lack of information on coal use in the other subregions. In addition, women appear to bear a higher proportion of the burden not only because they are likely to be more exposed, but because smoking attributable deaths (which are a higher proportion of male deaths) have been removed.

## 5. Discussion

### 5.1 Sources of uncertainty

Of a large number of sources of uncertainty, four major factors dominate these estimates.

- The choice of exposure variable, which, although necessary to match with current epidemiological studies, only roughly captures the population distribution of exposure and its variability in different populations.

Figure 18.6 Deaths from chronic obstructive pulmonary disease and lung cancer attributable to indoor smoke from use of solid fuels, 2000


- Distribution of the ventilation factor worldwide, i.e. what fractions of solid-fuel-using households do so in ways that vent some or all of the smoke outside and away from the householders.
- The different patterns of competing and confounding risks for ALRI in different circumstances, particularly those related to the severe forms affecting mortality.
- The relationship between the risks of indoor pollution and tobacco smoking, particularly for COPD and lung cancer in China where tobacco smoking is an important contributor (Liu et al. 1998).


### 5.2 Possible interventions

Although not included in the primary calculations here, as previously noted, there is growing evidence that other important health end-points can be attributed to exposure to indoor air pollution. Three of these, in particular, are of increasing concern worldwide: tuberculosis (because it is so closely related to the HIV epidemic); ischaemic heart disease (because of the shift in age and diet occurring in developing countries); and asthma (because of rising trends in diagnosed asthma in many parts of the world) (ISAAC 1998). There is some urgency that the associations of all potentially policy-sensitive risk factors (including use of solid household fuels) with these diseases be investigated.

There are four general categories of interventions that have been identified to reduce the health impacts of household use of solid fuel (Barnes et al. 1993; Ezzati and Kammen 2002; NCAER 2002; Smith and Desai 2002; Smith 1987, 1989).

- Behavioural changes to reduce exposure, for example, encouraging women to keep their young babies away from the fire.
- Changes in household ventilation, such as increasing the number of window openings in the kitchen, providing gaps between roof and wall, and moving the stove out of the living area.
- Improvements in stoves, either through venting by use of flues or hoods and/or improvements in stove combustion efficiency that reduce the emissions of toxic pollutants, nearly all of which are products of incomplete combustion.
- Shifts to higher-quality, low-emission liquid or gaseous fuels, such as kerosene and liquefied petroleum gas (which are based on petroleum) or biomass-based alcohol and biomass-based gaseous fuels derived either from biological processes (bio-gas) or thermochemical processing (producer gas).

Most research has focused on improvements in stoves and shifts to higher-quality, low-emission liquid or gaseous fuels; it seems that the efficacy of the interventions listed above generally increases as one moves down the list. The extent to which they can be successfully applied varies across different populations depending on income, housing, biomass availability, cultural factors and climate. It seems possible, however, that programmes can be designed to encourage many urban and peri-urban solid-fuel-using populations to move to using liquefied petroleum gas or kerosene, at lower incomes (i.e. sooner) than would occur without intervention. On the other hand, the poorest rural populations with nearly no cash income, but with access to wood and/or agricultural waste, are unlikely to move to clean fuels or use significantly improved stoves without large subsidies, which are usually not sustainable. There do seem to be large populations between these extremes, however, that can be targeted by efforts to introduce improved stoves. Although the fraction of improved-stove programmes that have succeeded is small, the total number of stoves successfully introduced is impressive because of the remarkable achievement of the Chinese programme, which has apparently been responsible for the introduction of nearly 200 million stoves since the early 1980s (Goldemberg et al. 2000; Smith 1993). More research and development work is needed, however, to learn how to successfully translate the lessons learned in China and elsewhere to other parts of the world in a sustainable cost-effective manner.

## 6. Exposure projections

The use of solid fuel will probably slowly decrease in absolute, as well as relative, terms, as economic development proceeds. This shift is occurring most rapidly in China and Latin America, at interim rates in south Asia, and slowest or not at all in sub-Saharan Africa (World Resources Institute 2000). Cooking outdoors, on the other hand, is likely to decrease with development, but as the number of separate kitchens may increase, it is not clear how exposures will change overall. Current trends in vented stoves are less certain outside China. The Indian national stove programme, for example, had mixed success (NCAER 2002) and was dismantled in 2002 (Mahapatra 2003). In China, however, nearly 90\% of the rural population seems to have adopted higher-efficiency vented stoves in recent years.

Table 18.19 Use of solid fuel and exposure to its smoke: estimates for 2000 and predictions for 2010

| Subregion | Estimated fuel use ${ }^{\text {a }}$ |  | Estimated exposure of adults ${ }^{\text {b }}$ |  |
| :---: | :---: | :---: | :---: | :---: |
|  | 2000 | 2010 | 2000 | 2010 |
| AFR-D | 73.4 | 69.0 | 55.1 | 52.0 |
| AFR-E | 85.8 | 83.0 | 64.3 | 62.0 |
| AMR-A | 1.5 | 1.0 | 1.1 | 1.0 |
| AMR-B | 24.6 | 20.0 | 18.4 | 15.0 |
| AMR-D | 52.9 | 52.0 | 39.7 | 39.0 |
| EMR-B | 6.1 | 5.0 | 4.6 | 4.0 |
| EMR-D | 55.2 | 50.0 | 41.4 | 37.0 |
| EUR-A | 0.2 | 0.2 | 0.0 | 0.0 |
| EUR-B | 41.5 | 35.0 | 20.5 | 19.0 |
| EUR-C | 22.8 | 21.0 | 6.4 | 6.0 |
| SEAR-B | 66.5 | 62.0 | 49.9 | 46.0 |
| SEAR-D | 83.5 | 77.0 | 62.6 | 58.0 |
| WPR-A | 0.2 | 0.0 | 0.1 | 0.0 |
| WPR-B ${ }^{\text {c }}$ | 78.1 | 70.0 | 41.8 | 23.0 |

[^66]Some insight can be gleaned about the potential for reduction in exposure by application of the model of solid-fuel use employed in this chapter. Estimates of income growth and shift of the population from rural to urban areas have different impacts on use of solid fuels in different subregions. Economic growth and urbanization over the next 10 years, for example, might substantially reduce the fraction of households that use solid fuel in the subregions that currently have the largest burdens. We examined changes that might occur over a 10 -year period in two major model parameters: GNP per capita and rural-urban population shift (World Bank 2001). Estimates based on changes in income and urbanization beyond 2010 would be highly unstable, since current trends are unlikely to be sustained over several decades. Countries for which data are lacking are assigned the global average values for GNP per capita (equivalent to a $1.3 \%$ annual growth rate) and global rate of urbanization (rural population decreases from around $58 \%$ to $51 \%$ of the total population. Among many other assumptions, of course, such an extrapolation supposes that the structure of the model remains valid over this period. Table 18.19 shows how predicted changes in GNP per capita and urbanization affect predictions of future household use of solid fuel and of future exposure in each subregion. The net impact of shifts in these factors seems to indicate that, globally, exposure to indoor smoke from use of solid fuel is likely to decrease. There are subregional variations in the pattern, however, with continuing large exposures in sub-Saharan Africa and south-east Asia (Indian subcontinent).

## Acknowledgements

We appreciate the excellent comments of many anonymous reviewers; assistance with extracting information from Chinese-language literature from Linwei Tian; advice on meta-analysis from Daniel Mäusezahl, Jack Colford and the Berkeley Meta-analysis Group; and the patience and thoughtfulness of the book's editors, particularly Majid Ezzati. The authors would also like to recognize the millions of poor women and children around the world who are exposed daily to toxic indoor air pollution produced by smoke from solid fuel, as well as to a range of other risks. We hope that these results will help to generate recognition of the potential magnitude of this problem and spark commensurate efforts to ameliorate the situation.

## Notes

1 Particulate matter, often abbreviated as PM, is categorized by size, specifically by aerodynamic diameter in microns (millionths of a meter or $\mu \mathrm{m}$ ). For example, $\mathrm{PM}_{2.5}$ refers to particulate matter with a diameter of less than $2.5 \mu \mathrm{~m}$. In general, small particles are thought to be more damaging to health.
2 See preface for an explanation of this term.

3 Seven urban and three rural areas were omitted because of missing data or likely errors in the government statistical publications, which suggested improbable levels of energy consumption per household (i.e. in provincial households, average levels of consumption that were more than one standard deviation from the mean).
4 All analysis was done using SPSS Version 8.0 (SPSS Inc., USA) and STATA 7.0 (Stata Corporation, USA).

5 Cross-sectional studies report odds ratios rather than relative risks. These terms are used interchangeably in this chapter.
6 Two hospital-based case-control studies in India came to our attention too late for inclusion in the meta-analysis. In New Delhi, Broor et al. (2001) found an adjusted odds ratio of 2.5 ( $95 \%$ CI 1.5-4.2) for ALRI in children aged $<5$ years in homes not using liquefied petroleum gas. In Calcutta, Mahalanabis et al. (2002) found an adjusted odds ratio of 4.0 ( $95 \%$ CI $2.0-7.9$ ) for pneumonia in children aged 2-35 months living in homes using solid fuels.
7 For further discussions, see reviews by Bruce et al. (2000), Chen et al. (1990) and Smith (1987).
8 For males, it did not seem appropriate to use the unadjusted estimate of risk, particularly when the adjusted estimates for both sexes were lower than either the unadjusted estimate for males only or the adjusted estimates for females only. Simple averaging of the risk chosen for males, 1.8, with the adjusted risk for females, 3.2 , results in the combined mean risk of 2.5 observed when analyses included both sexes. The lower bound of the confidence interval was set at 1.0 (no effect) and the higher bound only at the unadjusted risk for males, 3.2.

## References

Albalak R, Frisancho AR, Keeler GJ (1999) Domestic biomass fuel combustion and chronic bronchitis in two rural Bolivian villages. Thorax, 54:10041008.

Armstrong JR, Campbell H (1991) Indoor air pollution exposure and lower respiratory infections in young Gambian children. International Journal of Epidemiology, 20:424-429.
Awasthi S, Glick HA, Fletcher RH (1996) Effect of cooking fuels on respiratory diseases in preschool children in Lucknow, India. American Journal of Tropical Medicine and Hygiene, 55:48-51.

Azizi BH, Zulkifli HI, Kasim S (1995) Indoor air pollution and asthma in hospitalized children in a tropical environment. Journal of Asthma, 32:413-418.

Barker D (1997) Maternal nutrition, fetal nutrition, and disease in later life. Nutrition, 13:807-813.
Barnes DF, Openshaw K, Smith KR, van der Plas R (1993) The design and diffusion of improved cookstoves. 8:119-141.
Basu R, Samet J (1999) A review of the epidemiological evidence on health effects of nitrogen dioxide exposure from gas stoves. Journal of Environmental Medicine, 1:173-187.

Behera D, Dash S, Yadav SP (1991) Carboxyhaemoglobin in women exposed to different cooking fuels. Thorax, 46:344-346.

Bobak M (2000) Outdoor air pollution, low birth weight, and prematurity. Environmental Health Perspectives, 108:173-176.

Boy E, Bruce N, Delgado H (2002) Birth weight and exposure to kitchen wood smoke during pregnancy in rural Guatemala. Environmental Health Perspectives, 110:109-114.
Broor S, Pandey RM, Ghosh M et al. (2001) Risk factors for severe acute lower respiratory infection in under-five children. Indian Pediatrics, 38:1361-1369.
Bruce N, Perez-Padilla R, Albalak R (2000) Indoor air pollution in developing countries: a major environmental and public health challenge. Bulletin of the World Health Organization, 78:1078-1092.
Campbell H (1997) Indoor air pollution and acute lower respiratory infections in young Gambian children. Health Bulletin, 55:20-31.

Campbell H, Armstrong JR, Byass P (1989) Indoor air pollution in developing countries and acute respiratory infection in children. The Lancet, 1:1012.

Cerqueiro MC, Murtagh P, Halac A, Avila M, Weissenbacher M (1990) Epidemiologic risk factors for children with acute lower respiratory tract infection in Buenos Aires, Argentina: A matched case-control study. Reviews of Infectious Diseases, 12:S1021-1028.

Chen BH, Hong CJ, Pandey MR, Smith KR (1990) Indoor air pollution in developing countries. [World Health Statistics Quarterly] Rapport Trimestriel de Statistiques Sanitaires Mondiales, 43:127-138.
Collings DA, Sithole SD, Martin KS (1990) Indoor woodsmoke pollution causing lower respiratory disease in children. Tropical Doctor, 20:151-155.
Dai XD, Lin CY, Sun XW, Shi YB, Lin YJ (1996) The etiology of lung cancer in nonsmoking females in Harbin, China. Lung Cancer, 14:S85-91.
de Francisco A, Morris J, Hall AJ, Armstrong Schellenberg JR, Greenwood BM (1993) Risk factors for mortality from acute lower respiratory tract infections in young Gambian children. International Journal of Epidemiology, 22:1174-1182.

Dennis RJ, Maldonado D, Norman S, Baena E, Martinez G (1996) Woodsmoke exposure and risk for obstructive airways disease among women. Chest, 109:115-119.

Dhar SN, Pathania AGS (1991) Bronchitis due to biomass fuel burning in North India: "Guijjar lung" an extreme effect. Seminars in Respiratory Medicine, 12: 69-74.

Døssing M, Khan J, al-Rabiah F (1994) Risk factors for chronic obstructive lung disease in Saudi Arabia. Respiratory Medicine, 88:519-522.

Du E, Wang W, Zhang Q, Li Z, Wu Y, Zhao S (1988) Survey of indoor air quality and human health in rural Changchun, Jilin province, China. Journal of Norman Bethune University of Medical Science, 14:88-91.

Du YX, Ou XL (1990) Indoor air pollution and women's lung cancer. Proceedings of the Fifth International Conference on Indoor Air Quality and Climate, Toronto, 1:59-64.

Du YX, Cha Q, Chen XW et al. (1996) An epidemiological study of risk factors for lung cancer in Guangzhou, China. Lung Cancer, 14:S9-37.
Dutt D, Srinivasa DK, Rotti SB, Sahai A, Konar D (1996) Effect of indoor air pollution on the respiratory system of women using different fuels for cooking in an urban slum of Pondicherry. National Medical Journal of India, 9:113-117.

Ellegard A (1996) Cooking fuel smoke and respiratory symptoms among women in low-income areas in Maputo. Environmental Health Perspectives, 104: 980-985.

Ezzati M (2002) Indoor air pollution from biomass stoves as a risk factor for acute respiratory infections in Kenya. Proceedings of the Ninth International Conference on Indoor Air Quality and Climate, Monterey, 4:970-975.

Ezzati M, Kammen D (2001) Indoor air pollution from biomass combustion and acute respiratory infections in Kenya: an exposure-response study. The Lancet, 358:619-624.

Ezzati M, Kammen DM (2002) Household energy, indoor air pollution and health in developing countries: knowledge base for effective interventions. Annual Review of Energy and the Environment, 27:233-270.
FAO (1997a) Review of wood energy data in RWEDP member countries. (FAO Field Document No. 47.) Food and Agriculture Organization of the United Nations, Regional Wood Energy Development Program (RWEDP), Bangkok.
FAO (1997b) The role of wood energy in Asia. (FOPW/97/2.) Wood energy for today and tomorrow (WETT), Rome.
Finkelman RB, Belkin HE, Zheng B (1999) Health impacts of domestic coal use in China. Proceedings of the National Academy of Sciences of the United States of America, 96:3427-3431.

Franco EL, Kowalski LP, Oliveira BV et al. (1989) Risk factors for oral cancer in Brazil: a case-control study. International Journal of Cancer, 43:992-1000.
Fridley DG, Sinton JE, Lewis JI, Zhoui F, Li J (2001) China energy databook5th edition. LBL-47832. Lawrence Berkeley National Laboratory and Beijing Energy Research Institute, Berkeley.
Gao YT, Blot WJ, Zheng W et al. (1987) Lung cancer among Chinese women. International Journal of Cancer, 40:604-609.
García-Marcos L, Guillén JJ, Dinwiddie R, Guillén A, Barbero P (1999) The relative importance of socio-economic status, parental smoking and air pollution (SO2) on asthma symptoms, spirometry and bronchodilator response in 11-year-old children. Pediatric Allergy and Immunology, 10:96-100.

Gold JA, Jagirdar J, Hay JG et al. (2000) Hut lung. A domestically acquired particulate lung disease. Medicine, 79:310-317.

Goldemberg J, Reddy AKN, Smith KR, Williams RH (2000) Rural energy in developing countries. In: World energy assessment. Goldemberg JO, ed. United Nations Development Programme, New York.
Government of Botswana (1991) Population and housing census. Central Statistical Office, Gaborone.

Government of Brazil (1991) Censo demografico: 1991; familias e domicilios. [Demographic census: 1991: families and households] Instituto Brasileiro do Geografica e Estadistica (IBGE) [Brasilian Institute of National Statistics and Geography], Rio de Janeiro.
Government of Ecuador (1990a) V censo de poblacion y iv censo de vivienda [Census of population and households]. Instituto Nacional de Estadistica y Censo [National Institute of Statistics and Census], Ecuador.
Government of Ecuador (1990b) V censo de poblacion, iv censo de vivienda, 1990. vivienda [Census of population and households, 1990]. Instituto Nacional de Estatistica y Censo [National Institute of Statistics and Census], Ecuador.

Government of Ethiopia (1998) 1994 population and housing census of Ethiopia: results at the country level. Volume 1. Statistical Office, Addis Ababa.

Government of India (1991a) Census of India. Office of the Registrar General, New Delhi.
Government of India (1991b) Census of India: 1991. Office of the Registrar General, New Delhi.
Government of Indonesia (1995) Indonesia energy balance. Central Bureau of Statistics, Jakarta.
Government of Indonesia (1996a) Indonesia energy balance 1992-1996. Central Bureau of Statistics, Jakarta.
Government of Indonesia (1996b) Indonesia energy balance 1992-1996. Central Bureau of Statistics, Jakarta.

Government of Mexico (1990a) Censo general de poblacion y vivienda [General population and housing census]. Instituto Nacional de Estadistica, Geografica e Informatica [National Institute of Statistics, Geographics and Informatics], Mexico.

Government of Mexico (1990b) Censo general de poblacion y vivienda, 1990 [General population and household census, 1990]. Instituto Nacional de Estatistica, Geografica e Informatica [National Institute of Statistics, Geographics and Informatics], Mexico.
Government of Nigeria (1990) Social statistics in Nigeria, 1990. Federal Office of Statistics, Lagos.
Government of Pakistan (1997) Pakistan energy yearbook: 1996. Hydrocarbon Development Institute of Pakistan, Ministry of Petroleum and Natural Resources, Islamabad.

Government of China (1996) China energy statistical yearbook: 1991-1996. Department of Industrial and Transportation Statistics, Chinese Statistical Publishing House, and Department of Industrial and Transportation Statistics, China.
Government of Zimbabwe (1992) Census, Zimbabwe national report. CSO, Harare.

Gupta BN, Mathur N (1997) A study of household environmental risk factors pertaining to respiratory diseases. Energy Environment Monitor, 13: 61-67.

Huang Z-B (1999) A study on the risk factors and population attributable risk for primary lung cancer. Journal of Guangxi Medical University [in Chinese], 16:447-450.

International Energy Agency (1999) Energy statistics of non-OECD countries 1996-1997, The Organization for Economic Cooperation and Development (OECD), Paris.

International Energy Agency (2002) World energy outlook. International Energy Agency, Paris.

ISAAC (1998) Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The international study of asthma and allergies in childhood (ISAAC) steering committee. The Lancet, 351:1225-1232.

Johnson AW, Aderele WI (1992) The association of household pollutants and socio-economic risk factors with the short-term outcome of acute lower respiratory infections in hospitalized pre-school Nigerian children. Annals of Tropical Paediatrics, 12:421-432.
Johnson DL, Swank PR, Baldwin CD, McCormick D (1999) Adult smoking in the home environment and children's IQ. Psychological Reports, 84:149-154.

Jones G, Skeketee RW, Black RE, Bhutta ZA, Morris SS (2003) How many child deaths can we prevent this year? The Lancet, 362:65-71.

Ko Y (2000) Chinese food cooking and lung cancer in women nonsmokers. American Journal of Epidemiology, 151:140-147.
Ko YC, Lee CH, Chen MJ et al. (1997) Risk factors for primary lung cancer among non-smoking women in Taiwan. International Journal of Epidemiology, 26:24-31.

Kossove D (1982) Smoke-filled rooms and lower respiratory disease in infants. South Africa Medical Journal, 61:622-624.
Lan Q, Chapman RS, Schreinemachers DM, Tian L, He X (2002) Household stove improvement and risk of lung cancer in Xuanwei, China. Journal of the National Cancer Institute, 94:826-835.
Lei YX, Cai WC, Chen YZ, Du YX (1996) Some lifestyle factors in human lung cancer: a case-control study of 792 lung cancer cases. Lung Cancer, 14: S121-136.

Lin Z-G (1995) [China's climate and its extremes] Zhong Guo De Qi Hou Ji Qi Ji Zhi. The Commercial Press of China, Beijing.
Liu BQ, Peto R, Chen ZM et al. (1998) Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. British Medical Journal, 317:1411-1422.

Liu Q, Sasco AJ, Riboli E, Hu MX (1993) Indoor air pollution and lung cancer in Guangzhou, People's Republic of China. American Journal of Epidemiology, 137:145-154.

Liu ZY, He XZ, Chapman RS (1991) Smoking and other risk factors for lung cancer in Xuanwei, China. International Journal of Epidemiology, 20:26-31.

Loomis D, Castillejos M, Gold DR, McDonnell W, Borja-Aburto VH (1999) Air pollution and infant mortality in Mexico City. Epidemiology, 10: 118-123.

Luo RX, Wu B, Yi YN, Huang ZW, Lin RT (1996) Indoor burning coal air pollution and lung cancer-a case-control study in Fuzhou, China. Lung Cancer, 14:S113-119.

Mahalanabis D, Gupta S, Paul D et al. (2002) Risk factors for pneumonia in infants and young children and the role of solid fuel for cooking: a case-control study. Epidemiology and Infection, 29:65-71.

Mahapatra R (2003) Up in smoke: National programme on improved chulhas put on inexplicable hold. Down To Earth, 11:3-5.

Malik SK (1985) Exposure to domestic cooking fuels and chronic bronchitis. Indian Journal of Chest Diseases and Allied Sciences, 27:171-174.

Mavalankar DV, Trivedi CR, Grah RH (1991) Levels and risk factors for perinatal mortality in Ahmedabad, India. Bulletin of the World Health Organization, 69:435-442.

Mehta S, Smith KR (2002) Exposure atlas for household energy and indoor air pollution modeling component: predicting household pollution levels. World Bank Energy Sector Management Assistance Programme, Washington, DC.

Menezes AM, Victora CG, Rigatto M (1994) Prevalence and risk factors for chronic bronchitis in Pelotas, RS, Brazil: a population-based study. Thorax, 49:1217-1221.

Mishra VK, Retherford RD, Smith KR (1999a) Biomass cooking fuels and prevalence of blindness in India. Journal of Environmental Medicine, 1: 189-199.

Mishra VK, Retherford RD, Smith KR (1999b) Biomass cooking fuels and prevalence of tuberculosis in India. International Journal of Infections Diseases, 3:119-129.

Mohamed N, Ng'ang'a L, Odhiambo J, Nyamwaya J, Menzies R (1995) Home environment and asthma in Kenyan schoolchildren: a case-control study. Thorax, 50:74-78.

Mohan M, Sperduto RD, Angra SK et al. (1989) India-US case-control study of age-related cataracts. India-US case-control study group. Archives of Ophthalmology, 107:670-676.

Morris K, Morgenlander M, Coulehan JL, Gahagen S, Arena VC, Morganlander M (1990) Wood-burning stoves and lower respiratory tract infection in American Indian children. American Journal of Diseases of Children, 144: 105-108.

Mtango FD, Neuvians D, Broome CV, Hightower AW, Pio A (1992) Risk factors for deaths in children under 5 years old in Bagamoyo district, Tanzania. Tropical Medicine and Parasitology, 43:229-233.

Naeher LP, Leaderer BP, Smith KR (2000a) Particulate matter and carbon monoxide in highland Guatemala: indoor and outdoor levels from traditional and improved wood stoves and gas stoves. Indoor Air, 10:200-205.

Naeher LP, Smith KR, Leaderer BP, Mage D, Grajeda R (2000b) Indoor and outdoor PM2.5 and CO in high- and low-density Guatemalan villages. Journal of Exposure Analysis and Environmental Epidemiology, 10:544-551.
NCAER (2002) Household survey to evaluate national program for improved Chula. National Center for Applied Economic Research, Delhi.
NFHS (1995) National family health survey (MCH and family planning): India, 1992-93. International Institute for Population Sciences, Bombay.
NHLBI/WHO (2001) Global initiative for chronic obstructive lung disease. National Heart Lung Blood Institute/World Health Organization, Bethesda, MA.

Norboo T, Angchuk PT, Yahya M et al. (1991) Silicosis in a Himalayan village population: role of environmental dust. Thorax, 46:341-343.
O’Dempsey T, McArdle TF, Morris J et al. (1996) A study of risk factors for pneumococcal disease among children in a rural area of West Africa. International Journal of Epidemiology, 25:885-893.
Padmavati S, Pathak SN (1959) Chronic cor pulmonale in Delhi. Circulation, 20:343-352.

Pandey MR (1984a) Domestic smoke pollution and chronic bronchitis in a rural community of the hill region of Nepal. Thorax, 39:337-339.
Pandey MR (1984b) Prevalence of chronic bronchitis in a rural community of the hill region of Nepal. Thorax, 39:331-336.
Pandey MR, Regmi HN, Neupane RP, Gautam A, Bhandari DP (1985) Domestic smoke pollution and respiratory function in rural Nepal. Tokai Journal of Experimental and Clinical Medicine, 10:471-81.

Pandey MR, Basnyat B, Neupane RP (1988) Chronic bronchitis and cor pulmonale in Nepal. Mrigendra Medical Trust, Kathmandu.
Pandey MR, Boleij JS, Smith KR, Wafula EM (1989a) Indoor air pollution in developing countries and acute respiratory infection in children. The Lancet, 1:427-429.

Pandey MR, Neupane RP, Gautam A, Shrestha I (1989b) Domestic smoke pollution and acute respiratory infections in a rural community of the hill region of Nepal. Environment International, 15:337-340.
Pereira LA, Loomis D, Conceição GM et al. (1998) Association between air pollution and intrauterine mortality in São Paulo, Brazil. Environmental Health Perspectives, 106:325-329.
Perez-Padilla R, Perez-Guzman C, Baez-Saldana R, Torres-Cruz A (2001) Cooking with biomass stoves and tuberculosis: a case control study. International Journal of Tuberculosis and Lung Disease, 5:1-7.
Perez-Padilla R, Regalado J, Vedal S et al. (1996) Exposure to biomass smoke and chronic airway disease in Mexican women a case-control study. American Journal of Respiratory Critical Care Medicine, 154:701-706.

Pintos J, Franco EL, Kowalski LP, Oliveira BV, Curado MP (1998) Use of wood stoves and risk of cancers of the upper aero-digestive tract: a case-control study. International Journal of Epidemiology, 27:936-940.

Ponka A, Virtanen M (1996) Low-level air pollution and hospital admissions for cardiac and cerebrovascular diseases in Helsinki. American Journal of Public Health, 86:1273-1280.

Pope CA 3rd, Schwartz J, Ransom MR (1992) Daily mortality and PM10 pollution in Utah valley. Archives of Environmental Health, 47:211-217.
Prüss A, Mariotti S (2000) Preventing trachoma through environmental sanitation: a review of the evidence base. Bulletin of the World Health Organization, 78:258-265.

Qureshi KA (1994) Domestic smoke pollution and prevalence of chronic bronchitis/asthma in a rural area of Kashmir. Indian Journal of Chest Diseases and Allied Sciences, 36:61-72.

Ramage J, Roggli V, Bell D, Piantadosi C (1988) Interstitial lung disease and domestic wood burning. American Review of Respiratory Disease, 137: 1229-1232.

Ramakrishna J, Durgaprasad MB, Smith KR (1989) Cooking in India: the impact of improved stoves on indoor air quality. Environment International, 15: 341-352.

Rao C, Qin C, Robison W, Zigler J (1995) Effect of smoke condensate on the physiological integrity and morphology of organ cultured rate lenses. Current Eye Research, 14:295-301.

Ritz B, Yu F (1999) The effect of ambient carbon monoxide on low birth weight among children born in southern California between 1989 and 1993. Environmental Health Perspectives, 107:17-25.

Robin LF, Less PS, Winget M et al. (1996) Wood-burning stoves and lower respiratory illnesses in Navajo children. Pediatric Infectious Disease Journal, 15:859-865.

Rossi G, Vigotti MA, Zanobetti A, Repetto F, Gianelle V, Schwartz J (1999) Air pollution and cause-specific mortality in Milan, Italy, 1980-1989. Archives of Environmental Health, 54:158-164.
Sahlu T, Larson C (1992) The prevalence and environmental risk factors for moderate and severe trachoma in southern Ethiopia. Journal of Tropical Medicine and Hygiene, 95:36-41.
Saiyed HN, Sharma YK, Sadhu HG et al. (1991) Non-occupational pneumoconiosis at high altitude villages in central Ladakh. British Journal of Industrial Medicine, 48:825-829.

Saksena S, Prasad R, Pal RC, Joshi V (1992) Patterns of daily exposure to TSP and CO in the Garhwal, Himalaya. Atmospheric Environment, 26A: 2125-2134.

Sandoval J, Salas J, Martinez-Guerra ML et al. (1993) Pulmonary arterial hypertension and cor pulmonale associated with chronic domestic wood smoke inhalation. Chest, 103:12-20.

Schwartz J (1993) Air pollution and daily mortality in Birmingham, Alabama. American Journal of Epidemiology, 137:1136-1147.

Schwartz J, Dockery DW (1992) Increased mortality in Philadelphia associated with daily air pollution concentrations. American Review of Respiratory Disease, 145:600-604.

Schwartz J, Morris R (1995) Air pollution and hospital admissions for cardiovascular disease in Detroit, Michigan. American Journal of Epidemiology, 142:23-35.
Scram R (1999) Impact of air pollution on reproductive health. Environmental Health Perspectives, 107:A542-A543.
Shah N, Ramankutty V, Premila PG, Sathy N (1994) Risk factors for severe pneumonia in children in south Kerala: a hospital-based case-control study. Journal of Tropical Pediatrics, 40:201-206.
Shalini V, Lothra M, Srinivas L (1994) Oxidative damage to the eye lens caused by cigarette smoke and fuel smoke condensates. Indian Journal of Biochemistry and Biophysics, 31:261-266.
Shen XB, Wang GX, Huang YZ, Xiang LS, Wang XH (1996) Analysis and estimates of attributable risk factors for lung cancer in Nanjing, China. Lung Cancer, 14:S107-112.
Sinton JE, Smith KR, Hu H, Liu J (1995) Indoor air pollution database for China. (WHO/EHG/95.8.) World Health Organization, United Nations Environment Programme, Geneva.
Smith K, Desai M (2002) The contribution of global environmental factors to ill-health. In: Environmental change, climate, and health: issues and research methods. Martens P, McMichael A, eds. Cambridge University Press, Cambridge.

Smith KR (1987) Biofuels, air pollution, and health: a global review. Plenum, New York.

Smith KR (1989) The dialectics of improved stoves. Economic and Political Weekly, 24:517-522.
Smith KR (1993) Fuel combustion, air pollution exposure, and health: the situation in developing countries. Annual Review of Energy and the Environment, 18:529-566.
Smith KR (2000) National burden of disease in India from indoor air pollution. Proceedings of the National Academy of Sciences of the United States of America, 97:13286-13293.

Smith KR, Gu SH, Huang K, Qiu DX (1993) 100 million improved stoves in China: how was it done? World Development, 21:941-961.

Smith KR, Liu Y (1994) Indoor air pollution in developing countries. In: Epidemiology of lung cancer. Samet J, ed. Dekker, New York.
Smith KR, Mehta S (2003) Burden of disease from indoor air pollution in developing countries: comparison of estimates. International Journal of Hygiene and Environmental Health, 206:279-289.

Smith KR, Samet JM, Romieu I, Bruce N (2000) Indoor air pollution in developing countries and acute lower respiratory infections in children. Thorax, 55:518-532.

Sobue T (1990) Association of indoor air pollution and lifestyle with lung cancer in Osaka, Japan. International Journal of Epidemiology, 19:S62-66.
Spengler JD, Samet JM, McCarthy JF (2001) Indoor air quality handbook, McGraw-Hill, New York.
Steenland K, Sieber K, Etzel RA, Pechacek T, Maurer K (1998) Exposure to environmental tobacco smoke and risk factors for heart disease among never smokers in the third national health and nutrition examination survey. American Journal of Epidemiology, 147:932-939.

Strachan DP, Cook DG (1998) Health effects of passive smoking. 6. Parental smoking and childhood asthma: longitudinal and case-control studies. Thorax, 53:204-212.

Taylor H, West S (1989) Hygiene factors and increased risk of trachoma in central Tanzania. Archives of Ophthalmology, 107:1821-1825.
Thomas P, Zelikoff J (1999) Air pollutants: moderators of pulmonary host resistance against infection. In: Air pollution and health. Holgate ST et al., eds. Academic Press, San Diego, CA.
UN (1993) Energy statistics yearbook, United Nations, New York.
UN (1998) World population prospects. United Nations, New York.
UNCHS (1996) An urbanizing world: global report on buman settlements. United Nations Centre on Human Settlements (HABITAT), Oxford University Press, New York.
Velema JP, Ferrera A, Figueroa M et al. (2002) Burning wood in the kitchen increases the risk of cervical neoplasia in HPV-infected women in Honduras. International Journal of Cancer, 97:536-541.

Verma BK, Thakur DK (1995) Effect of stressful environmental factors upon neonatal immune system. Central European Journal of Public Health, 3:25-29.

Victora C, Fuchs S, Flores J, Fonseca W, Kirkwood B (1994) Risk factors for pneumonia among children in a Brazilian metropolitan area. Pediatrics, 93:977-985.

Victora CG, Kirkwood BR, Ashworth A et al. (1999) Potential interventions for the prevention of childhood pneumonia in developing countries: improving nutrition. American Journal of Clinical Nutrition, 70:309-320.
Walsh J (1993) Maternal and perinatal health. In: Disease control priorities in developing countries. Jamison D, ed. Oxford University Press, Oxford.
Wang TJ, Zhou BS, Shi JP (1996) Lung cancer in nonsmoking Chinese women: a case-control study. Lung Cancer, 14:S93-98.

Wang X, Ding H, Ryan L, Xu X (1997) Association between air pollution and low birth weight: a community-based study. Environmental Health Perspectives, 105:514-520.

West S (1992) Does smoke get in your eyes? Journal of the American Medical Association, 268:1025-1026.

West S, Lynch M (1989) Water availability and trachoma. Bulletin of the World Health Organization, 67:71-75.

Windham GC, Eaton A, Hopkins B (1999a) Evidence for an association between environmental tobacco smoke exposure and birthweight: a meta-analysis and new data. Paediatric and Perinatal Epidemiology, 13:35-57.
Windham GC, Von Behren J, Waller K, Fenster L (1999b) Exposure to environmental and mainstream tobacco smoke and risk of spontaneous abortion. American Journal of Epidemiology, 149:243-247.
Woodruff TJ, Grillo J, Schoendorf KG (1997) The relationship between selected causes of postneonatal infant mortality and particulate air pollution in the United States. Environmental Health Perspectives, 105:608-612.

World Bank (1988) Congo: issues and options in the energy sector. 6420-COB. The World Bank Energy Sector Management Assistance Programme (ESMAP), Washington, DC.
World Bank (1998) Chad household energy project staff appraisal report. 17780CD. The World Bank Energy Sector Management Assistance Programme (ESMAP), Washington, DC. (Also available at http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/1999/ 09/17/0000009265_3980624143048/Rendered/PDF/multi_page.pdf.)
World Bank (2000) Africa database CD-ROM. World Bank, Washington, DC.
World Bank (2001) World development indicators, World Bank, Washington, DC.

World Energy Council (1999) The challenge of rural energy poverty in developing countries, World Energy Council, London.
World Resources Institute (2000) World resources institute world resources report-2000-2001, Oxford University Press, New York.
World Resources Institute (2003) Earth-trends: the environmental information portal. Available on the Internet at: http://earthtrends.wri.org/

Wu AH, Henderson BE, Pike MC, Yu MC (1985) Smoking and other risk factors for lung cancer in women. Journal of the National Cancer Institute, 74: 747-751.
Wu Y-L, Cao K-j, Ma G-S et al. (1999) A case-control study of the risk factors of male lung cancer in Guangzhou. Cancer [in Chinese], 18:535-537.
Wu-Williams AH, Dai XD, Blot W et al. (1990) Lung cancer among women in north-east China. British Journal of Cancer, 62:982-987.
Xu X, Ding H, Wang X (1995) Acute effects of total suspended particles and sulfur dioxides on preterm delivery: a community-based cohort study. Archives of Environmental Health, 50:407-415.
Xu X, Niu T, Christian D (1996a) Occupational and environmental risk factors for asthma in rural communities in China. International Journal of Occupational and Environmental Health, 2:172-176.

Xu ZY, Brown L, Pan GW et al. (1996b) Lifestyle, environmental pollution and lung cancer in cities of Liaoning in northeastern China. Lung Cancer, 14: S149-160.

Yang X and Li YB (1990) A preliminary study of the impact of indoor air pollution on the health of the dwellers of Wuhan City. Proceedings of the Fifth International Conference on Indoor Air Quality and Climate (Toronto), 1:83-88.

Zelikoff J (1994) Woodsmoke emissions: effects on host pulmonary immune defence. CIAR Currents, 1:3.
Zelikoff JT, Chen LC, Cohen MD et al. (2003) Effects of inhaled ambient particulate matter on pulmonary antimicrobial immune defense. Inhalation Toxicology, 15:131-150.

Zhang J, Smith KR (2003) Indoor air pollution: a global health concern. British Medical Bulletin, 67:209-225.

Zhang J, Smith KR, Ma Y et al. (2000) Greenhouse gases and other airborne pollutants from household stoves in China: a database for emission factors. Atmospheric Environment, 34:4537-4549.
Zhong L, Goldberg MS, Gao YT, Jin F (1999a) A case-control study of lung cancer and environmental tobacco smoke among nonsmoking women living in Shanghai, China. Cancer Causes and Control, 10:607-616.
Zhong L, Goldberg MS, Gao YT, Jin F (1999b) Lung cancer and indoor air pollution arising from Chinese-style cooking among nonsmoking women living in Shanghai, China. Epidemiology, 10:488-494.

Zodpey S, Ughade S (1999) Exposure to cheaper cooking fuels and risk of agerelated cataract in women. Indian Journal of Occupational and Environmental Medicine, 3:159-161.

## Chapter I9

LEAD EXPOSURE<br>Annette Prüss-Üstün, Lorna Fewtrell, Philip J. Landrigan and José Luis Ayuso-Mateos

## Summary

Exposure to lead causes a number of diseases, including mild mental retardation resulting from loss of IQ points, as well as increased blood pressure, anaemia, and gastrointestinal effects. Several other disease outcomes have been associated with exposure to lead, but evidence is considered insufficient at this time for a quantitative assessment of their impact on health to be made here.

The exposure variable used was the population distribution of bloodlead concentrations. We compiled data on concentrations of lead in blood from general population samples in countries around the world, as reported in more than 700 published studies. Only recent studies (published in or after 1995) were considered, because of changes in lead exposure that have taken place since the 1970 s, mainly as a result of the implementation of lead reduction programmes (e.g. the phasing out of leaded petrol). Where current data were not available, we applied an adjustment of $39 \%$ reduction in blood-lead concentrations to allow for the effects of implementation of five-year lead reduction programmes. For countries for which data were not available, exposure reduction due to existing lead reduction programmes was accounted for by extrapolation.

Blood-lead concentrations of about $0.016 \mu \mathrm{~g} / \mathrm{dl}$ have been measured in pre-industrial humans, indicating that the contribution of natural sources of lead to human exposure is minimal. Estimates published recently suggest that the theoretical-minimum-risk of health effects may occur at blood-lead concentrations as low as $0-1 \mu \mathrm{~g} / \mathrm{dl}$.

The association of increased blood-lead concentrations with loss of IQ points has been described in a meta-analysis by Schwartz (1994). Hazards for blood lead and blood pressure were from a meta-analysis by Schwartz (1994) and a published analysis of data from the second National Health and Nutrition Examination Survey (NHANES II).

Hazards for anaemia and gastrointestinal effects were based on a large review of toxicological and epidemiological data (ATSDR 1999). Based on the results of a study by Schwartz et al. (1990), as a consequence of individual variation we considered that only $20 \%$ of the people with blood-lead concentrations above those indicated by the Agency for Toxic Substances and Disease Registry (ATSDR) would actually develop symptoms.

The number of people with mild mental retardation as a result of IQ loss was determined on the basis of a standardized intelligence curve. To account for the higher prevalence of other mild mental retardation risk factors (e.g. malnutrition) in some subregions, ${ }^{1}$ the prevalence of leadinduced mild mental retardation was adjusted for the known ratio of mental retardation caused by other factors. A number of health outcomes and social consequences of lead exposure (e.g. increased risk of violence and drug abuse) could not be quantified owing to insufficient evidence on hazard size.

In 2000 , an estimated 120 million people around the world had bloodlead concentrations of between 5 and $10 \mu \mathrm{~g} / \mathrm{dl}$, and about the same number had concentrations of $>10 \mu \mathrm{~g} / \mathrm{dl}$. Forty per cent of all children had blood-lead concentrations of $>5 \mu \mathrm{~g} / \mathrm{dl}$ and half of these children had blood-lead concentrations of $>10 \mu \mathrm{~g} / \mathrm{dl}$; of these children, $97 \%$ were living in developing countries. The burden of disease caused by mild mental retardation attributable to exposure to lead resulted in 9.8 million disability-adjusted life years (DALYs), and the burden from cardiovascular diseases caused by elevated blood pressure resulted in 229000 premature deaths and 3.1 million DALYs. In total, these two outcomes alone account for about $0.9 \%$ of the global burden of disease. Several health outcomes resulting from exposure to lead could not be quantified in this analysis, in particular, increased delinquent behaviour and its impact on injuries. Health impacts from anaemia and gastrointestinal effects caused by exposure to lead were relatively small. People affected by exposure to lead were concentrated mainly in developing countries. The burden of disease associated with exposure to lead could be virtually eliminated through interventions that have proven successful in developed countries, most importantly, the removal of lead from petrol.

## 1. Introduction

The toxic nature of lead has been recognized for millennia, with the earliest published reports dating back to 2000 BC (Needleman 1999). However, the range of health effects that exposure to lead can cause and the low concentrations of lead in blood at which these effects can occur is only now being fully appreciated. It is now understood that lead is toxic, especially to children, at levels that were previously thought to be safe.

Lead, due to its multiplicity of uses (e.g. leaded petrol, lead in paints, ceramics, food cans, make-up, traditional remedies, batteries), is present
in air, dust, soil and water to varying degrees. Each of these media can act as a route of human exposure, through ingestion or inhalation and, to a small degree for organic lead compounds, dermal absorption. Human exposure can be assessed directly, through body burden measurements (lead in blood, teeth or bone) or indirectly, by measuring levels of lead in the environment (air, dust, food or water).

Multiple health effects have been associated with lead exposure, including systemic effects (e.g. gastrointestinal effects, anaemia, hypertension and hearing loss), effects on the nervous system (e.g. on behaviour and cognition), on development, and on the reproductive system, as well as genotoxicity, carcinogenicity and social effects (ATSDR 1999). The strength of evidence supporting the association of these health effects with exposure to lead varies, and not all of these effects have been investigated sufficiently to permit quantification of their consequences in terms of disease burden.

### 1.1 Risk factor definition

In this analysis, exposure was characterized by the population mean and the population distribution of blood-lead concentrations. Occupational exposures or "hot spots" (i.e. areas of local relevance where lead levels are unusually high, such as around smelters) were excluded, unless they were assessed within general population samples.

### 1.2 ThEORETICAL-MINIMUM-RISK EXPOSURE DISTRIBUTION

The definition of an elevated concentration of lead in the blood according to the Centers for Disease Control and Prevention (CDC 1991) is $10 \mu \mathrm{~g} / \mathrm{dl}$. However, evidence indicating that some health effects can occur below this threshold is accumulating. Recent analyses suggest that health effects may become apparent at concentrations of $<5 \mu \mathrm{~g} / \mathrm{dl}$ (Lanphear 2000) and, indeed, that no evidence exists for a threshold, even at $1 \mu \mathrm{~g} / \mathrm{dl}$ (Schwartz 1994). For the purpose of this analysis, the concentration of blood-lead incurring the lowest population risk was considered to be $0-1 \mu \mathrm{~g} / \mathrm{dl}$, in the absence of scientific consensus and pending further investigation. The measurement of blood-lead concentrations in preindustrial humans has shown that the contribution of natural sources of lead to human exposure is small; Flegal and Smith (1992) have estimated that pre-industrial humans had blood-lead concentrations of only $0.016 \mu \mathrm{~g} / \mathrm{dl}$.

## 2. Estimating Risk Factor levels

### 2.1 Choice of exposure variable

The concentration of lead in blood was chosen as the measure of exposure of the population because:

- it is an objective physiological measure, which can be measured accurately;
- it is directly related to health outcome and can be expected to reflect exposure more closely than estimates derived from measurement of the concentration of lead in air, soil, dust or food; and
- it is the only parameter for which measurements are available from many parts of the world, thus making it preferable to other physiological measures, such as the concentration of lead in bone.

Blood-lead concentrations are indicative of recent lead exposure (within the preceding few weeks) rather than of cumulative long-term exposure. However, as exposure to lead varies relatively little over a time span of a year (WHO 1998), this measure can also be a sign of longerterm exposure.

### 2.2 Data on blood-lead concentrations

## Data sources

Exposure data were obtained from studies identified principally through Medline searches. The primary database used was compiled by the CDC (1999) and contained exposure data from over 700 articles published between 1965 and 1998. The search strategy used was the term "lead" paired with any of the following keywords: newborn, cord, adult, pregnancy, occupation, blood, tooth, hair, milk, placenta, urine, smelter, ceramics, pottery, petrol, cosmetics, kohl, surma, medicine, neurological deficit, cognitive function, pregnancy outcome, fertility and birth defect. Initial queries were followed up using the "related-articles" option of Medline and further searches on the basis of author. The reference list of each relevant article was also examined. Additionally, the authors conducted searches of the databases LILACS (Latin American and Caribbean Information System of Health Sciences), IMEMR (Index Medicus of the World Health Organization's [WHO] Regional Office for the Eastern Mediterranean-EMRO) and African Index Medicus, using the same keywords. Medline was also re-consulted to ensure coverage of the most recent publications (up to the end of 2000).

## Compilation and presentation of Data

Blood-lead concentrations in the population generally have a log-normal distribution, as reported in numerous countries and populations (Al-Saleh et al. 1999; Baghurst et al. 1995; Brody et al. 1994; CDC 2001; Harlan et al. 1985; Hense et al. 1992; Molla et al. 1997; Pocock et al. 1988; Schwartz 1991; Tong et al. 1998). Geometric means with standard deviations were therefore compiled to represent population exposures.

A few studies reported small differences (typically $<10 \%$ ) between blood-lead concentrations in males and females (e.g. CDC 2001; Nielsen et al. 1998; Omokhodion 1994; Paolielo et al. 1997). We therefore com-
bined exposure data for men and women in this analysis. Exposure data were compiled separately for children and adults where this information was available.

### 2.3 Country estimates and subregional aggregation

## AdJustments for data from countries currently phasing out lead

 IN PETROLBlood-lead concentrations can change dramatically over a few years in response to programmes to eliminate the use of lead in petrol. Therefore, data which were more than one year old from countries that had reduced exposure to lead required downward correction to avoid overestimating exposure.

Decreases in blood-lead concentrations correlate well with the removal of lead from petrol (Thomas et al. 1999). Therefore, progress in phasing out lead was used to adjust exposure levels. Although leaded petrol is not the only source of lead in the environment, it is a good indicator of reduction in exposure to lead (Landrigan et al. 2000). Multiple studies have shown reductions in blood-lead concentrations in parallel with decreases in levels of lead in petrol (Thomas et al. 1999). Full implementation of lead reduction programmes has produced decreases in blood-lead concentrations in children of $\geq 90 \%$ over 25 years (CDC 1997, 2000).

Figure 19.1 shows changes in population blood-lead concentrations in the United States of America over a five-year period, during the early stages of a lead reduction programme (Annest 1983; Annest et al. 1983). During this period, the mean blood-lead concentration dropped by $37 \%$. Other studies conducted in various countries have shown very similar results (Elinder et al. 1986; Schuhmacher et al. 1996; Wietlisbach et al. 1995), with decreases ranging from $30-48 \%$ over a five-year period. We chose the midpoint of $39 \%$ as a reduction factor for data that had been collected 5 years before the initiation of a leaded petrol phasing-out process or, for shorter periods, $7.8 \%$ decrease per year. Data which were $>6$ years old were not used. Only the most recent data were selected, provided that they were consistent with trends observed in older data.

Use of leaded petrol and the timing of any changes in the concentration of lead in petrol were assessed using data derived from the World Resources Institute (WRI 1998), the Earth Summit Watch (UNEP 2000), Car Lines (Walsh 2001) and M.P. Walsh (personal communication, 2002). A summary of the global situation with regard to reducing exposures to lead in petrol is provided in Figure 19.2 and below.

- AFR-D and AFR-E: With the exception of South Africa, countries in these subregions have not implemented lead reduction programmes.
- $A M R-A$ : Blood-lead concentrations are likely to be significantly higher in Cuba than in Canada or the United States, as Cuba started lead reduction programmes more recently.

Figure 19.1 Decrease in mean population blood-lead concentrations in relation to reduction of lead in petrol, ${ }^{a}$ in the United States


Source: Annest (1983).

Figure 19.2 Sales of leaded petrol as a percentage of total petrol sales, by country, end of 2001


[^67]- $A M R-B$ and $A M R-D$ : Most countries have now phased out leaded petrol. However, other sources, such as battery recycling and leadglazed ceramics, are of importance in these subregions. Some countries, including Mexico and Peru, are major producers of lead.
- EMR-B and EMR-D: Some countries have now started a lead reduction programme. Egypt and Saudi Arabia are almost lead-free with regard to petrol.
- EUR-A: Most countries have phased out lead in petrol, and lead reduction programmes have been in place for a considerable period of time.
- EUR-B and EUR-C: While lead reduction programmes in some countries are relatively advanced, the great majority of countries have made little progress. EUR-C is less advanced than EUR-B in terms of phasing out leaded petrol.
- SEAR-B and SEAR-D: While Thailand has phased out lead in petrol, significant amounts of leaded petrol are still used in other countries. Sources other than leaded petrol (especially lead-containing cosmetics and local medications) greatly contribute to exposure in certain parts of these subregions.
- WPR-A: All countries have undertaken lead reduction programmes, most of which have been fully implemented. Japan began implementing lead reduction programmes at a very early stage.
- WPR-B: Some countries in this subregion, such as China and the Philippines, have made considerable efforts to phase out leaded petrol. Other countries have made little or no progress.


## EXPOSURE IN RURAL POPULATIONS

Where leaded petrol is still in use, blood-lead concentrations in rural populations are generally lower than those in urban populations (Nriagu et al. 1997a; Piomelli et al. 1980; Rhainds and Levallois 1993; Vasilios et al. 1997).

Thomas et al. (1999) showed that soon after the elimination of leaded petrol, population blood-lead concentrations tend to converge on an average of $3.1 \mu \mathrm{~g} / \mathrm{dl}$ ( $\mathrm{SD} 2.3 \mu \mathrm{~g} / \mathrm{dl}$ ). This value is similar to the mean, $3.0 \mu \mathrm{~g} / \mathrm{dl}$, of the data from rural areas available from countries in which lead has not been totally or partially phased out $(3.4 \mu \mathrm{~g} / \mathrm{dl}$, Grobler et al. $1985 ; 2.3 \mu \mathrm{~g} / \mathrm{dl}$, Khwaja 2002; $3.8 \mu \mathrm{~g} / \mathrm{dl}$, Nriagu et al. 1997a; $3.4 \mu \mathrm{~g} / \mathrm{dl}$, Piomelli et al. 1980; $2.1 \mu \mathrm{~g} / \mathrm{dl}$, Vasilios et al. 1997). We therefore selected $3.1 \mu \mathrm{~g} / \mathrm{dl}$ as the mean blood-lead concentration for rural populations in countries where concentrations of blood-lead in urban areas were higher than this value. In Latin America and the Caribbean, however, it was assumed that rural populations would have higher concentrations of blood-lead as other sources, such as ceramics and recy-
cling of batteries, contribute significantly to lead exposure in these areas (Romieu 2001a). The mean of the most recently-reported blood-lead concentrations in urban Latin American countries that have phased out lead is $4.3 \mu \mathrm{~g} / \mathrm{dl}$ (Garcia and Mercer 2001; Sepulveda 2000), and we also used this value to represent rural blood-lead concentrations in this subregion.

After the complete phasing-out of leaded petrol, blood-lead concentrations continue to decrease, mainly as a result of additional efforts to reduce other sources of lead in the environment. Recent assessments from the United States have reported mean blood-lead concentrations of $1.6 \mu \mathrm{~g} / \mathrm{dl}$ for the total population and $2.0 \mu \mathrm{~g} / \mathrm{dl}$ for children aged $<5$ years (CDC 2001). In such cases, the same values were used to characterize urban and rural populations, as assuming higher exposures in urban environments did not seem justifiable.

## CALCULATING SUBREGIONAL MEANS

For each country with more than one source of data for blood-lead concentrations, geometric means were calculated (weighted by sample size) after the above adjustments. Means for urban and rural populations were estimated separately. Subregional means were calculated by weighting the mean for each country by the size of its urban population. For subregions with countries at different stages of lead phase-out (all subregions except EUR-A and AFR-D), urban means were estimated separately for countries which had made different degrees of progress in eliminating lead. In summary, we superimposed two or three log-normal distributions for each subregion to characterize the distribution of bloodlead concentrations in the population, each of which was weighted by the size of the population they represent. The urban/rural breakdown was based on data from UNDP (2000).

Fewer data on standard deviations were available than on mean blood-lead concentrations. In order to estimate standard deviations for areas for which data were sparse, we grouped subregions according to economic and lead-use patterns and calculated the average standard deviation for each grouping (AMR-A; EUR-A and WPR-A; AMR-B and D; remaining B and C subregions; remaining D and E subregions). Country averages (weighted by sample size) were estimated, and then averaged into the subregional standard deviations by weighting for the size of the urban population. For the grouping of D and E subregions, we did not weight by population size because the large countries contained within these two groupings were not representative of other countries. Table 19.1 shows the urban means for children and adults, the standard deviations, and the distribution of the population into categories of blood-lead concentration. As mentioned above, mean blood-lead concentrations in rural populations are assumed to be $3.1 \mu \mathrm{~g} / \mathrm{dl}$, or equal to blood-lead concentrations for urban populations in which levels have declined after lead reduction programmes.
Table I9.I Blood-lead concentrations in children and adults, by subregion

| Subregion | AFR-D ${ }^{\text {a,c }}$ | AFR-E ${ }^{\text {ac. }}$ | AMR-A | AMR-B | AMR-D | EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-A | WPR-B |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Mean blood-lead concentration, urban children $(\mu \mathrm{g} / \mathrm{d})^{\mathrm{b}}$ | 11.1 | 9.8 | 2.2 | 7.0 | 9.0 | 6.8 | 15.4 | 3.5 | 5.8 | 6.7 | 7.4 | 7.4 | 2.7 | 6.6 |
| Mean blood-lead concentration, urban adults $(\mu \mathrm{g} / \mathrm{d})^{\mathrm{b}}$ | 11.6 | 10.4 | 1.7 | 8.5 | 10.8 | 6.8 | 15.4 | 3.7 | 9.2 | 6.7 | 7.4 | 9.8 | 2.7 | 3.6 |
| Standard deviation ( $\mu \mathrm{g} / \mathrm{dl}$ ) | 5.6 | 5.6 | 2.9 | 3.9 | 3.9 | 3.9 | 5.6 | 1.9 | 3.0 | 3.0 | 3.0 | 5.6 | 1.9 | 3.0 |
| Percentage of urban population | 36 | 25 | 77 | 74 | 58 | 67 | 37 | 78 | 62 | 72 | 31 | 26 | 80 | 32 |
| Countries with recent data | Nigeria' | $\begin{aligned} & \text { South } \\ & \text { Africaª } \end{aligned}$ | Canada, USA ${ }^{4}$ | Argentina, ${ }^{5}$ Brazil, ${ }^{6}$ Chile, Jamaica, ${ }^{8}$ Mexico, ${ }^{9}$ Uruguay, ${ }^{10}$ Venezuela | $\begin{gathered} \text { Ecuador, }^{12} \\ \text { Nicaragua, }^{13} \\ \text { Peru }^{14} \end{gathered}$ | $\begin{gathered} \text { Saudi } \\ \text { Arabia }{ }^{15} \end{gathered}$ | Egypt, ${ }^{16}$ Morocco, ${ }^{17}$ Pakistan ${ }^{18}$ | Denmark, France, ${ }^{20}$ Germany, ${ }^{2}$ Greece, ${ }^{22}$ Israel, ${ }^{23}$ Sweden ${ }^{24}$ | Poland, ${ }^{25}$ Turkey, ${ }^{26}$ Former Yugoslavia ${ }^{27}$ | $\begin{gathered} \text { Hungary }{ }^{28} \\ \text { Russian } \\ \text { Federation }^{28} \\ \hline \end{gathered}$ | Indonesia, ${ }^{30}$ <br> Thailand ${ }^{3}$ | $\begin{gathered} \text { Bangladesh, }{ }^{32} \\ \text { India }^{33} \end{gathered}$ | Australia, ${ }^{34}$ <br> Japan, ${ }^{35}$ New <br> Zealand, ${ }^{36}$ <br> Singapore ${ }^{37}$ | China, ${ }^{38}$ <br> Micronesia, ${ }^{39}$ <br> Philippines, ${ }^{40}$ <br> Republic of <br> Korea ${ }^{41}$ |
| Percentage of children with $5-10 \mu \mathrm{~g} / \mathrm{d}$ | 18.6 | 19.1 | 12.4 | 21.2 | 23.2 | 23.3 | 18.1 | 22.7 | 22.7 | 23.6 | 21.8 | 19.2 | 14.1 | 21.8 |
| Percentage of children with $10-20 \mu \mathrm{~g} / \mathrm{dl}$ | 10.0 | 8.9 | 4.7 | 16.3 | 16.4 | 15.7 | 10.1 | 5.1 | 13.8 | 16.3 | 11.2 | 8.8 | 2.9 | 10.9 |
| Percentage of children with $>20 \mu \mathrm{~g} / \mathrm{dl}$ | 13.9 | 9.5 | 1.9 | 16.7 | 17.2 | 11.4 | 17.2 | 0.5 | 8.9 | 11.9 | 6.5 | 8.3 | 0.3 | 5.8 |

Table 19.1


Based on these methods, we estimated that globally 120 million people had blood-lead concentrations of between 5 and $10 \mu \mathrm{~g} / \mathrm{dl}$ in the year 2000, and about the same number of people had blood-lead concentrations of $>10 \mu \mathrm{~g} / \mathrm{dl}$. Forty per cent of all children had blood-lead concentrations of $>5 \mu \mathrm{~g} / \mathrm{dl}$, and $20 \%$ had concentrations of $>10 \mu \mathrm{~g} / \mathrm{dl}$, and $97 \%$ of the latter were living in developing countries. Nine per cent of children had blood-lead concentrations of $>20 \mu \mathrm{~g} / \mathrm{dl}$, and $99 \%$ of these children were living in developing countries.

## 3. Estimating Risk FACTOR-DISEASE <br> RELATIONSHIPS

### 3.1 Health outcomes

Exposure to lead affects multiple health outcomes and physiological systems (ATSDR 1999), including the following: hypertension, the gastrointestinal system, anaemia, nephropathy, vitamin D metabolism, decreased growth, the immune system, the nervous system, behavioural/cognitive/IQ effects (and as a result, multiple social effects, including increased risk of violence and drug abuse), nerve conductive effects, hearing loss, effects on reproduction and development and death from encephalopathy.

Evidence relating exposure to lead and various health effects has been reviewed extensively (ATSDR 1988, 1993, 1999; International Programme on Chemical Safety 1977, 1995; National Research Council 1993; Pocock et al. 1994; Schwartz 1994). The most recent comprehensive review of the evidence on the risk factor-disease relationship was conducted by ATSDR in the United States (1999). Health effects considered in this review included systemic effects (e.g. raised blood pressure, gastrointestinal effects and anaemia) and nervous system effects (IQ reduction). Nephropathy and encephalopathy were not included as they rarely occur after environmental exposures but rather as a result of highlevel exposure, such as ingestion of lead or lead salts, e.g. from local medication (Woolf 1990) or occupational exposure. A number of suggested health outcomes (e.g. developmental, reproductive and social effects) were not considered because of difficulties in quantifying the level of exposure at which a health outcome occurs, inadequate evidence of causality, or lack of information regarding baseline disease levels.

### 3.2 Evidence and exposure-Risk relationships

## GENERAL NERVOUS SYSTEM EFFECTS

The central and peripheral nervous systems are considered to be the principal targets affected by toxicity caused by the absorption of lead (Tsuchiya 1986; WHO 1996). Proposed mechanisms of toxicity (reviewed in ATSDR 1999) are based on the ability of lead to inhibit or mimic the action of calcium, and to interact with proteins. In terms of
general effects on the nervous system, the key mechanism is likely to be the substitution of lead for calcium as a "second messenger" in neurons. Lead blocks voltage-regulated calcium channels, inhibiting the influx of calcium and release of neurotransmitters and thus inhibiting synaptic transmission.

The most severe neurological effect of exposure to lead is encephalopathy. However, neurotoxic effects are apparent at much lower blood-lead concentrations than those that cause encephalopathy (i.e. $\leq 90 \mu \mathrm{~g} / \mathrm{dl}$ for children, $\leq 140 \mu \mathrm{~g} / \mathrm{dl}$ for adults, depending on the individual). Studies investigating occupational exposure to lead have reported symptoms such as loss of appetite, malaise, lethargy, headache, fatigue, forgetfulness and dizziness in workers with blood-lead concentrations of $40-120 \mu \mathrm{~g} / \mathrm{dl}$.

In the 1940s, Byers and Lord (1943) reported that children who had previously suffered from lead poisoning made poor progress at school, had a shorter attention span and exhibited behavioural disorders. Such observations were followed by epidemiological studies to determine the effects of low-level exposure to lead on children's intellectual abilities and behaviour. Low-level exposure to lead has been associated with failure to complete schooling, reading disability, longer reaction times, delinquent activity and other signs indicating effects on the central nervous system (Needleman et al. 1990, 1996). A large cohort study (Burns et al. 1999) showed that children exposed to relatively low levels of lead experienced an array of emotional and behavioural problems. Similarly, children who had experienced prenatal and postnatal exposure to lead were found to have an increased risk of cognitive deficit (Bellinger et al. 1990), problem behaviour and other dysfunctions, such as inappropriate approaches to tasks, or difficulty with simple directions and sequences of directions (Bellinger et al. 1994; Leviton et al. 1993). Nevin (2000) reported that long-term trends in population exposure to lead (indexed through use of leaded petrol and paint) were remarkably consistent with changes in violent crime; these findings are consistent with the reported link between IQ and social behaviour. Although it has been suggested that neurophysiological changes may be reversible (Ruff et al. 1993), the results of numerous studies indicate that this is unlikely (Schwartz et al. 2000b; Stokes 1998; Tong et al. 1998).

Effects on the nervous system other than loss of IQ and consequent mental retardation could not be quantified at the population level. This was either because of insufficient evidence linking effects to blood-lead concentrations, or because the outcome was not, in the strict sense, a quantifiable health effect (e.g. behavioural problems).

## Loss of IQ POINTS

Analyses of the body of evidence regarding the link between exposure to lead in early childhood and decrease in IQ score suggest that the relationship is causal (International Programme on Chemical Safety 1995;

Pocock et al. 1994; Schwartz 1994). In a meta-analysis, Schwartz (1994) estimated that a mean loss of 2.6 IQ points was associated with an increase in blood-lead concentration from 10 to $20 \mu \mathrm{~g} / \mathrm{dl}$. This result was robust to the inclusion or exclusion of the results of the strongest individual studies. This meta-analysis included eight cross-sectional and longitudinal studies, the largest longitudinal study being the Port Pirie cohort study in Australia (Baghurst et al. 1992) with about 500 participants and a follow-up of several years. Other meta-analyses report similar findings (International Programme on Chemical Safety 1995; Pocock et al. 1994). We selected the analysis by Schwartz (1994) to define the outcome as this study quantified the loss of IQ points and provided a point estimate with a confidence interval. At blood-lead concentrations of $>20 \mu \mathrm{~g} / \mathrm{dl}$, we assumed a loss of $2-5 \mathrm{IQ}$ points (midpoint of 3.5 IQ points) on the basis of the conclusions of the ATSDR report, which were derived from evidence from two studies (de la Burde and Choate 1972; Rummo et al. 1979).

Schwartz (1994) also reported that loss of IQ points is likely to be found between 5 and $10 \mu \mathrm{~g} / \mathrm{dl}$, with an even steeper relationship than in higher concentrations. For comparison, CDC currently defines bloodlead concentrations in excess of $10 \mu \mathrm{~g} / \mathrm{dl}$ as elevated, although acknowledging Schwartz' analysis as evidence for subtle effects at lower concentrations (CDC 2000). The existence of such effects has recently been confirmed by Lanphear et al. (2000), whose analysis of data from about 5000 children showed that children with blood-lead concentrations of between 5 and $10 \mu \mathrm{~g} / \mathrm{dl}$ had poorer cognitive skills.

In summary, to quantify IQ loss in the population, we assumed a linear relationship of 1.3 IQ points lost per $5 \mu \mathrm{~g} / \mathrm{dl}$ increase in blood lead, for blood-lead concentrations of between 5 and $20 \mu \mathrm{~g} / \mathrm{dl}$ (according to the analysis of Schwartz 1994, which showed a loss of 2.6 IQ points for an interval of $10 \mu \mathrm{~g} / \mathrm{dl})$. This linear relationship was divided into three segments, or increments, of $5 \mu \mathrm{~g} / \mathrm{dl}$, and the mean loss of IQ points in each increment was assigned to its mean blood-lead concentration-that is, 0.65 points for the increment $5-10 \mu \mathrm{~g} / \mathrm{dl}$ with a mean of $7.5 \mu \mathrm{~g} / \mathrm{dl}$; 1.95 points for the increment $10-15 \mu \mathrm{~g} / \mathrm{dl}$ with a mean of $12.5 \mu \mathrm{~g} / \mathrm{dl}$; and 3.25 points for the increment $15-20 \mu \mathrm{~g} / \mathrm{dl}$. The mean loss of IQ points is $0.65(1.3 / 2)$ for the first increment, $1.95(0.65+1.3)$ for the second, and $3.25(1.95+1.3)$ for the third. This is illustrated in Figure 19.3. A loss of 3.5 IQ points was assumed for blood-lead concentrations of $>20 \mu \mathrm{~g} / \mathrm{dl}$.

## Mild mental retardation as a consequence of loss of IQ points

As loss of IQ points per se is not considered to be a disease by the international classification of disease (ICD) system, we converted IQ loss into mild mental retardation. Although loss of IQ potentially increases the risk for other diseases, injuries and adverse outcomes, such as violence (Needleman et al. 1996; Nevin 2000), no quantification could be made

Figure 19.3 Decrease in IQ points per increment increase in blood-lead concentration ("best estimate")


Figure 19.4 Loss of IQ points resulting in mild mental retardation

owing to the very heterogeneous nature of these relationships in different populations. For the purpose of this study, a reduction in IQ points was considered to be a disease burden when resulting in mild intellectual impairment, which was defined as having an IQ score of 50-69 points (see Figure 19.4). Intelligence in human populations approximates a normal distribution (Lezak 1995), except for an excess below IQ 50 (representing brain damage and disorder). To estimate the incidence of mild mental retardation resulting from IQ reduction attributable to lead exposure, we first estimated the proportion of children who would have

Table 19.2 Proportion of the population having an IQ score of between 70 and 73.50, assuming a normal distribution ${ }^{2}$

| IQ (points) | \% of the population (assuming <br> a normal IQ distribution) |
| :--- | :---: |
| $70-70.65$ | 0.24 |
| $70-71.95$ | 0.80 |
| $70-73.25$ | 1.45 |
| $70-73.50$ | 1.59 |

a Mean IQ score of 100 , standard deviation of 15 .
Source: Lezak (1995).

IQ scores close to the threshold defined and for whom the loss of a few IQ points would result in a total score of $<70$ points. This included the fractions of the child population with IQ scores of between 70 and 70.65, $71.95,73.25$ and 73.5 points (i.e. the intervals of interest defined by the loss of IQ points as specified above, representing increments of 0.65 , $1.95,3.25$ and 3.5 points), assuming a normal distribution for IQ according to Lezak (mean IQ score of 100, standard deviation of 15; see Table 19.2).

The proportion of mild mental retardation attributable to exposure to lead was estimated as the proportion of children losing a number of IQ points (i.e. ratio of children with blood-lead concentrations within intervals $5-10 \mu \mathrm{~g} / \mathrm{dl}, 10-15 \mu \mathrm{~g} / \mathrm{dl}$ or $15-20 \mu \mathrm{~g} / \mathrm{dl}$; see Table 19.1), multiplied by the fraction of children within the interval $70+x$ IQ points (Table 19.2), for whom a loss of $x$ points results in a final IQ score of $<70$ points (Figure 19.4). This is similar to the method used by INSERM (1999).

This standard IQ distribution does not include additional risk factors for IQ loss, which may be more common than exposure to lead. Several diseases that occur more frequently in developing countries result in cognitive impairment or mental retardation. The detailed estimates of the global burden of disease listed in the World health report 2001 (WHO 2001) provide prevalences of cognitive impairment and mental retardation as a consequence of anaemia, meningitis and pertussis, Japanese encephalitis, ascaris, trichuriasis and infection with hookworm, as well as prevalences of cretinoidism and cretinism caused by iodine deficiency, for the 14 subregions studied here (Murray and Lopez 1996; WHO/EIP, unpublished data, 2001). The literature confirms that there are differences in the prevalence rates of mild mental retardation in developed countries compared to developing countries (Roeleveld et al. 1997), most of these differences being explained by noncongenital causes.

As the normal distributions of IQ scores were established on the basis of data from developed countries, the number of additional cases of mild
mental retardation that are likely to be observed in developing countries because of the additional risks mentioned above had to be estimated. This adjustment was based on the assumption that congenital causes of impaired cognitive function were separable and additive as compared to other risk factors. We thus assumed that the increase in the incidence of mild mental retardation (defined as an IQ score of $50-69$ points) caused by factors other than exposure to lead was proportional to the increase in frequency of the IQ scores of slightly more than 70 points (i.e. between 70 and 73.5 points).

The prevalence of mild mental retardation and cognitive impairment resulting from known, noncongenital causes (see above) was estimated, values from developed and developing countries were compared, and an "adjustment ratio" to account for the increased risk of mental retardation in developing countries was estimated:

$$
A R=\frac{P_{R}-P_{\text {bascline }}+P_{M R} \text { standard }}{P_{M R} \text { sandard }}
$$

AR Adjustment ratio
MR Mild mental retardation
$\mathrm{P}_{\mathrm{R}} \quad$ Region-specific prevalence of MR from known causes from the Global Burden of Disease (GBD) database (WHO/EIP, unpublished data, 2001)
$\mathrm{P}_{\text {bascline }} \quad$ Prevalence of MR from known, noncongenital causes in developed countries
$\begin{array}{ll}\mathrm{P}_{\text {MR standard }} & \begin{array}{l}\text { Prevalence of MR according to the standard distribution of } \\ \text { IQ score }\end{array}\end{array}$
The baseline prevalence of mild mental retardation caused by known, noncongenital factors of 420 per 100000 population (typical in developed countries), and the total rate of mental retardation of 2270 per 100000 population, as taken from the standard distribution, were used. The resulting subregion-specific adjustment ratios are summarized in Table 19.3. We assumed that the same adjustment ratio applied for the

Table 19.3 Adjustment ratios to account for excess prevalence rates of mild mental retardation caused by communicable diseases and iodine deficiency as compared to standardized rates

| Subregion | AFR-D | AFR-E | AMR-A | AMR-B | AMR-D | EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-A | WPR-B |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Adjustment ratio | 2.05 | 2.01 | 1.00 | 2.71 | 2.64 | 1.90 | 1.90 | 1.00 | 1.53 | 1.19 | 3.25 | 2.06 | 1.00 | 3.03 |

considered ranges of IQ (i.e. 70-73.5) in an attempt to take into account additional risks prevailing in certain developing subregions. It should be noted, however, that protein-energy malnutrition, potentially the most important risk factor for mild mental retardation, could not be taken into account here. The estimated prevalence rates per 1000 people affected by lead-induced mild mental retardation were multiplied by the subregion-specific adjustment ratios given in Table 19.3.

Lower mental ability has been associated with lower life expectancy. For example, a longitudinal cohort study of childhood IQ score and survival (Whalley and Deary 2000) found lower survival probabilities at age 76 for people who had a 15 -point disadvantage in IQ score at age 11 years $(\mathrm{RR}=0.76,95 \%$ CI $0.75-0.84)$. Various mechanisms to explain this association were proposed, including childhood IQ as a record of bodily insult (e.g. antenatal care), as an indicator of "system integrity" (the availability of a cerebral reserve capacity to deal with other risks of cognitive decline), as a predictor of healthy behaviour and of entry into safer environments. All but the first mechanism would be relevant in relating the lead-induced loss of IQ points to additional adverse health effects or to reduced life expectancy, rather than considering mild mental retardation in isolation. Additional evidence, however, is needed to quantify this association.

## Increased blood pressure

Schwartz (1995) conducted a meta-analysis examining the relationship between blood-lead concentration and systolic blood pressure in adult males (see Table 19.4). This analysis showed a significant association, with a reduction in blood-lead concentration from $10 \mu \mathrm{~g} / \mathrm{dl}$ to $5 \mu \mathrm{~g} / \mathrm{dl}$ being correlated with a decrease in blood pressure of $1.25 \mathrm{mmHg}(95 \%$ CI 0.87-1.63). It has been suggested that lead exerts an influence on calcium metabolism, which is linked to modulation of blood pressure through vascular tone. In vitro studies have reported increased blood pressure in isolated tail arteries and increased responsiveness to alphaadrenergic stimulation in response to exposure to lead. Raised blood pressure has been associated with increases in risk of cardiovascular and cerebrovascular disease.

A more recent meta-analysis by Nawrot et al. (2002) of data from 32000 men estimated that a two-fold increase in blood-lead concentration was associated with a 1.2 mmHg increase in systolic blood pressure. While this analysis considered a two-fold increase in blood-lead concentration as the measure of exposure, most of the studies in the metaanalysis fell within the interval of $5-10 \mu \mathrm{~g} / \mathrm{dl}$ considered by Schwartz (1995). The analysis of data from the second National Health and Nutrition Examination Survey (NHANES II) revealed decreases in blood pressure of 2 mmHg associated with a reduction in blood-lead from $20 \mu \mathrm{~g} / \mathrm{dl}$ to $15 \mu \mathrm{~g} / \mathrm{dl}$ and also from $15 \mu \mathrm{~g} / \mathrm{dl}$ to $10 \mu \mathrm{~g} / \mathrm{dl}$ (Pirkle et al. 1998; Schwartz 1988).

Table 19.4 The reduction in systolic blood pressure caused by a reduction in blood-lead concentration of $10 \mu \mathrm{~g} / \mathrm{dl}$ in adult males

|  | Reduction in <br> systolic blood <br> pressure (mmHg) | Standard <br> error | Age range <br> (years) | Study type |
| :--- | :---: | :---: | :---: | :--- |
| Reference | 1.74 | 0.73 | $24-55$ | Cross-sectional |
| Orssaud et al. (1985) | 2.24 | 0.86 | $20-74$ | Cross-sectional |
| Schwartz and Pitcher (1988) | 1.45 | 0.49 | $40-59$ | Cross-sectional |
| Pocock et al. (1988) | 3.15 | 1.20 | $57-76$ | Cross-sectional |
| Kromhout (1988) | 0.25 | 0.49 | $18-64$ | Cross-sectional |
| Elwood et al. (1988, Wales) | 0.39 | 0.63 | $49-65$ | Cross-sectional |
| Elwood et al. (1988, Caerphilly) | 1.05 | 0.70 | NA | Cross-sectional |
| Neri et al. (1988) | 1.50 | 0.76 | $23-57$ | Cross-sectional |
|  | 0.90 | 0.39 | $25-60$ | Cross-sectional |
| Moreau et al. (1988) | 0.80 | 1.25 | $28-64$ | Cross-sectional |
| de Kort and Zwennis (I988) | 3.17 | 1.59 | NA | Longitudinal |
| Sharp et al. (I988)) | 1.26 | 0.62 | NA | Cross-sectional |
| Morris et al. (I990) | 1.86 | 0.63 | $40-51$ | Cross-sectional |
| Egeland et al. (1992) | 0.90 | 0.74 | $40-51$ | Longitudinal |
| Møller and Kristensen (1992) | 1.45 | 0.51 | $28-67$ | Longitudinal |
| Møller and Kristensen (1992) |  |  |  |  |

NA Not applicable.
Source: Adapted from Schwartz (1995).

As a conservative estimate, we used the same change of 1.25 mmHg in systolic blood pressure for all three intervals in blood-lead concentration. The relationship was assumed to be linear between 5 and $20 \mu \mathrm{~g} / \mathrm{dl}$, with a 1.25 mmHg rise in blood pressure for each incremental increase of $5 \mu \mathrm{~g} / \mathrm{dl}$. As with loss of IQ points, we converted the linear increase into three equal increments, using the midpoints of the increments with the corresponding increase in blood pressure.

In women, the association between systolic blood pressure and bloodlead concentrations is weaker and less well documented. The most recent and comprehensive estimate, using data from 24000 women suggests that an increase of 0.8 mmHg in systolic blood pressure is associated with a doubling in blood-lead concentration (Nawrot et al. 2002). The association was not different from that of men in a statistically significant manner. We used an increase of 0.8 mmHg in systolic blood pressure for each $5 \mu \mathrm{~g} / \mathrm{dl}$ increase in blood-lead concentration for women, for the interval between 5 and $20 \mu \mathrm{~g} / \mathrm{dl}$.

The disease burden caused by exposure to lead and mediated through increased blood pressure was based on the method used in chapter 6.

## ANAEMIA

Absorbed lead inhibits the activity of a number of enzymes involved in haem biosynthesis. Several studies have shown that the activity of ALAD ( $\delta$-aminolevulinic acid dehydratase) is affected at very low blood-lead levels, with no apparent threshold (International Programme on Chemical Safety 1995). Typically, lead-induced anaemia arises from a combination of reduced haemoglobin formation (caused either by impaired haem synthesis or globin chain formation) and reduction in erythrocyte survival because of haemolysis (National Research Council 1993).

Adverse effects start to appear following decreases in concentrations of haemoglobin, which occurs at blood-lead concentrations of approximately $50 \mu \mathrm{~g} / \mathrm{dl}$ in adults and $40 \mu \mathrm{~g} / \mathrm{dl}$ in children, although there is increasing evidence that effects in children may occur at lower concentrations. Schwartz et al. (1990) studied the relationship between various levels of exposure to lead and anaemia in 579 children aged between 1 and 5 years living close to a primary lead smelter. The analysis related blood lead and hematocrit concentrations, and observed an increase in anaemia in children with blood-lead concentrations of $>20 \mu \mathrm{~g} / \mathrm{dl}$. However, ATSDR (1999) defines children with blood-lead concentrations of $\geq 70 \mu \mathrm{~g} / \mathrm{dl}$ and adults with $\geq 80 \mu \mathrm{~g} / \mathrm{dl}$ to be at risk of anaemia, and we chose these thresholds for estimating disease burden. The number of studies looking at more severe health effects and issues of individual variability is relatively limited; however, results suggest that only a proportion of those exposed become ill. In the study conducted by Schwartz et al. (1990), $20 \%$ of children with a blood-lead concentration of $\geq 60 \mu \mathrm{~g} / \mathrm{dl}$ exhibited signs of anaemia. This value was used in the present analysis to estimate the proportion of those exposed who became ill.

## GASTROINTESTINAL EFFECTS

Abdominal pain, constipation, cramps, nausea, vomiting, anorexia and weight loss, collectively known as colic, are early symptoms of lead poisoning in both adults and children. In adults, such symptoms occur at blood-lead concentrations of $>80 \mu \mathrm{~g} / \mathrm{dl}$ (International Programme on Chemical Safety 1995) and typically at concentrations of $100-200 \mu \mathrm{~g} / \mathrm{dl}$, while concentrations of $60-100 \mu \mathrm{~g} / \mathrm{dl}$ are more typical for children (ATSDR 1999). No dose-response relationship for blood-lead and gastrointestinal effects has been published, so the same correction factor $(20 \%)$ as that assumed for anaemia was used.

## OTHER HEALTH EFFECTS

A number of health effects, such as nephropathy and encephalopathy, are associated with higher exposures to lead. These effects were not quantified as they occur in extreme cases for which population-based data from assumed distributions are highly uncertain.

Other health effects associated with lead exposure, such as hearing loss, cognitive deficits and reproductive effects, were not included in this
estimate. Exclusion was based on a number of factors, including difficulty in determining the threshold at which an effect could be expected to occur, inadequate causal evidence, or an outcome that fell outside of those for which the disease burden had been estimated in the GBD project.

## SUMMARY

Table 19.5 summarizes blood-lead concentrations at which the population is considered to be at risk of the health outcomes discussed here and the quantitative relationship between exposure and outcome. The values given in this table do not necessarily indicate the lowest levels at which lead exerts an effect.

## Use of absolute versus relative risk ratios

The majority of studies investigating the relationship between lead exposure and disease have assessed incidence rates of disease for exposed and unexposed individuals in developed countries. These two rates are then generally combined into a relative risk. The exposure-risk relationships could therefore be applied to the populations of developing countries by transferring either the relative risks of disease, or the absolute disease

Table 19.5 Summary of health risks associated with blood-lead concentrations considered in this analysis

|  | Blood-lead concentration <br> threshold $(\mu \mathrm{g} / \mathrm{dl})$ |  |
| :--- | :---: | :---: | :--- |
| Outcome | Children | Adults |$\quad$| Description of relationship |
| :--- |

[^68]rates for those exposed at equivalent levels. A transfer of relative risk rates to a country with higher baseline rates of the considered disease would result in a higher burden of disease for lead-induced illness than if absolute rates were transferred.

In the case of lead-induced outcomes occurring at high concentrations of blood-lead, including gastrointestinal symptoms and anaemia, it may be argued that lead poisoning acts independently of the baseline rates of disease and the presence of other risk factors in the population. We therefore determined incidences for gastrointestinal symptoms and anaemia on the basis of absolute risks rather than of relative risks. However, the risk of anaemia from exposure to lead may be magnified by other risk factors, and therefore a relative risk approach may be envisaged when more solid exposure-risk relationships become available.

With regard to the effects of lead on cognitive functions, data reported in the literature mainly provide mean decreases in IQ points, rather than relative risks for selected decreases. However, exposure to lead may interact with other risk factors, resulting in a magnifying effect, as previously mentioned.

As the exposure distribution relies on a database containing limited information concerning very high concentrations of blood-lead, the estimation of the percentage of the population that experiences extreme levels of exposure becomes very uncertain. Although the number of individuals with lead-induced nephropathy and encephalopathy could, in principle, be estimated, calculating disease burden would be misleading.

### 3.3 Estimation of the number of people affected bY EXPOSURE TO LEAD

To estimate the number of people whose health was affected by exposure to lead, the exposure-risk relationships described in section 3.2 were applied to the fraction of the population having the blood-lead concentrations at which these health effects occur. Figure 19.5 shows schematically how this was applied to the distribution of blood-lead concentrations in a population.

## 4. Sources of uncertainty

We estimated upper and lower uncertainty bounds for the best estimates by selecting upper and lower values for those parameters which were most likely to contribute significantly to uncertainty and which could be quantified.

### 4.1 UnCERTAINTY IN EXPOSURE ASSESSMENT

Standard deviation of blood-Lead concentrations
Fewer data are available for standard deviations of population bloodlead distributions than for mean blood-lead concentrations. We selected

Figure 19.5 Schematic diagram of the distribution of blood-lead concentrations in the population and the number of individuals who are at risk of selected health effects

upper and lower values for the standard deviation by recalculating subregional values after eliminating the upper and lower $20 \%$ of the reported standard deviations. The calculated values varied by approximately $13 \%$.

## TEMPORAL CHANGE IN BLOOD-LEAD CONCENTRATIONS

Population blood-lead concentrations drop rapidly in countries that make substantial efforts to reduce exposure to lead, including the phasing-out of leaded petrol. Although we examined comprehensive sources documenting such efforts, and adjusted blood-lead concentrations to account for changes in exposure since the most recent assessment, uncertainty remained due to differences in the period of time needed to reduce the use of lead and the lack of data on timing of lead reduction programmes for some countries. In addition, lead reduction programmes other than the phasing-out of leaded petrol may have influenced blood-lead concentrations in the population since the year in
which the data were collected. The exposure level may therefore not always exactly correspond to the year 2000, but to a short time span around that year. To account for uncertainty in the timing of lead reduction, mean blood-lead concentrations were varied by $\pm 16 \%$, reflecting a two-year difference in progress in lead reduction programmes, assuming that blood-lead concentrations would change by $39 \%$ over a five-year period. Although not strictly applicable to the subregions in which leaded petrol has already been phased out for some time, we maintained this variation around the mean blood-lead concentration in these subregions to account for other factors that may have contributed to uncertainty.

## EXTRAPOLATION TO DATA-POOR COUNTRIES

Recent exposure information representative of parts of the general population was available for 41 countries. Many countries (or age groups) were, however, not represented. Exposure to lead in these countries was assumed to be similar to that in countries within the same subregion and which shared socioeconomic characteristics and similar implementation of lead reduction programmes.

## MEASUREMENTS OF CONCENTRATIONS OF BLOOD LEAD

Uncertainty in the accuracy of measurement of blood-lead concentrations can be due to a deviation of the measured sample from the study population (i.e. bias), or to contamination problems during sampling or laboratory analysis. Also, the use of blood-lead measurements taken at a single point in time does not capture temporal variation in exposure to lead. Many sources of exposure, however, are likely to occur virtually continuously (e.g. lead in air, in drinking water, in certain foods or through use of leaded ceramics), limiting temporal variation other than that already accounted for.

### 4.2 UnCERTAINTY IN THE EXPOSURE-RISK RELATIONSHIPS

Body burden of lead and associated health effects
Measurements of the concentration of lead in bone, which reflects longterm exposure, may be a better predictor of health effects than concentrations of lead in blood (Cheng et al. 2001; González-Cossío et al. 1997; Hu et al. 1996), but there are relatively few studies on which to base a global estimate. Also, the evidence on exposure-response relationships has not yet been quantified.

## HEALTH EFFECTS THRESHOLDS AND INDIVIDUAL VARIABILITY

Health effects are likely to occur at lower concentrations of blood-lead than have been considered in this study. There may be no threshold for $\mathrm{IQ} /$ cognitive effects, and both renal and cardiac effects have recently been
reported to occur at low concentrations of blood-lead (Cheng et al. 1998; Payton et al. 1994). Health-effect thresholds are linked to individual variability, for which a number of factors are known to be important, including dietary factors (such as calcium; Harlan et al. 1985), general level of health, and genetic differences (e.g. Glen et al. 2001; Kelada et al. 2001; Schwartz et al. 2000a). Current knowledge does not allow the influence of such factors to be assessed precisely.

THE RELATIONSHIP BETWEEN EXPOSURE AND EFFECT FOR BLOOD-LEAD CONCENTRATIONS AND BLOOD PRESSURE

To quantify uncertainty regarding the effects of exposure to lead on blood pressure, the confidence interval around the risks of $30 \%$ reported by the meta-analysis (Schwartz 1995) was used.

## POPULATION-SPECIFIC BACKGROUND MILD MENTAL RETARDATION

A limitation of the approach used to estimate the risk of mild mental retardation attributable to lead exposure was the lack of studies examining the distribution of IQ scores in different populations. Additionally, population distributions have been found to change over time. In our estimate, the same normal distribution was applied worldwide. For estimating the uncertainty in mild mental retardation rates, we varied the mean IQ score ( 98 and 102, instead of 100). The lower bound for mild mental retardation was estimated by assuming that health effects occurred at $>10 \mu \mathrm{~g} / \mathrm{dl}$ instead of at $>5 \mu \mathrm{~g} / \mathrm{dl}$ as used in the best estimate.

### 4.3 Estimation of upper and Lower uncertainty bounds

To derive an upper estimate for the proportion of people affected by lead exposure, the upper estimates for exposure assessment (i.e. upper estimates for standard deviation and mean of blood-lead concentrations) were multiplied by the upper rates in the risk estimates. A similar approach was used to obtain the lower estimates.

## 5. Results

### 5.1 Loss of IQ points and mild mental retardation

The incidence rates of mild mental retardation in children aged <5 years are summarized in Table 19.6. We assumed that loss of IQ and resulting mild mental retardation occurred only once, during the first year of life. Older age groups were assumed to have already experienced this health impact in previous years. Values presented in Table 19.6 are those estimated for the age group $0-1$ year, but are divided by a factor of 5 (as the age group 0-4 years includes five 1-year cohorts of children).

The highest rates of mild mental retardation caused by exposure to lead occurred in developing countries, where the mean blood-lead concentrations were estimated to be many times higher than those in
Table 19.6 Proportion of children aged 0-I year ${ }^{a}$ affected by loss of IQ points caused by exposure to lead, and incidence rates of mild mental retardation caused by exposure to lead in children aged 0-1 ${ }^{\text {a }}$ year, in the year 2000

| Proportion (number per 1000 children) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | AFR-D | AFR-E | AMR-A | AMR-B | AMR-D | EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-A | WPR-B |
| IQ loss category |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 0.65 IQ points | 186 | 191 | 124 | 222 | 232 | 233 | 181 | 227 | 227 | 236 | 218 | 192 | 141 | 218 |
| I.95 IQ points | 66 | 61 | 33 | 104 | 105 | 102 | 66 | 41 | 92 | 106 | 76 | 61 | 23 | 75 |
| 3.25 IQ points | 34 | 28 | 14 | 59 | 58 | 54 | 35 | 10 | 46 | 57 | 36 | 28 | 6 | 34 |
| 3.5 IQ points | 139 | 95 | 21 | 167 | 172 | 114 | 172 | 5 | 89 | 119 | 65 | 83 | 3 | 58 |
| Total | 425 | 375 | 192 | 552 | 567 | 503 | 454 | 283 | 454 | 518 | 395 | 364 | 173 | 385 |
| Mean incidence rate (number per 1000 children) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Mild mental retardation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Best estimate | 7.5 | 5.8 | 1.1 | 13.2 | 10.2 | 7.6 | 8.0 | 1.1 | 5.2 | 4.9 | 8.7 | 5.5 | 0.7 | 7.7 |
| Lower estimate | 4.2 | 3.0 | 0.5 | 7.0 | 5.3 | 3.6 | 4.6 | 0.3 | 2.4 | 2.3 | 3.9 | 2.8 | 0.2 | 3.4 |
| Upper estimate | 12.5 | 10.0 | 2.1 | 22.0 | 17.2 | 13.3 | 13.0 | 2.7 | 9.3 | 8.6 | 16.3 | 9.7 | 1.7 | 14.6 |

a GBD results were reported for children aged 0-4 years. As the effect is mostly irreversible, the entire cohort will remain affected.
developed countries. Latin American regions had relatively high incidence rates despite recent efforts to phase out lead.

### 5.2 Increased blood pressure

The proportion of adult men and women affected by increased blood pressure in age groups ranging from 20 to 79 years, are displayed in Table 19.7. To calculate the burden of disease for ischaemic heart disease, cerebrovascular disease, hypertensive disease and other cardiac diseases, these rates were converted into disease-specific relative risks, according to the methods used for the risk factor in chapter 6.

### 5.3 AnAEMIA AND GASTROINTESTINAL SYMPTOMS

Tables 19.8 and 19.9 summarize the proportion of people affected by anaemia and gastrointestinal symptoms, assuming that these people are not removed from the source of lead or treated in order to reduce their blood-lead concentrations.

As anaemia and gastrointestinal effects were not included in the list of diseases for which baseline global data were available, they could not be quantified in terms of the attributable fraction of total disease or DALYs.

## 6. Discussion

This estimate of the global burden of disease caused by exposure to lead suggests that lead had a significant impact on health in the year 2000, mainly in developing countries where lead reduction programmes have not yet been fully implemented or, in some cases, initiated. In many subregions, relatively large fractions of the population had significantly elevated blood-lead concentrations. In particular, blood-lead concentrations in many developing countries in 2000 were comparable to, or even higher than, concentrations reported in the United States and Europe in the 1970s. The main disease end-points considered in this analysis included mild mental retardation caused by cognitive impairment and reduction of IQ, ischaemic heart disease, cerebrovascular disease, hypertensive disease and other cardiac diseases induced by increased blood pressure. Several additional disease outcomes associated with exposure to lead could not be considered in this analysis, either because the evidence was considered insufficient for a quantitative assessment at this point in time, or because baseline global data were not available. We estimated that 120 million people around the world had blood-lead concentrations of between 5 and $10 \mu \mathrm{~g} / \mathrm{dl}$ in the year 2000 , and about the same number had concentrations of $>10 \mu \mathrm{~g} / \mathrm{dl}$. Forty per cent of all children had blood-lead concentrations of $>5 \mu \mathrm{~g} / \mathrm{dl}$ and $20 \%$ had concentrations of $>10 \mu \mathrm{~g} / \mathrm{dl} ; 97 \%$ of these children were living in developing countries. These exposures resulted in a burden of disease of 9.8 million DALYs caused by mild mental retardation and 229000
Table 19.7 Number of adult men and women ${ }^{\text {a }}$ (per 1000) affected by increased systolic blood pressure with increased blood-

| Incremental increase in blood-lead concentration | Number of adults (per 1000) |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | AFR-D | AFR-E | AMR-A | AMR-B | AMR-D | EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-A | WPR-B |
| 0.625 mmHg in males, 0.4 mmHg in females | 185 | 191 | 91 | 221 | 226 | 233 | 181 | 243 | 225 | 236 | 218 | 62 | 141 | 206 |
| 1.875 mmHg in males, <br> 1.2 mmHg in females | 66 | 61 | 23 | 108 | 106 | 102 | 66 | 46 | 106 | 106 | 76 | 29 | 17 | 61 |
| 3.125 mmHg in males, 2.0 mmHg in females | 34 | 28 | 9 | 63 | 60 | 54 | 35 | 11 | 61 | 57 | 36 | 97 | 3 | 24 |
| 3.75 mmHg in males, 2.4 mmHg in females | 143 | 98 | 11 | 199 | 201 | 114 | 172 | 6 | 155 | 119 | 65 | 97 | 1 | 28 |
| Total | 428 | 378 | 134 | 591 | 593 | 503 | 454 | 306 | 547 | 518 | 395 | 285 | 162 | 319 |
| a Aged 20-79 years. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Table 19.8 Number of people (per 1000) affected by anaemia caused by exposure to lead in the year 2000

| Number of people (per 1000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | AFR-D | AFR-E | AMR-A | AMR-B | AMR-D | EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-A | WPR-B |
| Children |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Best estimate | 10 | 6 | 0 | 6 | 7 | 2 | 14 | 0 | 2 | 2 | 1 | 5 | 0 | 1 |
| Lower estimate | 4 | 3 | 0 | 2 | 2 | 0 | 6 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| Upper estimate | 20 | 13 | 1 | 16 | 18 | 8 | 27 | 0 | 6 | 9 | 5 | 10 | 0 | 4 |
| Adults |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Best estimate | 9 | 6 | 0 | 7 | 8 | 2 | 13 | 0 | 3 | 2 | 1 | 6 | 0 | 0 |
| Lower estimate | 4 | 2 | 0 | 2 | 2 | 0 | 5 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| Upper estimate | 18 | 12 | 0 | 18 | 21 | 7 | 25 | 0 | 12 | 7 | 4 | 12 | 0 | 1 |

Table 19.9 Number of children (per 1000 children) affected by gastrointestinal effects ${ }^{\text {a }}$ caused by exposure to lead in the year 2000

| Number of children (per 1000 children) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | AFR-D | AFR-E | AMR-A | AMR-B | AMR-D | EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-A | WPR-B |
| Children |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Best estimate | 12 | 7 | 0 | 8 | 9 | 3 | 16 | 0 | 2 | 3 | 2 | 6 | 0 | I |
| Lower estimate | 5 | 3 | 0 | 2 | 3 | 0 | 7 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| Upper estimate | 22 | 14 | 1 | 20 | 22 | 11 | 30 | 0 | 8 | 11 | 6 | 11 | 0 | 5 |

The basis for lower and upper estimates is outlined in section 4.
premature deaths and 3.1 million DALYs caused by cardiovascular disease. These two health outcomes alone account for about $0.9 \%$ of the global burden of disease. ${ }^{2}$

Assuming that each cohort of children aged $0-1$ year is exposed to the same amount of lead year after year, the resulting prevalence of mild mental retardation attributable to lead exposure would be the following (obtained from Table 19.6):

- approximately $1-1.2 \%$ in AMR-B and AMR-D;
- approximately $0.5-0.8 \%$ for AFR, EMR, SEAR, WPR-B, EUR-B and EUR-C; and
- $<0.1 \%$ in developed countries.

Prevalences of mild mental retardation reported for developed countries are generally about 1-3\% (meta-analysis by Andersen et al. 1990; Baird and Sadovnick 1985; Murphy et al. 1995; Roeleveld et al. 1997; WHO 1985). Although it is commonly acknowledged that prevalence rates are higher in developing countries than in developed countries, data are scarce. Reported ranges vary greatly (0.4-9.5\%, Roeleveld et al. 1997), and the median of available rates gives a prevalence of mild mental retardation of $6.5 \%$ (Durkin et al. 1998; Roeleveld et al. 1997). While the contribution of lead to the total incidence of mild mental retardation in developed countries is small, in developing countries as much as $15-20 \%$ of mild mental retardation could be caused by exposure to lead. A study from Australia (Wellesley et al. 1991) estimated that $40 \%$ of mental retardation was of genetic origin, $20 \%$ was caused by environmental factors and $40 \%$ was of unknown etiology. Also, the metaanalysis by Roeleveld et al. (1997) concluded that the high prevalence of mild mental retardation observed in developing countries points towards a role for partly avoidable exogenous influences.

An alternative approach for considering the effects of IQ loss was offered by the Dutch Burden of Disease study (Stouthard et al. 1997). This study used a severity weighting of 0.06 for any loss of IQ of between 1 and 4 points, whether this resulted in mental retardation or remained within the normal IQ range. Such an approach would magnify the effects of exposure to lead, but was not employed in this analysis, since incremental IQ loss is not considered to be a disease in the strict sense. At the same time, although in most instances a loss of IQ points does not lead to a recognizable health condition, it can affect physical functioning and life achievement (e.g. survival and earning potential). The Dutch approach is also supported by recent studies relating reduced mental ability to survival and to reduced lifetime earning capacity (Grosse et al. 2002; Korten et al. 1999; Whalley and Deary 2000).

The burden of ischaemic heart disease, cerebrovascular disease, hypertensive disease and other cardiac disorders caused by exposure to lead amounts globally to 3.1 million DALYs, which is about $2 \%$ of the total burden of cardiovascular disease. Worldwide, exposure to lead causes 229000 deaths from these diseases. Ischaemic heart disease and cerebrovascular diseases are the two main contributors to the burden of disease in this group. Again, the burden is borne mainly by developing countries, owing to the higher exposures to lead in these areas. Together with mild mental retardation caused by exposure to lead, this brings the estimated total burden of disease in this analysis to $0.9 \%$, in terms of DALYs. With quantification of additional outcomes discussed in this chapter, in particular, increased delinquent behaviour and its impact on injuries, the burden would most probably exceed $1 \%$ of the global total.

Lead did not contribute significantly to the global burden of anaemia, because other causes, such as iron deficiency, accounted for much higher prevalences of anaemia. The burden of gastrointestinal symptoms caused by lead was also relatively small as compared to that provoked by major risk factors such as unclean water, poor sanitation and hygiene, or unsafe food.

To improve the accuracy of these estimates, more populationrepresentative blood-lead surveys from subregions for which little information has been published so far would be required. Also, additional information on the health impact of low lead levels would be needed in order to make estimates of the burden of disease caused by low doses of lead. The lack of quantitative information on health effects occurring at low exposure levels, the exclusion of data concerning exposure occurring around hot spots or at the workplace, combined with a number of conservative choices made throughout this study, all contribute to a probable underestimation of the burden of disease caused by lead. One particular health effect that could not be quantified but which has been associated in children with low levels of exposure to lead, and which may cause a significant disease burden, although indirectly, is violence.

Intentional injuries represent an important part of the burden of disease, a proportion of which may be attributable to low blood-lead concentrations encountered at high prevalences in many parts of the world.

In addition to the burden of disease, lead exposure may also contribute to socioeconomic burdens. Glotzer et al. (1995) estimated, for example, that in the United States, 45000 cases of reading disability could be prevented by lead reduction programmes, saving more than US $\$ 900$ million per year in overall costs of remedial education. This also has implications for inequalities in health, as exposure to lead tends to be higher in the lower socioeconomic groups of the population (Needleman 1994). These people often live in areas which are more exposed to industrial pollution or in degraded housing. Grosse et al. (2002) estimated that each IQ point raises worker productivity by $1.76-2.38 \%$, resulting in an economic benefit of US $\$ 110-319$ billion for each year's cohort of children.

All of the lead-induced disease burden is, in principle, preventable by phasing out the use of leaded petrol, reducing industrial emissions, removing lead from products such as ceramics, paint, "folk remedies" (traditional medicines) and food and drink cans, and replacing leaded pipes used for drinking-water. The phasing-out of leaded petrol is a particularly effective intervention, having the advantage of being a single action which permanently removes or reduces a health risk to current and future generations.

Although lead is one of the best-studied environmental pollutants, its full impact on population health is only now coming to light. The impacts of many other potentially harmful substances, such as heavy metals, pesticides or solvents, some of which are steadily accumulating in the environment, are as yet largely unknown.

## 7. Projections

Although exposure to lead can occur via a number of routes and from a range of sources, the level of lead in petrol is a key predictor of bloodlead concentrations at a country level. It has been shown that decreases in the use of leaded petrol are closely followed by parallel decreases in blood-lead concentrations (Annest 1983; Annest et al. 1983; Elinder et al. 1986; Schuhmacher et al. 1996; Thomas et al. 1999; Wietlisbach et al. 1995). We based the projected exposure estimates on predicted changes in transportation energy use by subregion (EIA 2001) and estimated completion dates of leaded petrol phase-out programmes (Walsh 2001). For countries that had not embarked upon lead reduction programmes, it was assumed that urban blood-lead concentrations would rise as a consequence of increases in transportation energy use, shown in Table 19.10. Where a lead reduction programme had been initiated, it was assumed that policy would not change and that the programme would be seen through to completion.

Table 19.10 Transportation energy use by subregion, 1990-2020

| Subregion of analysis | Energy Information Administration regional equivalent | Transportation energy consumption (million barrels of oil equivalent/day) |  |  |  | Average annual \% change |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 1990 | 1999 | 2010 | 2020 | 1999-2020 |
| AMR-A | North America | 13 | 15 | 19 | 23 | 2.0 |
| EUR-A | Western Europe | 6 | 7 | 8 | 9 | 1.0 |
| WPR-A | Industrialized Asia | 3 | 3 | 3 | 3 | 1.0 |
| EUR-B, EUR-C | Eastern Europe/ Former Soviet Union | 3 | 2 | 3 | 4 | 2.8 |
| SEAR-B, SEAR-D, WPR-B | Asia | 3 | 6 | 10 | 16 | 5.1 |
| EMR-B, EMR-D | Middle East | 1 | 2 | 3 | 5 | 4.8 |
| AFR-D, AFR-E | Africa | 1 | 1 | 2 | 2 | 3.0 |
| AMR-B, AMR-D | Central and South America | 2 | 2 | 4 | 6 | 4.6 |

Source: Adapted from EIA (2001).

The starting point for the projection was the estimated mean bloodlead concentration for each subregion in the year 2000. In cases where a national lead reduction programme was due to be completed prior to 2010, urban blood-lead concentrations were assumed to converge at $3.1 \mu \mathrm{~g} / \mathrm{dl}$, according to the calculations of Thomas et al. (1999). The exceptions to this assumption were AMR-B and AMR-D. As countries in these subregions possess other important sources of lead, such as leadglazed pottery, blood-lead concentrations were assumed to drop to $4.3 \mu \mathrm{~g} / \mathrm{dl}$, based on the mean of the most recently reported urban bloodlead concentrations in Latin American countries that have phased out leaded petrol (Garcia and Mercer 2001; Sepulveda et al. 2000). Energy use projections were available only until the year 2020; however, in the absence of other data, trends in 2000-2020 were assumed to continue until 2030. As urbanization is expected to increase steadily in most subregions (UN 1997), we also included a predicted change in urbanization in our projections of changes in blood-lead concentrations.

The effect of changes in energy use on blood-lead levels was calculated using the approach employed to adjust for the phasing-out of lead. Thus, as a $50 \%$ change in lead use (generally equivalent to the completion of a five-year leaded petrol reduction programme) is equal to a $39 \%$ change in blood-lead concentrations (see section 2.3 ), a $1 \%$ change in emissions would result in a $0.78 \%$ change in blood-lead concentrations. It was assumed that mean blood-lead concentrations would not increase beyond $30 \mu \mathrm{~g} / \mathrm{dl}$, as means exceeding this have rarely been reported.

Standard deviations were assumed to remain the same as in the year 2000. Projected blood-lead concentrations are presented in Table 19.11 for children and in Table 19.12 for adults. In subregions where many countries have not yet started to phase out lead (i.e. AFR, EMR, and EUR-B and C), it was estimated that blood-lead concentrations would increase steadily owing to the current widespread use of leaded petrol and lack of actions to reduce lead emissions. In most other subregions, lead emissions were predicted to decline gradually, with subsequent reductions in lead in the environment and in food. A large drop in bloodlead concentrations was projected to occur in many subregions between 2000 and 2010, as existing lead reduction programmes begin to take effect. After this period, rises in blood-lead concentrations would be observed due to continuing increases in the number of persons with elevated blood-lead concentrations living in countries and subregions which have not phased out lead.

Table 19.13 shows the projected incidence of mild mental retardation caused by exposure to lead for the years 2010, 2020 and 2030, using the methods described above and assuming constant prevalences of other diseases with cognitive impairment sequelae. As the proportion of people with elevated blood-lead concentrations will decrease in subregions where lead reduction programmes have recently been initiated (see Table 19.12), the incidence of mild mental retardation caused by exposure to lead is predicted to decline. Where no efforts to reduce exposure to lead are made, urbanization and increases in vehicle emissions are predicted to cause an increased incidence in lead-induced mild mental retardation. In the worst cases, in the African and Eastern Mediterranean subregions, exposure to lead could cause nearly $1.5 \%$ of cases of mild mental retardation.

The estimates for the incidence of anaemia and gastrointestinal symptoms caused by exposure to lead probably already carry the highest uncertainties, owing to the fact that these conditions appear at higher blood-lead concentrations, at which the distribution model is less accurate. Therefore projections for these outcomes have not been presented. In addition to the sources of uncertainty inherent in data collection and in the analysis for 2000, a number of additional uncertainties have been introduced into the projections we have made for future exposures, which are dominated by assumptions about policy and technology changes for lead reduction. Although it may well seem unacceptable that leaded petrol could still be in use in 10, 20 or 30 years' time, it must be remembered that lead has not been removed from the petrol supply of any country except by vigorous and concerted efforts by institutions concerned for the health of the public and especially of children.

For certain countries or subregions with significant additional sources of exposure to lead, the projections made here may be too optimistic. In general, however, we expect that appropriate lead-reduction measures will have been taken at a global level towards the end of the next decade.
Table 19.1I Projections of blood-lead concentrations in children for the years 2010, 2020 and 2030, by subregion

|  | Year | AFR-D | AFR-E | AMR-A | AMR-B | AMR-D | EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-A | WPR-B |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Subregional mean of urban blood-lead concentrations ( $\mu \mathrm{g} / \mathrm{dl}$ ) | 2010 | 14.1 | 9.7 | 2.2 | 4.7 | 4.3 | 7.7 | 17.0 | 3.1 | 6.3 | 8.3 | 3.2 | 3.7 | 2.7 | 3.1 |
|  | 2020 | 18.1 | 12.8 | 2.2 | 4.9 | 4.3 | 10.5 | 20.9 | 3.1 | 7.7 | 10.5 | 3.3 | 3.8 | 2.7 | 3.2 |
|  | 2030 | 23.5 | 17.4 | 2.2 | 5.1 | 4.3 | 14.7 | 20.6 | 3.1 | 9.2 | 13.1 | 3.5 | 3.7 | 2.7 | 3.1 |
| Subregional mean of rural blood-lead concentrations ( $\mu \mathrm{g} / \mathrm{dl}$ ) | $\begin{aligned} & 2010- \\ & 2030 \end{aligned}$ | 3.1 | 3.1 | 2.2 | 4.3 | 4.3 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 2.7 | 3.1 |
| Standard deviation ( $\mu \mathrm{g} / \mathrm{dl}$ ) | $\begin{aligned} & 2010- \\ & 2030 \end{aligned}$ | 5.6 | 5.6 | 2.9 | 3.8 | 3.8 | 3.0 | 5.6 | 1.9 | 3.0 | 3.0 | 3.0 | 5.6 | 1.9 | 3.0 |
| Percentage of children with $5-10 \mu \mathrm{~g} / \mathrm{d}$ | 2010 | 17.9 | 18.3 | 12.5 | 20.8 | 22.1 | 21.5 | 16.7 | 19.6 | 22.0 | 23.5 | 19.8 | 18.2 | 11.2 | 19.7 |
|  | 2020 | 16.9 | 17.7 | 12.5 | 20.3 | 21.7 | 19.1 | 15.7 | 19.6 | 21.5 | 22.5 | 19.6 | 17.9 | 11.3 | 19.6 |
|  | 2030 | 15.7 | 16.8 | 12.5 | 19.8 | 21.4 | 15.6 | 15.3 | 19.6 | 20.2 | 20.7 | 19.3 | 17.8 | 11.3 | 19.5 |
| Percentage of children with $10-20 \mu \mathrm{~g} / \mathrm{dl}$ | 2010 | 10.8 | 9.5 | 4.8 | 13.9 | 13.3 | 16.5 | 10.3 | 3.8 | 14.9 | 18.5 | 8.1 | 8.2 | 1.2 | 8.1 |
|  | 2020 | 11.4 | 10.1 | 4.8 | 13.9 | 13.4 | 17.4 | 10.6 | 3.8 | 16.5 | 20.5 | 8.5 | 8.5 | 1.2 | 8.3 |
|  | 2030 | 11.8 | 10.4 | 4.8 | 13.9 | 13.4 | 16.2 | 11.0 | 3.8 | 17.1 | 21.6 | 8.7 | 8.5 | 1.2 | 8.5 |
| Percentage of children with $>20 \mu \mathrm{~g} / \mathrm{dl}$ | 2010 | 19.2 | 13.1 | 1.6 | 12.1 | 9.2 | 16.3 | 22.9 | 0.4 | 11.7 | 16.5 | 2.8 | 6.9 | 0.0 | 2.7 |
|  | 2020 | 25.4 | 17.7 | 1.6 | 13.6 | 9.5 | 25.1 | 28.7 | 0.4 | 16.3 | 22.7 | 3.6 | 7.7 | 0.0 | 3.0 |
|  | 2030 | 32.2 | 23.1 | 1.6 | 14.5 | 9.9 | 36.0 | 31.2 | 0.4 | 21.0 | 29.8 | 4.6 | 7.8 | 0.0 | 3.3 |

Table 19.12 Projections of blood-lead concentrations in adults for the years 2010, 2020 and 2030, by subregion

|  | Year | AFR-D | AFR-E | AMR-A | AMR-B | AMR-D | EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-A | WPR-B |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Subregional mean of urban | 2010 | 14.8 | 10.1 | 1.7 | 4.8 | 4.3 | 7.7 | 17.0 | 3.1 | 8.6 | 8.3 | 3.2 | 3.8 | 2.4 | 2.7 |
| blood-lead concentrations | 2020 | 19.0 | 13.3 | 1.7 | 5.0 | 4.3 | 10.5 | 20.9 | 3.1 | 10.7 | 10.5 | 3.3 | 3.8 | 2.4 | 2.7 |
| ( $\mu \mathrm{g} / \mathrm{dl}$ ) | 2030 | 24.6 | 18.1 | 1.7 | 5.1 | 4.3 | 14.7 | 20.6 | 3.1 | 12.7 | 13.1 | 3.5 | 3.7 | 2.4 | 2.7 |
| Subregional mean of rural blood-lead concentrations ( $\mu \mathrm{g} / \mathrm{dl}$ ) | $\begin{aligned} & 2010- \\ & 2030 \end{aligned}$ | 3.1 | 3.1 | 2.2 | 4.3 | 4.3 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 2.7 | 3.1 |
| Standard deviation ( $\mu \mathrm{g} / \mathrm{dl}$ ) | $\begin{aligned} & 2010- \\ & 2030 \end{aligned}$ | 5.6 | 5.6 | 3.0 | 3.8 | 3.8 | 3.0 | 5.6 | 1.9 | 3.0 | 3.0 | 3.0 | 5.6 | 1.9 | 3.0 |
| Percentage of adults with | 2010 | 17.8 | 18.3 | 9.3 | 20.8 | 22.1 | 21.5 | 16.7 | 19.6 | 20.4 | 23.5 | 19.8 | 18.2 | 12.4 | 19.8 |
| $5-10 \mu \mathrm{~g} / \mathrm{dl}$ | 2020 | 16.8 | 17.6 | 9.3 | 20.2 | 21.7 | 19.1 | 15.7 | 19.6 | 18.9 | 22.5 | 19.6 | 17.8 | 12.4 | 19.6 |
|  | 2030 | 15.6 | 16.7 | 9.3 | 19.8 | 21.4 | 15.6 | 15.3 | 19.6 | 17.1 | 20.7 | 19.3 | 17.8 | 12.4 | 19.5 |
| Percentage of adults with$10-20 \mu \mathrm{~g} / \mathrm{dl}$ | 2010 | 10.8 | 9.5 | 3.2 | 13.9 | 13.3 | 16.5 | 10.3 | 3.8 | 16.0 | 18.5 | 8.1 | 8.2 | 1.4 | 8.0 |
|  | 2020 | 11.4 | 10.0 | 3.2 | 13.9 | 13.4 | 17.4 | 10.6 | 3.8 | 16.6 | 20.5 | 8.5 | 8.5 | 1.4 | 8.3 |
|  | 2030 | 11.7 | 10.4 | 3.2 | 13.9 | 13.4 | 16.2 | 11.0 | 3.8 | 16.1 | 21.6 | 8.7 | 8.5 | 1.4 | 8.5 |
| Percentage of adults with | 2010 | 19.7 | 13.3 | 0.9 | 12.3 | 9.2 | 16.3 | 22.9 | 0.4 | 18.3 | 16.5 | 2.8 | 6.9 | 0.1 | 2.6 |
| $>20 \mu \mathrm{~g} / \mathrm{dl}$ | 2020 | 26.0 | 18.0 | 0.9 | 13.8 | 9.5 | 25.1 | 28.7 | 0.4 | 24.3 | 22.7 | 3.6 | 7.8 | 0.1 | 2.9 |
|  | 2030 | 32.8 | 23.5 | 0.9 | 14.5 | 9.9 | 36.0 | 31.2 | 0.4 | 29.6 | 29.8 | 4.6 | 7.8 | 0.1 | 3.2 |

Table 19.13 Projected incidence rates of mild mental retardation in children (aged $0-1$ years) caused by exposure to lead in the years 2010, 2020 and $2030^{\text {a }}$

| Incidence rate (number per 1000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | AFR-D | AFR-E | AMR-A | AMR-B | AMR-D | EMR-B | EMR-D | EUR-A | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-A | WPR-B |
| 2010 | 9.4 | 7.0 | 1.0 | 10.4 | 6.9 | 9.2 | 9.7 | 0.9 | 6.0 | 6.1 | 5.6 | 4.9 | 0.4 | 5.2 |
| 2020 | 11.6 | 8.6 | 1.0 | 11.1 | 7.0 | 12.0 | 11.5 | 0.9 | 7.4 | 7.5 | 6.1 | 5.2 | 0.4 | 5.4 |
| 2030 | 13.8 | 10.4 | 1.0 | 11.4 | 7.1 | 14.9 | 12.3 | 0.9 | 8.6 | 9.0 | 6.7 | 5.2 | 0.4 | 5.6 |
| a GBD results reported for the 0-4 year age group. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

In countries where mean population blood-lead concentrations are currently low, it is expected that health effects caused by lead will no longer be a concern for most people. However, exposure to lead is likely to remain a hazard for a minority of people, in particular, the children of the socially disadvantaged, including those living in houses containing leaded paint, or lead piping, or in areas affected by industrial contamination containing lead. Control of these sources will require continuing efforts.

## Note

1 See preface for an explanation of this term.
2 Editorial note: The GBD mortality database includes a small number of deaths (approximately 5000) due to lead-induced mild mental retardation (MMR). These deaths are in fact deaths where MMR, regardless of being caused by lead or otherwise, has been specified as the underlying cause of death in the death registration data from some developed countries. The GBD has not attempted to make consistent estimates of MMR deaths for other regions, or to attribute some of these to lead. Because of the current GBD cause-of-death classification, these deaths are included in the Annex Tables (see the CD-ROM accompanying this book) and summary results in chapters 26 and 27. These deaths are however excluded from the results reported and discussed in this chapter.

## References

Al-Saleh I (1995) Lead exposure in Saudi Arabia and its relationship to smoking. Biometals, 8:243-245.

Al-Saleh I, Khalil MA, Taylor A (1995) Lead, erythrocyte protoporphyrin, and hematological parameters in normal maternal and umbilical cord blood from subjects of the Riyadh region, Saudi Arabia. Archives of Environmental Health, 50:66-73.
Al-Saleh I, Nester M, DeVol E, Shinwari N, Al-Shahria S (1999) Determinants of blood lead levels in Saudi Arabian schoolgirls. International Journal of Occupational and Environmental Health, 5:107-114.

Andersen E, Fledelius HC, Fons M, Haugsted R (1990) An epidemiological study of disability in 4-year-old children from a birth cohort in Frederiksborg County, Denmark. Danish Medical Bulletin, 37:182-185.

Annest JL (1983) Trends in blood lead levels of the US population. In: Lead versus health. Rutter M, Jones RR, eds. John Wiley and Sons, New York.
Annest JL, Pirkle JL, Makuc D, Neese JW, Bayse DD, Kovar MG (1983) Chronological trend in blood lead levels between 1976 and 1980. New England Journal of Medicine, 308:1373-1377.
ATSDR (1988) The nature and extent of lead poisoning in children in the United States: a report to Congress. (Agency for Toxic Substances and Disease Registry.) U.S. Department of Health and Human Services, Atlanta, GA.

ATSDR (1993) Toxicological profile for lead. (Agency for Toxic Substances and Disease Registry.) U.S. Department of Health and Human Services, Atlanta, GA.

ATSDR (1999) Toxicological profile for lead (update). (Agency for Toxic Substances and Disease Registry.) U.S. Department of Health and Human Services, Atlanta, GA.
Australian Institute of Health and Welfare (1996) Lead in Australian children: summary of the National Survey of Lead. Environment Protection Agency, Australia.
Awasthi S, Awasthi R, Pande VK, Srivastav RC, Frumkin H (1996) Blood lead in pregnant women in the urban slums of Lucknow, India. Occupational and Environmental Medicine, 53:836-840.

Azcona-Cruz MI, Rothenberg SJ, Schnaas-Arrieta L, Romero-Placeres M, Perroni-Hernandez E (2000) Relationship of blood lead levels with visualmotor and equilibrium disturbances in children aged 8 to 10 years [in Spanish]. Salúd Pública de México, 42:279-287.
Baghurst PA, McMichael AJ, Tong S, Wigg NR, Vimpani GV, Robertson EF (1995) Exposure to environmental lead and visual-motor integration at age 7 years. The Port Pirie cohort study. Epidemiology, 6: 104-109.
Baghurst PA, McMichael AJ, Wigg NR et al. (1992) Environmental exposure to lead and children's intelligence at the age of seven years. The Port Pirie Cohort Study. New England Journal of Medicine, 327:1279-1284.
Baird PA, Sadovnick AD (1985) Mental retardation in over half-a-million consecutive livebirths: an epidemiological study. American Journal of Mental Deficiency, 89:323-330.

Bashir R, Khan DA, Saleem M, Zaman KU, Malik IA (1995) Blood lead levels and anemia in lead exposed workers. Journal of the Pakistan Medical Association, 45:64-66.
Bellinger D, Leviton A, Allred E, Rabinowitz M (1994) Pre- and postnatal lead exposure and behavioural problems in school-aged children. Environmental Research, 66:12-30.
Bellinger D, Leviton A, Sloman J (1990) Antecedents and correlates of improved cognitive performance in children exposed in utero to low levels of lead. Environmental Health Perspectives, 89:5-11.

Bergdahl IA, Schutz A, Gerhardsson L, Jensen A, Skerfving S (1997) Lead concentrations in human plasma, urine and whole blood. Scandinavian Journal of Work, Environment and Health, 23:359-363.

Bitto A, Horvath A, Sarkany E (1997) Monitoring of blood lead levels in Hungary. Central European Journal of Public Health, 5:75-78.

Blanusa M, Telisman S, Hrsak J, Fugas M, Prpic-Majic D, Saric M (1991) Assessment of exposure to lead and cadmium through air and food in inhabitants of Zagreb. Archives of Industrial Hygiene and Toxicology, 42:257-266.
Bonilla CM, Mauss L, Mauss EA (1998) A community-initiated study of blood lead levels of Nicaraguan children living near a battery factory. American Journal of Public Health, 88:1843-1845.

Brody DJ, Pirkle JL, Kramer RA et al. (1994) Blood lead levels in the US population: phase 1 of the third National Health and Nutrition Examination Survey (NHANES III, 1988-1991). Journal of the American Medical Association, 272:277-283.

Burns JM, Baghurst PA, Sawyer MG, McMichael AJ, Tong S (1999) Life-time low-level exposure to environmental lead and children's emotional and behavioural development at ages 11-13 years. The Port Pirie Cohort Study. American Journal of Epidemiology, 149:740-749.
Byers RK, Lord EE (1943) Late effects of lead poisoning on mental development. American Journal of Diseases in Children, 66:471-494.
CDC (1991) Preventing lead poisoning in young children: a statement by the Centers for Disease Control. Centers for Disease Control and Prevention. Department of Human Services, Atlanta, GA.

CDC (1997) Update: blood lead levels—United States, 1991-1994. MMWR— Morbidity and Mortality Weekly Records, 46:141-146.

CDC (1999) Global lead literature database, Centers for Disease Control and Prevention, Atlanta, GA (unpublished).
CDC (2000) Blood lead levels in young children—United States and selected states, 1996-1999. MMWR—Morbidity and Mortality Weekly Records, 49:1133-1137.
CDC (2001) National report on human exposure to environmental chemicals. Centers for Disease Control and Prevention, Atlanta, GA.
Cheng Y, Schwartz J, Sparrow D, Aro A, Weiss ST, Hu H (2001) Bone lead and blood lead levels in relation to baseline blood pressure and the prospective development of hypertension. American Journal of Epidemiology, 153: 164-171.

Cheng Y, Schwartz J, Vokonas PS, Weiss St, Aro A, Hu H (1998) Electrocardiographic conduction disturbances in association with low level lead exposure. American Journal of Cardiology, 82:594-599.
Chia SE, Chia HP, Ong CN, Jeyaratnam J (1996) Cumulative concentrations of blood lead and postural stability. Occupational and Environmental Medicine, 53:264-268.

Chia SE, Chia HP, Ong CN, Jeyaratnam J (1997) Cumulative blood lead levels and neurobehavioral test performance. Neurotoxicology, 18:793-803.

Cordeiro R, Lima Filho EC, Salgado PE (1996) Neurological disorders in workers with low levels of lead in the blood I: peripheral neuropathy [in Portuguese]. Revista de Saude Pública, 30:248-255

Counter SA, Buchanan LH, Rosas HD, Ortega F (1998) Neurocognitive effects of chronic lead intoxication in Andean children. Journal of the Neurological Sciences, 160(1):47-53.
de Kort WLAM, Zwennis WCM (1988) Blood lead and blood pressure: some implications for the situation in the Netherlands. Environmental Health Perspectives, 78:6770.
de la Burde B, Choate Jr MS (1972) Does asymptomatic lead exposure in children have latent sequelae? Journal of Pediatrics, 81:1088-1091.

Deveaux P, Kibel MA, Dempster WS, Pocock F, Formenti K (1986) Blood lead levels in pre-school children in Cape Town. South African Medical Journal, 69:421-424.
dos Santos AC, Colacciopo S, Dal Bo CM, dos Santos NA (1994) Occupational exposure to lead, kidney function tests, and blood pressure. American Journal of Industrial Medicine, 26:635-643.
D'Souza SJ, Narurkar LM, Narurkar MV (1994) Effect of environmental exposures to lead and cadmium on human lymphocytic detoxifying enzymes. Bulletin of Environmental Contamination and Toxicology, 53:458-463.
Durkin MS, Hasan ZM, Hasan KZ (1998) Prevalence and correlates of mental retardation among children in Karachi, Pakistan. American Journal of Epidemiology, 147:281-288.
Dutkiewicz T, Sokolowska D, Kulka E (1993) Health risk assessment in children exposed to lead compounds in the vicinity of mine-smelter plant "Orzel Bialy". Polish Journal of Occupational Medicine and Environmental Health, 6:71-78.
Egeland GM, Burkhart GA, Schnorr TM et al. (1992) Effects of exposure to carbon disulfide on low density lipoprotein cholesterol concentration and diastolic blood pressure. British Journal of Industrial Medicine, 49:287-293.
EIA (2001) Transportation energy use. In: International energy outlook 2001. Energy Information Administration, Washington DC.
Elinder CG, Friberg L, Lind B, Nilsson B, Svartengren M, Overmark I (1986) Decreased blood lead levels in residents of Stockholm for the period 1980-1984. Scandinavian Journal of Work, Environment and Health, 12:114-120.

Elwood PC, Davey-Smith G, Oldham PD et al. (1988) Two Welsh surveys of blood lead and blood pressure. Environmental Health Perspectives, 78:119-121.
Factor-Litvak P, Kline JK, Popovac D et al. (1996) Blood lead and blood pressure in young children. Epidemiology, 7:633-637.
Factor-Litvak P, Slavkovich V, Liu X et al. (1998) Hyperproduction of erythropoietin in nonanemic lead-exposed children. Environmental Health Perspectives, 106:361-364.

Farias P, Hu H, Rubenstein E et al. (1998) Determinants of bone and blood lead levels among teenagers living in urban areas with high lead exposure. Environmental Health Perspectives, 106:733-737.

Fawcett JP Williams SM, Heydon JL, Walmsley TA, Menkes DB (1996) Distribution of blood lead levels in a birth cohort of New Zealanders at age 21. Environmental Health Perspectives, 104:1332-1335.
Feo O, Fernandez M, Santaella N, Valera L (1993) Plumbemia en madres y sus hijos recien nacidos en el Hospital Central de Maracay [Lead in mothers and newborn children in the Maracay Central Hospital]. Salud de los Trabajadores (Maracay), 1:69-76.
Flegal AR, Smith DR (1992) Lead levels in preindustrial humans. New England Journal of Medicine, 326:1293-1294.
Flurin V, Mauras Y, Le Bouil A, Krari N, Kerjan A, Allain P (1998) Lead blood levels in children under 6 years of age in the Le Mans region [in French]. Presse Médicale, 27:57-59.

Gao W, Li Z, Kaufmann RB, Jones RL et al. (2001) Blood lead levels among children aged 1 to 5 years in Wuxi City, China. Environmental Research, 87:11-19.

Garcia S, Mercer R (2001) La problematica del plomo y salud de la ninez en la Argentina [The problems related to lead and health in children in Argentina]. (Presented at: Latin American situation: lead poisoning in children: evaluation, prevention and treatment, May 07-10, WHO.) Lima, Peru.

Glen BS, Stewart WF, Schwartz BS, Bressler J (2001) Relation of alleles of the sodium-potassium adenosine triphosphate alpha 2 gene with blood pressure and lead exposure. American Journal of Epidemiology, 153: 537-545.

Glotzer DE, Freedberg KA, Bauchner H (1995) Management of childhood lead poisoning: clinical impact and cost-effectiveness. Medical Decision Making, 15:13-24.
Gogte ST, Basu N, Sinclair S, Ghai OP, Bhide NK (1991) Blood lead levels of children with pica and surma use. Indian Journal of Pediatrics, 58:513-519.
González-Cossío T, Peterson KE, Sanin LH et al. (1997) Decrease in birth weight in relation to maternal bone-lead burden. Pediatrics, 100:856-862.
Grobler SR, Maresky LS, Kotze TJ (1992) Lead reduction of petrol and blood lead concentrations of athletes. Archives of Environmental Health, 47: 139-142.

Grobler SR, Rossouw RJ, Maresky LS (1985) Blood lead levels in remote, unpolluted rural area in South Africa. South African Medical Journal, 68:323-324.

Grosse SD, Matte TD, Schwartz J, Jackson RJ (2002). Economic gains resulting from the reduction in children's exposure to lead in the United States. Environmental Health Perspectives, 110:563-569.
Hafeez A, Malik QU (1996) Blood lead levels in pre-school children in Rawalpindi. Journal of Pakistan Medical Association, 46:272-274.
Harlan WR, Landis JR, Schmouder RL, Goldstein NG, Harlan LC (1985) Blood lead and blood pressure. Relationship in the adolescent and adult

US population. Journal of the American Medical Association, 253:530534.

Heinze I, Gross R, Stehle P, Dillon D (1998) Assessment of lead exposure in schoolchildren from Jakarta. Environmental Health Perspectives, 106: 499-501.

Hense HW, Filipiak B, Keil U (1993) The association of blood lead and blood pressure in population surveys. Epidemiology, 4:173-179.
Hense HW, Filipiak B, Novak L, Stoeppler M (1992) Non occupational determinants of blood lead concentrations in a general population. International Journal of Epidemiology, 21:753-762.
Hernandez-Avila M, González-Cossío T, Palazuelos E et al. (1996) Dietary and environmental determinants of blood and bone lead levels in lactating postpartum women living in Mexico City. Environmental Health Perspectives, 104:1076-1082.

Hu H, Aro A, Payton M et al. (1996) The relationship of bone and blood lead to hypertension. The Normative Aging study. Journal of the American Medical Association, 275:1171-1176.
INSERM (1999) Plomb dans l'environnement. Quels risques pour la santé? INSERM, Paris.
International Programme on Chemical Safety (1977) Lead. Environmental health criteria 3. World Health Organization, Geneva.
International Programme on Chemical Safety (1995) Inorganic lead. Environmental health criteria 165. World Health Organization, Geneva.
Jacob B, Ritz B, Heinrich J, Hoelscher BB, Wichmann HE (2000) The effect of low-level blood lead on hematologic parameters in children. Environmental Research, 82:150-159.

Jacoby E (1998) Environmental lead is a problem in Lima, Peru. Environmental Health Perspectives, 10:A170.

Junco-Munoz P, Ottman R, Lee JH, Barton SA, Rivas F, Cerda-Flores RM (1996) Blood lead concentrations and associated factors in residents of Monterrey, Mexico. Archives of Medical Research, 27:547-551.
Kaiser R, Henderson AK, Daley WR et al. (2001) Blood lead levels of primary school children in Dhaka, Bangladesh. Environmental Health Perspectives, 109:563-566.

Kamal AA, Eldamaty SE, Faris R (1991) Blood lead level of Cairo traffic policemen. Science of the Total Environment, 105:165-170.

Karimi PG, Moodley J, Jinabhai CC, Nriagu, J (1999) Maternal and fetal blood lead levels. South African Medical Journal, 89:676-679.
Kelada SN, Shelton E, Kaufmann RB, Khoury MJ (2001) Delta aminolevulinic acid dehydratase genotype and lead toxicity: a HuGE review. American Journal of Epidemiology, 154:1-13.
Khan MH, Khan I, Shah SH, Rashid Q (1995) Lead poisoning-a hazard of traffic and industries in Pakistan. Journal of Environmental Pathology, Toxicology and Oncology, 14:117-120.

Khan DA, Malik IA, Saleem M, Hashim R, Bashir R (1994) Screening for chronic lead poisoning in lead factory workers. Journal of Pakistan Medical Association, 44:239-241.

Khassouani CE, Allain P, Soulaymani R (1997) Lead impregnation in inhabitants of the Rabat region [in French]. Presse Médicale, 26:1714-1716.
Khwaja MA (2002) Studies on blood lead levels in school children resulting from lead exposure due to lead petrol use and increasing road traffic in Pakistan. (Presented at: Environmental threats to the health of children: hazards and vulnerability, March 3-7). Bangkok, Thailand.
Koren G, Chang N, Gonen R et al. (1990) Lead exposure among mothers and their newborns in Toronto. Canadian Medical Association Journal, 142:1241-1244.

Korten AE, Jorm AF, Jiao Z, Leteneeur L, Jacomb PA, Henderson AS et al. (1999) Health, cognitive and psychosocial factors as predictors of mortality in an elderly community sample. Journal of Epidemiology and Community Health, 53:83-8.

Kostial K, Dekanic D, Telisman S et al. (1991) Dietary calcium and blood lead levels in women. Biological Trace Element Research, 28:181-185.
Kromhout D (1988) Blood lead and coronary heart disease risk among elderly men in Zutphen, the Netherlands. Environmental Health Perspectives, 78:43-46.

Lacasaña-Navarro M, Romieu I, Sanìn-Aguirre LH, Palazuelos-Rendón E, Hernández-Avila M (1996) Consumo de calcio y plomo en sangre de mujeres en edad reproductiva [Consumption of calcium and blood lead level in women of reproductive age]. Revista de Investgación Clinica, 48:425-430.

Lal M, Joseph D, Choudhury RK et al. (1991) Studies of blood lead levels in children by proton-induced X-ray emission (PIXE). Science of the Total Environment, 103:209-214.
Landrigan PJ, Boffetta P, Apostoli P (2000) The reproductive toxicity and carcinogenicity of lead: a critical review. American Journal of Industrial Medicine, 38:231-243.
Lanphear BP, Dietrich P, Auinger P, Cox C (2000) Subclinical lead toxicity in US children and adolescents. Public Health Reports, 115:521-529.

Levallois P, Lavoie M, Goulet L, Nantel AJ, Gingras S (1991) Blood lead levels in children and pregnant women living near a lead-reclamation plant. Canadian Medical Association Journal, 144:877-885.

Leviton A, Bellinger D, Allred EN, Rabinowitz M, Needleman H, Schoenbaum S (1993) Pre- and postnatal low-level lead exposure and children's dysfunction in school. Environmental Research, 60:30-43.

Lezak MD (1995) Neuropsychological assessment. Oxford University Press, New York.
López Lara B, Cantú Martinez PC, Hernández Arizpe H, Gómez-Gúzman LG (2000) Niveles de plomo en sangre en recién nacidos y su relación con el peso al nacer [Blood lead levels and its relation to birth weight in newborns]. Salud Pública y Nutrición, 1:1-7.

Maresky LS, Grobler SR (1993) Effect of the reduction of petrol lead on the blood lead levels of South Africans. Science of the Total Environment, 136:43-48.

Matte TD, Figueroa JP, Ostrowski S, Burr G, Jackson-Hunt L, Baker EL (1991) Lead exposure from conventional and cottage lead smelting in Jamaica. Archives of Environmental Contamination and Toxicology, 21: 65-71.

Molla AM, Akram S, Hussain A-K, Hussain NY (1997) Blood lead levels in Kuwaiti school children suspected of lead exposure. Saudi Medical Journal, 18:290-293.

Møller L, Kristensen TS (1992) Blood lead as cardiovascular risk factor. American Journal of Epidemiology, 136:1091-1100.

Moon CS, Zhang ZW, Shimbo S et al. (1995) Dietary intake of cadmium and lead among the general population in Korea. Environmental Research, 71:46-54.

Moreau T, Hannaert G, Orssaud G et al. (1988) Influence of membrane sodium transport upon the relationship between blood lead and blood pressure in a general male population. Environmental Health Perspectives, 78:47-52.
Morris C, McCarron DA, Bennet WM (1990) Low-level lead exposure, blood pressure, and calcium metabolism. American Journal of Kidney Diseases, 15:568-574.

Mujica N. (2001) Niveles de plomo en personas no expuestas ocupationalmente. Valores de referencia [in Spanish]. (Presented at: Latin American situation: lead poisoning in children: evaluation, prevention and treatment, May 7-10, WHO.) Lima, Peru.

Murata K, Araki S, Yokoyama K et al. (1995) Autonomic and central nervous system effects of lead in female glass workers in China. American Journal of Industrial Medicine, 28:233-244.

Murphy CC, Yeargin-Allsopp M, Decoufle P, Drews CD (1995) The administrative prevalence of mental retardation in 10-year-old children in metropolitan Atlanta, 1985 through 1987. American Journal of Public Health, 85:319-323.

Murray CJL, Lopez AD, eds. (1996) Global health statistics: a compendium of incidence, prevalence and mortality estimates for over 200 conditions. Global Burden of Disease and Injury, Vol. 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.
Nawrot TS, Thijs L, Den Hond EM, Roels HA, Staessen JA (2002) An epidemiological re-appraisal of the association between blood pressure and blood lead: a meta-analysis. Journal of Human Hypertension, 16:123-131.
Needleman HL (1994) Preventing childhood lead poisoning. Preventive Medicine, 23:634-637.
Needleman HL (1999) History of lead poisoning in the world. In: Lead poisoning prevention and treatment: implementing a national program in developing countries. George AM ed. The George Foundation, Bangalore.

Needleman HL, Riess JA, Tobin MJ, Biesecker GE, Greenhouse JB (1996) Bone lead levels and delinquent behaviour. Journal of the American Medical Association, 275:363-369.

Needleman HL, Schell A, Bellinger D, Leviton A, Allred EN (1990) The longterm effects of exposure to low doses of lead in childhood. An 11-year followup report. New England Journal of Medicine, 322:83-88.
Neo KS, Goh KT, Sam CT (2000) Blood lead levels of a population group not occupationally exposed to lead in Singapore. Southeast Asian Journal of Tropical Medicine and Public Health, 31:295-300.
Neri LC, Hewitt D, Orser B (1988) Blood lead and blood pressure: analysis of cross-sectional and longitudinal data from Canada. Environmental Health Perspectives, 78:123-126.

Nevin R (2000) How lead exposure relates to temporal changes in IQ, violent crime, and unwed pregnancy. Environmental Research, 83:1-22.
Nielsen JB, Grandjean P, Jorgensen PJ (1998) Predictors of blood lead concentrations in the lead-free gasoline era. Scandinavian Journal of Work, Environment and Health, 24:153-156.

National Research Council (1993) Measuring lead exposure in infants, children, and other sensitive populations. National Academy Press, Washington, DC.
Nriagu J, Jinabhai CC, Naidoo R, Coutsoudis A (1997a) Lead poisoning of children in Africa, II Kwaulu/Natal, South Africa. Science of the Total Environment, 197:1-11.
Nriagu J, Oleru NT, Cudjoe C, Chine A (1997b) Lead poisoning of children in Africa, III Kaduna, Nigeria. Science of the Total Environment, 197: 13-19.

Omokhodion FO (1994) Blood lead and tap water lead levels in Ibadan, Nigeria. Science of the Total Environment, 151:187-190.

Orssaud G, Claude JR, Moreau T et al. (1985) Blood lead concentration and blood pressure. British Medical Journal, 290:244.

Osman K, Elinder C, Schutz A, Grubb A (1999) Biomarkers of nephrotoxicity in children environmentally exposed to lead in Poland. Journal of Environmental Medicine, 1:33-38.

Osterberg K, Borjesson J, Gerhardsson L, Schutz A, Skerfving S (1997) A neurobehavioural study of long-term occupational inorganic lead exposure. Science of the Total Environment, 201:39-51.
Paolielo MM, Gutierrez PR, Turini CA et al. (1997) Reference values for lead levels in blood for the urban population [in Portuguese]. Revista de Saude Publica, 31:144-148.

Payton M, Hu H, Sparrow D, Weiss ST (1994) Low-level lead exposure and renal function in the normative aging study. American Journal of Epidemiology, 140:821-829.

Piomelli S, Corash L, Corash MB et al. (1980) Blood lead concentrations in a remote Himalayan population. Science, 210:1135-1137.

Pirkle JL, Kaufmann RB, Brody DJ, Hickman T, Gunter EW, Paschal DC (1998) Exposure of the US population to lead, 1991-1994. Environmental Health Perspectives, 106:745-750.

Pocock SJ, Shaper AG, Ashby D, Delves HT, Clayton BE (1988) The relationship between blood lead, blood pressure, stroke, and heart attacks in middle aged men. Environmental Health Perspectives, 78:23-30.
Pocock SJ, Smith M, Baghurst P (1994) Environmental lead and children's intelligence: a systematic review of the epidemiological evidence. British Medical Journal, 309:1189-1197.
Ramirez AV, Paucar JC, Medina JM (1997) Blood lead in the inhabitants of four Peruvian localities [in Spanish]. Revista Panamericana de Salud Pública, 1:344-348.
Rhainds M, Levallois P (1993) Umbilical cord blood lead levels in the Quebec City area. Archives of Environmental Health, 48:421-427.
Roeleveld N, Zielhuis GA, Gabreels F (1997) The prevalence of mental retardation: a critical review of recent literature. Developmental Medicine and Child Neurology, 39:125-132.
Romieu I (2001b) Estudios epidemiológicos: diseño, interpretación y uso de datos de plomo [Epidemiological studies: design, interpretation and use of lead data]. (Presented at: Latin American situation: lead poisoning in children: evaluation, prevention and treatment, May 7-10, WHO.) Lima, Peru.
Romieu I (2001a) Plomo en America Latina y el Caribe [Lead in Latin America and the Caribbean]. (Presented at: Latin American situation: lead poisoning in children: evaluation, prevention and treatment. May 7-10, WHO.) Lima, Peru.

Rothenberg SJ, Karchmer S, Schnaas L et al. (1996) Maternal influences on cord blood lead levels. Journal of Exposure Analysis and Environmental Epidemiology, 6:211-227.
Ruff HA, Bijur PE, Markowitz M, Ma YC, Rosen JF (1993) Declining blood lead and cognitive changes in moderately lead-poisoned children. Journal of the American Medical Association, 269:1641-1646.
Rummo JH, Routh DK, Rummo NJ (1979) Behavioural and neurological effects of symptomatic and asymptomatic lead exposure in children. Archives of Environmental Health, 34:120-125.

Sadaruddin A, Agha F, Khatoon N, Sultana K (1995) Blood lead levels in young children in Chakshahzad, Islamabad. Journal of Pakistan Medical Association, 45:215-218.

Saxena DK, Singh C, Murthy RC, Mathur N, Chandra SV (1994) Blood and placental lead levels in an Indian city: a preliminary report. Archives of Environmental Health, 49:106-110.
Schuhmacher M, Belles M, Rico A, Domingo JL, Corbella J (1996) Impact of reduction of lead in gasoline on the blood and hair lead levels in the population of Tarragona Province, Spain, 1990-1995. Science of the Total Environment, 184:203-209.

Schutz A, Barregard L, Sallsten G et al. (1997) Blood lead in Uruguayan children and possible source of exposure. Environmental Research, 74:17-23.

Schwartz BS, Stewart WF, Bolla KI et al. (2000b) Past adult exposure is associated with longitudinal decline in cognitive decline. Neurology, 55:1144-1150.
Schwartz BS, Stewart WF, Kelsey KT et al. (2000a) Associations of tibial lead levels with BsmI polymorphisms in the vitamin D receptor in former organolead manufacturing workers. Environmental Health Perspectives, 108: 199-203.

Schwartz J (1988) The relationship between blood lead and blood pressure in the NHANES II survey. Environmental Health Perspectives, 78:15-22.

Schwartz J, Pitcher H (1988) The relationship between blood lead in the United States. Environmental Health Perspectives, 78:23-30.

Schwartz J (1991) Lead, blood pressure and cardiovascular disease in men and women. Environmental Health Perspectives, 91:71-75.
Schwartz J (1994) Low-level lead exposure and children's IQ: a meta-analysis and search for a threshold. Environmental Research, 65:42-55.

Schwartz J (1995) Lead, blood pressure, and cardiovascular disease in men. Archives of Environmental Health, 50:31-37.
Schwartz J, Landrigan PJ, Baker EL Jr, Orenstein WA, von Lindern IH (1990) Lead-induced anemia: dose-response relationships and evidence for a threshold. American Journal of Public Health, 80:165-168.
Sepulveda AV, Vega Morales J, Delgado BI (2000) Childhood environmental lead exposure in Antofagasta, Chile [in Spanish]. Revista de Médica de Chile, 128:221-232.

Sharp DS, Osterloh J, Becker CE et al. (1988) Blood pressure and blood lead concentrations in bus drivers. Environmental Health Perspectives, 78: 131-137.

Shen XM, Rosen JF, Guo D, Wu SM (1996). Childhood lead poisoning in China. The Science of the Total Environment, 181:101-109.

Shen XM, Sheng SH, Yan CH (2001) Impacts of low-level lead exposure on development of children: recent studies in China. Clinica Chimica Acta, 313:217-220.
Shenoi RP, Khandekar RN, Jaykar AV, Raghunath R (1991) Sources of lead exposure in urban slum school children. Indian Journal of Pediatrics, 28:1021-1027.
Smith LF, Rea E (1995) Low blood lead levels in Northern Ontario-what now? Canadian Journal of Public Health, 86:373-376.

Stokes L (1998) Neurotoxicity in young adults 20 years after childhood exposure to lead: the Bunker Hill experience. Occupational and Environmental Medicine, 55:507-516.
Stouthard MEA, Essink-Bot ML, Bonsle GJ et al. (1997) Disability weights for diseases in the Netherlands. Department of Public Health, Erasmus University, Rotterdam, the Netherlands.

Tepferberg M, Almog S (1999) Prenatal lead exposure in Israel: an international comparison. Israel Medical Association Journal, 1:250-253.

Thomas VM, Socolow RH, Fanelli JJ, Spiro TG (1999) Effects of reducing lead in gasoline: an analysis of the international experience. Environmental Science and Technology, 33:3942-3948.
Tong S, Baghurst PA, Sawyer MG, Burns J, McMichael AJ (1998) Declining blood lead levels and changes in cognitive function during childhood: the Port Pirie cohort study. Journal of the American Medical Association, 280: 1915-1919.
Tsuchiya K (1986) Lead. In: Handbook on the toxicology of metals. Volume II. Specific metals. Friberg L, Nordberg GF, Vouk VB, eds. Elsevier, Amsterdam.
UN (1997) Urban and rural areas. (Publication ST/ESA/SER.A/166.) United Nations, New York.
UNDP (2000) Human development report 2000. United Nations Development Programme. Oxford University Press, New York.
UNEP (2000) The global phaseout of leaded gasoline: a successful initiative. UNEP Earth Summit Watch. Available at http://www.earthsummitwatch.org/gasoline.html.

Vasilios D, Theodor S, Konstantinos S, Evangelos PE, Fotini K, Dimitrios L (1997) Lead concentrations in maternal and umbilical cord blood in areas with high and low air pollution. Clinical and Experimental Obstetrics and Gynecology, 24:187-189.
von Schirnding Y, Mathee A, Robertson P, Strauss N, Kibel M (2001) Distribution of blood lead levels in schoolchildren in selected Cape Peninsula suburbs subsequent to reductions in petrol lead. South African Medical Journal, 91:870-872
Vural N, Gulvendik G (1988) Blood lead level distribution by age group in inhabitants of Ankara. Biological Trace Element Research, 18:8593.

Wahid A, Koul PA, Shah SU, Khan AR, Bhat MS, Malik MA (1997) Lead exposure in papier maché workers. Human and Experimental Toxicology, 16:281-283.
Walsh MP (2001) Car Lines newsletter, 2001(4). (Unpublished document available at http://www.walshcarlines.com
Wan BJ, Zhang Y, Tian CY, Cai Y, Jiang HB (1996) Blood lead dynamics of leadexposed pregnant women and its effects on fetus development. Biomedical and Environmental Sciences, 9:41-45.

Wananukul W, Sirivarasai J, Sriapha C et al. (1998) Lead exposure and accumulation in healthy Thais: assessed by lead levels, EDTA mobilization and heme synthesis-related parameters. Journal of the Medical Association of Thailand, 81:110-116.
Watanabe T, Nakatsuka H, Shimbo S et al. (1996) Reduced cadmium and lead burden in Japan in the past 10 years. International Archives of Occupational and Environmental Health, 68:305-314.

Wellesley D, Hockey A, Stanley F (1991) The aetiology of intellectual disability in Western Australia: a community-based study. Developmental Medicine and Child Neurology, 33:963-973.
Whalley LJ, Deary IJ (2000) Longitudinal cohort study of childhood IQ and survival up to age 76. British Medical Journal, 322:819-822.
White NW, Dempster WS, Pocock F, Kibel MA (1982) Lead absorption in Cape children: a preliminary report. South African Medical Journal, 62:799-802.
WHO (1985) Mental retardation: meeting the challenge. (WHO Offset Publication No. 86.) World Health Organization Geneva.
WHO (1996) Lead. In: Guidelines for drinking-water quality. Volume 2: Health criteria and other supporting information. World Health Organization, Geneva.
WHO (1998) Air management information system (AMIS) 2.0. World Health Organization, Geneva.
WHO (2001) World health report 2001. World Health Organization, Geneva.
Wietlisbach V, Rickenbach M, Berode M, Guillemin M (1995) Time trend and determinants of blood lead levels in a Swiss population over a transition period (1984-1993) from leaded to unleaded gasoline use. Environmental Research, 68:82-90.

Woolf MA (1990) Aetiology of acute lead encephalopathy in Omani infants. Journal of Tropical Pediatrics, 36:328-330.
World Resources Institute (1998) Resources 1998-1999. A guide to the global environment. WRI, United Nations Environment Programme, United Nations Development Programme/World Bank, New York.
Yan CH, Shen XM, Zhang YW et al. (1999) A study on relationship between blood level and physical growth and development of babies and young children in Shanghai. Chinese Journal of Preventive Medicine, 33:269-271.

Yang JS, Kang SK, Park IJ, Rhee KY, Moon YH, Sohn DH (1996) Lead concentrations in blood among the general population of Korea. International Archives of Occupational and Environmental Health, 68:199-202.

Zejda JE, Grabecki J, Krol B, Panasiuk Z, Jedrzejczak A, Jarkowski M (1997) Blood lead levels in urban children of Katowice Voivodship, Poland: results of the population-based biomonitoring and surveillance program. Central European Journal of Public Health, 5:60-64.
Zejda JE, Sokal A, Grabecki J, Panasiuk Z, Jarkowski M, Skiba M (1995) Blood lead concentrations in school children of Upper Silesian Industrial Zone, Poland. Central European Journal of Public Health, 3:92-96.
Zhang ZW, Moon CS, Watanabe T et al. (1997) Background exposure of urban populations to lead and cadmium: comparison between China and Japan. International Archives of Occupational and Environmental Health, 69: 273-281.

Zhang ZW, Subida RD, Agetano MG et al. (1998) Non-occupational exposure of adult women in Manila, the Philippines, to lead and cadmium. Science of the Total Environment, 215:157-165.

## Chapter 20

# Global Climate change 

Anthony J. McMichael, Diarmid Campbell-Lendrum, Sari Kovats, Sally Edwards, Paul Wilkinson, Theresa Wilson, Robert Nicholls, Simon Hales, Frank Tanser, David Le Sueur, Michael Schlesinger and<br>Natasha Andronova

## Summary

Accumulating evidence suggests that the global climate (i.e. conditions measured over 30 years or longer) is now changing as a result of human activities-most importantly, those which cause the release of greenhouse gases from fossil fuels. The most recent report (2001) from the United Nations' Intergovernmental Panel on Climate Change (IPCC) estimates that the global average land and sea surface temperature has increased by $0.6 \pm 0.2^{\circ} \mathrm{C}$ since the mid-19th century, with most change occurring since 1976. Patterns of precipitation have also changed: arid and semiarid regions are becoming drier, while other areas, especially mid-to-high latitudes, are becoming wetter. Where precipitation has increased, there has been a disproportionate increase in the frequency of the heaviest precipitation events. Based on a range of alternative development scenarios and model parameterizations, the IPCC concluded that if no specific actions were taken to reduce greenhouse gas emissions, global temperatures would be likely to rise between 1.4 and $5.8^{\circ} \mathrm{C}$ from 1990 to 2100 . Predictions for precipitation and wind speed were less consistent, but also suggested significant changes.

Risks to human health from climate change would arise through a variety of mechanisms. In this chapter, we have used existing or new models that describe observed relationships between climate variations, either over short time periods or between locations, and a series of health outcomes. These climate-health relationships were linked to alternative projections of climate change, related to unmitigated future emissions of greenhouse gases, and two alternative scenarios for greenhouse gas emissions. Average climate conditions during the period 1961-1990 were used as a baseline, as anthropogenic effects on climate are considered more significant after this period. The resulting models give estimates of the likely future effects of climate change on exposures to thermal extremes and weather disasters (deaths and injuries associated with
floods), the distribution and incidence of malaria, the incidence of diarrhoea, and malnutrition (via effects on yields of agricultural crops). As there is considerable debate over the extent to which such short-term relationships will hold true under the longer-term processes of climate change, we made adjustments for possible changes in vulnerability, either through biological or socioeconomic adaptation. Estimates of future effects were interpolated back to give an approximate measure of the effects of the climate change that have occurred since 1990 on the burden of disease in 2000.

The effects considered here represent only a subset of the ways in which climate change may affect health. Other potential consequences include influences of changing temperature and precipitation on other infectious diseases (including the possible emergence of new pathogens), the distribution and abundance of agricultural pests and pathogens, destruction of public health infrastructure, and the production of photochemical air pollutants, spores and pollens. Rising sea levels may cause salination of coastal lands and freshwater supplies, resulting in population displacements. Changes in the availability and distribution of natural resources, especially water, may increase risk of drought, famine and conflict.

Our analyses suggested that climate change will bring some health benefits, such as lower cold-related mortality and greater crop yields in temperate zones, but these will be greatly outweighed by increased rates of other diseases, particularly infectious diseases and malnutrition in developing regions. We estimated a small proportional decrease in cardiovascular and respiratory disease mortality attributable to climate extremes in tropical regions, and a slightly larger benefit in temperate regions, caused by warmer winter temperatures. As there is evidence that some temperature-attributable mortality represents small displacements of deaths that would occur soon in any case, no assessment was made of the associated increase or decrease in disease burden. Climate change was estimated to increase the relative risk of diarrhoea in regions made up mainly of developing countries to approximately 1.01-1.02 in 2000, and 1.08-1.09 in 2030. Richer countries (gross domestic product [GDP] >US $\$ 6000 /$ year), either now or in the future, were assumed to suffer little or no additional risk of diarrhoea. This modest change in relative risk relates to a major cause of ill-health, so that the estimated associated disease burden in 2000 is relatively large ( 47000 deaths and 1.5 million disability-adjusted life years [DALYs]). Effects on malnutrition varied markedly even across developing subregions, ${ }^{1}$ from large increases in SEAR-D ( $\mathrm{RR}=1.05$ in 2000, and 1.17 in 2030) to no change or an eventual small decrease in WPR-B. Again, these are small relative changes to a large disease burden, giving an estimated 77000 deaths and 2.8 million DALYs in 2000. We calculated much larger proportional changes in the numbers of people killed in coastal floods (RR in EUR-B of up to 1.8 in 2000, and 6.3 in 2030), and inland floods (RR in AMR-

A of up to 3.0 in 2000, and 8.0 in 2030). Although the proportional change is much larger than for other health outcomes, the baseline disease burden is much lower. The aggregate health effect in 2000 is therefore comparatively small (2000 deaths and 193000 DALYs). We estimated relatively large changes in the relative risk of falciparum malaria in countries at the edge of the current distribution. However, most of the estimated attributable disease burden (27000 deaths and 1 million DALYs) is associated with small proportional changes in regions that are already highly endemic, principally in Africa.

Overall, the effects of global climate change are predicted to be heavily concentrated in poorer populations at low latitudes, where the most important climate-sensitive health outcomes (malnutrition, diarrhoea and malaria) are already common, and where vulnerability to climate effects is greatest. These diseases mainly affect younger age groups, so that the total burden of disease due to climate change appears to be borne mainly by children in developing countries.

Considerable uncertainties surround these estimates. These stem partly from the complexity of climate models, partly from gaps in reliable data on which to base climate-health relationships, and, most importantly, from uncertainties around the degree to which current climate-health relationships will be modified by biological and socioeconomic adaptation in the future. These uncertainties could be reduced in subsequent studies by (i) applying projections from several climate models; (ii) relating climate and disease data from a wider range of climatic and socioeconomic environments; (iii) more careful validation against patterns in the present or recent past; and (iv) more detailed longitudinal studies of the interaction of climatic and non-climatic influences on health.

## 1. Introduction

### 1.1 Evidence for climate change in the recent past and PREDICTIONS FOR THE FUTURE

Humans are accustomed to climatic conditions that vary on daily, seasonal and inter-annual time-scales. Accumulating evidence suggests that in addition to this natural climate variability, average climatic conditions measured over extended time periods (conventionally 30 years or longer) are also changing, over and above the natural variation observed on decadal or century time-scales. The causes of this climate change are increasingly well understood. Climatologists have compared climate model simulations of the effects of greenhouse gas (GHG) emissions against observed climate variations in the past, and evaluated possible natural influences such as solar and volcanic activity. They concluded that ". . . there is new and stronger evidence that most of the warming observed over the last 50 years is likely to be attributable to human activities" (IPCC 2001b).

The Third Assessment Report of the IPCC (IPCC 2001b) estimates that globally the average land and sea surface temperature has increased by $0.6 \pm 0.2^{\circ} \mathrm{C}$ since the mid-19th century, with much of the change occurring since 1976 (Figure 20.1). Warming has been observed in all continents, with the greatest temperature changes occurring at middle and high latitudes in the Northern Hemisphere. Patterns of precipitation have also changed: arid and semi-arid regions are apparently becoming drier, while other areas, especially mid-to-high latitudes, are becoming wetter. Where precipitation has increased, there has also been a disproportionate increase in the frequency of the heaviest precipitation events (Karl and Knight 1998; Mason et al. 1999). The small amount of climatic change that has occurred so far has already had demonstrable effects on a wide variety of natural ecosystems (Walther et al. 2002).

Climate model simulations have been used to estimate the effects of past, present and likely future GHG emissions on climate changes. These models are primarily based on data on the heat-retaining properties of gases released into the atmosphere from natural and anthropogenic (man-made) sources, as well as the measured climatic effects of other natural phenomena, as described above. The models used by the IPCC have been validated by "back-casting"-that is, testing their ability to explain climate variations that already occurred in the past. In general, the models are able to give good approximations of past patterns only

Figure 20.1 Observed global average land and sea surface temperatures from 1860 to 2000


[^69]when anthropogenic emissions of non-GHG air pollutants (particulates, dust, oxides of sulfur, etc.) are included along with natural phenomena (IPCC 2001b). This emphasizes that (i) the models represent a good approximation of the climate system; (ii) natural variations are important contributors to climatic variations, but cannot adequately explain past trends on their own; and (iii) anthropogenic GHG emissions are an important contributor to climate patterns, and are likely to remain so in the future.

Considering a range of alternative economic development scenarios and model parameterizations, the IPCC concluded that if no specific actions were taken to reduce GHG emissions, global temperatures would rise between 1.4 and $5.8^{\circ} \mathrm{C}$ from 1990 to 2100 . The projections for precipitation and wind speed are less consistent in terms of magnitude and geographical distribution, but also suggest significant changes in both mean conditions and in the frequency and intensity of extreme events (Table 20.1).

### 1.2 Estimating the effects of climate change on health

Human health is sensitive to temporal and geographical variations in weather (short-term fluctuations in meteorological conditions) and climate (longer-term averages of weather conditions). Weather has not historically been considered as subject to alteration by human actions, although its effects may be lessened by adaptation measures (e.g. Kovats et al. 2000b). While adaptation is also a very important determinant of the health consequences of climate change, the effect of anthropogenic GHG emissions on climate means that climate change can in principle be considered a risk factor that could potentially be altered by human intervention, with associated effects on the burden of disease.

The effects of GHG emissions on human health differ somewhat from the effects of other risk factors in that they are mediated by a diversity of causal pathways (e.g. Figure 20.2; McMichael et al. 1996; Patz et al. 2000; Reiter 2000) and eventual outcomes, typically long delays between cause and effect, and great difficulties in eliminating or substantially reducing the risk factors. An additional challenge is that climate change occurs against a background of substantial natural climate variability, and its health effects are confounded by simultaneous changes in many other influences on population health (Kovats et al. 2001; Reiter 2001; Woodward et al. 1998). Empirical observation of the health consequences of long-term climate change, followed by formulation, testing and then modification of hypotheses would therefore require long timeseries (probably several decades) of careful monitoring. While this process may accord with the canons of empirical science, it would not provide the timely information needed to inform current policy decisions on GHG emission abatement, so as to offset possible health consequences in the future. Nor would it allow early implementation of policies for adaptation to climate changes, which are inevitable owing

Table 20.1 Estimates of confidence in observed and projected changes in extreme weather and climate events

|  | Confidence in observed <br> changes (latter half of 1900s) | Confidence in projected <br> changes (during <br> the 2 Ist century) |
| :--- | :--- | :--- |
| Changes in phenomenon | Likely ${ }^{\text {a }}$ | Very likely ${ }^{\text {a }}$ |


| a | Judgement estimates for confidence: virtually certain (greater than $99 \%$ chance that the result is true); very likely (90-99\% chance); likely (66-90\% chance); medium likelihood ( $33-66 \%$ chance); unlikely (IO-33\% chance); very unlikely (I-I0\% chance); exceptionally unlikely (less than I\% chance). |
| :---: | :---: |
| b | Past and future changes in tropical cyclone location and frequency are uncertain. |
| c | For other areas, there are either insufficient data or conflicting analyses. |
| d | Based on warm season temperature and humidity. |
| Source: | Adapted from IPCC (200 lb). |

to both natural variations and past GHG emissions. Therefore, the best estimation of the future health effects of climate change will necessarily come from modelling based on current understanding of the effects of climate (not weather) variation on health from observations made in the present and recent past, acknowledging the influence of a large range of mediating factors.

Since the early 1990s, IPCC Working Group II has collated some of the accumulating predictions of climate effects on health (IPCC 2001a). The health effects of climate variability and change have also been reviewed by a national committee in the United States of America

Figure 20.2 Pathways through which climate change may affect health


Source: Adapted from Patz et al. (2000).
(National Research Council 2001a) and, more recently, in a book by the World Health Organization (WHO) (McMichael et al. 2003). As yet, however, there has been no concerted attempt to integrate these various research findings into a single standardized estimate of the likely net health effects of climate change, nor to estimate the possible health gains associated with different mitigation and amelioration strategies. In addition to the uncertainties of future climate projections, there are several obstacles to achieving this aim.

- Not all of the probable health outcomes have been modelled, often due to lack of parameterization data and the complexity of causal pathways. Modelling efforts so far have tended to concentrate on those causal relationships that can more easily be modelled, rather than those with the potentially greatest effects (e.g. extreme temperatures on cardiovascular disease mortality, rather than sea-level rise on the health of displaced populations).
- Little emphasis has been given to the validation of models relating climate change to health. Validation would provide a basis for making uncertainty estimates around projections, and would afford an objective criterion for choosing between different models or modelling approaches.
- Adaptations to climate change (i.e. autonomous or planned responses that reduce the vulnerability of populations to the consequences of climate change) are often not addressed.
- Interactions between the effects of climate change and other changes to human populations (e.g. investment in health infrastructure, level and equity of distribution of wealth) are seldom explicitly estimated.
- The various disease-specific models invariably generate outputs in different units, which may be only indirectly related to disease burden (e.g. populations at risk of disease transmission, rather than disease incidence). This hampers estimation of the aggregated health impacts of different scenarios.
- Little effort has previously been directed to describing and understanding the geographical variations in likely impacts.

The first four obstacles are likely to be at least partially addressed in the future, as disease-specific models become more sophisticated and, perhaps more importantly, through the accumulation of greater quantities of reliable data for model parameterization and testing. In this chapter, we have estimated the relative risk of a series of health outcomes under a range of scenarios of climate change, variously mitigated by reducing GHG emissions. In all cases, care was taken to describe explicitly the scientific basis for our estimation, the assumptions that were built into the quantitative models, and to give realistic uncertainty estimates around projections. Later sections describe ways in which specific disease models may be improved.

## 2. Risk factor definition and measurement

### 2.1 Definitions of risk factor and exposure scenarios

The risk factor was defined as current and future changes in global climate attributable to increasing atmospheric concentrations of greenhouse gases (GHGs).

Composite climate scenarios are adopted instead of the (more preferable) continuous measurements of individual climate variables because (i) climate is a multivariate phenomenon, including temperature, precipitation, wind speed, etc., and therefore cannot be measured on a single scale; (ii) climate changes will vary significantly with geography and time: these are not fully captured in global averages of climate variables; and (iii) all aspects of climate are likely to be altered by GHG levels in the atmosphere.

The exposure categories considered here are global climate scenarios resulting at specified points in time over the coming half-century from:

1. unmitigated emissions trends, that is, approximately following the IPCC IS92a scenario;
2. emissions reduction, resulting in stabilization at 750 ppm CO 2 equivalent by the year 2210 (s750);
3. more severe emissions reduction, resulting in stabilization at 550 ppm $\mathrm{CO}_{2}$ equivalent by the year 2170 (s550);
4. average climate conditions for 1961-1990, the World Meteorological Organization (WMO) climate normal (baseline).

Although future GHG emissions are inherently uncertain, the unmitigated emissions scenario adopted here was, until recently, the IPCC mid-range projection, and was very widely used in climate impact modelling. The stabilization categories used here represent plausible, though economically and technically challenging, projections that are dependent on there being major efforts to curtail emissions. Estimated changes in $\mathrm{CO}_{2}$ concentrations, and associated changes in global temperature and sea level, are shown in Table 20.2 and Figure 20.3. Although alternative emissions scenarios for climate stabilization are available, they have not been applied to a wide range of impact models.

We do not attempt to estimate all health outcomes of specific policy/development pathways through which these, or other, GHG levels could be achieved: for example, compliance with the Kyoto Protocol of the United Nations Framework Convention on Climate Change (UNFCCC), or of the world following one or other of the IPCC Special Report on Emissions Scenarios (SRES)—both of which also include descriptions of alternative future global socioeconomic development scenarios. The costs or additional benefits of specific interventions to

Table 20.2 Successive measured and modelled $\mathrm{CO}_{2}$ concentrations, global mean temperature and sea-level rise associated with alternative emissions scenarios

|  | $1961-90$ | 1990 s | 2020 s | 2050 s |
| :--- | :---: | :---: | :---: | :---: |
| Carbon dioxide concentration (ppm) by volume |  |  |  |  |
| HadCM2 unmitigated emissions | 334 | 354 | 441 | 565 |
| S750 | 334 | 354 | 424 | 501 |
| S550 | 334 | 354 | 410 | 458 |
| Temperature $\left({ }^{\circ} \mathrm{C}\right.$ change) |  |  |  |  |
| HadCM2 unmitigated emissions | 0 | 0.3 | 1.2 | 2.1 |
| S750 | 0 | 0.3 | 0.9 | 1.4 |
| S550 | 0 | 0.3 | 0.8 | 1.1 |
| Sea-level (cm change) |  |  |  |  |
| HadCM2 unmitigated emissions | 0 | - | 12 | 25 |
| S750 | 0 | - | 11 | 20 |
| S550 | 0 | - | 10 | 18 |

- No data.

Source: McMichael et al. (2000a).

Figure 20.3 The global average temperature rise predicted from the unmitigated emissions scenario (upper trace), and emission scenario which stabilizes $\mathrm{CO}_{2}$ concentrations at 750 ppm (middle trace) and at 550 ppm (lower trace)


Note: All values are relative to mean values for the period 1961-1990, and may therefore be either positive or negative.
Source: Hadley Centre (1999).
achieve this reduction are artificially separated from the resulting health benefits. For such a distal risk factor, this separation of intervention from exposure and resulting health consequences may potentially introduce inconsistencies: for example, the economic changes necessary to achieve GHG stabilization are more consistent with some projections of levels and distribution of population and GDP than others. These socioeconomic factors are themselves likely to effect disease rates, potentially in interaction with climate. Integrated assessment of all effects of interventions would be conceptually more consistent, but this has not been attempted here, since it would introduce an additional layer of uncertainty and assumptions into the models and has previously only been explored for a few health outcomes (Tol and Dowlatabadi 2001).

The choice of baseline or "theoretical minimum" exposure follows the WMO and IPCC practice of using the observed global climate normal (i.e. averages) for 1961-1990 (New et al. 1999) as a reference point. Alternative baselines, such as pre-industrial climate, are not used,
because of the absence of a published consensus among climatologists on definitions of an appropriate time period and the relative roles of anthropogenic and natural influences before 1961-1990. This choice of a 1961-1990 baseline will therefore generate relatively conservative estimates of change in exposure and associated disease outcomes, as it does not address any human-induced climate change that occurred before this period. Indeed, as explained in section 2.6 below, the further choice of 1990 as the actual baseline year for linear-regression based estimates at current and future years heightens the conservative nature of these estimates.

The approach here treats climate change as a slowly evolving and continuous exposure, with the majority of disease models linked to those changes for which climate models make the most consistent predictions: gradual changes in temperature and, to a lesser extent, precipitation. This is again a limited approach. As shown in Table 20.1, it is very likely that climate change will also increase the frequency of extreme conditions, with likely effects on health. However, quantitative estimates of increased frequency have only very recently become available for some measures of extremes (e.g. wet winters; Palmer and Ralsanen 2002), and are not yet available for different GHG scenarios. The consequences of such changes are modelled here only in the context of inland flooding, but they could potentially be applied to other health end-points in the future.

Finally, there is some concern that disruption of the climate system may pass critical thresholds, resulting in abrupt rather than gradual changes (Broecker 1997; National Research Council 2001a) and associated rapid impacts on health. There is no consensus on the probability of such events, and they have therefore not been included in any published health outcome assessment studies. However, they should be borne in mind as a "worst-case" scenario.

### 2.2 Methods for estimating risk factor levels

Projections of the extent and geographical distribution of climate change were generated by applying the various emissions scenarios described above to the HadCM2 global climate model (GCM) of the Hadley Centre in the United Kingdom of Great Britain and Northern Ireland (Tett et al. 1997). This is one of several alternative GCMs used by the IPCC; it generates projections of changes in temperature and other climate properties which have been verified by back-casting (Johns et al. 2001), and which lie approximately in the middle of the range generated by alternative models.

The HadCM2 model generates estimates of the principal characteristics of climate, including temperature, precipitation and absolute humidity, for each cell of a global grid at resolution $3.75^{\circ}$ longitude by $2.5^{\circ}$ latitude. As for most climate change models, HadCM2 generates daily projections representing both long-term trends and the degree of natural
climate variability, but not necessarily its specific temporal pattern (i.e. the models do not accurately predict the climate of specific days or months). In order to account for such natural variability, the outputs that are most commonly used for modelling consequences of climate change are monthly means for average 30 -year periods centred on the 2020 s, $2050 \mathrm{~s}, 2080 \mathrm{~s}$, etc. The baseline climate (1961-1990) describes the same properties for the land surface of the world at $0.5^{\circ} \times 0.5^{\circ}$ resolution.

The climate model projections describe forecast changes in global climate conditions. Therefore, we did not attempt to estimate the "exposure prevalence": the entire world population was assumed to be exposed to one or other global climate scenario (i.e. exposure prevalence $=100 \%$ ). However, it should be noted that the climate scenarios incorporate geographical variation in both current climate (e.g. the lower temperatures in higher latitudes) and projected climate change (e.g. more rapid and intense warming is predicted in high northern latitudes than elsewhere). Different populations will therefore experience different climate conditions under any one climate-change scenario.

### 2.3 UnCERTAINTIES IN RISK FACTOR LEVELS

Two major sources of uncertainty surround the forecasting of future climate scenarios: (i) uncertainty in changes in factors such as population, economic growth, energy policies and practices on GHG emissions; and (ii) uncertainties over the accuracy of any climate model in predicting the effects of specified emissions scenarios on future climate in specific locations, against the background of substantial natural climate variability over time, and in space (i.e. downscaling). Climatologists have only very recently provided probabilistic measurements of uncertainty incorporating one or both of these sources (Knutti et al. 2002; Stott and Kettleborough 2002), and there is still debate over the reliability and utility of such measures (Schneider 2002). They have not previously been applied in impact studies (Katz 2002), and were therefore not used in this assessment. More importantly, however, there remains considerable uncertainty over the accuracy with which any single model can predict future climate. This is usually addressed by using outputs from a range of models, from independent groups. This was not possible here, as the particular GHG stabilization scenarios have only been applied, and fed through to estimates of likely consequences, for the HadCM2 model.

For this analysis, we did not address the first and last sources of uncertainty explicitly, and assumed that it is incorporated in the various alternative exposure scenarios, reflecting different trajectories of GHG emissions. The second source of uncertainty was partially addressed by using 30-year averages of climate conditions, which helps to "smooth out" the effects of natural climate variability. Further, the single model used was run with slight variations in initial conditions, allowing the calculation of an "ensemble mean". Four runs were used to generate an ensemble mean for the unmitigated emissions scenario. However, only
single climate runs were available for the stabilization scenarios. Therefore, the climate scenarios associated with those emissions scenarios are more uncertain.

Although it was not feasible to generate formal uncertainty ranges and feed these through the disease models, the approximate degree of uncertainty is illustrated in Appendix A. Here, the stabilization scenarios were run on a suite of 14 simple climate models, making different plausible assumptions about climate sensitivity to GHG emissions using the simplified COSMIC climate model described by Schlesinger and Williams (1997). This illustrates the degree of variation between the projections for future temperature and precipitation patterns. (Note that projections for precipitation vary more between models than does temperaturetherefore models that rely on precipitation estimates from single scenarios have an additional component of uncertainty.)

## 3. Risk factor-disease relationship

### 3.1 Outcomes included

The health outcomes addressed here were selected on the basis of observed sensitivity to temporal and geographical climate variation, importance in terms of mortality and/or burden of disease (Longstreth 1999; McMichael et al. 1996; Patz et al. 2000) and availability of quantitative global models (or feasibility of constructing informative models in the time available) (Table 20.3). More detail on evidence for causality and quantitative estimation methods for each health outcome is given below.

Climate change is by and large a relatively distal risk factor for illhealth, often acting through complex causal pathways which result in heterogeneous effects across populations. There is, therefore, a series of additional likely outcomes that have not yet been formally modelled. They include the potential health consequences of climate change on:

Table 20.3 Health outcomes considered in this analysis

| Outcome class | Incidence/ <br> prevalence | Outcome |
| :--- | :--- | :--- |
| Direct effects of heat and cold | Incidence | Cardiovascular disease deaths |
| Foodborne and waterborne diseases | Incidence | Diarrhoea episodes |
| Vector-borne diseases | Incidence | Malaria cases |
| Natural disasters ${ }^{\text {a }}$ | Incidence <br> Incidence | Deaths due to unintentional injuries <br> Other unintentional injuries (non-fatal) <br> Risk of malnutrition |
| Prevalence | Non-availability of recommended daily <br> calorie intake |  |

[^70]- changes in pollution and aeroallergen levels;
- the rate of recovery of the ozone hole, affecting exposure to ultraviolet radiation (Shindell et al. 1998);
- changes in the distribution and transmission of other infectious diseases, particularly other vector-borne diseases, geohelminths and rodent-borne diseases, and possible emergence of new pathogens;
- the distribution and abundance of plant and livestock pests and diseases, affecting agricultural production (Baker et al. 2000; Rosenzweig et al. 2001);
- the probability of crop failure through prolonged dry weather and famine, depending on location and crisis management;
- population displacement due to natural disasters, crop failure, water shortages; and
- destruction of health infrastructure in natural disasters.


### 3.2 Methods for estimating risk factor-disease RELATIONSHIPS

Various methods have been developed for the quantitative estimation of health outcomes of climatic change (reviewed by Martens and McMichael 2002; McMichael and Kovats 2000). It is not yet feasible to base future projections on observed long-term climate trends, for three reasons: (i) the lack of standardized long-term monitoring of climatesensitive diseases in many regions; (ii) methodological difficulties in measuring and controlling for non-climatic influences on long-term health trends; and (iii) the small (but significant) climate changes that have occurred so far are an inadequate proxy for the larger changes that are forecast for coming decades (Campbell-Lendrum et al. 2002).

Instead, estimates are based on observations of the effects of shorterterm climate variation in the recent past (e.g. the effects of daily or interannual climate variability on specific health outcomes) or the present (e.g. climate as a determinant of current disease distribution), or on specific processes that may influence health states (e.g. parasite and vector population dynamics in the laboratory, determining the transmission of infectious diseases). These quantitative relationships were then applied to future climate scenarios (Figure 20.3). Such an approach makes the important assumption that such associations will be maintained in the future, despite changes in mediating factors such as socioeconomic variables, infrastructure and technology. This introduces significant uncertainty, and possibly bias, in the estimates.

The extent and type of modelling applied to different health effects vary considerably. Consequently, several outcomes can only be estimated by crude adaptation of the outputs of available models. For example, some of the predictive models generate health-relevant outputs that do
not correspond directly to categories of disease states used in the Global Burden of Disease (GBD) study. These include the incidence of deaths and injuries due to, specifically, floods (rather than injuries due to all causes), or populations at risk of hunger or malaria infection (rather than prevalence of malnutrition, or incidence of clinical malaria). Currently, there are spatial resolution differences between models, which is not ideal. These relate to how the models account for geographical variation in: (i) baseline climate (potentially differentiated to $0.5^{\circ}$ globally, or higher resolutions for some regions); (ii) climate change (usually at the level of the GCM projections: $3.75^{\circ}$ longitude by $2.5^{\circ}$ latitude); and (iii) aggregation of the final results (occasionally according to regions other than subregions, depending on the purpose of the original model). Levels of spatial resolution for each disease model are described in section 3.6 and onwards. All of the above considerations are represented in the descriptions of strength of evidence and quantitative estimates of uncertainty for specific health outcomes.

## Assumptions

Simplifying assumptions have been made to facilitate clear definition of scenarios and associated consequences.

## Different mechanisms for reducing GHG emissions

The alternative GHG emissions scenarios outlined above could be achieved through an almost infinite variety of changes to economic and social development and energy use policies. As outlined in section 2.1, we did not attempt to estimate the secondary effects of GHG mitigation policies on health. These effects are potentially large. They include relatively direct mechanisms which may be negative, such as the potential negative effects of GHG emissions policies on economic development, personal wealth and vulnerability to disease ( Tol and Dowlatabadi 2001), or positive, such as reduction in the levels of ozone and other outdoor (Kunzli et al. 2000) and indoor air pollutants (Wang and Smith 1999). They may also act through more complex routes, for example, avoiding production of aeroallergens in $\mathrm{CO}_{2}$-enriched environments (Wayne et al. 2002; Ziska and Caulfield 2000).

## Population growth

The models described below estimated the relative per capita incidence of specific health outcomes under the different climate scenarios. The size and distribution of current and future populations therefore affect the relative risk estimates either (i) where the climate hazard is not evenly distributed geographically throughout the region, or (ii) where population is an integral part of the model-for example, in the risk-of-hunger model, where population size has an influence on food availability per capita. In adjusting the relative risks for these population effects, the distribution of future populations was estimated by applying the World

Bank mid-range estimate of population growth either at the national level (for malnutrition), or to a $1^{\circ} \times 1^{\circ}$ resolution grid map of population distribution (Bos et al. 1994) for all other outcomes.

## Modifying factors: adaptation and vulnerability

Factors such as physiological adaptation, technological and institutional innovation and individual and community wealth will influence not only the exposure of individuals and populations to climate hazards, but also the associated hazards (e.g. IPCC 2001a; Reiter 2000; Woodward et al. 1998). For some assessments, simpler modifying factors are integrated into the models for both present and future effects. For example, estimates of changes in the number of people at risk of hunger incorporate continental projections of economic growth, affecting capacity to buy food (Parry et al. 1999). Other models incorporate the effects of existing modifiers when defining current climate-disease relationships, such as estimates of the global distribution of malaria based on current climate associations (Rogers and Randolph 2000). Such models implicitly capture the current modifying effects of socioeconomic and other influences on climate effects, but do not attempt to model future changes in these modifiers. Finally, some models make no estimate of such modifying influences in either the present or future; for example, models that estimate future changes in the geographical range which is climatically suitable for malaria transmission, and associated populations at risk (Martens et al. 1999). To generate consistent estimates in this analysis:

- we attempted to account for current geographical variation in vulnerability to climate, where not already incorporated into the predictive models.
- we attempted to account for future changes in disease rates due to other factors (e.g. decreasing rates of infectious disease due to technological advances/improving socioeconomic status), and for changes in population size and age structure (e.g. potentially greater proportion of older people at higher risk of mortality related to cardiovascular disease in response to thermal extremes). This was addressed by calculating only relative risks under alternative climate change scenarios, which should be applied to GBD projections of disease rates and population size and age structure. The GBD projections take into account the effects of changing GDP, "human capital" (as measured by average years of female education), and time (to account for trends such as technological development) (Murray and Lopez 1996) on the overall "envelope" of cause-specific mortality and morbidity for diseases affected by climate change.
- all quantitative estimates of the health effects of climate change were based on observed effects of climate variations either over short time
periods, or between locations. They therefore made the important assumption that these relationships are also relevant to long-term climate change. To avoid unrealistic extrapolation of short-term relationships, we included consideration of mechanisms by which climate-health relationships may alter over time (i.e. adaptation). We considered whether each disease in turn was likely to be significantly affected by biological adaptation (i.e. either behavioural, immunological or physiological) and/or by generally improving socioeconomic conditions (i.e. increasing GDP) (see Table 20.4). In each case we defined appropriate adjustments to the relative risk estimates, in line with published studies. The different factors for each health outcome were then applied to the same projections of future GDP (WHO/ EIP/GPE, unpublished data, 2001) and changing climate (from our models), to adjust the relative risks over the time course of the assessment. There is, however, substantial uncertainty over the most likely degree of adaptation under different conditions. This was reflected in

Table 20.4 Assumptions on adaptation and vulnerability

|  | Biological ${ }^{\text {a }}$ adaptation <br> affecting RRs | Socioeconomic adaptation <br> affecting RRs |
| :--- | :--- | :--- |
| Direct physiological effects of <br> heat and cold | Yes. Temperature associated <br> with lowest mortality was <br> assumed to change directly <br> with temperature increases <br> driven by climate change | None |

the uncertainty estimates for the relative risks for each disease. Quoted uncertainty estimates therefore describe uncertainties around climate change predictions, about current exposure-response relationships, and around the degree to which these are likely to be maintained in the future. Since, to date, these have not been formally modelled, they were generated here by qualitative assessment in collaboration with the original modelling group. The uncertainty estimates should therefore be interpreted with caution.

- we ignored more complex aspects of future vulnerability. Whereas projected trends for average income (which is included in estimating baseline rates and, where possible, relative risks) are broadly positive, other factors may have opposite effects. These include income distribution, maintenance of disease surveillance, control and eradication programmes, technological change and secondary or threshold effects, such as the protective effect of forests in reducing the frequency and intensity of flooding (e.g. Fitzpatrick and Knox 2000).
- we made no attempt to estimate the effects of actions taken specifically to adapt to the effects of climate change (e.g. the upgrading of flood defences specifically to cope with sea-level rise attributable to climate change). Therefore, our estimates represent a "business-asusual" scenario of the health effects associated with global climate change.


### 3.3 Estimation for different time points

As stated above, climate model outputs are usually presented as averages over 30 -year periods, for example, centred on 2025 and 2055. In order to generate estimates for any specific year, we defined 1990 (i.e. the last year of the baseline climate period) as year 0 . Central, lower and upper estimates of relative risks were calculated for the 2020s (i.e. centred on 2025) and the 2050s, using models described elsewhere. Quoted estimates for the years 2001, 2005, 2010 and 2020 were estimated by linear regression against time between 1990 and 2025. Estimates for 2030 were estimated by linear regression between 2025 and 2055. Using this method, our choice of 1990 as the baseline year rather than the middle of the 1961-1990 period led to conservative estimates of health consequences, particularly in the near future.

### 3.4 Risk reversibility

In the context of this assessment, risk reversibility describes the proportion of the health consequences that would be avoided if the population were shifted to a different exposure scenario. Given the definition of exposure scenarios, complete avoidance of climate change (i.e. populations exposed to baseline climate conditions rather than unmitigated climate change) would avoid all of the health consequences that we have described. Risk reversibility would therefore be $100 \%$ in this case.

### 3.5 CRITERIA FOR IDENTIFYING RELEVANT STUDIES AND FOR estimating strength of evidence

In identifying relevant studies, we have considered all publications reviewed in the IPCC Third Assessment Report (IPCC 2001a), as well as those found in more recent literature searches (Appendix B). However, as the field is relatively new and expanding rapidly, we have also used some studies that are either in-press or submitted for publication (material available on request). As these may not be easily accessible to readers of this report, the major underlying assumptions are described.

There are still relatively few models that link climate change models to quantitative global estimates of health or health-relevant outcomes (e.g. numbers of people flooded or at risk of hunger). Where global models did not exist, we have extrapolated from models that make local or regional projections. Methods for extrapolating to the subregions are described separately for each disease. Where there was more than one published model for a particular outcome, decisions on model selection were based as much as possible on validation against historical/ geographical patterns. We excluded models: (i) that have been shown to be significantly less accurate than equivalent models in predicting historical or geographical trends; or (ii) that have been superseded by later models by the same group; or (iii) that are based on unrealistic biological assumptions; or (iv) that cannot be plausibly extrapolated to a wider area. Given the very limited means for model validation, these results involved choices made in this work among models and resulted in very large uncertainty.

As climate change impact assessment is, at this stage, predominantly a model-based exercise, the assessment of strength of evidence for causality is necessarily indirect. It is based on two considerations: (i) the strength of evidence for the current role of climate variability in affecting the health outcome, and (ii) the likelihood that this relationship between climate and health will be maintained through the process of long-term climate change.

Several of the uncertainties involved in this exercise, particularly those relating to climate modelling and those around the quantitative relationships between climate and health, should decrease through improvements in modelling, and more importantly, through better empirical research. However, estimation of climate change effects is currently a predictive exercise, based on some form of indirect modelling (e.g. analysis of temporal or geographical variation in climate), rather than direct experience (i.e. of a change process that has previously occurred). It is therefore possible that some of the health outcomes listed above may not respond in the predicted manner. On the other hand, we do not yet have direct experience of the full range of health outcomes that may be associated with exposure.

The analyses described here were specifically to estimate the future disease burden, which may be avoided by climate change mitigation policies. However, it must be emphasized that these burdens may also be reduced by adaptation interventions to reduce the vulnerability of populations. Given the "exposure commitment" (unavoidable climate change due to past GHG emissions), and the large gap between even plausible mitigation scenarios and the baseline climate, adaptation strategies are essential to the goal of reducing adverse health effects. These include early warning systems and defences for protection from natural disasters, and improved health infrastructure (e.g. water and sanitation, control programmes for vector-borne diseases) to reduce the baseline incidence of climate-sensitive diseases.

### 3.6 Direct effects of heat and cold on mortality

An association between weather and daily mortality has been shown consistently in many studies in diverse populations. The effect of extreme temperature events (heat-waves and cold spells) on mortality has also been well described in developed countries. However, temperatureattributable mortality has also been found at "moderate" temperatures (Curriero et al. 2000; Kunst et al. 1993).

Causality is supported by physiological studies of the effects of very high or very low temperatures in healthy volunteers. High temperatures cause some well described clinical syndromes such as heat-stroke (see review by Kilbourne 1989). Very few deaths are reported as directly attributed to "heat" (International Classification of Diseases, ninth revision [ICD-9 code 992.0]) in most countries. Exposure to high temperatures increases blood viscosity and it is therefore plausible that heat stress may trigger a vascular event such as heart attack or stroke (Keatinge et al. 1986a). Studies have also shown that elderly people have impaired temperature regulation (Drinkwater and Horvath 1979; Kilbourne 1992; Mackenbach et al. 1997; Vassallo et al. 1995). Clinical and laboratory studies indicate that exposure to low temperatures causes changes in haemostasis, blood viscosity, lipids, vasoconstriction and the sympathetic nervous system (e.g. Keatinge et al. 1986b, 1989; Khaw 1995; Woodhouse et al. 1993, 1994). The strongest physiological evidence is therefore for cardiovascular disease.

Population-based studies also provide evidence that environmental temperature affects mortality due to both cardiovascular and respiratory disease. The best epidemiological evidence is provided by time-series studies of daily mortality. These methods are considered sufficiently rigorous to assess short-term associations (days, weeks) between environmental exposures and mortality, if adjustment is made for longer-term patterns in the data series, particularly the seasonal cycle and any longterm trends, such as gradually decreasing mortality rates over decades (Schwartz et al. 1996).

The effect of a "hot" day is apparent for only a few days in the mortality series; in contrast, a "cold" day has an effect that lasts up to two weeks. Further, in many temperate countries, mortality rates in winter are $10-25 \%$ higher than death rates in summer. The causes of this winter excess are not well understood (Curwen and Devis 1988). It is therefore plausible that different mechanisms are involved and that cold-related mortality in temperate countries is related in some part to the occurrence of seasonal respiratory infections.

Although the physiological evidence for causality of an effect of temperature on mortality is greatest for cardiovascular, followed by respiratory, diseases, temperature has been shown to affect all-cause mortality in areas where cardiovascular disease rates are relatively low and infectious disease mortality is relatively high. Many studies report the seasonal patterns of infectious disease in developing countries, but the role of temperature is not well described. It is likely that seasonal rains influence the seasonal transmission of many infectious diseases. High temperatures encourage the growth of pathogens and are associated with an increased risk of diarrhoeal disease in poorer populations (see section 3.7).

Studies have also described mortality and morbidity during extreme temperature events (heat-waves). However, these events are, by definition, rare. It is therefore difficult to compare heat-waves in different populations and for different intensities. The studies that have been used to describe the effects of heat-waves (episode analyses) also use different methods for estimating the "expected" mortality, which makes comparison difficult (Whitman et al. 1997). An assessment of the health consequences of climate change on thermal stress requires estimation of past and future probabilities of extreme temperature events. The current methods of assessment use 30 -year averages of monthly data, and few scenarios consider change in the frequency or magnitude of extreme events. This is because: (i) suitable methods have not been developed, and (ii) climate model output at the appropriate spatial and temporal resolution is not readily available (Goodess et al. 2001).

There is little published evidence of an association between weather conditions and measures of morbidity such as hospital admissions or primary care consultations (Barer et al. 1984; Ebi et al. 2001; Fleming et al. 1991; McGregor et al. 1999; Rothwell et al. 1996; Schwartz et al. 2001). A study of general practitioner consultations among the elderly in Greater London found that temperature affected the rate of consultation for respiratory diseases but not that for cardiovascular diseases (Hajat et al. 2001). However, it is not clear how these end-points relate to quantitative measures of health burden.

## Estimating The TEMPERATURE-MORTALITY RELATIONSHIP

A review of the literature was undertaken to identify studies that report relationships between daily temperature and mortality (see Appendix B).

The following criteria were used to select studies for deriving the modelled estimates.

- A study that uses daily time-series methods to analyse the relationship between daily mean temperature and mortality.
- A study that reports a coefficient from linear regression which estimates the percentage changes in mortality per degree centigrade changes in temperature above and below a reported threshold temperature.

Several studies have estimated future temperature-related mortality for a range of climate scenarios (e.g. Guest et al. 1999; Kalkstein and Greene 1997; Langford and Bentham 1995; Martens 1998a). These methods were not considered appropriate for this project, as described in Appendix B.

The best characterized temperature-mortality relationships are those for total mortality in temperate countries. Fewer studies have also looked at the particular causes of death for which physiological evidence is strongest: cardiovascular disease, and to a lesser extent respiratory disease. In this study, we used the specific relationships for cardiovascular disease where these were available (temperate and cold-climate zones), and the general relationships for all-cause mortality for climatic regions where such disease-specific relationships could not be found in the literature review (all tropical populations) (Table 20.5).

As outlined in section 2.2, it was assumed that everybody is exposed to the ambient temperatures prevailing under the different climate scenarios. However, populations differ in their responses to temperature variability, which is partly explained by location or climate.

The global population distribution was divided into five climate zones (Table 20.6), according to definitions developed for urban areas by the

Table 20.5 Temperature-related mortality: summary of exposureresponse relationships, derived from the literature ${ }^{\text {a }}$

|  |  | Medical all-cause <br> mortality $^{\mathrm{b}}$ |  |  | Cardiovascular <br> mortality |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Heat | Cold |  | Heat | Cold |  |
| Climate zone | Threshold ( $\left.\mathrm{T}_{\text {cutoff }}\right)$ | 3.0 | 1.4 | - | - |  |
| Warm humid | 23 | 5.5 | 5.7 | - | - |  |
| Temperate | 29 | NA | NA | 2.6 | 2.9 |  |
| Cold | 16 | NA | NA | 1.1 | 0.5 |  |

- No data.

NA Not applicable.
a Change in mortality per $1^{\circ} \mathrm{C}$ change in mean daily temperature (\%).
b Excludes external causes (deaths by injury and poisoning).

Table 20.6 Climate zones

| Zone | Climate definition | \% of world population in zone (I990s) | City from which representative daily temperature distribution was derived | Mean temperature $\left({ }^{\circ} \mathrm{C}\right)$ (5th-95th percentile) |
| :---: | :---: | :---: | :---: | :---: |
| Hot/dry | Temperature of warmest month $>30^{\circ} \mathrm{C}$ | 17 | Delhi | $\begin{gathered} 25.0 \\ (13.5-35.2) \end{gathered}$ |
| Warm/humid | Temperature of the coldest month $>18^{\circ} \mathrm{C}$, warmest month $<30^{\circ} \mathrm{C}$ | 21 | Chiang Mai | $\begin{gathered} 26.3 \\ (21.6-29.5) \end{gathered}$ |
| Temperate | Average temperature of the coldest month $<18^{\circ} \mathrm{C}$ and $>-3^{\circ} \mathrm{C}$, and average temperature of warmest month $>10^{\circ} \mathrm{C}$ | 44 | Amsterdam | $\begin{gathered} 9.6 \\ (2.0-17.8) \end{gathered}$ |
| Cold | Average temperature of warmest month $>10^{\circ} \mathrm{C}$ and that of coldest month $<-3^{\circ} \mathrm{C}$ | 14 | Oslo | $\begin{gathered} 5 \\ (-6.3-16.5) \end{gathered}$ |
| Polar | Average temperature of the warmest month $<10^{\circ} \mathrm{C}$ | 0.2 | NA | NA |

NA Not applicable.

Australian Bureau of Meteorology (BOM 2001). The population in the polar zone is small ( $0.2 \%$ of world population) and was excluded.

It was necessary to estimate daily temperature distributions in order to calculate the number of attributable deaths. Daily temperature distributions clearly vary a great deal, even between localities within the same country. However, it was not feasible within this assessment to obtain sufficient meteorological data to estimate daily temperature distributions throughout the world at a fine spatial resolution. Therefore, a single distribution was chosen to represent each climate zone. New daily temperature distributions were then estimated for each climate scenario, by shifting the currently observed temperature distributions by the projected change in mean temperatures for each month, and of the variability of daily temperatures as well as changes in the mean.

## Estimating TEMPERATURE-ATTRIBUTABLE MORTALITY

An exposure-response relationship and threshold temperature ( $T_{\text {cutoff }}$ ) was applied within each climate zone (Table 20.7). The average temperature difference above (hot days) and below (cold days) this temperature was calculated for baseline climate and each of the climate scenarios.

The short-term relationships between daily temperature and mortality (Table 20.5) were used to estimate the annual attributable fraction of deaths due to hot days and cold days for each of the climate

Table 20.7 Threshold $T_{\text {cutoff }}$ for each scenario in each climate zone (original $T_{\text {cutoff }}+\mathrm{Dt}_{\text {summer }}$ rounded to integers)

|  |  | BaU | BaU | S550 | S550 | S750 | S750 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Climate zone | Baseline | 2020 s | 2050 s | 2020 s | 2050 s | 2020 s | 2050s |
| Cold | 16 | 17 | 18 | 17 | 17 | 17 | 17 |
| Temperate | 16 | 17 | 19 | 17 | 17 | 17 | 18 |
| Warm/humid | 29 | 29 | 31 | 29 | 29 | 29 | 30 |
| Hot/dry | 23 | 25 | 26 | 24 | 25 | 24 | 25 |

BaU Business-as-usual scenario.
scenarios (i.e. annual temperature distributions based on averages over 30 years). Deaths attributable to climate change were calculated as the change in proportion of temperature-attributable deaths (i.e. heat-attributable deaths plus cold-attributable deaths) for each climate scenario compared to the baseline climate. The $1^{\circ} \times 1^{\circ}$ resolution grid map of population distribution (Bos et al. 1994) was then overlaid on the maps of climate zones in a geographical information system (GIS), to estimate the proportion of the population in each subregion who live in each climate zone. The proportional changes in temperatureattributable deaths were therefore calculated by taking the average of the changes in each climate zone represented in the subregion, weighted by the proportion of the subregion's population living within that climate zone.

## Adaptation

Acclimatization includes autonomous adaptation in the individual (physiological adaptation, changes in behaviour) and autonomous and planned population-level adaptations (public health interventions and changes in built environment). Acclimatization to warmer climate regimes is likely to occur in individuals and populations, given the rate of change in mean climate conditions currently projected by climate models. However, it is uncertain whether populations are able to adapt to non-linear increases in the frequency or intensity of daily temperature extremes (heat-waves). Even small increases in average temperature can result in large shifts in the frequency of extremes (IPCC 2001b; Katz and Brown 1992).

Few studies have attempted to incorporate acclimatization into future projections of temperature-related mortality (Kalkstein and Greene 1997), but all studies report that acclimatization would reduce potential increases in heat-related mortality. Our estimates incorporated an assumption regarding acclimatization of the populations to the changing climate that describes this reduced effect. We assumed that the threshold temperature ( $T_{\text {cutoff }}$ ) is increased as populations adapt to a new
climate regime, reflecting physiological and behavioural acclimatization that can take place over the time-scale of decades. Changes in $T_{\text {cutoff }}$ are region and scenario specific, as they reflect the rate of warming experienced. Therefore, they were assumed to be proportional to the projected change in average summer temperature ( $\mathrm{Dt}_{\text {summer }}$, equal to the mean of the three hottest months) from the climate scenario. The temperature-mortality relationships were assumed not to change over time; that is, populations biologically adapt to their new average temperatures, but remain equally vulnerable to departures from these conditions. We made no explicit adjustment for an effect of socioeconomic development and technological change on temperature-related mortality. The resulting relative risks are given in Table 20.8.

## Short-term mortality displacement

Evidence suggests that the increase in mortality caused by high temperatures is partially offset by decreased deaths in a subsequent "rebound" period (Braga et al. 2001). This indicates that some of the observed increase in heat-related mortality may be displacement of deaths among those with pre-existing illness, which would have occurred soon in any case. However, this effect has not been quantified for temperature exposures and was not included in the model. The estimates are therefore used to calculate only attributable deaths, but not DALYs, as the estimate of attributable years of life lost was highly uncertain.

In subregions with predominantly temperate and cold climates, reductions in cold-related mortality are likely to be greater than increases in heat-related mortality. Therefore, all climate scenarios show a net benefit on mortality in these subregions, consistent with the IPCC conclusions described above. The effect of the adaptation assumption is to reduce relative risks and therefore net mortality.

## UnCERTAINTY ESTIMATES

The principal uncertainty in these estimates, and for all other health effects of climate change, relates to the extrapolation of a short-term climate-health relationship to the long-term effects of climate change. The degree to which this is a reasonable extrapolation relates to the degree to which populations will adapt to changing temperatures, both in terms of reducing the additional mortality attributable to heat, and the possible benefits of avoiding cold deaths. This outcome is unusual, in that the predicted health effects of climate change are negative in some regions, but positive in others. We therefore use slightly different terminology to describe the range of uncertainty around the estimates. The mid-range estimate was given by applying the model described above (i.e. making an adjustment for biological adaptation). The "high-impact" estimate assumes that there was no physiological or behavioural adaptation, and therefore no change in the dose-response relationship over time. This maximizes both positive and negative effects. The low-impact
Table 20.8
Central, low and high estimates of the relative risk of cardiovascular disease (all ages) for alternative climate

| Subregion | Climate ${ }^{\text {b }}$ | 2000 |  |  | 2001 |  |  | 2005 |  |  | 2010 |  |  | 2020 |  |  | 2030 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High |
| AFR-D | 2 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.002 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.006 | 1.004 | 1.000 | 1.008 |
|  | 3 | 1.001 | 1.000 | 1.003 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.005 | 1.004 | 1.000 | 1.008 | 1.005 | 1.000 | 1.009 |
|  | 4 | 1.002 | 1.000 | 1.004 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.005 | 1.004 | 1.000 | 1.007 | 1.005 | 1.000 | 1.011 | 1.007 | 1.000 | 1.013 |
| AFR-E | 2 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.002 | 1.000 | 1.003 | 1.002 | 1.000 | 1.005 | 1.003 | 1.000 | 1.006 |
|  | 3 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.006 | 1.003 | 1.000 | 1.007 |
|  | 4 | 1.001 | 1.000 | 1.003 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.005 | 1.004 | 1.000 | 1.008 | 1.005 | 1.000 | 1.010 |
| AMR-A | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 |
|  | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 |
|  | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 |
| AMR-B | 2 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.002 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 |
|  | 3 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.005 |
|  | 4 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.006 | 1.004 | 1.000 | 1.007 |
| AMR-D | 2 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.005 |
|  | 3 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.005 | 1.003 | 1.000 | 1.007 |
|  | 4 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.002 | 1.000 | 1.005 | 1.004 | 1.000 | 1.007 | 1.005 | 1.000 | 1.009 |
| EMR-B | 2 | 1.000 | 1.000 | 1.001 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 |
|  | 3 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.002 | 1.000 | 1.004 | 1.002 | 1.000 | 1.004 |
|  | 4 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.005 | 1.003 | 1.000 | 1.007 |
| EMR-D | 2 | 1.000 | 1.000 | 1.001 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 |
|  | 3 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.001 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.002 | 1.000 | 1.004 | 1.002 | 1.000 | 1.005 |
|  | 4 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.005 | 1.003 | 1.000 | 1.007 |


| EUR-A | 3 | 1.000 1.000 | 1.000 1.000 | 1.000 1.000 | 1.000 1.000 | 1.000 1.000 | 1.000 1.000 | 1.000 1.000 | 1.000 1.000 | 1.000 0.999 | 1.000 1.000 | 1.000 1.000 | 0.999 0.999 | 1.000 0.999 | 1.000 1.000 | 0.999 0.999 | 0.999 0.999 | 1.000 1.000 | 0.999 0.999 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 |
| EUR-B | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 |
|  | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 |
|  | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 1.000 | 0.998 |
| EUR-C | 2 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 1.000 | 0.998 | 0.999 | 1.000 | 0.998 |
|  | 3 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 1.000 | 0.998 | 0.999 | 1.000 | 0.998 |
|  | 4 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 1.000 | 0.998 | 0.999 | 1.000 | 0.997 | 0.998 | 1.000 | 0.997 |
| SEAR-B | 2 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.002 | 1.000 | 1.005 | 1.004 | 1.000 | 1.007 | 1.004 | 1.000 | 1.009 |
|  | 3 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.006 | 1.004 | 1.000 | 1.009 | 1.005 | 1.000 | 1.011 |
|  | 4 | 1.002 | 1.000 | 1.004 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.006 | 1.004 | 1.000 | 1.008 | 1.006 | 1.000 | 1.011 | 1.007 | 1.000 | 1.014 |
| SEAR-D | 2 | 1.001 | 1.000 | 1.002 | 1.001 | 1.000 | 1.002 | 1.002 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.006 | 1.004 | 1.000 | 1.008 |
|  | 3 | 1.001 | 1.000 | 1.003 | 1.001 | 1.000 | 1.003 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.005 | 1.004 | 1.000 | 1.008 | 1.005 | 1.000 | 1.009 |
|  | 4 | 1.002 | 1.000 | 1.004 | 1.002 | 1.000 | 1.004 | 1.003 | 1.000 | 1.005 | 1.004 | 1.000 | 1.007 | 1.005 | 1.000 | 1.011 | 1.007 | 1.000 | 1.013 |
| WPR-A | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 |
|  | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 |
|  | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 |
| WPR-B | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
|  | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
|  | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

[^71]estimate assumes complete adaptation to a changing climate, and therefore no change in the relative risk as the climate changes.

There is therefore a need for further time-series studies, applying a standard approach to populations living in as wide a range of climates as possible. Previous analyses have focused mainly on temperate zones, where the winter effects are greatest, potentially causing over-estimation of the reduced burden attributable to climate change (e.g. Martens 1998a). More analyses of temperature-mortality relationships are therefore required in tropical developing countries.

Such studies should attempt to estimate formally the degree to which adaptation may decrease mortality, and to which observed associations between climate and mortality reflect displaced rather than additional deaths. Finally, there is a need for greater investigation of the health burden of morbidity associated with temperature extremes, including, for example, inability to work in extreme temperatures.

### 3.7 Diarrhoeal disease

Diarrhoeal diseases are highly sensitive to climate, showing seasonal variations in numerous sites (Drasar et al. 1978). This observation is supported by regression analyses of the effects of seasonal and longer-term variation in a limited number of sites (Checkley et al. 2000; Singh et al. 2001). The results of a literature search on the associations between diarrhoeal disease and climate are shown in Appendix B, Table B.2. The climate-sensitivity of diarrhoeal disease is consistent with observations of the direct effects of climate variables on the causative agents. Temperature and relative humidity have a direct influence on the rate of replication of bacterial and protozoan pathogens, and on the survival of enteroviruses in the environment (Blaser et al. 1995). Rainfall may affect the frequency and level of contamination of drinking water (Curriero et al. 2001).

Quantitative relationships can be defined between climate variations and incidence, which can in turn be directly linked to the outputs of global climate change models. There are, however, challenges and uncertainties in estimating the magnitude of effects.

- The sites from which these relationships were defined cover only a small part of the spectrum of global climate variation. Different relationships may apply at higher or lower temperatures.
- The relative importance of different pathogens and modes of transmission (e.g. via water, food, insects or human-human contact) varies between locations, and is heavily influenced by level of sanitation (Black and Lanata 1995). As the pathogens are known to vary in their response to climate (e.g. Cook et al. 1990; Chaudhury et al. 1996), this will cause uncertainty in extrapolating temperature relationships from local studies to other regions with different levels of development.
- Pathogens vary in the severity of clinical symptoms, and the likelihood that they will be reported to health services (e.g. Wheeler et al. 1999). Therefore, climate-disease relationships derived only from passive reporting may differ from those based on other methods of surveillance.
- While several studies describe climate effects on particular diarrhoea pathogens (e.g. Eberhard et al. 1999; Konno et al. 1983; Purohit et al. 1998), these cannot be used directly to estimate effects on diarrhoeal disease without information on: (i) their relative contribution to overall disease incidence, and (ii) equivalent data on climate sensitivity and relative prevalence for all other diarrhoea pathogens.
- Despite convincing evidence on the effect of extreme rainfall on waterborne outbreaks of diarrhoea, even in highly developed countries (Curriero et al. 2001), this cannot easily be generalized to the total burden of diarrhoeal disease without information on the relative contribution of such outbreaks to overall diarrhoea incidence.
- Rainfall effects on overall diarrhoea (where observed) are non-linear, and cannot easily be extrapolated to other regions.

In order to minimize these uncertainties, we restricted our estimates to the effect of increasing temperatures on the incidence of all-cause diarrhoea reported to health services (i.e. without attempting to make separate estimates for different pathogens, transmission routes, severities, or for the more complex associations with rainfall). There are important residual uncertainties related to the extrapolation of temperature relationships from specific study sites to others with different temperature regimes and levels of development, and climate effects on reported diarrhoea compared to the true burden of disease. Projections of increasing frequency of extreme wet seasons are very large (e.g. two to five times increase in the regions analysed; Palmer and Ralsanen 2002), and extreme precipitation is associated with increased diarrhoea in both developed (Curriero et al. 2001) and developing (Singh et al. 2001) countries.

Graded maps of $3.75^{\circ}$ longitude by $2.5^{\circ}$ latitude resolution showing the change in temperature under the alternative scenarios were overlaid with $0.5^{\circ} \times 0.5^{\circ}$ resolution maps of predicted population distributions for the 2020s and 2050s in a GIS. The GIS was used to calculate the average change in exposure (temperature) for each population grid-cell.

Although seasonality of diarrhoeal disease is well recognized, the quantitative relationship between climate and overall diarrhoea incidence has only been explicitly measured in two studies. Both studies described relationships with all-cause diarrhoea, that is, specific pathogens were not differentiated. Checkley et al. (2000) used time-series analysis to correlate measurements of temperature and relative humidity against daily hospital admissions at a single paediatric diarrhoeal

Figure 20.4 Hospitalizations for diarrhoea (upper line) correspond closely with temperature (lower line) at a clinic in Lima, Peru


Note: Shaded region corresponds to the 1997-1998 El Niño event.
Source: Checkley et al. (2000).
disease clinic in Lima, Peru (Figure 20.4) for just under 6 years. A total of 57331 admissions were recorded during the study. The analysis showed a $4 \%(95 \%$ CI $2-5 \%)$ increase in admissions for each degree centigrade increase in temperature during the hotter months, and a $12 \%$ ( $95 \%$ CI $10-14 \%$ ) per degree centigrade increase in the cooler months, averaging at an $8 \%(95 \%$ CI $7-9 \%)$ per degree centigrade increase over the course of the study. During the 1997-1998 El Niño period there was an additional increase in admissions above that expected on the basis of pre-El Niño temperature relationships, but no association with relative humidity independent of temperature. No rainfall during that period.

In the Checkley et al. study, exposure (climate) data were recorded at local meteorological stations, and can be considered to have negligible measurement error at the population level. The analysis independently controlled for seasonal variations and long-term trends, imparting high confidence to the observed effect of temperature on the outcome recorded. The positive correlation with temperature is also biologically plausible, as a high proportion of diarrhoea cases in many tropical developing countries are caused by bacteria, entamoeba and protozoa (Black
and Lanata 1995), which are favoured by high temperatures. The principle limitations of the study by Checkley et al. are that the outcome recorded may not be representative of climate effects on: (i) less severe disease (i.e. not requiring hospitalization) or more severe disease (diarrhoeal deaths), or (ii) disease in adults rather than children.

Singh et al. (2001) used similar time-series analyses to correlate monthly reported incidence of diarrhoea throughout Fiji against variations in temperature and rainfall, after allowing for the effects of seasonal variation and long-term trend. The study covered the period between 1978 and 1998, with an average of some 1000 reported cases for each of the 228 study months. Reported incidence increased by approximately $3 \%$ ( $95 \%$ CI $1.2-5.0 \%$ ) for each degree centigrade increase in temperature, by $2 \%$ ( $95 \%$ CI $1.5-2.3 \%$ ) per unit increase in rainfall above average rainfall conditions $\left(5 \times 10^{-5} \mathrm{~kg} / \mathrm{m}^{2}\right.$ per min), and by $8 \%$ per unit decrease below average conditions. The pattern is supported by a positive geographical correlation between temperature and incidence in 18 Pacific Island countries (Singh et al. 2001). Climate measurements were from a $2.5^{\circ} \times 2.5^{\circ}$ cell of a global gridded data set corresponding to Fiji. The use of monthly averages of climate data from this large geographical area may not have reflected the full range of climate exposures of the population; this would have introduced random error and decreased sensitivity. Low rainfall may force use of contaminated water, while high rainfall may contaminate water through flooding. The major limitations of this study are lack of a clear clinical or laboratory definition for diarrhoea, and lack of information on the age distribution of cases.

We are not aware of any similar studies of climate effects on all-cause diarrhoea in developed regions, although studies have been carried out on some subsets of total incidence. Bentham (1997) showed that the incidence of food poisoning, usually caused by bacteria, increased by approximately 9\% per degree centigrade in England and Wales. Konno et al. (1983) demonstrated a non-linear inverse relationship between rotavirus infection and temperature in Japan. The relative importance of pathogens that thrive at lower temperatures appears to be greater in populations with higher standards of living, who have access to clean water and sanitation and for whom there is no clear and consistent evidence for peaks in all-cause diarrhoea in warmer months. This is in contrast to the situation of less well-off populations, where diarrhoea is usually more common in warmer, wetter months, as well demonstrated by clear summer peaks of diarrhoea in black, but not white, infants in Johannesburg during the 1970s (Robins-Browne 1984).

For this assessment we defined developing countries as those with per capita incomes lower than the richer of the two study countries (Fiji) in the year 2000—approximately US\$ 6000/year in 1990 US dollars. For such countries, we applied a dose-response relationship of $5 \%$ increase in diarrhoea incidence per degree centigrade increase in temperature, to
both sexes and all age groups. This is consistent with the relationships derived from the two studies described above. We chose $5 \%$ rather than the arithmetic mean of the constants from the two studies ( $5.5 \%$ ) for two reasons: (i) to avoid giving a false impression of precision based on only two estimates, each with their own confidence intervals, and (ii) in order to be conservative.

Although the confidence intervals around the estimates from the individual studies are relatively small, these clearly cover only a small range of climatic and socioeconomic environments. As described above, Checkley et al. (2000) showed that even in a single socioeconomic setting the temperature dependence of diarrhoeal disease may vary across the temperature range or, less plausibly, with non-climatic seasonal variations. This introduces uncertainty when extrapolating such a single relationship. There is also potential bias: if the temperature-responsiveness is indeed greater at low temperatures, extrapolation of an average value will tend to underestimate effects in areas that are on average colder, and overestimate in hotter regions. However, the average annual temperatures experienced by the residents of the areas to which we extrapolated this relationship was $20.3^{\circ} \mathrm{C}$, as calculated by averaging the temperatures in each $1^{\circ} \times 1^{\circ}$ grid-cell weighted by the population. This is in the mid-range of the temperatures experienced throughout the year in Lima $\left(16-26^{\circ} \mathrm{C}\right)$, and cooler than those in the capital of Fiji $\left(23-27^{\circ} \mathrm{C}\right)$ (World Climate 2002). Our extrapolation tended towards being conservative, but with significant uncertainty. Therefore, we place a wide uncertainty range ( $0-10 \%$ ) on this value.

As there are no such studies published for developed regions, we made the assumption that overall diarrhoea incidence in richer countries is insensitive to climate change, that is, $0 \%(-5$ to $+5 \%$ ) change per degree centigrade temperature change. Relative risks for each country under each scenario were calculated by multiplying the projected increase in temperature by the relevant exposure-response value. The quoted estimate for each subregion is the population-weighted average of the relative risks for each country in the subregion (Table 20.9). In addition to changes in baseline diarrhoeal disease over time, we assumed that the climate sensitivity of diarrhoea in developing countries will decrease as they become better off. For projections of relative risks for years after 2000, we used projections of future changes in GDP (WHO/EIP/GPE, unpublished data, 2000) to apply the relationship used above-that is, overall diarrhoea incidence does not respond to temperature in any country that attains a per capita GDP of at least US $\$ 6000 /$ year. Relative risks for each time point were then calculated as above.

The quantitative estimates in this analysis were highly sensitive to the exposure-response relationship, around which there is substantial uncertainty due to the very small number of analysed time-series. The estimates could be rapidly improved by analysis of climate exposure-response relationships from sites from a wider climatic and
Table 20.9

| Subregion | Climate ${ }^{\text {a }}$ | 2000 |  |  | 2001 |  |  | 2005 |  |  | 2010 |  |  | 2020 |  |  | 2030 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High |
| AFR-D | 2 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.06 | 1.04 | 1.00 | 1.08 | 1.05 | 0.99 | 1.10 |
|  | 3 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.05 | 1.04 | 1.00 | 1.07 | 1.05 | 1.00 | 1.11 | 1.06 | 0.99 | 1.13 |
|  | 4 | 1.02 | 1.00 | 1.05 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.07 | 1.05 | 1.00 | 1.09 | 1.07 | 1.00 | 1.14 | 1.08 | 0.99 | 1.16 |
| AFR-E | 2 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.06 | 1.04 | 0.99 | 1.08 | 1.05 | 0.99 | 1.11 |
|  | 3 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.07 | 1.05 | 0.99 | 1.10 | 1.06 | 0.99 | 1.13 |
|  | 4 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.05 | 1.03 | 1.00 | 1.07 | 1.04 | 0.99 | 1.09 | 1.06 | 0.99 | 1.13 | 1.08 | 0.99 | 1.16 |
| AMR-A | 2 | 1.00 | 0.99 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.97 | 1.03 | 1.00 | 0.96 | 1.05 | 1.00 | 0.95 | 1.06 |
|  | 3 | 1.00 | 0.99 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.97 | 1.03 | 1.00 | 0.96 | 1.05 | 1.00 | 0.94 | 1.06 |
|  | 4 | 1.00 | 0.98 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.97 | 1.03 | 1.00 | 0.96 | 1.04 | 1.00 | 0.94 | 1.06 | 1.00 | 0.93 | 1.08 |
| AMR-B | 2 | 1.00 | 0.99 | 1.02 | 1.00 | 0.99 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.97 | 1.03 | 1.00 | 0.96 | 1.04 | 1.00 | 0.95 | 1.05 |
|  | 3 | 1.00 | 0.99 | 1.02 | 1.00 | 0.99 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.97 | 1.03 | 1.00 | 0.96 | 1.05 | 1.00 | 0.94 | 1.06 |
|  | 4 | 1.00 | 0.98 | 1.02 | 1.00 | 0.98 | 1.03 | 1.00 | 0.97 | 1.03 | 1.00 | 0.96 | 1.04 | 1.00 | 0.94 | 1.06 | 1.00 | 0.92 | 1.08 |
| AMR-D | 2 | 1.01 | 1.00 | 1.03 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.05 | 1.02 | 0.98 | 1.06 | 1.02 | 0.96 | 1.07 |
|  | 3 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.06 | 1.02 | 0.97 | 1.07 | 1.02 | 0.96 | 1.08 |
|  | 4 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.06 | 1.04 | 1.00 | 1.08 | 1.03 | 0.97 | 1.09 | 1.02 | 0.95 | 1.10 |
| EMR-B | 2 | 1.01 | 1.00 | 1.03 | 1.01 | 1.00 | 1.03 | 1.02 | 0.99 | 1.04 | 1.02 | 0.99 | 1.06 | 1.00 | 0.95 | 1.05 | 1.00 | 0.94 | 1.06 |
|  | 3 | 1.01 | 1.00 | 1.03 | 1.01 | 1.00 | 1.03 | 1.02 | 0.99 | 1.04 | 1.02 | 0.99 | 1.05 | 1.00 | 0.95 | 1.05 | 1.00 | 0.94 | 1.06 |
|  | 4 | 1.02 | 0.99 | 1.04 | 1.02 | 0.99 | 1.05 | 1.03 | 0.99 | 1.06 | 1.04 | 0.99 | 1.08 | 1.00 | 0.93 | 1.08 | 1.00 | 0.91 | 1.09 |
| EMR-D | 2 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.06 | 1.05 | 1.00 | 1.10 | 1.06 | 1.00 | 1.12 |
|  | 3 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.05 | 1.03 | 1.00 | 1.07 | 1.05 | 1.00 | 1.10 | 1.06 | 1.00 | 1.13 |
|  | 4 | 1.03 | 1.00 | 1.05 | 1.03 | 1.00 | 1.06 | 1.04 | 1.00 | 1.08 | 1.05 | 1.00 | 1.10 | 1.08 | 1.00 | 1.15 | 1.09 | 1.00 | 1.19 |

Table 20.9
Central, low and high estimates of the relative risk of diarrhoea for alternative climate scenarios relative to baseline climate (continued)

| Subregion | Climate | 2000 |  |  | 2001 |  |  | 2005 |  |  | 2010 |  |  | 2020 |  |  | 2030 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High |
| EUR-A | 2 | 1.00 | 0.98 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.97 | 1.03 | 1.00 | 0.95 | 1.05 | 1.00 | 0.94 | 1.06 |
|  | 3 | 1.00 | 0.98 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.97 | 1.03 | 1.00 | 0.97 | 1.03 | 1.00 | 0.95 | 1.05 | 1.00 | 0.94 | 1.06 |
|  | 4 | 1.00 | 0.98 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.97 | 1.03 | 1.00 | 0.96 | 1.04 | 1.00 | 0.94 | 1.06 | 1.00 | 0.92 | 1.08 |
| EUR-B | 2 | 1.01 | 0.99 | 1.03 | 1.01 | 0.99 | 1.03 | 1.01 | 0.99 | 1.04 | 1.02 | 0.98 | 1.05 | 1.02 | 0.97 | 1.07 | 1.01 | 0.94 | 1.07 |
|  | 3 | 1.01 | 0.99 | 1.03 | 1.01 | 0.99 | 1.03 | 1.01 | 0.99 | 1.04 | 1.02 | 0.98 | 1.05 | 1.02 | 0.97 | 1.07 | 1.01 | 0.94 | 1.08 |
|  | 4 | 1.01 | 0.99 | 1.03 | 1.01 | 0.99 | 1.04 | 1.02 | 0.98 | 1.05 | 1.02 | 0.98 | 1.06 | 1.02 | 0.96 | 1.09 | 1.01 | 0.93 | 1.09 |
| EUR-C | 2 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.01 | 0.98 | 1.03 | 1.01 | 0.98 | 1.04 | 1.01 | 0.96 | 1.06 | 1.00 | 0.94 | 1.07 |
|  | 3 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.01 | 0.98 | 1.03 | 1.01 | 0.98 | 1.04 | 1.01 | 0.96 | 1.06 | 1.00 | 0.94 | 1.07 |
|  | 4 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.05 | 1.01 | 0.98 | 1.04 | 1.01 | 0.97 | 1.06 | 1.01 | 0.95 | 1.08 | 1.00 | 0.92 | 1.08 |
| SEAR-B | 2 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.03 | 1.00 | 0.98 | 1.03 | 1.00 | 0.97 | 1.04 | 1.00 | 0.95 | 1.05 |
|  | 3 | 1.01 | 1.00 | 1.03 | 1.01 | 1.00 | 1.03 | 1.02 | 0.99 | 1.04 | 1.00 | 0.97 | 1.04 | 1.00 | 0.95 | 1.05 | 1.00 | 0.94 | 1.06 |
|  | 4 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.04 | 1.02 | 0.99 | 1.05 | 1.00 | 0.96 | 1.04 | 1.00 | 0.94 | 1.07 | 1.00 | 0.92 | 1.08 |
| SEAR-D | 2 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.05 | 1.03 | 1.00 | 1.07 | 1.05 | 1.00 | 1.10 | 1.06 | 1.00 | 1.13 |
|  | 3 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.06 | 1.04 | 1.00 | 1.08 | 1.06 | 1.00 | 1.12 | 1.07 | 1.00 | 1.15 |
|  | 4 | 1.03 | 1.00 | 1.05 | 1.03 | 1.00 | 1.06 | 1.04 | 1.00 | 1.08 | 1.05 | 1.00 | 1.10 | 1.07 | 1.00 | 1.15 | 1.09 | 1.00 | 1.19 |
| WPR-A | 2 | 1.00 | 0.99 | 1.01 | 1.00 | 0.99 | 1.01 | 1.00 | 0.98 | 1.02 | 1.00 | 0.97 | 1.03 | 1.00 | 0.96 | 1.04 | 1.00 | 0.95 | 1.05 |
|  | 3 | 1.00 | 0.99 | 1.01 | 1.00 | 0.99 | 1.01 | 1.00 | 0.98 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.96 | 1.04 | 1.00 | 0.95 | 1.05 |
|  | 4 | 1.00 | 0.98 | 1.02 | 1.00 | 0.98 | 1.02 | 1.00 | 0.97 | 1.03 | 1.00 | 0.96 | 1.04 | 1.00 | 0.94 | 1.06 | 1.00 | 0.93 | 1.07 |
| WPR-B | 2 | 1.01 | 1.00 | 1.03 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.06 | 1.00 | 0.96 | 1.05 | 1.00 | 0.95 | 1.06 |
|  | 3 | 1.01 | 1.00 | 1.03 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.06 | 1.00 | 0.96 | 1.05 | 1.00 | 0.95 | 1.06 |
|  | 4 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.05 | 1.03 | 1.00 | 1.07 | 1.05 | 1.00 | 1.09 | 1.00 | 0.93 | 1.08 | 1.01 | 0.92 | 1.09 |

[^72]socioeconomic spectrum. Future studies should also explicitly measure the degree to which economic development and improved levels of sanitation influence vulnerability to the effects of climate variation on diarrhoeal disease.

### 3.8 Malnutrition

Multiple biological and social factors affect the incidence of malnutrition, but one of the fundamental determinants is the availability of staple foods. Climate change may affect this through the balance of the (broadly negative) effects of changes in temperature and precipitation, and (broadly positive) effects of higher $\mathrm{CO}_{2}$ levels on yields of major food crops (e.g. IPCC 1996; Rosenzweig and Parry 1994; see also Appendix B, Table B.4). These effects are likely to vary markedly with geography: productivity is projected to increase in higher-latitude producers such as Canada and the United States, but to decrease closer to the equator. The global food trade system may be able to absorb these effects at the global level. However, climate change can be expected to have significant effects on food poverty in some regions, owing to variation in both productivity and in economic capacity to cope (Parry et al. 1999).

Crop models have been validated at 124 sites in 18 countries over a wide range of environments (e.g. Otter-Nacke et al. 1986). However, estimation of changes in food availability (and by inference malnutrition) requires additional analyses of the geographical variation in effects on food production, and of food trade patterns. Only one modelling group (Parry et al. 1999) has integrated a basic physiological model of climate change effects on region-specific crop production with food trade models, in order to make projections of the numbers of people actually at risk of hunger. All results presented here were based on the model described in Parry et al. (1999) and other work by the same group.

As for other potential effects of climate change, there is considerable uncertainty over the degree to which current relationships, such as those between climate and crops, and food trade systems will remain constant over time. The most important uncertainties probably relate to the ability of the world food trade system to adapt to changes in production (Dyson 1999; Waterlow et al. 1998). Although these are the most complete models currently available, they do not describe the likely effect of climate change on more complex pathways, such as animal husbandry, or the relative importance of fruit and vegetable production. These in turn may affect micronutrient (e.g. vitamin A , iodine, iron and zinc) deficiency.

Global distribution of temperature, rainfall and $\mathrm{CO}_{2}$ were mapped for each of the alternative scenarios, as described above. Climate dose-response relationships have been defined for yields of major grain cereals and soybean, which account for $85 \%$ of world cereal exports.

Effects of temperature and precipitation, and the beneficial effects of higher $\mathrm{CO}_{2}$ levels, have been defined using the IBSNAT-ICASA dynamic crop growth models (IBSNAT 1989). The exposure distributions described above were applied to these crop growth models, and the derived yield functions were extrapolated to other crops and regions on the basis of agro-climatic similarity.

These crop yield estimates are used as inputs for the Basic Linked System world food trade model. This consists of a linked series of 40 national and regional food models, representing food production, the effects of market forces and government policies on prices and trade, and trends in agricultural, economic and technological conditions over time (see Fischer et al. 1988 for a full description).

The model is represented schematically in Figure 20.5. Principal characteristics of this model are:

- no major changes in the political or economic context of world food trade or in food production technology;
- population growth to occur following the World Bank mid-range estimate (World Bank 1994) i.e. 10.7 billion by the 2080s;
- GDP to accumulate as projected by EMF14 (Energy Modeling Forum 1995);
- a $50 \%$ trade liberalization in agriculture is introduced gradually by 2020.

The model results in an estimation of national food availability. This is used to generate an estimate of per capita food availability in each country, assuming that this food is distributed among the population following a skewed (beta) distribution. The final model output is the number or proportion of the population in each subregion who do not have access to sufficient food to maintain a basal metabolic rate of 1.4, which is the Food and Agriculture Organization of the United Nations' (FAO) definition of undernourishment (FAO 1987).

The model generates outputs for continents made up principally of vulnerable developing countries (thus excluding China and the former Soviet Union, countries in North America, and in western and eastern Europe). As these model continents do not map directly on to the subregions, we generated estimates for each subregion by calculating the proportion of the population that lives within each continent in the food availability model. Where more than $90 \%$ of the subregion population live within a single model continent, we quoted the model estimate for that continent. Otherwise an average was calculated, weighted by the distribution of the subregion population among the various continents. While the subregions mapped reasonably well on to the climate change/food availability model continents, the aggregation meant that some of the geographical variation in vulnerability was lost. Most notably, despite severe problems in some countries, EUR-C was assumed

Figure 20.5 Key elements of the crop yields and world food trade study


ET Evapotranspiration.
Source: Rosenzweig et al. (1993).
not to suffer from malnutrition as it lies within the "developed" European continent of the food availability model.

Preliminary analysis correlating estimates for the 1990s at the level of the model regions (data not shown) indicated that the model output was positively related to more direct measures of malnutrition, such as the incidence of underweight, and stunting and wasting in children aged <5 years, as measured by the WHO Global Database on Child Growth and Malnutrition (WHO 2002). The aggregation of the food availability model means that this correlation was based on only a small number of independent data points. We therefore did not attempt to make any quantitative estimate of these relationships. Instead, the relative risk of the incidence of energy shortfall (Table 20.10) was interpreted as being directly proportional to the relative risk of suffering from the risk factor "underweight". The relative risk estimates were therefore applied to all diseases affected by the underweight risk factor (see chapter 2). These include diarrhoea and malaria. We therefore assumed that these diseases are affected in two distinct ways by climate change-through meteorological effects on the pathogens and vectors, and through increased susceptibility of the human population due to undernutrition.

In common with most climate change impact assessments, the published studies do not quote uncertainties around the various relationships in the model, either separately or aggregated. Small variations in the initial conditions of a single climate model (the HadCM2 ensemble) generated only slight variations in projections of crop production and incidence of food shortfall. Using a different climate model (HadCM3), however, generated markedly different projections (Parry et al. 1999). These comparisons relate to only part of the possible error, as they do not address uncertainties in the trade and social components of the model.

From the published descriptions of the model, there is no reason to assume that the estimates generated were systematically biased either upward or downward. In the absence of formal sensitivity analyses of the complete model, however, uncertainty estimates are arbitrary. We presented the relative risks generated above as mid-range estimates, with the upper and lower range covering a complete adaptation to any changes in agricultural output (i.e. no change in risk), to a doubling of the estimate of the relative risk calculated above. However, this uncertainty range should be treated with caution. Priorities for the future are investigations of:

- variation in output when using a wider range of food production models as inputs to the food trade/availability models;
- sensitivity of estimates to the various climate scenarios;
- sensitivity analyses to estimate uncertainty around the exposureresponse relationships;
Table 20.10 Central, low and high estimates of the relative risk of malnutrition for alternative climate scenarios relative to

| Subregion | Climate ${ }^{\text {a }}$ | 2000 |  |  | 2001 |  |  | 2005 |  |  | 2010 |  |  | 2020 |  |  | 2030 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High |
| AFR-D | 2 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.06 | 1.03 | 1.00 | 1.06 |
|  | 3 | 1.01 | 1.00 | 1.03 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.05 | 1.04 | 1.00 | 1.08 | 1.04 | 1.00 | 1.09 |
|  | 4 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.04 |
| AFR-E | 2 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.06 | 1.03 | 1.00 | 1.06 |
|  | 3 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.05 | 1.04 | 1.00 | 1.07 | 1.04 | 1.00 | 1.08 |
|  | 4 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.05 |
| AMR-A | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| AMR-B | 2 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.06 | 1.05 | 1.00 | 1.09 | 1.05 | 1.00 | 1.10 |
|  | 3 | 1.03 | 1.00 | 1.07 | 1.04 | 1.00 | 1.07 | 1.05 | 1.00 | 1.10 | 1.07 | 1.00 | 1.13 | 1.10 | 1.00 | 1.20 | 1.11 | 1.00 | 1.22 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.00 |
| AMR-D | 2 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.06 | 1.05 | 1.00 | 1.09 | 1.05 | 1.00 | 1.10 |
|  | 3 | 1.03 | 1.00 | 1.07 | 1.04 | 1.00 | 1.07 | 1.05 | 1.00 | 1.10 | 1.07 | 1.00 | 1.13 | 1.10 | 1.00 | 1.20 | 1.11 | 1.00 | 1.22 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.00 |
| EMR-B | 2 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.06 | 1.03 | 1.00 | 1.06 |
|  | 3 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.06 | 1.04 | 1.00 | 1.08 | 1.06 | 1.00 | 1.12 | 1.06 | 1.00 | 1.13 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.00 |
| EMR-D | 2 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.07 | 1.04 | 1.00 | 1.09 | 1.07 | 1.00 | 1.13 | 1.07 | 1.00 | 1.15 |
|  | 3 | 1.03 | 1.00 | 1.07 | 1.04 | 1.00 | 1.07 | 1.05 | 1.00 | 1.10 | 1.07 | 1.00 | 1.13 | 1.10 | 1.00 | 1.20 | 1.11 | 1.00 | 1.22 |
|  | 4 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.05 | 1.04 | 1.00 | 1.07 | 1.05 | 1.00 | 1.10 | 1.07 | 1.00 | 1.15 | 1.08 | 1.00 | 1.16 |

Table 20.10 Central, low and high estimates of the relative risk of malnutrition for alternative climate scenarios relative to baseline climate (continued)

| Subregion | Climate ${ }^{\text {a }}$ | 2000 |  |  | 2001 |  |  | 2005 |  |  | 2010 |  |  | 2020 |  |  | 2030 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High |
| EUR-A | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| EUR-B | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| EUR-C | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| SEAR-B | 2 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.06 | 1.05 | 1.00 | 1.09 | 1.05 | 1.00 | 1.10 |
|  | 3 | 1.03 | 1.00 | 1.06 | 1.03 | 1.00 | 1.07 | 1.04 | 1.00 | 1.09 | 1.06 | 1.00 | 1.12 | 1.09 | 1.00 | 1.18 | 1.10 | 1.00 | 1.19 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.01 |
| SEAR-D | 2 | 1.04 | 1.00 | 1.07 | 1.04 | 1.00 | 1.08 | 1.06 | 1.00 | 1.11 | 1.07 | 1.00 | 1.15 | 1.11 | 1.00 | 1.22 | 1.12 | 1.00 | 1.25 |
|  | 3 | 1.05 | 1.00 | 1.10 | 1.06 | 1.00 | 1.12 | 1.08 | 1.00 | 1.16 | 1.10 | 1.00 | 1.21 | 1.16 | 1.00 | 1.31 | 1.17 | 1.00 | 1.35 |
|  | 4 | 1.05 | 1.00 | 1.10 | 1.06 | 1.00 | 1.11 | 1.08 | 1.00 | 1.15 | 1.10 | 1.00 | 1.20 | 1.15 | 1.00 | 1.31 | 1.17 | 1.00 | 1.33 |
| WPR-A | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| WPR-B | 2 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.01 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.02 |
|  | 3 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.03 | 1.03 | 1.00 | 1.05 | 1.03 | 1.00 | 1.05 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 |

[^73]- back-casting of climate/hunger relationships for verification of model accuracy;
- a finer geographical (e.g. country level) breakdown of the outputs of malnutrition models;
- correlations between model outputs and health outcomes at a high spatial resolution; and
- investigation of the synergistic effects of water availability and poverty on malnutrition.


### 3.9 DISASTERS CAUSED BY EXTREME WEATHER EVENTS: COASTAL FLOODS, INLAND FLOODS AND LANDSLIDES

Natural disasters are ultimately a function of both the average and degree of variability of weather conditions, which are further modulated by multiple aspects of population vulnerability such as topography, housing quality and early warning systems (Alexander 1993; McMichael et al. 1996). They are therefore likely to be directly affected by the observed and predicted trend towards increasingly variable weather.

In addition to the axiomatic link between extreme weather events and weather-related deaths and injuries, there is strong statistical evidence that long-term weather cycles (e.g. the ENSO quasi-periodic cycle) correlate with incidence of deaths and injuries attributable to natural disasters (Bouma et al. 1997a; Kovats et al. 1999). The evidence for increased frequency of different categories of extreme events in the past, and the likelihood of changes in the future (Table 20.1), has been strengthened by recent demonstrations of increases in the frequency of large floods during the 20th century (Milly et al. 2002). This evidence is reinforced by projections of several-fold increases in the frequency of what are currently considered extreme wet seasons, for various regions over the world, using a range of climate models (Palmer and Ralsanen 2002).

Based on this, we presented estimates of consequences of increasing frequency of coastal flooding caused by sea-level rise, and inland flooding and landslides caused by increased frequency of extreme precipitation events, which are described by IPCC as very likely (90-99\% probability) to increase in many areas under climate change. We did not attempt to estimate the effects of changing frequency and intensity of wind storms, owing to inconsistencies between models (IPCC 2001b) and a lack of quantitative projections of changes in exposure under different climate change scenarios. Our estimates excluded the direct effect of thermal extremes (e.g. heatstroke, increased risk of cardiovascular disease), which are dealt with elsewhere. We also excluded any potential longer-term health consequences arising through mechanisms such as population displacement, economic damage to public health infrastructure, increased risk of infectious disease epidemics and mental illness (Jovel 1989; Menne et al. 1999; WHO 1992).

Climate change is likely to have different effects on the frequency of coastal versus inland floods. Changes in the frequency of coastal floods were defined using published models (Hoozemans and Hulsburgen 1995; Nicholls et al. 1999) that estimate change in sea level for each climate scenario. These changes were applied to topographical and population distribution maps to estimate the change in incidence of exposure to flooding by subregion. The predictions did not account for changes in frequency of storm surges. The global model used here has been shown to be relatively accurate in validations against more detailed assessments at the national level (summarized in Nicholls et al. 1999).

Inland floods and landslides are not affected by sea-level rise, but will be influenced by any increase in the frequency of intense precipitation. Despite the clear causal link, this relationship is poorly researched (Pielke 1999), and has not previously been modelled as a health exposure. There are no published analyses of global relationships between intensity of precipitation, the likelihood of a declared disaster, or the magnitude of health consequences. Clearly, at the local level the frequency of health effects will be determined by the temporal distribution of rainfall (i.e. not only by the average amount of rain over an extended period, but by the peak amount falling in a week, day or hour), and modulated by topography and social aspects of vulnerability (Kundzewicz and Kaczmarek 2000). However, in the absence of detailed data on these variables and their effects, we made the a priori assumption that flood frequency is proportional to the frequency with which monthly rainfall exceeds the 1 in 10 year limit (i.e. upper $99.2 \% \mathrm{CI}$ ) of the baseline climate. We also assumed that determinants of vulnerability are distributed evenly throughout the population of a subregion-so that the change in relative risk of health consequences is proportional to the per capita change in risk of experiencing such an extreme event. For each $1^{\circ} \times 1^{\circ}$ population grid-cell, we estimated the $99.2 \%$ upper CI for the "baseline climate" using means and standard deviations derived from the 1961-1990 averages for each month of the year. Using equivalent data for future scenarios and time points, we estimated the change in frequency with which such a " 1 in 10 year event" occurs.

The difference (in standard deviates of the new distribution) between the new mean and the previously defined " 1 in 10 year limit" is given by:

$$
\left(\left(X_{1}+2.41 \times u_{1}\right)-X_{2}\right) / u_{2}
$$

where $X_{1}, u_{1}=$ mean and intra-annual deviation from 1961-1990 and $X_{2}$, $u_{2}=$ mean and intra-annual deviation under new scenario.

The probability that this difference will be exceeded in any one month under the new distribution was taken from probability tables. This was divided by the frequency of occurrence under the baseline scenario $(=0.008)$ to give the relative frequency of exceeding the 1 in 10 year limit
for each future scenario. Results were weighted by the population in each cell, and averaged across the countries in each subregion. The final measure of exposure was therefore the relative frequency with which each person in a subregion experiences 1 in 10 year rainfall events.

The process of estimating the disease burden of this change in frequency differs from that for other climate-sensitive health outcomes, as flood effects do not have a specific GBD code. Relative risks should therefore be applied to estimates of health consequences such as deaths and injuries attributable to these climate events under baseline climate (i.e. rather than change in total incidence of this outcome).

The EM-DAT database records numbers of deaths and injuries attributed to each natural disaster reported by the media or aid agencies in the last 100 years. Disasters are defined as events that resulted in at least one of the following conditions: (i) >10 people killed, (ii) $>200$ injured, (iii) a call for international assistance. Although this is the best comprehensive data source available on the current health consequences of natural disasters, all such sources may be subject to underreporting (Noji 1997). Estimates of attributable burden of disease derived from these figures are therefore conservative.

Individual events in the database were classified as inland or coastal floods on the basis of geography, or descriptions of events in the database. Total numbers of deaths for each class of event were summed for subregions. Although the EM-DAT database also records the number of injuries in flooding events, these are not included in this assessment, as they are considered particularly unreliable for floods (Guha-Sapir, personal communication, 2002). The effects of events that could not be identified as inland or coastal were assigned in proportion to the distribution of consequences of classified events in each subregion. The annual incidence of flood death under baseline climate conditions was estimated by dividing the annual average over the last 20 years by the subregional population in 1990.

These baseline incidence rates alter over time, depending on the balance between factors that decrease vulnerability (particularly improving flood defences as populations become richer), and those which increase vulnerability (particularly increasing population density in coastal zones and other flood-prone areas). Baseline estimates for future years were therefore adjusted as far as possible for these effects. For coastal flooding, the effects of projected changes in population distribution in relation to coastline and improving coastal defences in line with GDP were incorporated in the model of Nicholls et al. (1999). The baseline estimates of the incidence of deaths and injuries in years after 1990 were therefore scaled by the ratio of the model projections for numbers of people flooded in each year compared to that in 1990 (Table 20.11).

Such vulnerability effects have not been explicitly modelled for inland flooding. However, Yohe and Tol (2002) have carried out a cross-

Table 20.1I Annual incidence of deaths per 10000000 population caused by coastal floods, in the absence of climate change

| Subregion | $1980-1999$ | 2000 | 2001 | 2005 | 2010 | 2020 | 2030 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| AFR-D | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AFR-E | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AMR-A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AMR-B | 2.00 | 1.59 | 1.56 | 1.43 | 1.29 | 1.08 | 0.96 |
| AMR-D | 0.40 | 0.41 | 0.41 | 0.41 | 0.41 | 0.41 | 0.34 |
| EMR-B | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EMR-D | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EUR-A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EUR-B | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EUR-C | 0.10 | 0.11 | 0.11 | 0.12 | 0.12 | 0.14 | 0.15 |
| SEAR-B | 0.10 | 0.11 | 0.11 | 0.11 | 0.12 | 0.12 | 0.11 |
| SEAR-D | 1.20 | 1.39 | 1.40 | 1.47 | 1.54 | 1.69 | 1.78 |
| WPR-A | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.11 | 0.11 |
| WPR-B | 0.90 | 0.98 | 0.99 | 1.03 | 1.07 | 1.15 | 1.22 |

Source: EM-DAT (2002) for the period I980-I999. Estimates are based on changing GDP for other time points (see text).
sectional analysis of the effect of per capita income on the incidence of death due to all natural disasters (as reported in the EM-DAT database) for the period 1990-2000. They conclude that increasing wealth has a protective effect, best described by:
$\operatorname{Ln}$ (proportion of population killed/decade)
$=4.7271-0.3858(\mathrm{Ln}$ GDP per capita)

The effect of income is marginally non-significant at the $5 \%$ level ( $P<0.07$ ), generic to all natural disasters and does not take account of the magnitude of the physical hazard. However, this is likely to introduce noise rather than bias in the relationship, and the relationship represents the only published basis for projection of the protective effects of economic development. The relationship is therefore applied to future projections of GDP (WHO/EIP/GPE, unpublished data, 2000). Our projected baseline incidence of deaths for years after 1990 were scaled by the ratio of the projections of deaths due to all natural disasters in that year divided by the estimate for 1990 (Table 20.12).

Some evidence from studies of a small number of earthquakes (Beinin 1981) and famines (Rivers 1982) suggests that women and young children are more vulnerable than men to the acute effects of natural disasters. However, there are insufficient data to derive subregional estimates

Table 20.12 Annual incidence of deaths per 10000000 population caused by inland floods and landslides, in the absence of climate change

| Subregion | $1980-1999$ | 2000 | 2001 | 2005 | 2010 | 2020 | 2030 |
| :--- | :---: | ---: | ---: | ---: | ---: | ---: | ---: |
| AFR-D | 2.7 | 2.7 | 2.7 | 2.7 | 2.6 | 2.4 | 2.2 |
| AFR-E | 6.5 | 6.6 | 6.6 | 6.5 | 6.4 | 6.0 | 5.4 |
| AMR-A | 2.2 | 2.1 | 2.1 | 2.0 | 1.9 | 1.8 | 1.6 |
| AMR-B | 52.2 | 48.4 | 48.1 | 46.6 | 44.8 | 41.0 | 36.9 |
| AMR-D | 52.1 | 49.0 | 48.7 | 47.2 | 45.4 | 41.6 | 37.4 |
| EMR-B | 14.9 | 13.8 | 13.7 | 13.4 | 13.1 | 12.1 | 11.0 |
| EMR-D | 32.2 | 30.9 | 30.6 | 29.5 | 28.2 | 25.6 | 23.0 |
| EUR-A | 1.3 | 1.2 | 1.2 | 1.1 | 1.1 | 0.9 | 0.8 |
| EUR-B | 8.9 | 9.2 | 9.1 | 8.7 | 8.2 | 7.4 | 6.6 |
| EUR-C | 1.2 | 1.4 | 1.4 | 1.3 | 1.2 | 1.1 | 1.0 |
| SEAR-B | 9.9 | 8.2 | 8.1 | 7.5 | 6.8 | 5.8 | 5.1 |
| SEAR-D | 20.3 | 17.7 | 17.5 | 16.5 | 15.4 | 13.5 | 12.0 |
| WPR-A | 3.7 | 3.3 | 3.3 | 3.0 | 2.7 | 2.3 | 2.0 |
| WPR-B | 13.8 | 10.6 | 10.4 | 9.6 | 8.7 | 7.4 | 6.5 |

Source: EM-DAT (2002) for the period 1980-1999. Estimates are based on changing GDP for other time points (see text).
of the relative vulnerability of different age groups and sexes to the consequences of flooding: we therefore made the assumption, that all age groups are equally at risk.

The models presented here for coastal flooding assumed that protection evolves over time in proportion to projected increases in GDP. The mid-range estimates presented therefore incorporated an effect of increasing wealth, not only in the baseline estimates, but assumed the same proportional change in the relative risks (i.e. as described by Yohe and Tol 2002). This accounted for the effect of increasing wealth not only in reducing the likely health consequences of "baseline" (i.e. climate change independent floods), but also providing better adaptive capacity for increases driven by climate change. The mid-range estimates did not include any further adjustments for biological/behavioural adaptation to increased flood risk (Table 20.13).

As for other health outcomes of climate change, the only published sensitivity analyses relate to an unmitigated emissions scenario applied to the HadCM2 model with four slightly varying sets of initial conditions, and a comparison with the same emissions scenario run on the HadCM3 model. These resulted in almost no difference in model outputs over the time-scale considered in this assessment. Uncertainties in the model relate to the degree and manner to which individuals respond to
Table 20.13 Central, low and high estimates of the relative risk of death in coastal floods for alternative climate scenarios

|  |  | 2000 |  |  | 2001 |  |  | 2005 |  |  | 2010 |  |  | 2020 |  |  | 2030 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Subregion | Climate ${ }^{\text {a }}$ | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High |
| AFR-D | 2 | 1.07 | 1.04 | 1.14 | 1.07 | 1.04 | 1.15 | 1.09 | 1.05 | 1.18 | 1.11 | 1.05 | 1.21 | 1.13 | 1.06 | 1.26 | 1.44 | 1.22 | 1.89 |
|  | 3 | 1.07 | 1.04 | 1.15 | 1.08 | 1.04 | 1.16 | 1.10 | 1.05 | 1.19 | 1.11 | 1.06 | 1.23 | 1.14 | 1.07 | 1.27 | 1.48 | 1.24 | 1.96 |
|  | 4 | 1.10 | 1.05 | 1.20 | 1.11 | 1.05 | 1.21 | 1.13 | 1.06 | 1.25 | 1.15 | 1.07 | 1.30 | 1.18 | 1.09 | 1.36 | 1.64 | 1.32 | 2.29 |
| AFR-E | 2 | 1.06 | 1.03 | 1.12 | 1.06 | 1.03 | 1.13 | 1.07 | 1.04 | 1.15 | 1.08 | 1.04 | 1.17 | 1.09 | 1.05 | 1.19 | 1.12 | 1.06 | 1.25 |
|  | 3 | 1.06 | 1.03 | 1.13 | 1.07 | 1.03 | 1.14 | 1.08 | 1.04 | 1.16 | 1.09 | 1.04 | 1.18 | 1.10 | 1.05 | 1.20 | 1.13 | 1.07 | 1.27 |
|  | 4 | 1.09 | 1.04 | 1.17 | 1.09 | 1.04 | 1.18 | 1.10 | 1.05 | 1.21 | 1.12 | 1.06 | 1.23 | 1.13 | 1.07 | 1.27 | 1.18 | 1.09 | 1.35 |
| AMR-A | 2 | 1.03 | 1.02 | 1.06 | 1.04 | 1.02 | 1.07 | 1.06 | 1.03 | 1.11 | 1.08 | 1.04 | 1.16 | 1.12 | 1.06 | 1.25 | 1.13 | 1.06 | 1.25 |
|  | 3 | 1.03 | 1.02 | 1.07 | 1.04 | 1.02 | 1.08 | 1.06 | 1.03 | 1.12 | 1.09 | 1.04 | 1.17 | 1.13 | 1.07 | 1.27 | 1.14 | 1.07 | 1.27 |
|  | 4 | 1.05 | 1.03 | 1.10 | 1.06 | 1.03 | 1.12 | 1.09 | 1.04 | 1.17 | 1.12 | 1.06 | 1.24 | 1.18 | 1.09 | 1.37 | 1.19 | 1.09 | 1.38 |
| AMR-B | 2 | 1.24 | 1.12 | 1.48 | 1.27 | 1.13 | 1.54 | 1.38 | 1.19 | 1.75 | 1.52 | 1.26 | 2.04 | 1.84 | 1.42 | 2.69 | 1.90 | 1.45 | 2.81 |
|  | 3 | 1.26 | 1.13 | 1.51 | 1.29 | 1.14 | 1.57 | 1.40 | 1.20 | 1.80 | 1.55 | 1.28 | 2.11 | 1.90 | 1.45 | 2.80 | 1.96 | 1.48 | 2.93 |
|  | 4 | 1.34 | 1.17 | 1.68 | 1.38 | 1.19 | 1.75 | 1.53 | 1.26 | 2.05 | 1.73 | 1.36 | 2.46 | 2.18 | 1.59 | 3.37 | 2.27 | 1.64 | 3.54 |
| AMR-D | 2 | 2.07 | 1.53 | 3.14 | 2.15 | 1.58 | 3.30 | 2.45 | 1.73 |  | 2.77 | 1.89 | 4.55 | 3.27 | 2.14 | 5.55 | 3.58 | 2.29 | 6.17 |
|  | 3 | 2.14 | 1.57 | 3.28 | 2.23 | 1.61 | 3.45 | 2.55 | 1.77 | 4.10 | 2.89 | 1.94 | 4.78 | 3.42 | 2.21 | 5.84 | 3.76 | 2.38 | 6.52 |
|  | 4 | 2.50 | 1.75 | 4.00 | 2.62 | 1.81 | 4.23 | 3.04 | 2.02 | 5.08 | 3.49 | 2.24 | 5.97 | 4.19 | 2.59 | 7.38 | 4.64 | 2.82 | 8.28 |
| EMR-B | 2 | 1.14 | 1.07 | 1.29 | 1.16 | 1.08 | 1.32 | 1.22 | 1.11 | 1.45 | 1.31 | 1.16 | 1.63 | 1.52 | 1.26 | 2.03 | 1.53 | 1.27 | 2.06 |
|  | 3 | 1.15 | 1.08 | 1.31 | 1.17 | 1.08 | 1.34 | 1.24 | 1.12 | 1.48 | 1.33 | 1.17 | 1.67 | 1.55 | 1.28 | 2.10 | 1.57 | 1.28 | 2.13 |
|  | 4 | 1.20 | 1.10 | 1.40 | 1.22 | 1.11 | 1.45 | 1.31 | 1.16 | 1.63 | 1.44 | 1.22 | 1.88 | 1.72 | 1.36 | 2.45 | 1.75 | 1.37 | 2.50 |
| EMR-D | 2 | 1.53 | 1.27 | 2.07 | 1.57 | 1.28 | 2.14 | 1.68 | 1.34 | 2.36 | 1.79 | 1.39 | 2.58 | 1.94 | 1.47 | 2.88 | 3.01 | 2.01 | 5.02 |
|  | 3 | 1.57 | 1.28 | 2.14 | 1.61 | 1.30 | 2.21 | 1.73 | 1.36 | 2.45 | 1.84 | 1.42 | 2.68 | 2.00 | 1.50 | 3.00 | 3.18 | 2.09 | 5.36 |
|  | 4 | 1.75 | 1.38 | 2.50 | 1.80 | 1.40 | 2.59 | 1.96 | 1.48 | 2.91 | 2.11 | 1.55 | 3.21 | 2.32 | 1.66 | 3.63 | 3.91 | 2.46 | 6.82 |


| EUR-A | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.09 | 1.04 | 1.18 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.10 | 1.05 | 1.20 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.14 | 1.07 | 1.29 |
| EUR-B | 2 | 1.57 | 1.28 | 2.13 | 1.64 | 1.32 | 2.27 | 1.95 | 1.48 | 2.91 | 2.45 | 1.73 | 3.91 | 4.05 | 2.52 | 7.09 | 4.78 | 2.89 | 8.55 |
|  | 3 | 1.60 | 1.30 | 2.21 | 1.68 | 1.34 | 2.36 | 2.02 | 1.51 | 3.03 | 2.55 | 1.77 | 4.10 | 4.24 | 2.62 | 7.49 | 5.02 | 3.01 | 9.05 |
|  | 4 | 1.79 | 1.40 | 2.59 | 1.89 | 1.45 | 2.79 | 2.34 | 1.67 | 3.68 | 3.04 | 2.02 | 5.08 | 5.27 | 3.14 | 9.54 | 6.31 | 3.65 | 11.61 |
| EUR-C | 2 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.02 | 1.01 | 1.01 | 1.02 | 1.02 | 1.01 | 1.03 | 1.02 | 1.01 | 1.04 | 1.03 | 1.01 | 1.06 |
|  | 3 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.02 | 1.01 | 1.01 | 1.03 | 1.02 | 1.01 | 1.03 | 1.02 | 1.01 | 1.05 | 1.03 | 1.02 | 1.06 |
|  | 4 | 1.01 | 1.01 | 1.02 | 1.01 | 1.01 | 1.03 | 1.02 | 1.01 | 1.03 | 1.02 | 1.01 | 1.04 | 1.03 | 1.02 | 1.06 | 1.04 | 1.02 | 1.08 |
| SEAR-B | 2 | 1.11 | 1.06 | 1.22 | 1.12 | 1.06 | 1.24 | 1.15 | 1.08 | 1.31 | 1.19 | 1.09 | 1.38 | 1.24 | 1.12 | 1.49 | 1.28 | 1.14 | 1.56 |
|  | 3 | 1.12 | 1.06 | 1.24 | 1.13 | 1.06 | 1.26 | 1.16 | 1.08 | 1.33 | 1.20 | 1.10 | 1.40 | 1.26 | 1.13 | 1.52 | 1.30 | 1.15 | 1.59 |
|  | 4 | 1.16 | 1.08 | 1.32 | 1.17 | 1.09 | 1.34 | 1.22 | 1.11 | 1.43 | 1.26 | 1.13 | 1.53 | 1.34 | 1.17 | 1.68 | 1.39 | 1.20 | 1.78 |
| SEAR-D | 2 | 1.01 | 1.00 | 1.02 | 1.01 | 1.01 |  | 1.01 | 1.01 | 1.03 |  | 1.01 | 1.03 | 1.02 | 1.01 | 1.04 | 1.03 | 1.01 | 1.05 |
|  | 3 | 1.01 | 1.01 | 1.02 | 1.01 | 1.01 | 1.02 | 1.01 | 1.01 | 1.03 | 1.02 | 1.01 | 1.03 | 1.02 | 1.01 | 1.04 | 1.03 | 1.01 | 1.05 |
|  | 4 | 1.01 | 1.01 | 1.03 | 1.01 | 1.01 | 1.03 | 1.02 | 1.01 | 1.04 | 1.02 | 1.01 | 1.04 | 1.03 | 1.01 | 1.05 | 1.04 | 1.02 | 1.07 |
| WPR-A | 2 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.02 | 1.01 | 1.01 | 1.03 | 1.02 | 1.01 | 1.04 | 1.03 | 1.01 | 1.06 |
|  | 3 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.01 | 1.01 | 1.01 | 1.02 | 1.02 | 1.01 | 1.03 | 1.02 | 1.01 | 1.05 | 1.03 | 1.02 | 1.06 |
|  | 4 | 1.01 | 1.00 | 1.02 | 1.01 | 1.01 | 1.02 | 1.02 | 1.01 | 1.03 | 1.02 | 1.01 | 1.04 | 1.03 | 1.02 | 1.06 | 1.04 | 1.02 | 1.09 |
| WPR-B | 2 | 1.01 | 1.01 | 1.02 | 1.01 | 1.01 | 1.03 | 1.02 | 1.01 | 1.04 | 1.02 | 1.01 | 1.05 | 1.03 | 1.02 | 1.06 | 1.04 | 1.02 | 1.07 |
|  | 3 | 1.01 | 1.01 | 1.03 | 1.01 | 1.01 | 1.03 | 1.02 | 1.01 | 1.04 | 1.02 | 1.01 | 1.05 | 1.03 | 1.02 | 1.07 | 1.04 | 1.02 | 1.08 |
|  | 4 | 1.02 | 1.01 | 1.04 | 1.02 | 1.01 | 1.04 | 1.03 | 1.01 | 1.05 | 1.03 | 1.02 | 1.07 | 1.04 | 1.02 | 1.09 | 1.05 | 1.03 | 1.10 |

the increased risk (Hoozemans et al. 1993). The lower estimates therefore assumed that $90 \%$ of the risk could be avoided either by highly efficient coastal defences or individual adaptations. The higher estimates assumed no adaptation either with increasing GDP or individual level measures.

The model for inland flooding was subject to the same uncertainty over adaptive responses. Both the baseline and relative risks were assumed to change with GDP, as described for coastal flooding. As outlined above, however, the uncertainty is also greater for a hazard driven by the magnitude and temporal variation of precipitation (which varies considerably between climate models), rather than the more predictable process of temperature-driven sea level rise. This consideration is particularly important as only one climate model was used in this assessment. Although the relative risks estimated by our method (e.g. 5.53 for our estimates for EUR-A by the year 2030 under an unmitigated emissions scenario) were broadly comparable to estimates of changes in frequency of extreme wet seasons generated using multi-climate model analyses (fivefold for northern Europe for the period 2060-2080; Palmer and Ralsanen 2002), more formal analyses would be necessary to give more accurate estimates. We therefore gave a larger uncertainty range around these predictions than for coastal flooding, by assuming a $50 \%$ greater exposure and no adaptation with GDP for the high estimate, and no increase in risk under any scenario for the lower estimate (Table 20.14). As for the other outcomes, this uncertainty range should be interpreted with caution.

The potential health consequences of changing frequency and intensity of extreme weather events are surprisingly poorly researched. Substantial improvements in assessment could be made through better estimates of the current health impacts of natural disasters, which suffer from poor baseline data and probably severe underreporting, particularly in developing countries (Noji 1997). Analyses could also be greatly improved by geo-referencing and more detailed descriptions of disasters to allow differentiation of inland and coastal events, detailed analysis of the relationships between intensity of precipitation and health effects, and projections of future precipitation at higher temporal and spatial resolution, using output from a range of climate models. In order to generate better uncertainty estimates, formal sensitivity analyses of the contributions of each model parameter to the final uncertainty estimates are required.

Finally, it should be stressed that the estimates given here represent the immediate acute consequences of natural disasters, which are likely to be only one component of the total attributable disease burden. Other outcomes of these natural disasters need to be considered, such as the probability of outbreaks of water, vector- and rodent-borne diseases, the effects of sequential disasters on both public health defences and stability of natural ecosystems limiting disease outbreaks (Epstein 1999), and
Table 20.14 Central, low and high estimates of the relative risk of death in inland floods/landslides for alternative climate scenarios relative to baseline climate

| Subregion | Climate ${ }^{\text {a }}$ | 2000 |  |  | 2001 |  |  | 2005 |  |  | 2010 |  |  | 2020 |  |  | 2030 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High |
| AFR-D | 2 | 1.40 | 1.00 | 1.60 | 1.44 | 1.00 | 1.66 | 1.60 | 1.00 | 1.89 | 1.77 | 1.00 | 2.19 | 2.09 | 1.00 | 2.79 | 2.30 | 1.00 | 3.13 |
|  | 3 | 1.30 | 1.00 | 1.45 | 1.32 | 1.00 | 1.50 | 1.44 | 1.00 | 1.68 | 1.58 | 1.00 | 1.89 | 1.83 | 1.00 | 2.34 | 1.99 | 1.00 | 2.64 |
|  | 4 | 1.18 | 1.00 | 1.27 | 1.20 | 1.00 | 1.30 | 1.26 | 1.00 | 1.41 | 1.35 | 1.00 | 1.54 | 1.50 | 1.00 | 1.81 | 1.66 | 1.00 | 2.08 |
| AFR-E | 2 | 1.37 | 1.00 | 1.54 | 1.40 | 1.00 | 1.60 | 1.53 | 1.00 | 1.83 | 1.71 | 1.00 | 2.10 | 2.03 | 1.00 | 2.64 | 2.30 | 1.00 | 3.18 |
|  | 3 | 1.26 | 1.00 | 1.41 | 1.30 | 1.00 | 1.45 | 1.39 | 1.00 | 1.60 | 1.53 | 1.00 | 1.81 | 1.75 | 1.00 | 2.22 | 1.99 | 1.00 | 2.65 |
|  | 4 | 1.22 | 1.00 | 1.33 | 1.24 | 1.00 | 1.36 | 1.33 | 1.00 | 1.50 | 1.43 | 1.00 | 1.66 | 1.62 | 1.00 | 1.99 | 1.86 | 1.00 | 2.44 |
| AMR-A | 2 | 4.19 | 1.00 | 6.00 | 4.66 | 1.00 | 6.51 | 5.76 | 1.00 | 8.50 | 7.33 | 1.00 | 11.0 | 10.5 | 1.00 | 16.00 | 11.5 | 1.00 | 18.69 |
|  | 3 | 3.62 | 1.00 | 5.10 | 4.00 | 1.00 | 5.50 | 4.90 | 1.00 | 7.15 | 6.19 | 1.00 | 9.19 | 8.77 | 1.00 | 13.29 | 9.66 | 1.00 | 15.61 |
|  | 4 | 3.02 | 1.00 | 4.18 | 3.34 | 1.00 | 4.50 | 4.03 | 1.00 | 5.77 | 5.03 | 1.00 | 7.36 | 7.03 | 1.00 | 10.54 | 7.99 | 1.00 | 12.79 |
| AMR-B | 2 | 1.43 | 1.00 | 1.69 | 1.50 | 1.00 | 1.75 | 1.66 | 1.00 | 2.04 | 1.88 | 1.00 | 2.37 | 2.26 | 1.00 | 3.06 | 2.60 | 1.00 | 3.67 |
|  | 3 | 1.59 | 1.00 | 1.98 | 1.72 | 1.00 | 2.07 | 1.93 | 1.00 | 2.46 | 2.25 | 1.00 | 2.94 | 2.78 | 1.00 | 3.91 | 3.18 | 1.00 | 4.65 |
|  | 4 | 1.56 | 1.00 | 1.89 | 1.66 | 1.00 | 1.98 | 1.87 | 1.00 | 2.34 | 2.13 | 1.00 | 2.79 | 2.63 | 1.00 | 3.67 | 3.03 | 1.00 | 4.39 |
| AMR-D | 2 | 1.53 | 1.00 | 1.83 | 1.60 | 1.00 | 1.92 | 1.79 | 1.00 | 2.25 | 2.06 | 1.00 | 2.65 | 2.52 | 1.00 | 3.48 | 2.92 | 1.00 | 4.20 |
|  | 3 | 1.32 | 1.00 | 1.50 | 1.36 | 1.00 | 1.56 | 1.48 | 1.00 | 1.75 | 1.63 | 1.00 | 2.01 | 1.92 | 1.00 | 2.50 | 2.26 | 1.00 | 3.10 |
|  | 4 | 1.36 | 1.00 | 1.57 | 1.42 | 1.00 | 1.62 | 1.54 | 1.00 | 1.84 | 1.73 | 1.00 | 2.13 | 2.03 | 1.00 | 2.70 | 2.40 | 1.00 | 3.33 |
| EMR-B | 2 | 1.61 | 1.00 | 1.99 | 1.71 | 1.00 | 2.10 | 1.98 | 1.00 | 2.49 | 2.29 | 1.00 | 2.98 | 2.83 | 1.00 | 3.97 | 3.20 | 1.00 | 4.63 |
|  | 3 | 1.85 | 1.00 | 2.40 | 2.01 | 1.00 | 2.53 | 2.37 | 1.00 | 3.09 | 2.82 | 1.00 | 3.79 | 3.57 | 1.00 | 5.19 | 4.04 | 1.00 | 6.03 |
|  | 4 | 1.89 | 1.00 | 2.44 | 2.05 | 1.00 | 2.58 | 2.41 | 1.00 | 3.15 | 2.88 | 1.00 | 3.87 | 3.64 | 1.00 | 5.31 | 4.04 | 1.00 | 6.01 |
| EMR-D | 2 | 2.32 | 1.00 | 3.09 | 2.51 | 1.00 | 3.28 | 3.01 | 1.00 | 4.12 | 3.66 | 1.00 | 5.16 | 4.78 | 1.00 | 7.24 | 5.29 | 1.00 | 8.17 |
|  | 3 | 2.07 | 1.00 | 2.68 | 2.23 | 1.00 | 2.85 | 2.62 | 1.00 | 3.52 | 3.14 | 1.00 | 4.36 | 4.05 | 1.00 | 6.04 | 4.56 | 1.00 | 6.94 |
|  | 4 | 2.13 | 1.00 | 2.77 | 2.29 | 1.00 | 2.94 | 2.70 | 1.00 | 3.64 | 3.26 | 1.00 | 4.53 | 4.20 | 1.00 | 6.30 | 4.68 | 1.00 | 7.15 |

Table 20.14 Central, low and high estimates of the relative risk of death in inland floods/landslides for alternative climate

| Subregion | Climate | 2000 |  |  | 2001 |  |  | 2005 |  |  | 2010 |  |  | 2020 |  |  | 2030 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High |
| EUR-A | 2 | 2.13 | 1.00 | 2.83 | 2.34 | 1.00 | 3.01 | 2.67 | 1.00 | 3.75 | 3.44 | 1.00 | 4.66 | 3.99 | 1.00 | 6.48 | 5.30 | 1.00 | 8.28 |
|  | 3 | 2.13 | 1.00 | 2.83 | 2.34 | 1.00 | 3.03 | 2.69 | 1.00 | 3.76 | 3.44 | 1.00 | 4.68 | 4.01 | 1.00 | 6.51 | 5.27 | 1.00 | 8.20 |
|  | 4 | 2.22 | 1.00 | 2.98 | 2.44 | 1.00 | 3.18 | 2.82 | 1.00 | 3.97 | 3.64 | 1.00 | 4.95 | 4.24 | 1.00 | 6.93 | 5.53 | 1.00 | 8.65 |
| EUR-B | 2 | 1.31 | 1.00 | 1.44 | 1.32 | 1.00 | 1.48 | 1.42 | 1.00 | 1.66 | 1.55 | 1.00 | 1.89 | 1.79 | 1.00 | 2.32 | 2.32 | 1.00 | 3.22 |
|  | 3 | 1.64 | 1.00 | 1.92 | 1.67 | 1.00 | 2.01 | 1.88 | 1.00 | 2.38 | 2.15 | 1.00 | 2.83 | 2.66 | 1.00 | 3.75 | 3.16 | 1.00 | 4.65 |
|  | 4 | 1.35 | 1.00 | 1.53 | 1.38 | 1.00 | 1.57 | 1.50 | 1.00 | 1.78 | 1.66 | 1.00 | 2.04 | 1.94 | 1.00 | 2.56 | 2.46 | 1.00 | 3.45 |
| EUR-C | 2 | 1.33 | 1.00 | 1.42 | 1.30 | 1.00 | 1.45 | 1.39 | 1.00 | 1.62 | 1.52 | 1.00 | 1.83 | 1.75 | 1.00 | 2.25 | 2.45 | 1.00 | 3.42 |
|  | 3 | 2.05 | 1.00 | 2.37 | 2.00 | 1.00 | 2.50 | 2.26 | 1.00 | 3.04 | 2.68 | 1.00 | 3.72 | 3.49 | 1.00 | 5.08 | 4.31 | 1.00 | 6.46 |
|  | 4 | 1.70 | 1.00 | 1.90 | 1.66 | 1.00 | 1.99 | 1.84 | 1.00 | 2.37 | 2.11 | 1.00 | 2.82 | 2.67 | 1.00 | 3.72 | 3.45 | 1.00 | 5.04 |
| SEAR-B | 2 | 1.33 | 1.00 | 1.59 | 1.43 | 1.00 | 1.65 | 1.56 | 1.00 | 1.89 | 1.71 | 1.00 | 2.19 | 2.01 | 1.00 | 2.77 | 2.51 | 1.00 | 3.60 |
|  | 3 | 1.61 | 1.00 | 2.11 | 1.81 | 1.00 | 2.23 | 2.04 | 1.00 | 2.67 | 2.34 | 1.00 | 3.22 | 2.89 | 1.00 | 4.33 | 3.57 | 1.00 | 5.37 |
|  | 4 | 1.30 | 1.00 | 1.54 | 1.40 | 1.00 | 1.60 | 1.50 | 1.00 | 1.81 | 1.65 | 1.00 | 2.08 | 1.92 | 1.00 | 2.62 | 2.39 | 1.00 | 3.37 |
| SEAR-D | 2 | 1.21 | 1.00 | 1.36 | 1.26 | 1.00 | 1.41 | 1.34 | 1.00 | 1.56 | 1.45 | 1.00 | 1.74 | 1.65 | 1.00 | 2.10 | 1.73 | 1.00 | 2.22 |
|  | 3 | 1.10 | 1.00 | 1.20 | 1.14 | 1.00 | 1.21 | 1.19 | 1.00 | 1.29 | 1.24 | 1.00 | 1.38 | 1.33 | 1.00 | 1.57 | 1.39 | 1.00 | 1.68 |
|  | 4 | 1.05 | 1.00 | 1.11 | 1.08 | 1.00 | 1.11 | 1.09 | 1.00 | 1.15 | 1.13 | 1.00 | 1.20 | 1.18 | 1.00 | 1.30 | 1.21 | 1.00 | 1.36 |
| WPR-A | 2 | 1.46 | 1.00 | 1.80 | 1.58 | 1.00 | 1.87 | 1.73 | 1.00 | 2.19 | 1.95 | 1.00 | 2.59 | 2.35 | 1.00 | 3.37 | 2.91 | 1.00 | 4.29 |
|  | 3 | 1.20 | 1.00 | 1.33 | 1.24 | 1.00 | 1.38 | 1.31 | 1.00 | 1.50 | 1.40 | 1.00 | 1.68 | 1.56 | 1.00 | 2.01 | 2.04 | 1.00 | 2.80 |
|  | 4 | 1.30 | 1.00 | 1.50 | 1.36 | 1.00 | 1.54 | 1.45 | 1.00 | 1.75 | 1.59 | 1.00 | 1.99 | 1.85 | 1.00 | 2.49 | 2.32 | 1.00 | 3.28 |
| WPR-B | 2 | 1.17 | 1.00 | 1.35 | 1.26 | 1.00 | 1.38 | 1.31 | 1.00 | 1.51 | 1.42 | 1.00 | 1.69 | 1.58 | 1.00 | 2.04 | 1.88 | 1.00 | 2.50 |
|  | 3 | 1.22 | 1.00 | 1.41 | 1.29 | 1.00 | 1.45 | 1.39 | 1.00 | 1.62 | 1.49 | 1.00 | 1.83 | 1.70 | 1.00 | 2.23 | 2.00 | 1.00 | 2.70 |
|  | 4 | 1.40 | 1.00 | 1.60 | 1.44 | 1.00 | 1.66 | 1.60 | 1.00 | 1.89 | 1.77 | 1.00 | 2.19 | 2.09 | 1.00 | 2.79 | 2.30 | 1.00 | 3.13 |

[^74]longer-term effects such as post-traumatic stress after floods (e.g. Phifer 1990) and of population displacement through coastal flooding. More importantly, it is necessary to expand the range of natural disasters beyond those considered in this chapter. There is an obvious causal chain between apparently increasing variability in precipitation frequency of droughts and their associated health consequences, particularly food shortages and possible famine (UNDMTP 1990; WHO/PTC 1995). While these impacts may be expected to increase under climate change, no models have yet been developed to link climate scenarios, frequency of drought, and associated health effects, so that quantitative estimation of climate change effects is not currently possible. Given the global importance of drought-related disease (WHO/EHA 1998) and the effectiveness of early preparation and rapid response for avoiding health impacts (Gupta 2000; WHO/EHA 2002), this is clearly a priority area for research.

### 3.10 Vector-borne diseases

Viruses, bacteria, protozoa and helminths transmitted by biting insects and other intermediate hosts are among the most important causes of ill-health in tropical regions (WHO 2000b). The climate sensitivity of such diseases has long been recognized (e.g. Celli 1933), and knowledge of the relationships has been used to help predict epidemics of vectorborne diseases since at least the early years of the last century (e.g. Christophers 1911; MacDonald 1957). More recently, there have been many quantitative studies of the effects of climate variables on the population biology of vectors and pathogens in the laboratory (e.g. reviews by Martens 1998b; Massad and Forattini 1998), and on spatial and temporal variations in vector abundance and disease incidence in the field (review by Kovats et al. 2000a). The results of a literature review are shown in Appendix B, Tables B. 5 and B.6.

This climate sensitivity has prompted several studies correlating longterm changes in vector distribution (Lindgren et al. 2000) or disease incidence (e.g. Bouma et al. 1996; Loevinsohn 1994) with local climate trends, which apparently reflect global climate change. However, many of the inferences about resulting health consequences have been called into question, due to concurrent local changes in crucial non-climatic factors, such as human behaviour, disease reporting or control programmes (Hay et al. 2002; Mouchet 1998; Randolph 2001; Reiter 2001).

As for other health effects, it is unlikely that simply correlating longterm trends in disease against trends in climate will soon (or perhaps ever) give unequivocal evidence of the effects of gradual climate change. As climate varies naturally between years, long time-series (i.e. two to three decades) will be needed for statistical tests of association in longterm trends. It is almost inevitable that non-climatic determinants of risk will also change over such long periods, either obscuring or offering
alternative explanations for any effects of climate change. In addition, reliable multi-decadal time-series for vector-borne diseases in developing countries are rare, and it is problematic to interpret analyses of a few data sets as evidence of a general global pattern. Recent reviews therefore suggest that a general effect of climate change on vector-borne disease, while suspected and possible, is highly dependent on other nonclimatic factors that act at a small scale (IPCC 2001a; Kovats et al. 2001, Reiter 2001, Reiter et al. 2003).

Despite the practical problems in making direct correlations with recent trends, the extreme climate sensitivity of vector-borne diseases means that it is almost inevitable that they will respond in some way to climate change, in some settings and to some degree. The most reliable basis for estimating such changes should come from information on the relationships between variations in climate and disease in either the past or present. Several studies have used such data to model the effect of predicted climate change on either the distribution of vector-borne diseases, and/or measures of risk within existing or predicted newly endemic areas. The data, techniques and assumptions used in the various analyses are reviewed in detail in later sections.

Whatever the quality of the data and modelling techniques used for predictions, they will remain contingent upon other determinants of disease. Socioeconomic conditions, control programmes, human immunity and the specific combinations of climate variables required by particular vector species or transmission cycles also affect disease incidence, and may be more important than global climate trends, particularly at small spatial scales (Mouchet and Manguin 1999; Randolph et al. 2000; Reiter 2001; Rogers and Randolph 2000; Sutherst 1998).

The IPCC has reviewed the observed and predicted effects of climate variability and change in the context of the other factors listed above (IPCC 2001a). On balance, the IPCC concludes that climate change is likely to expand the geographical distribution of several vector-borne diseases, including malaria, dengue and leishmaniasis to higher altitudes (high confidence) and higher latitudes with limited public health defences (medium/low confidence), and to extend the transmission seasons in some locations (medium/high confidence). For some vector-borne diseases in some locations, climate change may decrease transmission by reductions in rainfall or temperatures too high for transmission (medium/low confidence).

The associations between climate and infectious diseases have also been reviewed by the United States National Research Council (National Research Council 2001b). The review highlights the climate sensitivity of vector-borne diseases, and describes some studies modelling the potential consequences of future climate change on vector-borne diseases. The report re-emphasizes the need for caution when making future projections, and stresses the importance of public health provision in mitigating increases in incidence driven by climate change. No judgement is
made on the likely effects of future climate change on specific diseases, owing to the limited evidence.

Aside from analyses of the potential effect of gradual changes in average climate conditions, predicted increases in climate variability, including possible increases in the frequency and intensity of El Niño events (IPCC 2001b), may also affect vector-borne diseases. There is evidence that the El Niño cycle causes inter-annual variations in disease incidence in several areas (e.g. Bouma et al. 1997b). However, even in these sites, other determinants such as seasonal variations and control programmes exert a larger influence, and epidemic cycles also occur in areas where El Niño has little or no effect on climate, apparently driven by gradual post-epidemic waning of herd immunity (Hay et al. 2000). As the effect of climate variability differs so greatly between sites, global projections of the associations described so far are unlikely to be informative.

## Modelling of specific diseases

Most modelling of the effects of climate change has focused on malaria, and to a lesser extent dengue. These are therefore the only vector-borne diseases considered here. Some preliminary modelling work has been carried out on schistosomiasis (Martens et al. 1997), but this is based on a relatively small data set and has not been validated against current distributions. Randolph and Rogers (2000) have also modelled the potential effects of climate change on tick-borne encephalitis (TBE) in Europe, demonstrating that increased temperatures are likely to reduce the endemic range. Although TBE is a relatively small public health problem in global terms, and was not covered in this assessment, this study does demonstrate that climate change may potentially decrease, rather than increase, the transmission of some diseases. Effects on other major vector-borne diseases have been investigated either qualitatively (e.g. Carcavallo and Curto de Casas 1996 for American trypanosomiasis) or in terms of distribution of vectors rather than human disease (e.g. Rogers and Packer 1993 for African trypanosomiasis).

## Falciparum malaria

Falciparum malaria is unusual in that several research groups have independently modelled the relationships between climate and disease distribution (Appendix B, Table B.5). The models used can be broadly classified as structural/biological, based on aggregating the effect of climate on the individual components of the disease transmission cycle, or "statistical", derived from direct correlations between geographic or temporal variations in climate, and associated variations in disease incidence or distribution, either in the present or recent past.

Published biological models for the global distribution of falciparum malaria use laboratory data to define the relationship between temperature and the extrinsic incubation period of the parasite, and therefore
the probability of completing development during the lifetime of the mosquito and completing the transmission cycle (Martin and Lefebvre 1995). Later models incorporate temperature effects on the survival probability and biting frequency of mosquitoes (Jetten et al. 1996; Martens et al. 1995a, 1995b, 1999). In these later models, the various temperature-dependent relationships are aggregated into the entomological version of the equation for $R_{o}$, the number of cases arising from each new case in a completely susceptible population (Anderson and May 1991; Dye 1992; Garrett-Jones 1964). Because of the lack of data on several key parameters, these are set as biologically plausible constants, allowing the calculation of the critical vector density required for sustainable disease transmission (i.e. $R_{0}>1$ ). This threshold is lower under more suitable (generally warmer) climate conditions. The inverse of the critical density threshold, the "transmission potential", is used as a relative measure of transmission intensity under different climatic conditions. The models also assume a threshold level of transmission potential required to sustain transmission, which allows the identification of areas that are climatically suitable for transmission under both observed and projected climate scenarios.

These studies highlight the extreme climate sensitivity of several stages of the malaria transmission cycle. By aggregating these effects into a single measure related to $R_{o}$, they demonstrate that even small temperature increases could potentially cause large relative increases in risk, particularly at the edges of the distribution where temperature may be a limiting factor. They also suggest that those areas climatically suitable for Plasmodium falciparum transmission could expand substantially.

While valid for their original purpose as sensitivity analyses for relative changes in risk, these models are not ideal for defining the most probable changes in either geographical distribution or disease burden within endemic areas. Both outputs require the calculation of absolute rather than relative values of $R_{\mathrm{o}}$, so as to identify areas where $R_{0}>1$, allowing disease transmission to persist. In these incomplete biological models, such calculations are partly dependent on parameter values that are arbitrarily defined in the absence of empirical data (Rogers and Randolph 2000). As the models are based on temperature relationships derived from the laboratory, they also rely on the assumption that meteorological station data accurately represent the climatic conditions that mosquitoes and parasites experience in the field-which disregards the possibility that vectors might exploit microhabitats that are very different from those in meteorological stations. Since the outputs from these models have not been validated against current disease distributions, they were not used in this assessment.

In the absence of data to generate complete biological models of all stages in the transmission cycle, an alternative approach is to use statistical relationships to define only the distributional limits of disease. Although this approach does not allow disaggregation of the specific
mechanisms driving the climate-sensitivity of vector-borne diseases, it is generally considered more objective than the use of incomplete biological models, in that model outputs are not dependent on arbitrarily defined parameter values.

The international MARA (Mapping malaria risk in Africa) collaboration generated a model that used a combination of biological and statistical approaches to define the limits of climate suitability for falciparum malaria in Africa (Craig et al. 1999). Laboratory data on the rate of development of falciparum parasites (Detinova 1962) and laboratory and localized field observations of temperature effects on mosquito survival (Haddow 1943; Jepson et al. 1947; Le Sueur 1991; Maharaj 1995) were used to define upper and lower thresholds for mean monthly temperatures, and winter minima, which would allow both mosquito survival and the completion of the parasite extrinsic incubation period during the lifetime of mosquitoes, thereby permitting transmission. Rainfall thresholds were defined by comparing regions with and without stable malaria transmission, which have similar temperature conditions but different precipitation profiles. In order to take account of uncertainty about the precise values of the upper and lower bounds of temperature and rainfall necessary for transmission, climatic conditions near to the thresholds were not defined as either entirely suitable or unsuitable. Instead, they were assigned a probability of suitability between 0 and 1, defined by a "fuzzy membership curve", which is assumed to follow a pre-specified sigmoidal shape between the plausible values for the upper and lower thresholds for each climate variable (e.g. a decreasing probability of suitability between mean temperatures of 32 and $40^{\circ} \mathrm{C}$ ). These relationships were applied to high-resolution interpolated maps of climate throughout Africa (Hutchinson et al. 1996) to define areas that meet all suitability conditions (i.e. both temperature and rainfall) throughout the continent. For validation, model outputs were visually compared with independent high-resolution maps of the edges of the distribution, based either on field surveys or expert opinion. The model showed a good fit to the observed distributions in both southern Africa and East Africa.

The main advantages of this approach are that the model:

- describes only the cut-offs for any level of transmission, rather than quantitative estimates of transmission risk: it therefore does not rely on arbitrarily defined parameter values to complete the $R_{o}$ equation;
- represents uncertainty around the edges of the distribution;
- allows description of seasonal patterns of transmission, based on the suitability of individual months; and
- most importantly, has been compared with current and historical distribution maps which are apparently independent of the model building process.

The main caveats are:

- the reliance on laboratory data and a small number of field studies to define climate cut-offs;
- apparent subjectivity in at least one parameter estimate (the proportion of mosquitoes that need to survive the sporogonic cycle in order to maintain transmission, which defines the precise value of the lower temperature cut-off);
- the need to make an assumption about the shape of the "fuzzy membership curve"; and
- lack of systematic empirical validation (validation by visual comparison, rather than calculating diagnostic statistics).
While each of the assumptions can legitimately be questioned, the visual validation suggests that the data and assumptions used are at least reasonably accurate. Again, there are further caveats in using the model to try to describe the true global distribution of falciparum malaria (rather than just climatically suitable areas), either now or in the future. To use such models, it is necessary to make the assumptions that distributions vary directly with climate, without any interactive effect of control programmes, or socioeconomic conditions. The comparisons of model outputs with current data suggest that this is a reasonable assumption for most of sub-Saharan Africa, although control programmes have altered distributions in South Africa. The assumption is much less secure for other endemic regions, which are invariably richer.

The relationship between climate variables and the global distribution of malaria can also be defined in statistical terms. Rogers and Randolph (2000) converted WHO maps (WHO 1997) of the limits of reported malaria distribution in the 1990 s into $0.5^{\circ} \times 0.5^{\circ}$ resolution grids, coding each cell as either endemic or non-endemic. These grid maps were overlaid on $0.5^{\circ}$ grid maps describing climate surfaces for the period 1961-1990. A statistical model was generated by randomly selecting a subsample of $50 \%$ of the observations of disease presence or absence. First, grid-cells were assigned by applying $k$-means clustering to six groups based on climatic similarity, thereby allowing for potentially different climate-disease relationships in different ecological zones. Stepwise discriminant analysis was applied to find the combination of temperature, rainfall and humidity parameters that gave the greatest statistical differentiation between positive vs negative grid cells within each cluster. The model was then applied to the remaining $50 \%$ of the observations to assess model accuracy. The fit of the predicted distributions to the observed WHO malaria maps was found to be significantly better than for previous falciparum malaria models: $77.71 \%$ of grid cells correctly predicted for this model vs $75.79 \%$ for Martin and Lefebvre (1995) and $67.26 \%$ for Martens et al. (1999).

The principal advantage of this approach is that it is entirely data driven. The relationships between global climate and malaria distributions are defined using transparent statistical techniques, applied to both disease and climate data from throughout the globe. They therefore do not require either unsupported assumptions of parameter values in order to complete a biological model, or global extrapolation of data from a limited number of observations. In addition, by mapping the observed distribution rather than climatically suitable areas, such analyses do not require the assumption that distributions are defined only by climate.

Despite these advantages, the quality of the data available at the global scale places several limitations on global statistical models. WHO maps of observed distributions are based on a combination of field observations and expert opinion to draw "inclusive" boundaries of the extremes of distributions. In many cases the maps define large areas as endemic (e.g. all of sub-Saharan Africa north of South Africa, Namibia and Botswana), although significant areas are actually disease free, often apparently due to unsuitable climate. WHO maps (and therefore the statistical relationships) also do not differentiate between malaria caused by P. falciparum and Plasmodium vivax, which have quite different sensitivities to temperature (Detinova 1962; MacDonald 1957). It is unclear what effect these two simplifications may have on the climate sensitivity of the model.

For this assessment, projected changes in temperature and rainfall under each of the alternative climate scenarios relative to the baseline (1961-1990) climate were mapped at the resolution of the HadCM2 climate model ( $3.75^{\circ}$ longitude by $2.5^{\circ}$ latitude). Maps of future climate were then generated by adding these values to maps of baseline climate for the 1961-1990 climate at $0.5^{\circ}$ resolution ( $0.05^{\circ}$ resolution for the MARA malaria model in Africa; Hutchinson et al. 1996).

An adapted version of the MARA climate model described in detail above (Tanser et al. 2003) was used to generate the mid-range estimates for this assessment. This decision was based on the independent (though continental rather than global) verification, plus the fact that the model was developed and tested using data from throughout Africa (where the overwhelming majority of the burden of malaria currently occurs, and where accuracy is therefore most important), and the potential for developing projections of increased force of infection and disease incidence in the near future.

The model was applied to the global climate maps for baseline climate, and for the unmitigated emissions scenario for the 2020s and 2050s. Relative risk estimates presented here were the ratios of the projected population at risk (i.e. living in areas climatically suitable for $>1$ month falciparum malaria transmission per year) in each subregion under climate change, relative to the population at risk under the 1961-1990 climate. As an approximation, estimates for the s750 and s550
scenarios were derived proportionately by multiplying the relative risks under unmitigated climate change by the ratio of global temperature change under each scenario/change under unmitigated emissions (Table 20.15). This model gave considerably larger estimates of changes in population at risk than the statistical model of Rogers and Randolph (2000), which predicted approximately no overall change under an unmitigated emissions scenario by the 2050s. In the absence of further comparisons and formal uncertainty assessments, our lower range estimate therefore included the possibility of no change in risk in any subregion. The upper range estimate is a doubling of the mid-range estimate from the MARA model. We emphasize that given the difficulties in validating any specific model and its suitability for extrapolation to other subregions, the choice of MARA climate model over the other possible models was somewhat arbitrary.

In addition, the calculation of disease burdens requires estimates of change in incidence within each subregion, rather than population at risk. In the absence of models for changes in malaria incidence within endemic regions, we therefore made the assumption that relative changes in incidence will vary in direct relation to predicted changes in population at risk-that is, a doubling of the population at risk within a region will lead to a doubling of the clinical disease incidence.

These measures are related in broad terms: countries or regions with higher populations at risk tend to have higher incidence and disease burdens. However, this relationship is a crude generalization, as it assumes that these relationships will remain constant as the population at risk expands or contracts. This may lead to underestimation of effects, if there is an increase in transmission within already at-risk populations, driving up infection incidence. Alternatively, this relationship may overestimate risk, depending on the extent to which increasing vectorial capacity promotes herd immunity (Rogers et al. 2002), and causes first infections to occur earlier in life, when patients suffer less severe clinical symptoms for some diseases, potentially conferring immunity on the more clinically vulnerable older age-groups (Coleman et al. 2001; Snow and Marsh 1995). In addition, socioeconomic conditions and control programmes clearly influence vector-borne diseases. Future changes in these factors are likely to affect (and hopefully reduce) transmission, as they have done in some regions in the past (e.g. Jetten and Takken 1994; Reiter 2001). The role of adaptation was discussed in section 2.5.

By applying relative risks to baseline incidences of zero in some subregions, our assessment did not allow for the spread of disease from endemic subregions to non-endemic subregions. This is a reasonable, but conservative assumption: non-endemic subregions have better developed health systems and a less amenable socioeconomic environment, in addition to usually being cooler. These factors may protect against reestablishment of vector-borne disease transmission (IPCC 2001b; Kuhn et al. 2003), providing they are maintained.
Table 20.15 Central, low and high estimates of the relative risk of falciparum malaria for alternative climate scenarios relative to baseline climate

| Subregion | Climate ${ }^{\text {a }}$ | 2000 |  |  | 2001 |  |  | 2005 |  |  | 2010 |  |  | 2020 |  |  | 2030 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High |
| AFR-D | 2 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.01 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.02 |
|  | 3 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.01 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.03 |
|  | 4 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.05 |
| AFR-E | 2 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.06 | 1.04 | 1.00 | 1.08 | 1.06 | 1.00 | 1.12 | 1.07 | 1.00 | 1.15 |
|  | 3 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.05 | 1.04 | 1.00 | 1.07 | 1.05 | 1.00 | 1.10 | 1.07 | 1.00 | 1.15 | 1.09 | 1.00 | 1.18 |
|  | 4 | 1.04 | 1.00 | 1.08 | 1.04 | 1.00 | 1.09 | 1.06 | 1.00 | 1.12 | 1.08 | 1.00 | 1.16 | 1.12 | 1.00 | 1.23 | 1.14 | 1.00 | 1.28 |
| AMR-A | 2 | 1.08 | 1.00 | 1.15 | 1.08 | 1.00 | 1.17 | 1.11 | 1.00 | 1.23 | 1.15 | 1.00 | 1.30 | 1.23 | 1.00 | 1.46 | 1.27 | 1.00 | 1.53 |
|  | 3 | 1.09 | 1.00 | 1.19 | 1.10 | 1.00 | 1.20 | 1.14 | 1.00 | 1.28 | 1.19 | 1.00 | 1.37 | 1.28 | 1.00 | 1.56 | 1.33 | 1.00 | 1.65 |
|  | 4 | 1.15 | 1.00 | 1.29 | 1.16 | 1.00 | 1.32 | 1.22 | 1.00 | 1.44 | 1.29 | 1.00 | 1.59 | 1.44 | 1.00 | 1.88 | 1.51 | 1.00 | 2.03 |
| AMR-B | 2 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.07 | 1.04 | 1.00 | 1.09 | 1.07 | 1.00 | 1.13 | 1.08 | 1.00 | 1.16 |
|  | 3 | 1.03 | 1.00 | 1.05 | 1.03 | 1.00 | 1.06 | 1.04 | 1.00 | 1.08 | 1.05 | 1.00 | 1.11 | 1.08 | 1.00 | 1.16 | 1.10 | 1.00 | 1.19 |
|  | 4 | 1.04 | 1.00 | 1.09 | 1.05 | 1.00 | 1.09 | 1.06 | 1.00 | 1.13 | 1.09 | 1.00 | 1.17 | 1.13 | 1.00 | 1.26 | 1.15 | 1.00 | 1.30 |
| AMR-D | 2 | 1.01 | 1.00 | 1.02 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.07 | 1.04 | 1.00 | 1.09 |
|  | 3 | 1.01 | 1.00 | 1.03 | 1.02 | 1.00 | 1.03 | 1.02 | 1.00 | 1.04 | 1.03 | 1.00 | 1.06 | 1.04 | 1.00 | 1.08 | 1.05 | 1.00 | 1.10 |
|  | 4 | 1.02 | 1.00 | 1.04 | 1.02 | 1.00 | 1.05 | 1.03 | 1.00 | 1.07 | 1.04 | 1.00 | 1.09 | 1.07 | 1.00 | 1.13 | 1.08 | 1.00 | 1.17 |
| EMR-B | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| EMR-D | 2 | 1.03 | 1.00 | 1.07 | 1.04 | 1.00 | 1.07 | 1.05 | 1.00 | 1.10 | 1.07 | 1.00 | 1.13 | 1.10 | 1.00 | 1.20 | 1.15 | 1.00 | 1.30 |
|  | 3 | 1.04 | 1.00 | 1.08 | 1.04 | 1.00 | 1.09 | 1.06 | 1.00 | 1.12 | 1.08 | 1.00 | 1.16 | 1.12 | 1.00 | 1.24 | 1.19 | 1.00 | 1.37 |
|  | 4 | 1.06 | 1.00 | 1.13 | 1.07 | 1.00 | 1.14 | 1.10 | 1.00 | 1.19 | 1.13 | 1.00 | 1.26 | 1.19 | 1.00 | 1.38 | 1.29 | 1.00 | 1.59 |

Table 20.15 Central, low and high estimates of the relative risk of falciparum malaria for alternative climate scenarios relative to baseline climate (continued)

| Subregion | Climate | 2000 |  |  | 2001 |  |  | 2005 |  |  | 2010 |  |  | 2020 |  |  | 2030 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High | Mid | Low | High |
| EUR-A | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| EUR-B | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| EUR-C | 2 | 1.07 | 1.00 | 1.13 | 1.07 | 1.00 | 1.15 | 1.10 | 1.00 | 1.20 | 1.13 | 1.00 | 1.27 | 1.20 | 1.00 | 1.40 | 1.25 | 1.00 | 1.50 |
|  | 3 | 1.08 | 1.00 | 1.16 | 1.09 | 1.00 | 1.18 | 1.12 | 1.00 | 1.25 | 1.16 | 1.00 | 1.33 | 1.25 | 1.00 | 1.49 | 1.31 | 1.00 | 1.61 |
|  | 4 | 1.13 | 1.00 | 1.26 | 1.14 | 1.00 | 1.29 | 1.19 | 1.00 | 1.39 | 1.26 | 1.00 | 1.52 | 1.39 | 1.00 | 1.78 | 1.48 | 1.00 | 1.97 |
| SEAR-B | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| SEAR-D | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.01 |
|  | 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.01 | 1.01 | 1.00 | 1.01 |
|  | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.01 | 1.01 | 1.00 | 1.01 | 1.01 | 1.00 | 1.02 |
| WPR-A | 2 |  | 1.00 | 1.14 | 1.08 | 1.00 | 1.15 | 1.11 | 1.00 | 1.21 | 1.14 | 1.00 | 1.28 | 1.21 | 1.00 | 1.42 | 1.25 | 1.00 | 1.49 |
|  | 3 | 1.09 | 1.00 | 1.17 | 1.09 | 1.00 | 1.19 | 1.13 | 1.00 | 1.26 | 1.17 | 1.00 | 1.34 | 1.26 | 1.00 | 1.52 | 1.30 | 1.00 | 1.60 |
|  | 4 | 1.14 | 1.00 | 1.27 | 1.15 | 1.00 | 1.30 | 1.20 | 1.00 | 1.41 | 1.27 | 1.00 | 1.54 | 1.41 | 1.00 | 1.81 | 1.48 | 1.00 | 1.95 |
| WPR-B | 2 | 1.06 | 1.00 | 1.12 | 1.07 | 1.00 | 1.14 | 1.09 | 1.00 | 1.18 | 1.12 | 1.00 | 1.25 | 1.18 | 1.00 | 1.37 | 1.22 | 1.00 | 1.43 |
|  | 3 | 1.08 | 1.00 | 1.15 | 1.08 | 1.00 | 1.17 | 1.11 | 1.00 | 1.23 | 1.15 | 1.00 | 1.30 | 1.23 | 1.00 | 1.45 | 1.26 | 1.00 | 1.53 |
|  | 4 | 1.12 | 1.00 | 1.24 | 1.13 | 1.00 | 1.26 | 1.18 | 1.00 | 1.36 | 1.24 | 1.00 | 1.48 | 1.36 | 1.00 | 1.71 | 1.42 | 1.00 | 1.83 |

[^75]
## Dengue

Several studies have explored the relationship between climate and the distribution of intensity of dengue transmission (see Appendix B, Table B.6). Most are derived from a series of biological models that relate climate variables to determinants of the population biology of Aedes vectors (Focks et al. 1993a, 1993b) and dengue transmission (Focks et al. 1995). Adaptations of these models to map climate relationships at the global scale (Jetten and Focks 1997; Martens et al. 1997; Patz et al. 1998) have used field and laboratory data to define the relationships between temperature and the length of the gonotrophic cycle (and the associated feeding frequency), larval weight (and therefore the need to take multiple feeds within a single cycle), and the extrinsic incubation period of dengue virus within the vector. Mosquito survival, human biting habit and the duration of human infectiousness are set as tem-perature-independent constants, with parameter values defined using field data from a number of sites.

As for the biological models of malaria, these relationships are aggregated into a simplified version of the $R_{o}$ equation, excluding measures of vector abundance. This equation is again used to define the critical density threshold (the number of mosquitoes which would be required to maintain $R_{0}>1$ ), and its inverse, "transmission potential" or "transmission intensity". The model has been applied to local climate data in a series of sites, and the rank order of monthly predicted values of transmission potential showed good correspondence with the observed seasonal distribution of dengue cases (e.g. Pearson's $R=0.837$ in San Juan, Puerto Rico).

These models have similar characteristics to the biological models for malaria. By aggregating temperature effects from various stages of the transmission cycle, they demonstrate that overall dengue risk is likely to be extremely sensitive to even small changes in temperature. Sensitivity analyses show that a $2^{\circ} \mathrm{C}$ increase in global temperature would lead to large (commonly 1-5 times) increases in transmission intensity in many regions of the world. However, these models are also subject to the same caveats. As there are no data available on mosquito abundance throughout the globe, or the relationship between vector abundance and climate variables, the transmission potential remains a relative rather than absolute measure of $R_{0}$. It is consequently unsuitable for defining the conditions under which transmission can persist, and therefore mapping the distributional limits of dengue either now or in the future. In addition, although the validation exercise demonstrates that transmission potential is correlated to the incidence of infection and clinical disease, in that months with higher transmission potential have more cases, the quantitative relationships have not been explicitly defined. It is therefore difficult to interpret exactly what effect a doubling of transmission potential (for example) would have on the burden of disease caused by dengue.

Methods similar to those used by Rogers and Randolph (2000) for malaria have recently been applied to define the relationship between multiple climate parameters and the reported geographic distribution of dengue. Hales et al. (2002) used WHO data (WHO 2000a) to map reports of dengue transmission during the period 1975-1996 at the level of the country, or smaller administrative area when this was specified in the report. These were converted into grid maps at $0.5^{\circ}$ resolution. Logistic regression analysis was then used to correlate presence or absence of reported transmission against the average values of various climate parameters (monthly average rainfall, vapour pressure, and maximum, minimum and mean temperature) for the corresponding gridcells throughout 1961-1990. Vapour pressure (approximately equivalent to absolute humidity, and reflecting both temperature and precipitation) gave the best discrimination between areas with and without reported dengue transmission, with no other climate variables adding significant explanatory power. The accuracy of the model was assessed by crossvalidation, repeatedly using $95 \%$ of the data to generate predictions for the remaining $5 \%$. The model gave correct predictions of observed presence or absence for an average of $89-92 \%$ of grid-cells, with the precise accuracy depending on the radius over which vapour pressure was considered. The limitations of this model are similar to the statistical model for malaria. To be consistent with the choice and exclusion of models for malaria, the model of Hales et al. (2002), which is similar to the excluded model of Rogers and Randolph (2000), was not used in the estimates of disease burden.

## Future research

A comprehensive analysis would require the generation of predictive models for other important vector-borne diseases. There is also a need to investigate the effect of the predicted increases in climate variability (e.g. more frequent or more extreme El Niño events), especially on the frequency and intensity of epidemics.

Most importantly, more reliable estimation would require models that cover all stages of the causal chain from climate change to clinical outcome. These should explicitly address the role of socioeconomics and control in determining absolute, rather than relative, measures of $R_{0}$. This would allow better definition both of the geographical limits of transmission, and of variations in the incidence within existing or potentially newly endemic regions. They should also address the role of hostimmunity in protecting individuals and populations from clinical disease, as exposure to infection changes. Such models would represent a considerable advance over the assumption made here, that changes in the proportion of the populations at risk within a subregion will be reflected in proportional changes in the burden of disease.

For the purpose of health assessments, the nature of the model (i.e. whether each of the biological processes are modelled separately and
then aggregated, or whether climate is statistically related directly to empirical measures of disease burden) is probably less important than model accuracy. There is a clear need for greater model validation, by comparison of predictions with geographical and temporal patterns of infection and disease in the present and recent past. The most urgent requirement for all of these objectives is the availability of better quality surveillance data on infection and disease, from a wide variety of geographical and socioeconomic settings.

Finally, the potential effects of future climate change on vector-borne diseases (or other diseases) should not divert attention or resources from current control efforts. On the contrary, they provide the additional argument that disease control now should also reduce vulnerability to climate change in the future.

## 4. Results

Summary measures of the effects of climate change on health are presented in this chapter only for the estimated current effects, using the relative risks obtained by extrapolation of the future predictions, as described in section 2.6. The estimates presented here did not attribute DALYs to cardiovascular deaths due to thermal extremes, and excluded any increase due to dengue (see sections 3.6 and 3.10 ). They should be interpreted with caution as, in contrast to most other risk factors, they relied on modelled rather than directly observed outcomes. They did, however, indicate the estimated distribution of impacts both among geographical regions and among the various causes of disease (Tables 20.16 and 20.17). The models may also be useful for the secondary purpose of indicating the magnitude of health impacts that might already be caused by climate change, but which may not be detected by direct observation using current surveillance systems.

The various causes considered here differed markedly in their contribution to the estimates of the overall burden of disease. In our analysis, climate-change effects on malnutrition, diarrhoea and vector-borne diseases appeared considerably more important than effects on flooding, or on deaths attributable to thermal extremes. It should be noted that, with the exception of malaria, these outcomes are relatively poorly studied in comparison with the direct effects of thermal stress.

The health consequences of climate change are distributed very unevenly among regions. Estimated DALY burdens per capita are several hundred times greater in the poorer regions of Africa, parts of the Eastern Mediterranean region and South-East Asia than in western Europe, North America and the more developed regions of the Western Pacific. This is largely a reflection of the much higher baseline incidence of the most important climate-sensitive diseases (malaria, diarrhoea and malnutrition) in these poorer regions, but also of greater vulnerability to climate change effects. Because these major climate-sensitive diseases

Table 20.16 Estimated mortality (000s) attributable to climate change in the year 2000, by cause and subregion

| Subregion | Malnutrition | Diarrhoea | Malaria | Floods | CVD | All causes | Total deaths/million population |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | 8 | 5 | 5 | 0 | I | 19 | 66.83 |
| AFR-E | 9 | 8 | 18 | 0 | 1 | 36 | 109.40 |
| AMR-A | 0 | 0 | 0 | 0 | 0 | 0 | 0.15 |
| AMR-B | 0 | 0 | 0 | 1 | 1 | 2 | 3.74 |
| AMR-D | 0 | 1 | 0 | 0 | 0 | 1 | 10.28 |
| EMR-B | 0 | 0 | 0 | 0 | 0 | I | 5.65 |
| EMR-D | 9 | 8 | 3 | 1 | 1 | 21 | 61.30 |
| EUR-A | 0 | 0 | 0 | 0 | 0 | 0 | 0.07 |
| EUR-B | 0 | 0 | 0 | 0 | 0 | 0 | 1.04 |
| EUR-C | 0 | 0 | 0 | 0 | 0 | 0 | 0.29 |
| SEAR-B | 0 | I | 0 | 0 | 1 | 2 | 7.91 |
| SEAR-D | 52 | 22 | 0 | 0 | 7 | 80 | 65.79 |
| WPR-A | 0 | 0 | 0 | 0 | 0 | 0 | 0.09 |
| WPR-B | 0 | 2 | 1 | 0 | 0 | 3 | 2.16 |
| World | 77 | 47 | 27 | 2 | 12 | 166 | 27.82 |

CVD Cardiovascular disease. As described in section 3.6, the estimated cardiovascular deaths represent temperature-related mortality displacement. Therefore no disease burden is estimated for deaths from this cause in Table 20.I7.
mainly affect younger age groups, the health burden associated with climate change appears to be borne mainly by children rather than adults.

## 5. Discussion

The collective scientific evidence indicates that anthropogenic climate change has already begun and will continue, with potential consequences for human health. Global warming over the past quarter-century was of the order of half a degree centigrade. Such a gradual change is partly obscured by natural climate variability and affects health through complex causal pathways. These characteristics, coupled with considerably larger effects of other factors in the most vulnerable populations, mean that it is inherently difficult to measure directly net health losses or gains attributable to the climate change that have occurred until now.

However, climate change differs from most other health determinants in that considerable effort has been devoted to generating and evaluating formal models to forecast future climate, in response to likely trajectories of atmospheric gas composition. These models are in general agreement that over the next $50-100$ years global warming will be approximately five times greater than has been experienced in the last

Table 20.17 Estimated disease burden (000s of DALYs) attributable to climate change in the year 2000, by cause and subregion

|  |  |  |  |  |  | Total DALYs/million <br> population |
| :--- | :---: | ---: | :---: | ---: | :---: | ---: |
| Subregion | Malnutrition | Diarrhoea | Malaria | Floods | All causes | 2185.78 |
| AFR-D | 293 | 154 | 178 | 1 | 626 | 3839.58 |
| AFR-E | 323 | 260 | 682 | 3 | 1267 | 11.85 |
| AMR-A | 0 | 0 | 0 | 4 | 4 | 166.62 |
| AMR-B | 0 | 0 | 3 | 67 | 71 | 324.15 |
| AMR-D | 0 | 17 | 0 | 5 | 23 | 147.57 |
| EMR-B | 0 | 14 | 0 | 6 | 20 | 2145.91 |
| EMR-D | 313 | 277 | 112 | 46 | 748 | 6.66 |
| EUR-A | 0 | 0 | 0 | 3 | 3 | 48.13 |
| EUR-B | 0 | 6 | 0 | 4 | 10 | 14.93 |
| EUR-C | 0 | 3 | 0 | 1 | 4 | 117.19 |
| SEAR-B | 0 | 28 | 0 | 6 | 34 | 2080.84 |
| SEAR-D | 1918 | 612 | 0 | 8 | 2538 | 8.69 |
| WPR-A | 0 | 0 | 0 | 1 | 1 | 111.36 |
| WPR-B | 0 | 89 | 43 | 37 | 169 | 925.35 |
| World | 2846 | 1459 | 1018 | 193 | 5517 |  |

25 years, with associated changes in other potentially hazardous climate characteristics, such as the frequency of extreme precipitation events.

Such modelling is at a relatively early stage. Few modelling studies have estimated health effects at the global scale, and not all of these directly estimate incidence or prevalence of GBD outcomes. However, they provide the best current basis for making indicative forecasts in order to inform policy decisions. These models nevertheless make only crude adjustments for the effects of other variables (such as decreasing poverty), which may both determine the vulnerability of populations to potential health effects of climate change, and exert much larger independent effects on health. Taking each disease in turn:

1. We estimated a small proportional decrease in cardiovascular and respiratory disease mortality attributable to climate extremes in tropical regions and a slightly larger benefit in temperate regions, caused by warmer winter temperatures. Although these proportional changes are modest, they apply to significant causes of death. Uncertainties around these estimates are largely due to lack of knowledge of the degree to which populations physiologically and behaviourally adapt to increasing temperatures.
2. The relative risk for diarrhoea in 2030 in developing regions was estimated to be between 1 and 1.1 under unmitigated emissions compared with baseline climate. Richer countries (GDP > US\$ 6000/year),
either now or in the future, were assumed to suffer little or no additional risk of diarrhoea. Again, these small changes in relative risk relate to a major cause of ill-health. Uncertainties were mainly due to poor characterization of variations in the relationship between climate and diarrhoea in more or less developed regions, which have different balances between pathogens preferring higher or lower temperatures.
3. Estimated effects on malnutrition varied markedly across subregions. By 2030, the relative risks for unmitigated emissions relative to no climate change varied from a large increase ( $\mathrm{RR}=1-1.33$ ) in SEAR-D to a small decrease ( $\mathrm{RR}=1-0.99$ ) in WPR-B. Developed countries were assumed to be immune to climate change effects on malnutrition. There was no consistent pattern of reduction in relative risk with intermediate levels of climate change stabilization. Apparent inconsistency in the estimates may be due to the high sensitivity of the models to regional variations in precipitation, for which future projections are much more uncertain than for temperature. Although these estimates are somewhat unstable, they are relatively large, and again relate to a major disease burden.
4. We estimated much larger proportional changes in the numbers of people killed in coastal floods (RR of up to 6.3 in EUR-B for unmitigated emissions compared to baseline conditions in 2030), but applied to a very low burden of disease. Consequences of inland floods were predicted to increase by a similar order of magnitude $(R R=$ $1-18.5$ in AMR-A), and are generally more common. In contrast to most other outcomes, the increase in relative risk tended to be at least as high in developed as developing subregions. However, these apply to baseline rates that are much higher in developing than developed countries. Both estimates are subject to uncertainty around the likely effectiveness of adaptation measures. Inland floods are subject to additional large uncertainties around the quantitative relationships between changes in the intra-annual variation in precipitation (on which our model is based), the magnitude and geographical distribution of extreme precipitation events, and in turn the frequency of flooding and its health consequences. The suggestion of a trend towards decreasing incidence with increasing GHG emissions in some regions is probably due to the uncertainties in predicting precipitation trends. As projections for precipitation are less secure than for temperature, mid-range estimates and uncertainties around the effects of inland floods could be much better described using multiple climate models, rather than the single model used in this assessment.
5. We estimated relatively large changes in the relative risk of falciparum malaria in countries at the edge of the current distribution, for example, increases in relative risk of falciparum malaria of between

1 and 1.83 in WPR-B by 2030. Relative changes were much smaller in areas that are already highly endemic for these diseases. The principal uncertainties specific to these estimates related to the reliability of extrapolations made between subregions, the relationship between changes in the population at risk of these diseases, and incidence (and therefore disease burden), and over the degree to which changes in the non-climatic influences on vector-borne diseases could affect not only the baseline rates of disease, but also interact with climate change to affect the relative risks.

The estimates for the year 2000 (Tables 20.16 and 20.17) suggest that, according to models summarizing our current knowledge of the relationships between climate and health, past climate change may already be causing some health consequences.

These relative risk estimates are much greater for projections into the future, as climate change continues. Both current and future estimates show extreme variations in the estimated effects among geographical regions. Negative consequences are overwhelmingly concentrated in the developing regions of the world (particularly in Africa and the poorer regions of the Eastern Mediterranean and South-East Asia). This is partly a function of variation in baseline climate (hot regions suffer more from increases in temperature), but more importantly due to population vulnerability (e.g. developed countries are assumed to be completely immune from some diseases, such as malnutrition).

Global models have not yet addressed all of the likely effects of climate change on health. The potential omissions are many infectious diseases, the health consequences of drought and famine (beyond those included in current estimates of malnutrition), population displacement, destruction of health infrastructure in natural disasters, increased pollution and aeroallergen levels, effects of plant pests and diseases on agriculture, and risk of conflict over declining natural resources. It is likely that these health consequences will be larger than those estimated in this chapter.

Although incomplete and encompassing a wide uncertainty range, the results of these analyses suggested that the attributable burden of climate change is likely to be significant, even under the relatively short (in climatological terms) time-scale considered for the comparative risk assessment (CRA). The effect of plausible reductions in climate change was estimated to be relatively small over this assessment period. However, the health gains would clearly be much greater over longer time periods. Given the long time-lag and apparent irreversibility of climate change, early mitigation should therefore result in long-term health benefits. Our results therefore indicate the urgent need for: (i) consideration of optimal policies to reduce climate change; (ii) strengthening of current actions to control climate-sensitive diseases, both as ends in themselves, and as adaptation to future climate change; and (iii) con-
tinued research to revise and narrow the uncertainty range around the estimates of disease burdens.

This assessment has highlighted the most important gaps in data and understanding that should be addressed for the next global burden of disease assessment. Marked improvements in our assessment of this risk factor would come from:

- the use of multiple climate models;
- climate-health relationships derived from a greater range of climatic and socioeconomic environments;
- more explicit and routine validation of the accuracy of disease models in the present or recent past;
- formal analyses to aggregate uncertainty arising from multiple causes (i.e. GHG emissions scenarios, climate models, climate-health relationships, and effect modifiers);
- efforts to formally model climate change effects through to disease burden, rather than intermediate indicators such as population at risk;
- a greater emphasis on investigating the consequences of increased climate variability, rather than gradual changes in mean conditions; and
- the development of analytical tools to assess outcomes acting through more complex causal mechanisms.


## Acknowledgements

To Matthew Livermore, Tim Mitchell, David Viner, Pim Martens for provision of data and advice; Richard Tol for provision of unpublished data; Debarati Guha-Sapir for advice on EM-DAT disaster data; Clive Davies, Paul Coleman, Sandy Cairncross, Bo Draser, Alessandro Loretti, Charles Delacollette, Monika Blössner, Mercedes de Onis, Rosalie Woodruff, Will Checkley; and nine referees, for helpful review comments.

## Steering committee

Carlos Corvalan, Annette Prüss-Ustün, Roberto Bertollini, Bettina Menne, Andy Haines and Alistair Woodward.

## Note

1 See preface for an explanation of this term.

## References

Alexander D (1993) Natural disasters. University College London Press, London.
Anderson RM, May RM (1991) Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford.

Baker RHA, Sansford CE, Jarvis CH, Cannon RJC, MacLeod A, Walters KFA (2000) The role of climatic mapping in predicting the potential geographical distribution of non-indigenous pests under current and future climates. Agriculture Ecosystems and Environment, 82:57-71.
Barer D, Ebrahim S, Smith C (1984) Factors affecting day to day incidence of stroke in Nottingham. British Medical Journal, 289:662.
Beinin C (1981) An examination of health data following two major earthquakes in Russia. Disasters, 5:142-146.
Bentham G (1997) Health. In: Economic impacts of the hot summer and unusually warm year of 1995. Palutikof JP, Subak S, Agnew MD, eds. University of East Anglia, Norwich.
Black RE, Lanata CF (1995) Epidemiology of diarrhoeal diseases in developing countries. In: Infections of the gastrointestinal tract. Blaser MJ, Smith PD, Ravdin JI, Greenberg HP, Guerrant RI, eds. Raven Press, New York.

BOM (2001) Climate zones for urban design. Available at http://www.bom.gov.au/climate/environ/design/climzone.shtml.
Bos ET, Vu MT, Massiah E, Bulatao RA (1994) World population projections 1994-1995: estimates and projections with related demographic statistics. World Bank, Johns Hopkins University Press, Baltimore, MD.
Bouma MJ, Dye C, van der Kaay HJ (1996) Falciparum malaria and climate change in the Northwest Frontier Province of Pakistan. American Journal of Tropical Medicine and Hygiene, 55:131-137.
Bouma MJ, Kovats RS, Goubet SA, Cox JSH, Haines A (1997a) Global assessment of El Niño's disaster burden. The Lancet, 350:1435-1438.

Bouma MJ, Poveda G, Rojas W et al. (1997b) Predicting high-risk years for malaria in Colombia using parameters of El Niño southern oscillation. Tropical Medicine and International Health, 2:1122-1127.

Braga A, Zanobetti A, Schwartz J (2001) A tridimensional estimate of the lag structure between temperature and total daily deaths in 14 US cities. Epidemiology, 12:180.
Broecker WS (1997) Thermohaline circulation, the Achilles heel of our climate system: will man-made $\mathrm{CO}_{2}$ upset the current balance? Science, 278: 1582-1588.
Campbell-Lendrum DH, Wilkinson P, Kuhn K et al. (2002) Monitoring the health impacts of global climate change. In: Environmental change, climate and health: issues and research methods, Martens P, McMichael AJ, eds. Cambridge University Press, Cambridge.

Carcavallo RU, Curto de Casas SI (1996) Some health effects of global warming in South America: vector-borne diseases. Journal of Epidemiology, 6:S153-S157.
Celli A (1933) A bistory of malaria in the Roman Campagna from ancient times. AMS Press Inc., New York.
Chaudhury A, Nath G, Shukla B, Panda S, Singh TB (1996) Diarrhoea associated with Candida spp.: incidence and seasonal variation. Journal of Diarrhoeal Diseases Research, 14:110-112.

Checkley W, Epstein LD, Gilman RH et al. (2000) Effect of El Niño and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. The Lancet, 355:442-450.

Christophers RS (1911) Malaria in the Punjab: scientific memoirs by officers of the Medical and Sanitary Department of the Government of India. (New Series, No. 46.) Medical and Sanitary Department of the Government of India. Government Press, Calcutta.
Coleman PG, Perry BD, Woolhouse MEJ (2001) Endemic stability—a veterinary idea applied to human public health. The Lancet, 357:1284-1286.
Cook SM, Glass RI, LeBaron CW, Ho MS (1990) Global seasonality of rotavirus infections. Bulletin of the World Health Organization, 68:171-177.

Craig MH, Snow RW, Le Sueur D (1999) A climate-based distribution model of malaria transmission in sub-Saharan Africa. Parasitology Today, 15:105-111.

Curriero FC, Patz JA, Rose JB, Lele S (2001) The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948-1994. American Journal of Public Health, 91:1194-1199.
Curwen M, Devis T (1988) Winter mortality, temperature and influenza: has the relationship changed in recent years? Population Trends, 54:17-20.
Detinova TS (1962) Age grouping methods in diptera of medical importance, with special reference to some vectors of malaria. World Health Organization, Geneva.
Drasar BS, Tomkins AM, Feacham RG (1978) Seasonal aspects of diarrhoeal disease. Seasonal dimensions to rural poverty. (Report to UK Overseas Development Association.) London School of Hygiene and Tropical Medicine, London.

Drinkwater BL, Horvath SM (1979) Heat tolerance and aging. Medicine and Science in Sports and Exercise 11:49-55.

Dye C (1992) The analysis of parasite transmission by bloodsucking insects. Annual Review of Entomology, 37:1-19.
Dyson T (1999) Prospects for feeding the world. British Medical Journal, 319:988-990.

Eberhard ML, Nace EK, Freeman AR, Streit TG, Da Silva AJ, Lammie PJ (1999) Cyclospora cayetanensis infections in Haiti: a common occurrence in the absence of watery diarrhea. American Journal of Tropical Medicine and Hygiene, 60:584-586.
Ebi KL, Exuzides KA, Lau E, Kelsh M, Barsnton A (2001) Association of normal weather conditions and El Niño events with viral pneumonia hospitalizations in females, California, 1983-1998. American Journal of Public Health, 91:1200-1208.

EM-DAT (2002) The OFDA/CRED international disaster database. Université Catholique de Louvain, Brussels.

Energy Modeling Forum (1995) Second round study design for EMF14. (EMF Working Paper No. 14.1.) Energy Modeling Forum, Stanford, CA.
Epstein PR (1999) Climate and health. Science, 285:347-348.

FAO (1987) Fifth world food survey. Food and Agriculture Organization of the United Nations, Rome.

Fischer G, Frohberg K, Keyzer MA, Parikh KS (1988) Linked national models. A tool for international food policy analysis. Kluwer, Dordrecht.

Fitzpatrick FA, Knox JC (2000) Spatial and temporal sensitivity of hydrogeomorphic response and recovery to deforestation, agriculture, and floods. Physical Geography, 21:89-108.
Fleming DM, Norbury CA, Crombie DL (1991) Annual and seasonal variation in the incidence of common diseases. The Royal College of General Practitioners, London.
Focks DA, Daniels E, Haile DG, Keesling JE (1995) A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. American Journal of Tropical Medicine and Hygiene, 53:489-506.
Focks DA, Haile DG, Daniels E, Mount GA (1993a) Dynamic life table model for Aedes aegypti (diptera: culicidae) analysis of the literature and model development. Journal of Medical Entomology, 30:1003-1017.
Focks DA, Haile DG, Daniels E, Mount GA (1993b) Dynamic life table model for Aedes aegypti (diptera: culicidae) simulation results and validation. Journal of Medical Entomology, 30:1018-1028.
Garrett-Jones C (1964) Prognosis for interruption of malaria transmission through assessment of the mosquito's vectorial capacity. Nature, 204: 1173-1175.
Goodess C, Hulme M, Osborn T (2001) The identification and evaluation of suitable scenario development methods of the estimation of future probabilities of extreme weather events. School of Environmental Sciences, University of Norwich, Norwich.
Guerrant RL, Kosek M, Lima AAM, Lorntz B, Guyatt HL (2002) Updating the DALYs for diarrhoeal disease. Trends in Parasitology, 18:191-193.
Guest CS, Willson K, Woodward AJ et al. (1999) Climate and mortality in Australia. Retrospective study, 1979-1990, and predicted impacts in five major cities in 2030. Climate Research, 13:1-15.
Hadley Centre (1999) Climate change and its impacts: stabilisation of $\mathrm{CO}_{2}$ in the atmosphere. Hadley Centre, Reading.
Haddow AJ (1943) Measurements of temperature and light in artificial pools with reference to the larval habitat of Anopheles (Myzomyia) gambiae Giles and A. (M.) funestus Giles. Bulletin of Entomological Research, 34:89.

Hajat S, Haines A, Atkinson RW, Bremner SA, Anderson HR, Emberlin J (2001) Association between air pollution and daily consultations with general practitioners for allergic rhinitis in London, United Kingdom. American Journal of Epidemiology, 153:704-714.
Hales S, de Wet N, Maindonald J, Woodward A (2002) Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. The Lancet, 360:830-834.

Halstead SB (1988) Pathogenesis of dengue: challenges to molecular biology. Science, 239:476-481.

Hay SI, Cox J, Rogers DJ et al. (2002) Climate change and the resurgence of malaria in the East African highlands. Nature, 415:905-909.

Hay SI, Myers MF, Burke DS et al. (2000) Etiology of interepidemic periods of mosquito-borne disease. Proceedings of the National Academy of Sciences of the United States of America, 97:9335-9339.
Hoozemans FMJ, Hulsburgen CH (1995) Sea-level rise: a worldwide assessment of risk and protection costs. In: Climate change: impact on coastal habitation. Eisma D, ed. Lewis Publishers, London.
Hoozemans FMJ, Marchand M, Pennekamp HA (1993) A global vulnerability analysis: vulnerability assessment for population, coastal wetlands and rice production on a global scale. 2nd edn. Delft Hydraulics, Delft.
Hutchinson MF, Nix HA, McMahon JP, Ord KD (1996) The development of a topographic and climate database for Africa. In: Proceedings of the third international conferencelworkshop on integrating GIS and environmental modeling, NCGIA, Santa Barbara, CA.
IBSNAT (1989) International benchmark sites network for agrotechnology transfer. Decision Support System for Agrotechnology Transfer Version 2.1 (DSSAT V2.1), Department of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, University of Hawaii, HI.
IPCC (1996) Climate change 1995: impacts, adaptations and mitigation of climate change: contribution of Working Group II to the second assessment report of the Intergovernmental Panel on Climate Change. Watson RT, Zinyowera MC, Moss RH, eds. Cambridge University Press, Cambridge.
IPCC (2001a) Climate change 2001: impacts, adaptation and vulnerability: contribution of Working Group II to the third assessment report of the Intergovernmental Panel on Climate Change. McCarthy JJ, Canziani OF, Leary NA, Dokken DJ, White KS, eds. Cambridge University Press, Cambridge.
IPCC (2001b) Climate change 2001: the scientific basis: contribution of Working Group I to the third assessment report of the Intergovernmental Panel on Climate Change. Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Xiaosu D, eds. Cambridge University Press, Cambridge.

Jepson WF, Moutia A, Courtois C (1947) The malaria problem in Mauritius: the bionomics of Mauritian anophelines. Bulletin of the Entomological Research, 38:177-208.
Jetten TH, Focks DA (1997) Potential changes in the distribution of dengue transmission under climate warming. American Journal of Tropical Medicine and Hygiene, 57:285-287.
Jetten TH, Martens WJM, Takken W (1996) Model simulations to estimate malaria risk under climate change. Journal of Medical Entomology, 33:361371.

Jetten TH, Takken W (1994) Anophelism without malaria in Europe: a review of the ecology and distribution of the genus Anopheles in Europe. Wageningen Agricultural University Press, Wageningen.

Johns TC, Gregory JM, Stott PA, Mitchell JFB (2001) Correlations between patterns of 19th and 20th century surface temperature change and HadCM2 climate model ensembles. Geophysical Research Letters, 28:1007-1010.

Jovel JR (1989) Natural disasters and their economic and social impact. Economic Commissions, Latin America and the Caribbean, Santiago de Chile.
Kalkstein LS, Greene JS (1997) An evaluation of climate/mortality relationships in large US cities and the possible impacts of a climate change. Environmental Health Perspectives, 105:84-93.
Karl TR, Knight RW (1998) Secular trends of precipitation amount, frequency, and intensity in the United States. Bulletin of the American Meteorological Society, 79:231-241.

Katz RW (2002) Techniques for estimating uncertainty in climate change scenarios and impact studies. Climate Research, 20:167-185.

Katz RW, Brown BG (1992) Extreme events in a changing climate: variability is more important than averages. Climatic Change, 21:289-302.
Keatinge WR, Coleshaw SRK, Easton JC, Cotter F, Mattock MB, Chelliah R (1986a) Increased platelet and red-cell counts, blood-viscosity, and plasmacholesterol levels during heat-stress, and mortality from coronary and cerebral thrombosis. American Journal of Medicine, 81:795-800.
Keatinge WR, Coleshaw SRK, Holmes J (1989) Changes in seasonal mortalities with improvement in home heating in England and Wales from 1964 to 1984. International Journal of Biometeorology, 33:71-76.
Keatinge WR, Coleshaw SRK, Millard CE, Axelsson J (1986b) Exceptional case of survival in cold water. British Medical Journal, 292:171-172.
Khaw KT (1995) Temperature and cardiovascular mortality. The Lancet 345:337-338.

Kilbourne EM (1989) Heat waves. In: The public health consequences of disasters. Gregg MB, ed. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Atlanta, GA.
Kilbourne EM (1992) Illness due to thermal extremes. In: Public health and preventative medicine. Last JM, Wallace RB, eds. Appleton Lang, Norwalk, CT.
Knutti R, Stocker TF, Joos F, Plattner G-K (2002) Constraints on radiative forcing and future climate change from observations and climate model ensembles. Nature, 416:719-723.
Konno T, Suzuki H, Katsushima N et al. (1983) Influence of temperature and relative humidity on human rotavirus infection in Japan. Journal of Infectious Diseases, 147:125-128.
Kovats RS, Bouma MJ, Haines A (1999) El Niño and health, World Health Organization, Geneva.
Kovats RS, Campbell-Lendrum DH, McMichael AJ, Woodward A, Cox JS (2001) Early effects of climate change: do they include changes in vectorborne disease? Philosophical Transactions of the Royal Society, Series B, 356:1057-1068.

Kovats RS, Campbell-Lendrum DH, Reid C, Martens P (2000a) Climate and vector-borne disease: an assessment of the role of climate in changing disease patterns. ICIS/LSHTM/UNEP, Maastricht.
Kovats RS, Menne B, McMichael AJ, Corvalán C, Bertollini R (2000b) Climate change and human health: impact and adaptation. (WHO Technical Report WHO/SDE/OEH/00.4.) World Health Organization, Geneva.
Kuhn KG, Campbell-Lendrum DH, Armstrong B, Davies CR (2003) Malaria in Britain: past, present and future. Proceedings of the National Academy of Sciences, USA, 100:9997-10001.
Kundzewicz ZW, Kaczmarek Z (2000) Coping with hydrological extremes. Water International, 25:66-75.

Kunst A, Looman C, Mackenbach J (1993) Outdoor air temperature and mortality in the Netherlands-a time series analysis. American Journal of Epidemiology, 137:331-341.

Kunzli N, Kaiser R, Medina S et al. (2000) Public-health impact of outdoor and traffic-related air pollution: a European assessment. The Lancet, 356: 795-801.

Le Sueur D (1991) The ecology, over-wintering and population dynamics of the Anopheles gambiae (diptera: culicidae) complex in northern Natal, South Africa. [Dissertation.] University of Natal, Pietermaritzburg.
Lindgren E, Talleklint L, Polfeldt T (2000) Impact of climatic change on the northern latitude limit and population density of the disease-transmitting European tick Ixodes ricinus. Environmental Health Perspectives, 108: 119-123.

Loevinsohn ME (1994) Climatic warming and increased malaria incidence in Rwanda. The Lancet, 343:714-718.

Longstreth J (1999) Public health consequences of global climate change in the United States-some regions may suffer disproportionately. Environmental Health Perspectives, 107:169-179.
MacDonald G (1957) The epidemiology and control of malaria. Oxford University Press, London.
Mackenbach JP, Borst V, Schols JM (1997) Heat-related mortality among nursing home patients. The Lancet, 349:1297-1298.
Maharaj R (1995) Effects of temperature on members of the Anopheles gambiae complex (diptera: culicidae) in South Africa-implications for malaria transmission and control. [Dissertation.] University of Natal, Pietermaritzburg.
Martens P, Kovats RS, Nijhof S et al. (1999) Climate change and future populations at risk of malaria. Global Environmental Change-Human and Policy Dimensions, 9:S89-S107.

Martens P, McMichael AJ (2002) Environmental change, climate and health. Cambridge University Press, Cambridge.
Martens W (1998a) Climate change, thermal stress and mortality changes. Social Science Medicine, 46:331-334.

Martens WJ, Jetten TH, Rotmans J, Niessen LW (1995a) Climate change and vector-borne diseases: a global modelling perspective. Global Environmental Change, 5:195-209.
Martens WJ, Niessen LW, Rotmans J, Jetten TH, McMichael AJ (1995b) Potential impact of global climate change on malaria risk. Environmental Health Perspectives, 103:458-464.
Martens WJM (1998b) Health impacts of climate change and ozone depletion: an ecoepidemiologic modeling approach. Environmental Health Perspectives, 106:241-251.
Martens WJM, Jetten TH, Focks DA (1997) Sensitivity of malaria, schistosomiasis and dengue to global warming. Climatic Change, 35:145-156.
Martin PH, Lefebvre MG (1995) Malaria and climate—sensitivity of malaria potential transmission to climate. Ambio, 24:200-207.
Mason SJ, Waylen PR, Mimmack GM, Rajaratnam B, Harrison JM (1999) Changes in extreme rainfall events in South Africa. Climatic Change, 41:249-257.

Massad E, Forattini OP (1998) Modelling the temperature sensitivity of some physiological parameters of epidemiologic significance. Ecosystem Health, 4:119-129.
McGregor GR, Walters S, Wordley J (1999) Daily hospital respiratory admissions and winter air mass types, Birmingham, UK. International Journal of Biometeorology, 43:21-30.
McMichael AJ, Githeko A (2001) Human health. In: Climate change 2001: impacts, adaptation and vulnerability. McCathy JJ, Canziani OF, Leary NA, Dokken DJ, White KS, eds. Cambridge University Press, Cambridge.
McMichael AJ, Haines A, Slooff R, Kovats RS (1996) Climate change and buman health: an assessment by a task group on behalf of the World Health Organization, the World Meteorological Organization and the United Nations Environment Programme. (WHO/EHG/96.7.) World Health Organization, Geneva.
McMichael AJ, Kovats RS (2000) Strategies for assessing health impacts of global environmental change. In: Implementing ecological integrity: restoring regional and global environmental and human health. Westra L, ed. Kluwer Academic Publishers, Dordrecht.
McMichael AJ, Kovats RS, Cawthorne A et al. (2000a) Fast-track: climate change and human health. Department of Environment, Transport and the Regions Global Atmosphere Division, London.
McMichael AJ, Campbell-Lendrum DH, Corvalán CF et al. eds. (2003) Climate change and human health: risks and responses. World Health Organization, Geneva.
Menne B, Pond K, Noji EK, Bertollini R (1999) Floods and public health consequences, prevention and control measures. WHO European Centre for Environment and Health, Rome.
Milly PCD, Wetherald RT, Dunne KA, Delworth TL (2002) Increasing risk of great floods in a changing climate. Nature, 415:514-517.

Mouchet J (1998) Malaria epidemics on the Highlands of Madagascar and of East and South Africa. Bulletin de la Societe de Pathologie Exotique, 91:64-66.

Mouchet J, Manguin S (1999) Global warming and malaria expansion. Annales de la Societe Entomologique de France, 35:549-555.
Murray CJL, Lopez AD, eds. (1996) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, Vol. 1. Harvard School of Public Health on behalf of the WHO, Cambridge, MA.

National Research Council (2001a) Abrupt climate change: inevitable surprises. National Academy Press, Washington, DC.

National Research Council (2001b) Under the weather: climate, ecosystems, and infectious disease. National Academy Press, Washington, DC.
New M, Hulme M, Jones P (1999) Representing twentieth-century space-time climate variability. Part I. Development of a 1961-90 mean monthly terrestrial climatology. Journal of Climate, 12:829-856.
Nicholls RJ, Hoozemans FMJ, Marchand M (1999) Increasing flood risk and wetland losses due to global sea-level rise: regional and global analyses. Global Environmental Change-Human and Policy Dimensions, 9:S69-S87.
Noji EK (1997) The public health consequences of disasters. Oxford University Press. New York.

Otter-Nacke S, Godwin DC, Ritchie JT (1986) Testing and validating the CERES-wheat model in diverse environments. (AGGRISTARS YM-1500407). Johnson Space Center, Houston, TX.

Palmer TN, Ralsanen J (2002) Quantifying the risk of extreme seasonal precipitation events in a changing climate. Nature, 415:512-514.

Parry M, Rosenzweig C, Iglesias A, Fischer G, Livermore M. (1999) Climate change and world food security: a new assessment. Global Environmental Change-Human and Policy Dimensions, 9:S51-S67.
Patz JA, Martens WJM, Focks DA, Jetten TH (1998) Dengue fever epidemic potential as projected by general circulation models of global climate change. Environmental Health Perspectives, 106:147-153.
Patz JA, McGeehin MA, Bernard SM et al. (2000) The potential health impacts of climate variability and change for the United States: executive summary of the report of the health sector of the US National Assessment. Environmental Health Perspectives, 108:367-376.

Phifer JF (1990) Psychological distress and somatic symptoms after natural disaster: differential vulnerability among older adults. Psychology and Aging, 5:412-420.

Pielke RA (1999) Nine fallacies of floods. Climatic Change, 42:413-438.
Purohit SG, Kelkar SD, Simha V (1998) Time series analysis of patients with rotavirus diarrhoea in Pune, India. Journal of Diarrhoeal Diseases Research, 16:74-83.

Randolph S (2001) Tick-borne encephalitis in Europe. The Lancet, 358: 1731-1732.

Randolph SE, Green RM, Peacey MF, Rogers DJ (2000) Seasonal synchrony: the key to tick-borne encephalitis foci identified by satellite data. Parasitology, 121:15-23.
Randolph SE, Rogers DJ (2000) Fragile transmission cycles of tick-borne encephalitis virus may be disrupted by predicted climate change. Proceedings of the Royal Society of London Series B, 267:1741-1744.
Reiter P (2001) Climate change and mosquito-borne disease. Environmental Health Perspectives, 109:141-161.
Reiter P et al. (2003) Texas lifestyle limits transmission of dengue virus. Emerging Infectious Diseases, 9:86-89.
Rivers JPW (1982) Women and children last: an essay on sex discrimination in disasters. Disasters, 6:256-267.
Robins-Browne RM (1984) Seasonal and racial incidence of infantile gastroenteritis in South Africa. American Journal of Epidemiology, 119:350355.

Rogers DJ, Packer MJ (1993) Vector-borne diseases, models and global change. The Lancet, 342:1282-1284.
Rogers DJ, Randolph SE (2000) The global spread of malaria in a future, warmer world. Science, 289:1763-1766.
Rogers DJ, Randolph SE, Snow RW, Hay SI (2002) Satellite imagery in the study and forecast of malaria. Nature, 415:710-715.
Rosenzweig C, Iglesias A, Yang XB, Epstein PR, Chivian E (2001) Climate change and extreme weather events: implications for food production, plant diseases, and pests. Global Change and Human Health, 2:90-104.
Rosenzweig C, Parry ML (1994) Potential impact of climate-change on world food-supply. Nature, 367:133-138.
Rosenzweig C, Parry ML, Fischer G, Frohberg K (1993) Climate change and world food supply. Environmental Change Unit, University of Oxford, Oxford.

Rothwell PM, Wroe SJ, Slattery J, Warlow CP (1996) Is stroke incidence related to season or temperature? The Lancet, 347:934-936.
Schlesinger ME, Williams LJ (1997) COSMIC—COuntry Specific Model for Intertemporal Climate, computer software. Electric Power Research Institute, Palo Alto, CA.
Schneider SH (2002) Can we estimate the likelihood of climatic changes at 2100? Climatic Change, 52:441-451.
Schwartz J, Samet J, Patz J (2001) The effects of temperature and humidity on hospital admissions for heart disease. Epidemiology, 12:61.
Schwartz J, Spix C, Touloumi G et al. (1996) Methodological issues in studies of air pollution and daily counts of death or hospital admissions. Journal of Epidemiology and Community Health, 50:S3-11.

Shindell DT, Rind D, Lonergan P (1998) Increased polar stratospheric ozone losses and delayed eventual recovery owing to increasing greenhouse-gas concentrations. Nature, 392:589-592.

Singh RB, Hales S, de Wet N, Raj R, Hearnden M, Weinstein P (2001) The influence of climate variation and change on diarrheal disease in the Pacific Islands. Environmental Health Perspectives, 109:155-159.
Snow RW, Marsh K (1995) Will reducing Plasmodium falciparum transmission alter malaria mortality among African children? Parasitology Today, 11: 188-190.

Stott PA, Kettleborough JA (2002) Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. Nature, 416:723726.

Sutherst RW (1998) Implications of global change and climate variability for vectorborne diseases: generic approaches to impact assessments. International Journal for Parasitology, 28:935-945.
Sutherst RW, Yonow T, Chakraborty S, O’Donnell C, White N (1996) A generic approach to defining impacts of climate change on pests, weeds and diseases in Australia. In: Greenhouse: coping with climatic change. Bouma WJ, Pearman GJ, Manning MR, eds. CSIRO, Melbourne.
Tanser FC, Sharp B, Le Sueur D (2003) Potential effect of climate change on malaria transmission in Africa. The Lancet, 362:1792-1798.

Tett SFB, Johns TC, Mitchell JFB (1997) Global and regional variability in a coupled AOGCM. Climate Dynamics, 13:303-323.

Tol RSJ, Dowlatabadi H (2001) Vector-borne diseases, development and climate change. Integrated Assessment, 2:173-181.

UNDMTP (1990) DMTP training modules, United Nations Disaster Management Training Programme. Available at http://www.undmtp.org/.

Vassallo M, Navarro K, Allen S (1995) Factors associated with high risk of marginal hyperthermia in elderly patients living in an institution. Postgraduate Medical Journal, 71:213-216.
Walther GR, Post E, Convey P et al. (2002) Ecological responses to recent climate change. Nature, 416:389-395.
Wang XD, Smith KR (1999) Secondary benefits of greenhouse gas control: health impacts in China. Environmental Science and Technology, 33:3056-3061.
Waterlow J, Armstrong DG, Fowden L, Riley R (1998) Feeding a world population of more than eight billion people. Oxford University Press. Oxford.

Wayne P, Foster S, Connolly J et al. (2002) Production of allergenic pollen by ragweed (Ambrosia artemisiifolia L.) is increased in $\mathrm{CO}_{2}$-enriched atmospheres. Annals of Allergy, Asthma and Immunology, 88:279-282.

Wheeler JG, Sethi D, Cowden JM et al. (1999) Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. British Medical Journal, 318: 1046-1050.

Whitman S, Good G, Donoghue ER, Benbow N, Shou W, Mou S (1997) Mortality in Chicago attributed to the July 1995 heat wave. American Journal of Public Health, 87:1515-1518.

WHO (1992) Psychological consequences of disasters. World Health Organization, Geneva.
WHO (2000a) WHO report on global surveillance of epidemic-prone infectious diseases. (WHO/CDS/CSR/ISR/2000.1.) Anker M, Schaaf D, eds. World Health Organization, Geneva.
WHO (2000b) The world health report 2000. World Health Organization, Geneva.
WHO (2002) Global database on child growth and malnutrition. Available at http://www.who.int/nutgrowthdb/.
WHO/EHA (1998) Emergency health training programme for Africa. Panafrican Emergency Training Centre, Addis Ababa.

WHO/EHA (2002) Technical hazard sheet—drought. World Health Organization, Department of Emergency and Humanitarian Action, Geneva.
WHO/PTC (1995) Drought: preparedness and response in the bealth sector. World Health Organization/Panafrican Emergency Training Centre, Addis Ababa.

Woodhouse PR, Khaw KT, Plummer M (1993) Seasonal variation of blood pressure and its relationship to ambient temperature in an elderly population. Journal of Hypertension, 11:1267-1274.
Woodhouse PR, Khaw KT, Plummer M, Foley A, Meade TW (1994) Seasonal variations of plasma fibrinogen and factor VII activity in the elderly: winter infections and death from cardiovascular disease. The Lancet, 343:435-439.
Woodward A, Hales S, Weinstein P (1998) Climate change and human health in the Asia Pacific region: who will be most vulnerable? Climate Research, 11:31-38.

World Bank (1994) World population projections 1994-1995. Johns Hopkins University Press, Baltimore, MD.

World Climate (2002) Average temperature at Laucala Bay, Fiji. Available at http://www.worldclimate.com/cgi-bin/data.pl?ref=S18E178+1102+91690W.
Yohe G, Tol RSJ (2002) Indicators for social and economic coping capacitymoving toward a working definition of adaptive capacity. Global Environmental Change, 12:25-40.
Ziska LH, Caulfield F (2000) The potential influence of rising atmospheric carbon dioxide $\left(\mathrm{CO}_{2}\right)$ on public health: pollen production of common ragweed as a test case. World Resource Review, 12:449-457.

## Appendix A: Uncertainty around climate predictions

Results presented here were generated using COSMIC (Country Specific Model for Intertemporal Climate) (Schlesinger and Williams 1997). This contains simplified versions of the 14 different climate models used by the IPCC, and gives output at the country level. Each model was run specifying three different plausible sensitivities (1.5, 2.5 and $4.5^{\circ} \mathrm{C}$ ) to the effects of a doubling of atmospheric $\mathrm{CO}_{2}$, generating 42 alternative future climate scenarios for each GHG emissions scenario. Estimates for changes in annual average temperature and precipitation were generated for each subregion by taking population weighted averages of the country level estimates. Means, ranges and $95 \%$ confidence intervals assume independence, and equal probability for each future model and climate sensitivity. All values in figures and tables are for changes relative to 1990.

Figure A.I Mean and $95 \% \mathrm{Cl}$ of COSMIC model outputs for temperature change in the 2020s (relative to 1990) for each subregion under each GHG emissions scenario


Table A.I Mean and range of COSMIC model outputs, for temperature change in the 2020 s , relative to $1990^{\circ}$
(a) S 550

| Subregion | $\begin{gathered} \text { Mean dT }\left({ }^{\circ} \mathrm{C}\right) \\ 95 \% \mathrm{Cl} \\ \hline \end{gathered}$ | Min-max range | COSMIC version HadCM2 ( $2.5^{\circ} \mathrm{C}$ sensitivity) |
| :---: | :---: | :---: | :---: |
| AFR-D | 0.53 (0.46-0.59) | 0.24-0.97 | 0.53 |
| AFR-E | 0.50 (0.44-0.56) | 0.23-0.96 | 0.50 |
| AMR-A | 0.68 (0.60-0.77) | 0.32-1.21 | 0.70 |
| AMR-B | 0.50 (0.44-0.56) | 0.25-0.87 | 0.47 |
| AMR-D | 0.45 (0.39-0.51) | $0.21-0.86$ | 0.42 |
| EMR-B | 0.67 (0.59-0.76) | 0.28-1. 25 | 0.67 |
| EMR-D | 0.63 (0.55-0.71) | 0.28-1.19 | 0.62 |
| EUR-A | 0.64 (0.56-0.72) | 0.31-1.12 | 0.64 |
| EUR-B | 0.66 (0.58-0.74) | $0.28-1.18$ | 0.73 |
| EUR-C | 0.73 (0.65-0.82) | $0.37-1.23$ | 0.72 |
| SEAR-B | 0.40 (0.35-0.45) | 0.17-0.77 | 0.35 |
| SEAR-D | 0.51 (0.44-0.57) | 0.22-0.98 | 0.50 |
| WPR-A | 0.58 (0.5I-0.64) | 0.27-0.97 | 0.54 |
| WPR-B | 0.64 (0.56-0.71) | $0.31-1.25$ | 0.61 |

## (b) S 750

| Subregion | Mean $\mathrm{dT}\left({ }^{\circ} \mathrm{C}\right)$ <br> $95 \% \mathrm{Cl}$ | Min-max range | COSMIC version HadCM2 <br> $\left(2.5^{\circ} \mathrm{C}\right.$ sensitivity $)$ |
| :--- | :---: | :---: | :---: |
| AFR-D | $0.59(0.52-0.66)$ | $0.28-\mathrm{I} .08$ | 0.61 |
| AFR-E | $0.56(0.49-0.63)$ | $0.27-1.06$ | 0.57 |
| AMR-A | $0.77(0.68-0.86)$ | $0.37-1.35$ | 0.79 |
| AMR-B | $0.56(0.49-0.62)$ | $0.29-0.97$ | 0.53 |
| AMR-D | $0.51(0.45-0.57)$ | $0.24-0.95$ | 0.48 |
| EMR-B | $0.76(0.67-0.85)$ | $0.33-1.39$ | 0.76 |
| EMR-D | $0.71(0.63-0.80)$ | $0.33-1.32$ | 0.71 |
| EUR-A | $0.72(0.64-0.81)$ | $0.36-1.25$ | 0.72 |
| EUR-B | $0.74(0.65-0.83)$ | $0.32-1.31$ | 0.83 |
| EUR-C | $0.82(0.73-0.92)$ | $0.43-1.37$ | 0.81 |
| SEAR-B | $0.45(0.39-0.51)$ | $0.20-0.86$ | 0.40 |
| SEAR-D | $0.57(0.50-0.64)$ | $0.25-1.09$ | 0.57 |
| WPR-A | $0.65(0.58-0.72)$ | $0.31-1.08$ | 0.61 |
| WPR-B | $0.72(0.63-0.80)$ | $0.36-1.39$ | 0.69 |

$\begin{array}{ll}\text { Table A.I } & \begin{array}{l}\text { Mean and range of COSMIC model outputs, for } \\ \text { temperature change in the } 2020 \text { s, relative to } 1990^{\text {a }} \\ \text { (continued) }\end{array}\end{array}$
(c) Unmitigated emissions

| Subregion | $\begin{gathered} \text { Mean dT }\left({ }^{\circ} \mathrm{C}\right) \\ 95 \% \mathrm{Cl} \end{gathered}$ | Min-max range | COSMIC version HadCM2 <br> ( $2.5^{\circ} \mathrm{C}$ sensitivity) |
| :---: | :---: | :---: | :---: |
| AFR-D | 0.83 (0.73-0.92) | 0.41-1.47 | 0.85 |
| AFR-E | 0.78 (0.70-0.87) | 0.39-1.45 | 0.80 |
| AMR-A | 1.07 (0.96-I.19) | 0.54-1.83 | 1.11 |
| AMR-B | 0.78 (0.69-0.86) | 0.42-1. 32 | 0.75 |
| AMR-D | 0.71 (0.63-0.79) | 0.35-1.29 | 0.67 |
| EMR-B | 1.06 (0.94-I.I7) | 0.48-1.88 | 1.08 |
| EMR-D | 0.99 (0.88-1.11) | 0.48-1.79 | 1.00 |
| EUR-A | 1.01 (0.90-I.12) | 0.52-1.69 | 1.02 |
| EUR-B | 1.03 (0.92-I.15) | 0.46-1.78 | 1.17 |
| EUR-C | 1.15 (1.03-1.27) | $0.63-1.86$ | 1.15 |
| SEAR-B | 0.63 (0.55-0.70) | 0.29-1.17 | 0.56 |
| SEAR-D | 0.80 (0.70-0.89) | 0.36-1.49 | 0.80 |
| WPR-A | 0.91 (0.8I-I.00) | 0.45-1.46 | 0.86 |
| WPR-B | 1.00 (0.89-I.II) | 0.53-1.89 | 0.97 |

a Results of the simplified version of the HadCM2 model at medium $\left(2.5^{\circ} \mathrm{C}\right)$ sensitivity are shown for comparison.

Figure A. 2 Mean and $95 \% \mathrm{Cl}$ of COSMIC model outputs for precipitation change in the 2020s (relative to 1990), for each subregion under each GHG emissions scenario


Table A. 2 Mean and range of COSMIC model outputs, for change in annual precipitation in the 2020s, relative to $1990^{\text {a }}$
(a) S 550

| Subregion | $\begin{gathered} \text { Mean dPrecipn (mm) } \\ 95 \% \mathrm{Cl} \end{gathered}$ | Min-max range | COSMIC version HadCM2 <br> ( $2.5^{\circ} \mathrm{C}$ sensitivity) |
| :---: | :---: | :---: | :---: |
| AFR-D | 8.18 (4.36-11.99) | -15.8-43.88 | 4.92 |
| AFR-E | 6.71 (1.29-12.14) | -31.83-57.55 | 14.88 |
| AMR-A | 12.01 (8.53-15.48) | -10.38-54.00 | 13.56 |
| AMR-B | 9.17 (3.12-15.22) | -41.47-60.79 | 5.28 |
| AMR-D | 20.03 (13.87-26.18) | -31.17-71.40 | 33.48 |
| EMR-B | 3.45 (1.38-5.52) | -7.10-25.61 | 2.52 |
| EMR-D | 9.05 (6.71-11.38) | -1.83-32.97 | 5.76 |
| EUR-A | 9.94 (7.10-12.77) | -9.58-43.69 | 14.88 |
| EUR-B | 3.33 (1.18-5.47) | -9.55-15.18 | 7.56 |
| EUR-C | 11.50 (9.14-13.86) | 0.72-37.68 | 12.48 |
| SEAR-B | 20.11 (11.05-29.16) | -35.40-135.16 | 7.20 |
| SEAR-D | 25.87 (20.74-31.01) | 4.40-71.19 | 17.64 |
| WPR-A | 14.39 (8.93-19.84) | -33.56-68.61 | 1.32 |
| WPR-B | 14.82 (10.82-18.82) | -7.36-55.37 | 13.44 |

(b) S 750
$\left.\begin{array}{lccc}\hline & \text { Mean dPrecipn (mm) } \\ 95 \% \mathrm{Cl}\end{array}\right)$

Table A. 2 Mean and range of COSMIC model outputs, for change in annual precipitation in the 2020s, relative to $1990^{\text {a }}$ (continued)
(c) Unmitigated emissions

| Subregion | Mean dPrecipn (mm) <br> $95 \% ~ C I$ | Min-max range | COSMIC version HadCM2 <br> $\left(2.5{ }^{\circ} \mathrm{C}\right.$ sensitivity $)$ |
| :--- | :---: | :---: | :---: |
| AFR-D | $13.39(7.44-19.35)$ | $-23.85-66.26$ | 7.96 |
| AFR-E | $11.67(3.69-19.64)$ | $-37.40-86.89$ | 23.74 |
| AMR-A | $18.81(13.49-24.13)$ | $-16.60-81.56$ | 21.78 |
| AMR-B | $14.76(5.64-23.88)$ | $-52.87-92.27$ | 8.42 |
| AMR-D | $31.84(22.38-41.31)$ | $-47.05-107.81$ | 53.44 |
| EMR-B | $6.54(2.75-10.34)$ | $-10.49-54.18$ | 4.69 |
| EMR-D | $15.10(11.37-18.82)$ | $-2.55-49.78$ | 9.14 |
| EUR-A | $15.58(11.24-19.92)$ | $-15.30-65.98$ | 23.79 |
| EUR-B | $5.22(1.90-8.53)$ | $-15.27-22.92$ | 12.13 |
| EUR-C | $18.04(14.50-21.58)$ | $1.22-56.88$ | 19.99 |
| SEAR-B | $32.75(17.95-47.54)$ | $-53.44-220.93$ | 11.62 |
| SEAR-D | $40.72(33.00-48.44)$ | $7.38-107.89$ | 28.12 |
| WPR-A | $22.67(14.23-31.11)$ | $-50.64-104.49$ | 2.09 |
| WPR-B | $23.33(17.18-29.47)$ | $-11.13-86.26$ | 21.45 |

[^76]
## Appendix B: Literature review

For all impacts, we attempted to find all models that directly related climate change to the selected health effects, either at the global or largeregional level. We describe all such models, and outline reasons for using or not using them in the assessment. Note that the comments given relate to their relative suitability for generating the estimates required for this assessment, rather than a general judgement on their merits. For the effects of thermal extremes, we also include a summary of the study populations that were used to derive the heat and cold estimates used in this assessment.

Where existing global or large scale regional models are not considered appropriate (thermal extremes), or do not exist at all (diarrhoea and floods), the literature search describes the procedure for identifying relevant studies from which to derive new quantitative climate-health relationships and generate models.
THERMAL EXTREMES
Information sources:

- Medline search for all references (1966-2002) containing the terms "temperature"; "mortality"; "cardiovascular disease"
(or CVD); "respiratory"; "weather"
- references cited in these papers.
Table B.I Results of literature search for effects of climate change on deaths due to thermal extremes ${ }^{\text {a }}$

| Reference | Method | Inputs | Region | Key findings | Suitability for generating estimates for this assessment |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Langford and Bentham (1995) | Regression model: monthly mortality and temperature series (flu included) | Seasonal average dT from United Kingdom scenarios (Warrick and Barrow 1991) | England and Wales | Winter deaths avoided: <br> 2010: all cause 3301, IHD I308, CVD 429 <br> 2030: all cause 6353, IHD 2550, CVD 836 <br> 2050: all cause 8922, IHD 363I, CVD 1187 | No population growth, ageing. No seasonal adjustment. <br> Local rather than global |
| Martens (1998a) | MIASMA vI.0—empiricalstatistical model, meta-analysis. A single temperature-mortality relationship was applied to all cities | ECHAMI-A, UKTR, GFDL89 (IPCC scenarios) | Cause+age specific. Global20 cities | Changes in mortality rates for CVD ( $<65$ ), CVD (>65), respiratory, and total mortality Net reductions in mortality in all cities except respiratory mortality in a few cities | Lack of control for seasonal variations, perhaps over-estimating heat effects |
| McMichael et al. (2000b) | MIASMA v.I.0—as for Martens (1998a) | HadCM2, ensemble mean+HadCM3 | Global- <br> 20 cities | Net reductions in mortality in all cities except Athens, due to decreases in winter mortality | As for Martens (1998a) |

Table B.I Results of literature search for effects of climate change on deaths due to thermal extremes ${ }^{\text {a }}$ (continued)

| Reference | Method | Inputs | Region | Key findings | Suitability for generating estimates for this assessment |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Kalkstein and Smoyer (1993) | Model derived from observed relationship between synoptic air masses and mortality | GCM scenarios, no downscaling | Global- <br> 27 cities | Significant increases in heat-related mortality under various climate change scenarios, with or without acclimatization, projected for all cities | Methods used rely on synoptic classification of local air masses; therefore cannot be directly related to changes in temperature described by climate scenarios |
| Kalkstein and Greene (1997) | Model derived from observed relationship between synoptic air masses and mortality | GCM scenarios, no downscaling | 44 cities in the USA | Increases in heat-related mortality are much larger overall than decreases in cold-related mortality | Methods used rely on synoptic classification of local air masses; therefore cannot be directly related to changes in temperature described by climate scenarios |
| Duncan et al. (1997) | Model derived from observed relationship between synoptic air masses and mortality | Climate scenarios for Canada | 10 cities <br> in Canada | 240-1140 additional heat-related deaths/year in Montreal by 2050; 230-1220 additional deaths in Toronto, assuming no acclimatization. No climate/mortality relationship in some cities | Country specific |
| Guest et al. (1999) | Regression model: daily mortality and synoptic indices | CSIRO Mark 2 model, $\mathrm{CO}_{2}$ doubling; high and low scenarios estimated | 5 cities in Australia | Net decrease in heat-related mortality, particularly in the age group $\geq 65$ years; range of 47-62 fewer deaths/year in all 5 cities. Significant increase in summer deaths in Sydney (76-239) | Country specific |
| Dessai (2003) | Empirical statistical model, observedexpected | 2 regional climate models-PROMES and HadRM2 | Lisbon, Portugal | Heat-related death rates increase by $57-113 \%$ by 2020 s, by $97-255 \%$ by 2050s. Acclimatization assumptions, reduce estimates | Country specific |

Key: CVD, cardiovascular disease; IHD, ischaemic hearth disease; GCM, global climate model.
A total of 76 other publications were not used, either because they were analyses of specific episodes ( 35 studies), or because they did not match the criteria outlined above ( 41 studies).

## Estimates of temperature-MORTALITY relationship

Table B. 2 below lists all studies $(n=4)$ that meet the following criteria for inclusion in the meta-analysis. The study:

- uses daily time-series methods to analyse the relationship between daily mean temperature and mortality;
- Reports a coefficient from log linear regression that estimates the percentage changes in mortality per degree centigrade change in temperature, above a reported threshold temperature;
- has controls for the following confounders: season, air pollution and influenza;
- is published in English language only; and
- reports confidence intervals around the coefficient and is within the range of other reported estimates.

The studies are classified according to climate zone. The Netherlands population is used to approximate a population in the "cold" zone as there are no other appropriate time-series studies.

Note that some studies are included that are not yet published. However,

- they are the only studies available that provide estimates for developing country populations; and
- the results will be submitted to peer reviewed journals and the methods are at least as rigorous as those applied in previous published studies.

All estimates have been rounded to one decimal place. Where more than one estimate was available, the average estimate was calculated.
Table B. 2 Results of literature search for temperature-mortality relationship

| Population/ climate zone | Age | Cause of death | \% change ${ }^{\text {a }}$ per <br> $1{ }^{\circ} \mathrm{C}$ decrease | Cut-point ( ${ }^{\circ} \mathrm{C}$ ) | Lag | \% change ${ }^{\text {a }}$ per <br> $1{ }^{\circ} \mathrm{C}$ increase | Cut-point ( ${ }^{\circ}$ C) | Lag | Reference |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | COLD |  |  | HEAT |  |  |  |
| Netherlands | All | All cause | 0.41 | 16.5 | 7-15 | 1.23 | 16.5 | 1-2 | Kunst et al. (1993) |
| Netherlands | All | CVD | 0.46 | 16.5 | 7-15 | 1.13 | 16.5 | 1-2 | Kunst et al. (1993) |
| Netherlands | All | Respiratory | 1.43 | 16.5 | 7-15 | 3.11 | 16.5 | 1-2 | Kunst et al. (1993) |
| COLD |  | CVD | 0.5 |  |  | 1.1 |  |  |  |
| Bulgaria Sofia | All | All | 2.69 (0.88-4.54) | -2 | 0-13 | 1.93 (1.41-2.45) | 17 | 0-2 | ISOTHURM (forthcoming) |
| Chile Santiago | All | All | 5.21 (3.55-6.89) | 11 | 0-13 | 0.92 (0.44-1.31) | 16 | 0-2 | ISOTHURM (forthcoming) |
| Slovenia Ljubljana | All | All | 0.77 (-0.16-1.70) | 7 | 0-13 | 2.25 (1.09-3.42) | 17 | 0-2 | ISOTHURM (forthcoming) |
| South Africa Cape Town | All | All | 3.32 (2.89-3.75) | 19 | 0-13 | 1.02 (-0.32-2.38) | 21 | 0-2 | ISOTHURM (forthcoming) |
| Spain |  |  |  |  |  |  |  |  |  |
| Madrid | All | All | 1.7 | 20 | 7 | 0.97 | 20 | 0-1 | Alberdi (1998) |
| Valencia | All | All | 3.2 (1.8-4.6) | 15 | 7-14 | 3.6 (1.2-6.0) | 24 | I-2 | Ballester et al. (1997) |
| Valencia | >70 | All | 3.7 (2.1-5.4) | 15 | 7-14 | 5.0 (2.1-8.0) | 24 | I-2 | Ballester et al. (1997) |
| Valencia | All | CVD | 4.3 (2.1-6.4) | 15 | 7-14 | 2.3 (-1.5-4.5) | 24 | 1-2 | Ballester et al. (1997) |
| Valencia | All | CVD | 1.5 (-0.3-3.3) | 15 | 3-6 | 2.9 (-0.4-7.4) | 24 | 3-6 | Ballester et al. (1997) |
| Valencia | All | Respiratory | I. $7(-0.4-6.0)$ | 15 | 7-14 | 5.7 (-2.9-8.2) | 24 | I-2 | Ballester et al. (1997) |


| TEMPERATE |  | CVD | 2.9 |  |  | 2.6 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Brazil |  |  |  |  |  |  |  |  |  |
| São Paulo | All | All cause | 3.92 (3.43-4.40) | 19 | 0-13 | 2.28 (2.11-3.66) | 23 | 0-2 | ISOTHURM (forthcoming) |
| São Paulo | $\geq 65$ | All cause | 5.5 | 20 | 0-21 | 2.5 (2.1-2.8) | 20 | 0-1 | Gouveia et al. (2003) |
| El Salvador |  |  |  |  |  |  |  |  |  |
| San Salvador | All | All cause | -13.5 (-32.35-10.9) | 23 | 0-13 | 1.59 (0.86-2.32) | 23 | 0-2 | ISOTHURM (forthcoming) |
| Mexico |  |  |  |  |  |  |  |  |  |
| Mexico City | All | All cause | 8.60 (7.86-9.34) | 15 | 0-13 | 0.6 (0.2I-1.00) | 18 | 0-2 | ISOTHURM (forthcoming) |
| Mexico |  |  |  |  |  |  |  |  |  |
| Monterrey | All | All cause | 5.54 (4.52-6.58) | 17 | 0-13 | 19.85 (14.69-25.25) | 31 | 0-2 | ISOTHURM (forthcoming) |
| Thailand |  |  |  |  |  |  |  |  |  |
| Bangkok | All | All cause | 4.13 (1.71-6.61) | 28 | 0-13 | 7.66 (5.87-9.47) | 30 | 0-2 | ISOTHURM (forthcoming) |
| WARM HUMID | All | All cause | 5.5 |  |  | 5.7 |  |  |  |
| India |  |  |  |  |  |  |  |  |  |
| Delhi | All | All cause | 1.36 (0.56-2.16) | 19 | 0-13 | 3.03 (2.48-3.58) | 28 | 0-2 | ISOTHURM (forthcoming) |
| HOT DRY | All | All cause | 1.4 |  |  | 3.0 |  |  |  |
| CVD Cardiovascular disease. |  |  |  |  |  |  |  |  |  |
| a \% change = | -1) $\times$ |  |  |  |  |  |  |  |  |

DIARRHOEA
Information sources:
references cited in these papers.
Table B. 3 Results of literature search for effects of climate change on diarrhoea ${ }^{\text {a }}$

| Global models relating climate/climate change to diarrhoea | Local studies quantitatively relating all-cause diarrhoea to temperature | Local studies showing seasonal patterns in all- or multiple-cause diarrhoea (non-quantitative) | Local studies showing pathogen-specific associations with climate |
| :---: | :---: | :---: | :---: |
| None | Checkley et al. (2000); Singh et al. (2001) | Anjaneyulu et al. (1975); Becker (I98I); Brewster and Greenwood (1993); Dean and Jones (1972); Hoge et al. (1996); Jin et al. (1996); Ling and Cheng (1993); Merlin et al. (1986); Parashar et al. (I999); Pinfold et al. (I99\|); Robins-Browne (1984); Rousham and Mascietaylor (1995); Saidi et al. (1997); Sawchuk (I993); Shaikh et al. (1990); Sutra et al. (I990); Tsukamoto et al. (I978); Van den Broeck et al. (I993); Williams et al. (I986); Yang et al. (1990) | Showing association: <br> Adegbola et al. (1994); Adkins et al. (1987); Aggarwal et al. (I983); Ansari et al. (1991); Armah et al. (1994); Beards and Graham (1995); Bockemuhl (1976); Callejas et al. (I999); Chakravarti et al. (1992); Chan et al. (I998); Chaudhury et al. (1996); Cook et al. (1990); Cunliffe et al. (1998); da Rosa e Silva et al. (2001); Douglas and Kurien (1997); Eberhard et al. (I999); Fujita (1990); Henry and Bartholomew (1990); Hirschl et al. (1987); ljaz et al. (1985); Konno et al. (1983); Laursen et al. (1994); LeBaron et al. (I990); Malakooti et al. (1997); Muhuri (1996); Musa et al. (I999); Nchito et al. (1998); Parashar et al. (1998a, 1998b); Pazzaglia et al. (I993); Purohit et al. (1998); Qiao et al. (1999); Ram et al. (1990a, 1990b); Reyes et al. (1996); Rytlewska et al. (2000); Sallon et al. (1991); Sethi et al. (1984); Shahid et al. (1987); Shkarin et al. (1983); Sorvillo et al. (1998); Stewien et al. (I991); Stintzing et al. (1981); Tswana et al. (1990); Utsalo et al. (I991); Vlasov et al. (I983); Wilcox et al. (1992); Wuhib et al. (I994) <br> Failing to find association: <br> Bishop et al. (2001); Conteas et al. (1998); Varoli et al. (1989) |

[^77]MALNUTRITION
Information sources:

- references cited in these papers.
Table B. 4 Results of literature search for effects of climate change on malnutrition ${ }^{\text {a }}$

| Study | Model type | Model output | Area <br> covered | Conclusions |
| :--- | :--- | :--- | :--- | :--- |

Coastal and inland flooding
Information sources:

## all relevant papers cited in IPCC (2001a)

- ISI-Web of Science search for "Floods (or flooding) and climate (or climatic) change"; "Floods (or flooding) and health (or death or injury or mortality)"; "Extreme precipitation and health (or death or injury or mortality)"


## Medline search for "Floods (or flooding) and climate (or climatic) change"

References cited in these papers

- Inspection of descriptions of all flood events listed in EM-DAT enhanced database 1980-1999 (EM-DAT 2002).
Table B. 5 Results of literature search for effects of climate change on deaths and injuries due to floods and landslides ${ }^{\text {a }}$

| Large area studies quantitatively relating precipitation or sea-level rise to deaths/injuries | Global or large regional models predicting climate change effects on frequency of extreme precipitation events or inland flooding | Global models predicting climate change effects on frequency of coastal flooding | Causes of flood events which caused deaths and/or injuries reported in the EM-DAT enhanced database (1980-1999) | Causes of landslides which caused deaths and/or injuries reported in the EM-DAT enhanced database (1980-1999) |
| :---: | :---: | :---: | :---: | :---: |
| None | Booij (2002), western Europe | Baarse (1995) | Associated with precipitation | 63 associated with precipitation |
|  | Fowler and Hennessy (1995), specific locations worldwide | Hoozemans and Hulsburgen (1995) | 2 with high tides and storm surges | 10 associated with other causes (e.g. landslides at mines, cliffs collapsing) |
|  | Jones and Reid (2001) United Kingdom | Nicholls and Mimura (1998) | 19 other causes (e.g. cyclones, ice melt, dams bursting) | II5 without specific information on cause |
|  | Milly et al. (2002) specific locations worldwide | Nicholls et al. (1999) | 559 without specific information on cause |  |
|  | Palmer and Ralsanen (2002) northern Europe, South-East Asia |  |  |  |

a A total of 342 other publications were rejected as they are either specific to particular regions, or irrelevant to quantitative assessment (mainly reviews, descriptions of specific events).
MALARIA
Information sources:
Table B. 6 Results of literature search for effects of climate change on malaria

| Study | Model type | Model output | Area covered | Conclusions |
| :--- | :--- | :--- | :--- | :--- |

Table B. 6 Results of literature search for effects of climate change on malaria (continued)

| Study | Model type | Model output | Area covered | Conclusions | Suitability for generating estimates for this assessment |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Bryan et al. (1996) | CLIMEX model | Vector distributions only | Northern <br> Australia | Change in distribution under CC | Single vector, local rather than global predictions |
| Jetten et al. (1996) | Biological modelMIASMA vI. 0 | Maps of TP and change in TP | Global | Expansion of distribution, increasing TP under CC | Prediction of climate suitability for transmission, rather than actual transmission. Uncertainty over cut-off values to define endemic/non-endemic areas. Vector distributions not considered |
| Martens (1997) | MIASMA vi. 0 | Maps of TP and change in TP, P. vivax and P. faliiparum | Global | Expansion of distribution, increasing TP under CC | Prediction of climate suitability for transmission, rather than actual transmission. Uncertainty over cut-off values to define endemic/non-endemic areas. Vector distributions not considered |
| Lindsay and Martens (1998) | MIASMA vI.0— biological model based on vectorial capacity | Maps-epidemic potential | African Highlands | Increase in latitude under CC | Regional projections only |
| Rogers (1996) | Empirical statistical mapping | Maps of distribution of Anopheles gambiae | Southern Africa | Change in distribution under CC | Regional projections only |
| Martens et al. (1999) | Biological model, overlaying a population grid. MIASMA v2.0 | Maps of TP, maps of changes in seasonal transmission, additional population at risk | Global | Expansion of distribution, increasing TP under CC | Prediction of climate suitability for transmission, rather than actual transmission |


| Arnell et al. (2002) | Biological model, overlaying a population grid. MIASMA v2.0 | Maps of changes in $P$. falciparum TP, population at risk, seasons suitable for transmission relative to 1961-1990 | Global | Expansion of distribution, increasing TP under CC | Prediction of climate suitability for transmission, rather than actual transmission |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Rogers and Randolph (2000) | Empirical-statistical model | Changes in populations at risk: P. falciparum | Global | Approximately no change in distribution under climate change | Additional population at risk $\pm 25$ million. Validation of model derived from a subset of the data against the remaining observations. (Used to inform uncertainty range around the projections) |
| Tanser et al. (2003) | MARA biological/ statistical model (Craig et al. 1999) | Changes in populations at risk-months suitable for transmission | Africa | Expansion in population at risk under CC | Biological model based on localized field studies, used to predict population at risk and numbers of months at risk throughout Africa, but not elsewhere. Potential for developing into predictions of incidence. Validated against detailed independent data set. (Used to determine mid-range estimates of population at risk, as a relative measure of change in incidence) |

Key: CC, climate change; EIP, extrinsic incubation period; TP, transmission potential (also called EP, epidemic potential).
${ }^{\text {a }} \quad$ A further 41 publications were rejected due to lack of relevance for making global projections (mainly reviews, local studies).
DENGUE
Information sources:

## - all relevant papers cited in IPCC (2001a)

- Medline search for all references (1966-2002) containing the terms "Dengue and climate (or climatic) change"
- Secondary references cited in these papers.
Table B. 7 Results of literature search for projected effects of climate change on dengue ${ }^{\text {a }}$

| Study | Model type | Model output | Area covered | Conclusions |
| :--- | :--- | :--- | :--- | :--- |


| Patz et al. (1998) | Biological model summarizing climate effects on vector- and parasite population dynamics | Maps of change in epidemic potential under climate change | City specific (Athens, Bangkok, Mexico City, Philadelphia, Puerto Rico, San Juan) | Large increase in transmission potential with small temperature increase | Model is most accurate for nonendemic areas bordering endemic areas and may underestimate changes in transmission in temperate zones. Relationship between TP and disease incidence not characterized |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Jetten and Focks (1997) | Biological model summarizing climate effects on vector- and parasite population dynamics | Maps of changes in Critical Density Threshold | Global | Large increase in areas suitable for transmission and length of transmission season under climate change | Relationship between TP and disease incidence not characterized. <br> Uncertainty over setting of cut-offs for transmission |
| Hales et al. (2002) | Empirical statistical model | Maps of population at risk for dengue transmission | Global | Absolute humidity accurately explains current distribution of dengue. Large increases in population at risk predicted under climate change | Validation of model derived from a subset of the data against the remaining observations. (Used to determine population at risk, as a relative measure of change in incidence) |

## Appendix references

Adegbola RA, Demba E, de Veer G, Todd J (1994) Cryptosporidium infection in Gambian children less than 5 years of age. Journal of Tropical Medicine and Hygiene, 97:103-107.

Adkins HJ, Escamilla J, Santiago LT, Ranoa C, Echeverria P, Cross JH (1987) Two-year survey of etiologic agents of diarrheal disease at San Lazaro Hospital, Manila, Republic of the Philippines. Journal of Clinical Microbiology, 25:|I43-II47.

Aggarwal P, Sarkar R, Gupta JP, Ahuja S, Ray K, Rai Chowdhuri AN (I983) Current epidemiological aspects of cholera in Delhi. Journal of Communicable Diseases, 15:26-32.

Alberdi JC (1998) Daily mortality in Madrid community, I986-I992: relationship with meteorological variables. European Journal of Epidemiology, 14:571-578.

Anjaneyulu G, Banerji SC, Indrayan A (1975) A study of fly density and meteorological factors in occurrence of diarrhoea in a rural area. Indian Journal of Public Health, 19:115-121.

Ansari SA, Springthorpe VS, Sattar SA (I99I) Survival and vehicular spread of human rotaviruses-possible relation to seasonality of outbreaks. Reviews of Infectious Diseases, I3:448-461.

Armah GE, Mingle JAA, Dodoo AK et al. (1994) Seasonality of rotavirus infection in Ghana. Annals of Tropical Paediatrics, I4:223-229.

Arnell N, Cannell MG, Hulme M et al. (2002) The consequences of $\mathrm{CO}_{2}$ stabilization for the impacts of climate change. Climatic Change, 53:4I3-446

Baarse G (1995) The development of an operational tool for global vulnerability assessment (GVA) update of the number of people at risk due to sea level rise and increased flooding probabilities. Ministry of Transport, Public Works and Water Management, The Hague.
Ballester DF, Corella D, Perez-Hoyos S, Saez M, Hervas A (1997) Mortality as a function of temperature. A study in Valencia, Spain, 1991-1993. International Journal of Epidemiology, 26:55I-56I.
Beards G, Graham C (1995) Temporal distribution of rotavirus G-serotypes in the West Midlands region of the United Kingdom, I983-I994. Journal of Diarrhoeal Diseases Research, 13:235-237.

Becker S (I98I) Seasonality of deaths in Matlab, Bangladesh. International Journal of Epidemiology, I0:27I-280.

Bishop RF, Masendycz PJ, Bugg HC, Carlin JB, Barnes GL (200I) Epidemiological patterns of rotaviruses causing severe gastroenteritis in young children throughout Australia from 1993 to 1996. Journal of Clinical Microbiology, 39:1085-1091.
Bockemuhl J (1976) Salmonellosis and shigellosis in Togo (West Africa), I97I-I973. I. Carrier rates in the rural population. Tropenmedizin und Parasitologie, 27:112-I20.
Booij MJ, (2002) Extreme daily precipitation in Western Europe with climate change at appropriate spatial scales. International Journal of Climatology, 22:69-85.
Brewster DR, Greenwood BM (1993) Seasonal-variation of pediatric diseases in the Gambia, West-Africa. Annals of Tropical Paediatrics, I3:I33-I46.

Bryan JH, Foley DH, Sutherst RW (1996) Malaria transmission and climate change in Australia. Medical Journal of Australia, I64:345-347.

Callejas D, Estevez J, Porto-Espinoza L et al. (I999) Effect of climatic factors on the epidemiology of rotavirus infection in children under 5 years of age in the city of Maracaibo, Venezuela [in Spanish]. Investigacion Clinica, 40:8I-94.

Chakravarti A, Broor S, Natarajan R, Setty VS, Mittal SK (1992) Epidemiological and clinical characteristics of acute diarrhoea in children due to human rotavirus. Journal of Tropical Pediatrics, 38:192-193.
Chan PKS, Tam JS, Nelson EAS et al. (1998) Rotavirus infection in Hong Kong: epidemiology and estimates of disease burden. Epidemiology and Infection, I 20:32I-325.
Chaudhury A, Nath G, Shukla B, Panda S, Singh TB (1996) Diarrhoea associated with Candida spp.: incidence and seasonal variation. Journal of Diarrhoeal Diseases Research, 14:110-1I2.

Checkley W, Epstein LD, Gilman RH et al. (2000) Effect of El Niño and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. The Lancet, 355:442-450.

Conteas CN, Berlin OG, Lariviere MJ et al. (1998) Examination of the prevalence and seasonal variation of intestinal microsporidiosis in the stools of persons with chronic diarrhea and human immunodeficiency virus infection. American Journal of Tropical Medicine and Hygiene, 58:559-561.
Cook SM, Glass RI, LeBaron CW, Ho MS (1990) Global seasonality of rotavirus infections. Bulletin of the World Health Organization, 68:171-I77.
Cunliffe NA, Kilgore PE, Bresee JS et al. (1998) Epidemiology of rotavirus diarrhoea in Africa: a review to assess the need for rotavirus immunization. Bulletin of the World Health Organization, 76:525-537.
da Rosa e Silva ML, Naveca FG, Pires de Carvalho I (200I) Epidemiological aspects of rotavirus infections in Minas Gerais, Brazil. Brazilian Journal of Infectious Diseases, 5:215-222.

Dean AG, Jones TC (1972) Seasonal gastroenteritis and malabsorption at an American military base in the Philippines. I. Clinical and epidemiologic investigations of the acute illness. American Journal of Epidemiology, 95: I II-I27.
Dessai SR (2003) Heat stress and mortality in Lisbon. Part II. An assessment of the potential impacts of climate change. International Journal of Biometeorology, 48:37-44.
Douglas AS, Kurien A (1997) Seasonality and other epidemiological features of haemolytic uraemic syndrome and E-coli OI57 isolates in Scotland. Scottish Medical Journal, 42:166-171.

Duncan K, Guidotti TL, Cheng K et al. (1997) The Canadian Country study: impacts and adaptation, Health Sector. In: The Canadian Country study: climate impacts and adaptation, National Sectoral Volume VII. Environment Canada. Toronto.

Eberhard ML, Nace EK, Freeman AR, Streit TG, Da Silva AJ, Lammie PJ (I999) Cyclospora cayetanensis infections in Haiti: a common occurrence in the absence of watery diarrhea. American Journal of Tropical Medicine and Hygiene, 60:584-586.
Fischer G, Frohberg K, Parry ML, Rosenzweig C (1994) Climate-change and world food-supply, demand and trade-who benefits, who loses. Global Environmental Change-Human and Policy Dimensions, 4:7-23.

Fowler AM, Hennessy KJ (1995) Potential impacts of global warming on the frequency and magnitude of heavy precipitation. Natural Hazards, I I:283-303.

Fujita Y (1990) Rotavirus infection-clinical symptoms and influence of climate [in Japanese]. Kansenshogaku Zasshi, 64:I255-I263.

Gouveia N, Hajat S, Armstrong B (2003) Socio-economic differentials in the temper-ature-mortality relationship in Sao Paulo, Brazil. International Journal of Epidemiology, 32:390-397.

Guest CS, Willson K, Woodward AJ et al. (1999) Climate and mortality in Australia. Retrospective study, 1979-1990, and predicted impacts in five major cities in 2030. Climate Research, 13:1-15.

Hales S, de Wet N, Maindonald J, Woodward A (2002) Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. The Lancet, 360:830-834.

Henry FJ, Bartholomew RK (1990) Epidemiology and transmission of rotavirus infections and diarrhoea in St. Lucia, West Indies. West Indian Medical Journal, 39:205-2I2.

Hirschl AM, Lior H, Wolf D et al. (1987) Occurrence, serotypes and biotypes of thermophilic Campylobacters isolated in Vienna. Zentralblatt für Bakteriologie, Mikrobiologie, und Hygiene, Serie A, 266:94-I03.

Hoge CW, Shlim DR, Echeverria P, Rajah R, Herrmann JE, Cross JH (1996) Epidemiology of diarrhea among expatriate residents living in a highly endemic environment. Journal of the American Medical Association, 275:533-538.
Hoozemans FMJ, Marchand M, Pennekamp HA (1993) A global vulnerability analysis: vulnerability assessment for population, coastal wetlands and rice production on a global scale. 2nd edn. Delft Hydraulics, Delft.
ljaz MK, Sattar SA, Johnson-Lussenburg CM, Springthorpe VS, Nair RC (I985) Effect of relative humidity, atmospheric temperature, and suspending medium on the airborne survival of human rotavirus. Canadian Journal of Microbiology, 3 I:68I-685.

ISOTHURM (forthcoming) ISOTHURM Study Group: international study of temperature and heat-waves on urban mortality in low and middle income countries. The Lancet,

Jetten TH, Focks DA (1997) Potential changes in the distribution of dengue transmission under climate warming. American Journal of Tropical Medicine and Hygiene, 57:285-287.

Jetten TH, Martens WJM, Takken W (I996) Model simulations to estimate malaria risk under climate change. Journal of Medical Entomology, 33:36I-37I.
Jin SX, Kilgore PE, Holman RC, Clarke MJ, Gangarosa EJ, Glass RI (1996) Trends in hospitalizations for diarrhea in United States children from 1974 through 1992: estimates of the morbidity associated with rotavirus. Pediatric Infectious Disease Journal, 15:397-404.

Jones PD, Reid PA (2001) Assessing future changes in extreme precipitation over Britain using regional climate model integrations. International Journal of Climatology, 21:I337-I356.

Kalkstein LS, Greene JS (1997) An evaluation of climate/mortality relationships in large US cities and the possible impacts of a climate change. Environmental Health Perspectives, 105:84-93.

Kalkstein LS, Smoyer KE (1993) The impact of climate change on human health: some international implications. Experientia, 49:969-979.

Konno T, Suzuki H, Katsushima N et al. (1983) Influence of temperature and relative humidity on human rotavirus infection in Japan. Journal of Infectious Diseases, 147:125-I28.

Kunst A, Looman C, Mackenbach J (1993) Outdoor air temperature and mortality in the Netherlands-a time series analysis. American Journal of Epidemiology, 137:331-341.
Langford IH, Bentham G (1995) The potential effects of climate-change on winter mortality in England and Wales. International Journal of Biometeorology, 38:14I-I47.
Laursen E, Mygind O, Rasmussen B, Ronne T (1994) Gastroenteritis: a waterborne outbreak affecting 1600 people in a small Danish town. Journal of Epidemiology and Community Health, 48:453-458.
LeBaron CW, Lew J, Glass RI, Weber JM, Ruiz-Palacios GM (1990) Annual rotavirus epidemic patterns in North America. Results of a 5 -year retrospective survey of 88 centers in Canada, Mexico, and the United States. Rotavirus Study Group. Journal of the American Medical Association, 264:983-988.
Lindsay SW, Martens WJM (1998) Malaria in the African highlands: past, present and future. Bulletin of the World Health Organization, 76:33-45.
Ling JM, Cheng AF (1993) Infectious diarrhea in Hong-Kong. Journal of Tropical Medicine and Hygiene, 96:107-I I2.
Malakooti MA, Alaii J, Shanks GD, Phillips-Howard PA (1997) Epidemic dysentery in western Kenya. Transactions of the Royal Society of Tropical Medicine and Hygiene, 91:54I-543.
Martens P (1997) Health impacts of climate change and ozone depletion: an eco-epidemiological modelling approach. Maastricht University, Maastricht.
Martens W (1998a) Climate change, thermal stress and mortality changes. Social Science Medicine, 46:33I-334.
Martens WJ, Jetten TH, Rotmans J, Niessen LW (1995a) Climate change and vectorborne diseases: a global modelling perspective. Global Environmental Change, 5:195-209.

Martens WJ, Niessen LW, Rotmans J, Jetten TH, McMichael AJ (I995b) Potential impact of global climate change on malaria risk. Environmental Health Perspectives, 103:458-464.
Martens WJM, Jetten TH, Focks DA (1997) Sensitivity of malaria, schistosomiasis and dengue to global warming. Climatic Change, 35:145-156.
Martens P, Kovats RS, Nijhof S et al. (1999) Climate change and future populations at risk of malaria. Global Environmental Change-Human and Policy Dimensions, 9:S89-SIO7.

Martin PH, Lefebvre MG (1995) Malaria and climate-sensitivity of malaria potential transmission to climate. Ambio, 24:200-207.

Matsuoka Y, Kai K (I994) An estimation of climatic change effects on malaria. Journal of Global Environment Engineering, I:I-I5.

McMichael AJ, Kovats S, Martens P et al. (2000b) Climate change and human health: final report to the Department of Environment, Transport and the Regions. London School of Hygiene and Tropical Medicine/ICIS, London/Maastricht.

Merlin M, Roure C, Kollo B et al. (I986) Evaluation of morbidity, mortality and therapeutic procedures in diarrheal diseases of young children in Cameroon. Médecine Tropicale (Marseilles), 46:355-357.

Milly PCD, Wetherald RT, Dunne KA, Delworth TL (2002) Increasing risk of great floods in a changing climate. Nature, 415:5I4-5I7.
Muhuri PK (1996) Estimating seasonality effects on child mortality in Matlab, Bangladesh. Demography, 33:98-1 10.
Musa HA, Shears P, Kafi S, Elsabag SK (1999) Water quality and public health in northern Sudan: a study of rural and peri-urban communities. Journal of Applied Microbiology, 87:676-682.

Nchito M, Kelly P, Sianongo S et al. (1998) Cryptosporidiosis in urban Zambian children: an analysis of risk factors. American Journal of Tropical Medicine and Hygiene, 59:435-437.

Nicholls RJ, Mimura N (I998) Regional issues raised by sea level rise and their policy implications. Climate Research, I I:5-I8.

Nicholls RJ, Hoozemans FMJ, Marchand M (1999) Increasing flood risk and wetland losses due to global sea-level rise: regional and global analyses. Global Environmental Change-Human and Policy Dimensions, 9:S69-S87.
Palmer TN, Ralsanen J (2002) Quantifying the risk of extreme seasonal precipitation events in a changing climate. Nature, 4I5:5I2-5I4.
Parashar UD, Chung MA, Holman RC, Ryder RW, Hadler JL, Glass RI (I999) Use of state hospital discharge data to assess the morbidity from rotavirus diarrhea and to monitor the impact of a rotavirus immunization program: a pilot study in Connecticut. Pediatrics, 104:489-494.

Parashar UD, Holman RC, Bresee JS et al. (1998a) Epidemiology of diarrheal disease among children enrolled in four west coast health maintenance organizations. Pediatric Infectious Disease Journal, 17:605-61I.

Parashar UD, Holman RC, Clarke MJ, Bresee JS, Glass RI (1998b) Hospitalizations associated with rotavirus diarrhea in the United States, 1993 through 1995: surveillance based on the new ICD-9-CM rotavirus-specific diagnostic code. Journal of Infectious Diseases, 177:13-17.
Parry M, Rosenzweig C, Iglesias A, Fischer G, Livermore M. (1999) Climate change and world food security: a new assessment. Global Environmental Change-Human and Policy Dimensions, 9:S5I-S67.

Parry ML, Arnell NW, McMichael A et al. (200I) Millions at risk: defining critical climate change threats and targets. Global Environmental Change, II:I8I-I83.

Patz JA, Martens WJM, Focks DA, Jetten TH (1998) Dengue fever epidemic potential as projected by general circulation models of global climate change. Environmental Health Perspectives, 106:147-I53.
Pazzaglia G, Bourgeois AL, Araby I et al. (1993) Campylobacter-associated diarrhea in Egyptian infants-epidemiology and clinical manifestations of disease and high-frequency of concomitant infections. Journal of Diarrhoeal Diseases Research, I I:6-I3.

Pinfold JV, Horan NJ, Mara DD (199I) Seasonal effects on the reported incidence of acute diarrhoeal disease in northeast Thailand. International Journal of Epidemiology, 20:777-786.

Purohit SG, Kelkar SD, Simha $V$ (1998) Time series analysis of patients with rotavirus diarrhoea in Pune, India. Journal of Diarrhoeal Diseases Research, 16:74-83.

Qiao HP, Nilsson M, Abreu ER et al. (1999) Viral diarrhea in children in Beijing, China. Journal of Medical Virology, 57:390-396.
Ram S, Khurana S, Khurana SB, Sharma S, Vadehra DV (1990a) Seasonal fluctuations in the occurrence of enteroinvasive Escherichia coli diarrhoea. Indian Journal of Medical Research, 91:258-262.

Ram S, Khurana S, Khurana SB, Sharma S, Vadehra DV, Broor S (I990b) Bioecological factors and rotavirus diarrhoea. Indian Journal of Medical Research, 91:167170.

Reyes M, Hedlund KO, Lorenzana I, Ehrnst A (1996) Respiratory infection and iatrogenic diarrhea in Honduras and El Salvador during the 1991-1992 season. American Journal of Tropical Medicine and Hygiene, 54:260-264.

Robins-Browne RM (I984) Seasonal and racial incidence of infantile gastroenteritis in South Africa. American Journal of Epidemiology, I 19:350-355.
Rogers DJ (1996) Regional impacts of climate change; changes in disease vector distributions. In: Climate change and Southern Africa: an exploration of some potential impacts and implications in the SADC region. Hulme M, ed. University of East Anglia, Norwich.
Rogers DJ, Randolph SE (2000) The global spread of malaria in a future, warmer world. Science, 289:I763-I766.
Rosenzweig C, Parry ML (1994) Potential impact of climate-change on world foodsupply. Nature, 367:I33-138.
Rousham EK, Mascietaylor CGN (I995) Seasonality and child morbidity in rural Bangladesh. American Journal of Human Biology, 7:369-379.
Rytlewska M, Bako W, Ratajczak B et al. (2000) Epidemiological and clinical characteristics of rotaviral diarrhoea in children from Gdansk, Gdynia and Sopot. Medical Science Monitor 6:117-I22.

Saidi SM, lijima Y, Sang WK et al. (1997) Epidemiological study on infectious diarrheal diseases in children in a coastal rural area of Kenya. Microbiology and Immunology, 41:773-778.
Sallon S, el Showwa R, el Masri M, Khalil, M, Blundell N, Hart CA (1991) Cryptosporidiosis in children in Gaza. Annals of Tropical Paediatrics, I I:277-28I.
Sawchuk LA (1993) Societal and ecological determinants of urban health: a case study of pre-reproductive mortality in 19th century Gibraltar. Social Science and Medicine, 36:875-892.

Sethi SK, Al-Nakib W, Khuffash FA, Majeed HA (1984) Acute diarrhoea and rotavirus infections in young children in Kuwait. Annals of Tropical Paediatrics, 4:117-I21.
Shahid NS, Rahman AS, Sanyal SC (1987) Cryptosporidium as a pathogen for diarrhoea in Bangladesh. Tropical and Geographical Medicine, 39:265-270.

Shaikh K, Wojtyniak B, Mostafa G, Khan MU (1990) Pattern of diarrheal deaths during 1966-1987 in a demographic surveillance area in rural Bangladesh. Journal of Diarrhoeal Diseases Research, 8:147-I54.

Shkarin VV, Ouchfoun A, Minaev VI, Naceur D (1983) Epidemiology of bacillary dysentery in Algeria. II. The seasonality of dysentery. Zhurnal Mikrobiologii, Epidemiologii i Immunobiologii, 53-58.

Singh RB, Hales S, de Wet N, Raj R, Hearnden M, Weinstein P (200I) The influence of climate variation and change on diarrheal disease in the Pacific Islands. Environmental Health Perspectives, 109:155-159.
Sorvillo F, Beall G, Turner PA et al. (1998) Seasonality and factors associated with cryptosporidiosis among individuals with HIV infection. Epidemiology and Infection, I2I:197-204.

Stewien KE, Dacunha LCF, Alvim AD et al. (1991) Rotavirus associated diarrhea during infancy in the city of S Luis (Ma), Brazil-a 2-year longitudinal-study. Revista do Instituto de Medicina Tropical de São Paulo, 33:459-464.

Stintzing G, Back E, Tufvesson B, Johnsson T, Wadstrom T, Habte D (I98I) Seasonal fluctuations in the occurrence of entero-toxigenic bacteria and rotavirus in pediatric diarrhea in Addis-Ababa. Bulletin of the World Health Organization, 59:67-73.
Sutra S, Srisontrisuk S, Panpurk W et al. (1990) The pattern of diarrhea in children in Khon Kaen, northeastern Thailand: I. The incidence and seasonal variation of diarrhea. Southeast Asian Journal of Tropical Medicine and Public Health, 21:586-593.
Tanser FC, Sharp B, Le Sueur D (2003) Potential effect of climate change on malaria transmission in Africa. The Lancet, 362:I792-I798.

Tsukamoto T, Kinoshita Y, Shimada T, Sakazaki R (1978) Two epidemics of diarrhoeal disease possibly caused by Plesiomonas shigelloides. Journal of Hygiene, 80:275-280.
Tswana SA, Jorgensen PH, Halliwell RW, Kapaata R, Moyo SR (I990) The incidence of rotavirus infection in children from 2 selected study areas in Zimbabwe. Central African Journal of Medicine, 36:24I-246.

Utsalo SJ, Eko FO, Antia-Obong OE (199I) Cholera and Vibrio parahaemolyticus diarrhoea endemicity in Calabar, Nigeria. West African Journal of Medicine, 10:I75-180.
Van den Broeck J, Eeckels R, Devlieger H (1993) Child morbidity patterns in two tropical seasons and associated mortality rates. International Journal of Epidemiology, 22:1104-1110.
Varoli O, Gatti M, Piscolla F (1989) A one-year study on thermophilic campylobacters isolated from fecal specimens. Microbiologica, 12:263-265.
Vlasov VI, Zamotin BA, Burykh VM (1983) Method for the short-term prognosis of annual morbidity indices in Sonne-type dysentery. Zhurnal Mikrobiologii, Epidemiologii i Immunobiologii, 53-56.
Warrick RA, Barrow EM (199I) Climate change scenarios for the UK. Transactions of the Institute of British Geographers, 16:387-399.
Wilcox MH, Cook AM, Eley A, Spencer RC (1992) Aeromonas spp. as a potential cause of diarrhoea in children. Journal of Clinical Pathology, 45:959-963.

Williams EH, Hayes RJ, Smith PG (1986) Admissions to a rural hospital in the West Nile District of Uganda over a 27 year period. Journal ofTropical Medicine and Hygiene, 89:193-211.

WuhibT, Silva TM, Newman RD et al. (1994) Cryptosporidial and micro-sporidial infections in human immunodeficiency virus-infected patients in northeastern Brazil. Journal of Infectious Diseases, I70:494-497.

Yang CR, Meng ZD, Wang X, Li YL, Zhang YX, Zhao QP (I990) Diarrhea surveillance in children aged under 5 years in a rural area of Hebei Province, China. Journal of Diarrhoeal Diseases Research, 8:155-159.

## Chapter $2 I$

# Selected occupational risk factors 

Marisol Concha-Barrientos, Deborah Imel Nelson, Timothy Driscoll, N. Kyle Steenland, Laura Punnett, Marilyn A. Fingerhut, Annette Prüss-Üstün, James Leigh, SangWoo Tak and Carlos Corvalan

## Summary

Many of the 2.9 billion workers across the globe are exposed to hazardous risks at their workplaces. This chapter examines the disease and injury burden produced by selected occupational risk factors: occupational carcinogens, airborne particulates, noise, ergonomic stressors and risk factors for injuries. Owing primarily to lack of data in developing countries, we were unable to include important occupational risks for some cancers, reproductive disorders, dermatitis, infectious diseases, ischaemic heart disease, musculoskeletal disorders (MSDs) of the upper extremities, and other conditions such as workplace stress. Mesothelioma and asbestosis due to asbestos exposure, silicosis and coal workers' pneumoconiosis are almost exclusively due to workplace exposure, but limitations in global data precluded a full analysis of these outcomes.

The economically active population (EAP) aged $\geq 15$ years, which includes people in paid employment, the self-employed, and those who work to produce goods and services for their own household consumption, were considered the group at risk of exposure to occupational hazards. Both formal and informal sectors of employment are included in the EAP, but child labour was excluded. Exposure was quantified based on the economic sector (where people do the work) and on occupation (what people do). Our sources of data to delineate categories of exposed workers included economic databases and publications of the International Labour Organization (ILO) and the World Bank and the published scientific literature. For most risk factors the workers were grouped into high- and low-exposure categories, and the exposed population was distributed by age, sex and subregion. ${ }^{1}$ Risk estimates for the occupational hazards were obtained from the published epidemiological literature, particularly from studies of large populations, reviews and meta-analyses when available.

The occupational risk factors in our study accounted for an estimated $37 \%$ of back pain, $16 \%$ of hearing loss, $13 \%$ of chronic obstructive
pulmonary disease (COPD), $11 \%$ of asthma, $8 \%$ of injuries, $9 \%$ of lung cancer and $2 \%$ of leukaemia. These work-related risks caused 775000 deaths worldwide in 2000. There were five times as many deaths in males as in females ( 647000 vs 128000 ). The leading occupational cause of death among the six risk factors was unintentional injuries ( $41 \%$ ) followed by COPD $(40 \%)$ and cancer of the trachea, bronchus or lung (13\%). Workers who developed outcomes related to the occupational risk factors lost about 22 million years of healthy life. By far the main cause of years of healthy life lost (measured in disability-adjusted life years [DALYs]), within occupational diseases, was unintentional injuries with $48 \%$ of the burden. This was followed by hearing loss due to occupational noise ( $19 \%$ ) and COPD due to occupational agents ( $17 \%$ ). Males experienced almost five times greater loss of healthy years (DALYs) than females. Low back pain and hearing loss have in common the fact that they do not directly produce premature mortality, but they cause substantial disability and have multiple consequences for the individual and society, particularly for workers suffering the outcomes at an early age.

The major source of uncertainty in our analysis was characterizing exposure, which was based solely on economic subsectors and/or occupations and involved a large number of extrapolations and assumptions. High-quality exposure data are lacking, especially in developing countries, and European and American exposure estimates were thus applied in many instances in developing regions. This extrapolation could have substantial impact on the accuracy of analysis for the developing regions if exposures, as usually occur, vary from place to place and over time. Diseases with long latency (e.g. cancers) are more susceptible to the assumptions and extrapolations. In addition to problems produced by the length of the latency period, the magnitude of the excess risk may vary depending on the age of the person when exposure began, the duration and strength of exposure and other concomitant exposures. The turnover of workers is another issue that affects both exposure and risk assessment. Sources of uncertainty in hazard estimates (relative risk and mortality rates) include variations determined from the literature (once again caused by the use of different exposure proxies), extrapolations to regions with different working conditions, the application to females of risk measures from male cohorts, and the application of the same relative risk values to all age groups (e.g. carcinogens). Restricting the analysis to persons aged $\geq 15$ years excludes the quantification of child labour. The exclusion of children in the estimation was due to the wide variation in the youngest age group for which countries reported economic activity rates (EARs). In addition to inconsistent data on EARs for children, there were virtually no data available on their exposure to occupational risk factors or the relative risks of such exposures. Specific, focused research on children is needed to quantify the global burden of disease due to child labour and the resulting implications.

## 1. Introduction

Throughout the world, most adults-and many children-spend much of their waking hours at work. Work provides a number of economic and other benefits. At the same time, people at work face a variety of hazards owing to chemicals, biological agents, physical factors, adverse ergonomic conditions, allergens, a complex network of safety risks, and many and varied psychosocial factors. In addition to injuries, more than 100 occupational diseases have been classified according to the tenth revision of the International Classification of Diseases and Related Health Problems (ICD-10). Broadly, these include respiratory, musculoskeletal, cardiovascular, reproductive, neurotoxic, skin and psychological disorders, hearing loss and cancers.

Of the wide variety of work-related exposures, only the most widespread are evaluated here. Other criteria for selection of risk factors include adequacy of exposure information and the applicability of health outcome data to all regions of the globe, and the inclusion of the relevant health outcomes in the global burden of disease (GBD) database of diseases and injuries.

Exposure to occupational hazards can adversely affect the human body. Adverse effects range from asymptomatic physiological and biochemical changes to symptoms of illness, to diagnosed diseases and, finally, to death. For some risk factors there is a very clear connection between the exposure and the disease. For example, the primary route of exposure to airborne particulates, gases and vapours is inhalation, whereby these agents gain access to the respiratory system and are either deposited (in the case of particulates) or enter the circulatory system (gases and vapours). Many risk factors cause more than one type of outcome of interest. For example, exposure to asbestos can result in malignant conditions of the lung and the pleura, malignant conditions of the peritoneum, and nonmalignant conditions of the lung (asbestosis). Some exposures, such as occupational noise, are well characterized. Others have not been well characterized or are multi-faceted, but the condition they cause is clear (such as occupational injuries).

Following a general description of methods and data sources, individual sections provide details of specific aspects of methodology and results for each of the selected occupational health risk factors that were analysed: occupational carcinogens, occupational airborne particulates, occupational noise, occupational ergonomic stressors and occupational risk factors for injuries.

In this study, the term "occupational risk factor" is defined as a chemical, physical, biological or other agent that may cause harm to an exposed person in the workplace and is potentially modifiable. Figure 21.1 shows the selected risk factors along with related health outcomes. Owing to complex etiology and lack of data, a different approach was developed for some conditions such as asthma and low back pain, using

Figure 21.I Relationship between occupational risk factors and outcomes ${ }^{\text {a }}$

anly selected relationships were quantified.
occupation as a proxy for exposure to the causative agents. The utility of this work as a risk-based framework has thus been limited.

### 1.1 EXCLUDED EXPOSURES AND OUTCOMES

No effects specific to the hazards associated with child labour are addressed in this report owing to a lack of data. Other excluded risks or outcomes include respiratory diseases other than COPD and asthma; some infectious diseases; less widespread cancers and carcinogens (e.g. bladder cancer and cancer of the liver); MSDs such as carpal tunnel syndrome; intentional injuries in the workplace; organ and systemic diseases resulting from occupational exposure to solvents, pesticides and heavy metals such as lead or mercury; maternal and perinatal conditions resulting from occupational exposures; skin disorders, including dermatitis, dermatosis and melanoma; ischaemic heart disease and other outcomes associated with work-related stress.

Malignant mesothelioma of the pleura and peritoneum is virtually uniquely due to asbestos exposure. Occupational dusts can also result in
nonmalignant respiratory diseases other than asthma and COPD. The most important of these are silicosis, asbestosis and coal workers' pneumoconiosis, which are caused by exposure to silica, asbestos and coal dust, respectively. While evidence for a causal relationship is strong, lack of data on accumulated exposure, especially in developing countries, restricted the ability to provide a detailed assessment of attributable mortality and disease burden for these outcomes. Preliminary estimates are provided in the note under Table 21.62 in Section 7.

Because of lack of available data and difficulties in quantification, it was not possible to conduct a global quantitative analysis for the health consequences of stress at work. Overall, the evidence indicates that incidence of stress-related cardiovascular disease is likely to be higher in the blue-collar occupations when the following factors are present: restricted discretion, shiftwork (particularly nightshift), effort-reward imbalance, high demands, poor psychosocial work environment, social isolation, physical inactivity or occupational violence. These risk factors may be interactive. Nurminen and Karjalainen (2001) estimated for Finland an attributable fraction of $16.9 \%$ ( $18.9 \%$ for men and $9.1 \%$ for women) for ischaemic heart disease due to the combined occupational risk factors of shiftwork, noise, and exposure to engine exhaust and environmental tobacco smoke. For ischaemic heart disease, Steenland et al. (2003) used an attributable fraction of $6-18 \%$ for individuals in the United States of America aged 24-64 years, based on the combined effects of noise, job strain (stress), shiftwork and environmental tobacco smoke. Occupational dermatitis accounts for about $10 \%$ of all occupational disease in the United States (Emmett 2002) but exposure data are lacking at global level.

Although there were adequate global data to analyse the risks to health care workers from contaminated sharps (e.g. syringe needles and scalpels), the full analysis has been omitted from this chapter. Since health care workers make up only $0.6 \%$ of the global population, the contribution to hepatitis B, hepatitis C and HIV/AIDS infections on a global level was close to zero. However, health care workers are at high risk of preventable infection from bloodborne pathogens, owing to occupational exposure to infected blood and body fluids.

### 1.2 Choice of theoretical-Minimum-Risk EXPOSURE DISTRIBUTIONS

For some occupational hazards, a theoretical minimum exposure of zero is not possible, as there is some low-level environmental exposure. Two occupational risk factors (carcinogens and airborne particulates) involve workplace exposure at concentrations higher than the environmental or background levels of these substances. For noise, the theoretical minimum was defined as less than 80 dBA , a level found not to have an increased risk of causing hearing loss (NIOSH 1998). For the other risk factors (ergonomic stressors and work-related risk factors for injuries), a category of workers with the lowest risks was identified as the com-
parison group for occupational categories of workers with higher risks. Thus, the theoretical minimum risk corresponds to "no occupational exposure above levels found in the defined comparison group". Selection of a defined comparison group provides a realistic basis for a theoretical minimum, but it does not establish the lowest rate of adverse outcome that could ever be experienced. While it is not expected that occupational exposures will be eliminated in the foreseeable future, it is possible to control exposures through recognized industrial hygiene practices. Engineering controls (including prevention, substitution of materials, process automation, enclosure, process elimination, isolation of workers and process change) constitute effective methods of minimizing exposures (Burton 1997). Administrative controls (such as education and training, work practice controls, worker rotation, maintenance and housekeeping) provide another means of risk reduction.

### 1.3 Data sources

A systematic assessment of the literature was carried out to identify studies on occupational exposures and health outcomes. This included searching Medline, occupational health and safety databases such as OSHROM and NIOSHTIC and databases of various organizations; reviewing relevant references cited in publications identified through the initial literature search and of references cited in these secondary references; communicating with relevant experts; and seeking other information recommended by referees following the initial review of the draft manuscript. PubMed was searched using keywords for exposures and health outcomes, including (separately and in combination, with no limit on year of publication): exposure, occupational, cancer, carcinogen, silica, silicosis, benzene, asbestos, asbestosis, pneumoconiosis and developing country. Names of regions (e.g. Africa, Asia) and specific countries were also used as keywords. A systematic search was conducted using Ovid Healthstar and the former HealthSTAR databases, covering the period 1975-2001. Keywords included: asbestos; asthmagens; chronic obstructive lung disease; cancer and diesel exhaust; arsenic; benzene and leukaemia; ionizing radiation and leukaemia; back (for low back pain); injury; accidents; ergonomics; and hearing impairment and noise.

Studies of large populations, reviews and meta-analyses were specifically sought. Reports and publications were critically assessed to determine their methodology, validity and the characteristics of the population studied.

### 1.4 Estimating risk factor levels

In general, since the types of risk factor to which workers are exposed are primarily influenced by where the work is performed (economic sector) and the type of work they do (occupation), the assessment of proportion of population exposed in each subregion was based on (Figure 21.2):

Figure 21.2 Exposure assessment overview


- economic sector distribution (total nine sectors), used for carcinogens and agents leading to COPD) (Equation 1);
- occupational distribution (occupation within economic sector) (total seven occupations), used for asthmagens, noise and ergonomic stressors (Equation 2); and
- exposure could not be estimated for injury risk factors, and thus estimates of disease burden were made based on the reported rates of the outcome (injury mortality) rather than on exposure.

$$
\begin{align*}
& \operatorname{PEP}(\mathrm{r}, \mathrm{~g}, \mathrm{a})=\operatorname{EAR}(\mathrm{r}, \mathrm{~g}, \mathrm{a}) \times \mathrm{OT}(\mathrm{r}) \\
& \quad \times \operatorname{EPF}(\mathrm{r}) \sum_{i=1}^{9}(\operatorname{PW}(\mathrm{es}(\mathrm{r}, \mathrm{~g}) \mathrm{i}) \times \operatorname{PEW}(\mathrm{es}(\mathrm{r}, \mathrm{~g}) \mathrm{i})  \tag{1}\\
& \operatorname{PEP}(\mathrm{r}, \mathrm{~g}, \mathrm{a})=\operatorname{EAR}(\mathrm{r}, \mathrm{~g}, \mathrm{a}) \times \mathrm{OT}(\mathrm{r}) \\
& \quad \times \operatorname{EPF}(\mathrm{r}) \sum_{i=1}^{7}(\operatorname{PW}(\mathrm{oc}(\mathrm{r}, \mathrm{~g}) \mathrm{i}) \times \operatorname{PEW}(\mathrm{oc}(\mathrm{r}, \mathrm{~g}) \mathrm{i}) \tag{2}
\end{align*}
$$

where
$\operatorname{PEP}(\mathrm{r}, \mathrm{g}, \mathrm{a})=$ proportion of the population occupationally exposed to a specific risk factor in that subregion, by sex and age, at low or high level
$\operatorname{EAR}(\mathrm{r}, \mathrm{g}, \mathrm{a})=$ economic activity rate, by subregion, sex and age
$\mathrm{OT}(\mathrm{r})=$ occupational turnover, if applicable, to account for workers exposed in the past, by subregion
$\operatorname{EPF}(\mathrm{r})=$ exposure partitioning factor, by subregion, to delineate proportion exposed at low or at high levels
$\operatorname{PW}(\mathrm{es}(\mathrm{r}, \mathrm{g}) \mathrm{i})=$ proportion of the population working in economic subsector (i), by subregion and sex

PEW $(e s(r, g)$ i) $=$ proportion of workers in economic subsector (i) with exposure to the specific risk factor, by subregion and sex
$\mathrm{PW}(\mathrm{oc}(\mathrm{r}, \mathrm{g}) \mathrm{i})=$ proportion of the population working in occupational category (i), by subregion and sex

PEW $(\mathrm{oc}(\mathrm{r}, \mathrm{g}) \mathrm{i})=$ proportion of workers in occupational category (i) with exposure to the specific risk factor, by subregion and sex

The differences between the two equations are the term $\operatorname{PW}(\mathrm{es}(\mathrm{r}, \mathrm{g}) \mathrm{i})$ in Equation 1, which is used when exposure data are available by economic sector, and the term PW (oc(r,g)i) in Equation 2, which is used when exposure data are available by occupational category. Occupational turnover (OT), defined as "the rate of replacement of workers due to departures from the workplace", was utilized for carcinogens because health effects due to these risk factors occur many years after exposure (latency) and it was therefore necessary to know how many persons had been exposed in the past to these risk factors. The effects of noise, ergonomic stressors and risk factors for injuries are relatively immediate; latent effects were therefore not a consideration for these risk factors. Additional detail on each term is provided below in the text, and is also summarized in Table 21.1.

The primary data sources used for the exposure assessment and the determination of some of the risk measures (see Table 21.2) included: the World Bank, ILO, the European Union carcinogen exposure (CAREX) database, published literature on prevalence and level of exposure to occupational risk factors, and published literature on epidemiology of health outcomes linked to occupational risk factors, as cited in the relevant sections for each risk factor.

## ECONOMIC ACTIVITY RATE

EAR is defined as the proportion of the economically active population (EAP) among the overall population. EAR was calculated for each region and sex in persons aged $\geq 15$ years, and used to estimate the proportion of the population potentially exposed to occupational risks. EAR provides the most comprehensive accounting of persons who may be exposed to occupational risks, as it includes people in paid employment, the self-employed, and people who work to produce goods and services
Table 2l.I Summary of determinants of population exposure to occupational risk factors

| Term | Comments | Application | Primary data sources |
| :---: | :---: | :---: | :---: |
| EAP | Economically active population is calculated by application of the EAR to the national population | Used for injuries | ILO (2002a) |
| EAR | Economic activity rate, calculated as the EAP in each age group compared to the number of people in that age group, males and females $\geq 15$ years | Used in exposure assessments of all risk factors | ILO (2002a) |
| PW(es) | Proportion working, i.e. fraction of EAP in economic sector. Data on distribution of EAP into three economic sectors (agriculture, industry, service) or nine economic subsectors. Country-level data were weighted by working-age population to develop subregional averages | Used for carcinogens, selected airborne particulates (agents leading to COPD) | World Bank (2001) |
| PW(oc) | Proportion working, i.e. fraction of EAP in occupational category. Country-level data for about 31 countries were weighted by working-age population to develop subregional averages. Owing to lack of country-level data, EMR-B was based on EMR-D data, EUR-C was based on EUR-B data and WPR-A was based on AMR-A data | Used for asthmagens, noise, ergonomic stressors | ILO (1995a); World Bank (200I) |
| PEW | Proportion exposed working, i.e. fraction of population working in economic sector (or in an occupational category) with exposure to risk factor. Owing to data limitations, data from developed countries were usually applied to developing countries, verified where possible by data on specific risk factors from specific countries | PEW(es): carcinogens PEW(es): selected airborne particulates (agents leading to COPD) <br> PEW(oc): selected airborne particulates (asthmagens) <br> PEW (oc): noise <br> PEW(oc): ergonomic stressors | FIOH (1999); Kauppinen et al. (2000) FIOH (1999); Kauppinen et al. <br> (2000); Korn et al. (1987); USEIA (2001) <br> Karjalainen et al. (2002); Kogevinas et al. (1999) <br> NIOSH (1998) <br> Leigh and Sheetz (1989) |
| EPF | Exposure partitioning factor, i.e. proportion of PEW with low- or high-level exposure to risk factor | Carcinogens, selected airborne particulates, noise | NIOSH (1998, 1999, 2000a); Pearce et al. (1994); Yin et al. (1987) |
| OT | Occupational turnover factor. Used only for risk factors for which latent effects must be considered (carcinogens, selected airborne particulates). A factor of 4 was estimated on the basis of published data on labour turnover rates, published cohort data and modelling of cohorts with various mean lengths of exposure. Higher value used for specific regions for coal mining | Carcinogens | K. Steenland, personal communication, 2002 |

Table 21.2 Key sources, data supplied and special characteristics of the sources used to estimate exposure

| Source | Data supplied | Comments |
| :--- | :--- | :--- |
| ILO (1995a, 2000, 2002b) | Employment in economic <br> sectors and subsectors, and <br> in occupations within economic <br> sectors; EARs by age and sex <br> for selected countries | Collected by national EAP <br> surveys. Differences among <br> and within countries (e.g. <br> applicable ages, time period <br> covered) limit international <br> comparability |
| World Bank (200I) | Distribution of EAP (males and <br> females) in agriculture, industry <br> and services; participation of <br> females in the EAP | Based on ILO data |
| FIOH (1999); Kauppinen | Proportion of the working <br> population with occupational <br> exposure to carcinogens in <br> the European Union, by <br> economic sector and subsector, <br> at the 3-digit classification level | Applicable to A subregions, <br> extrapolated to B, C, D <br> and E subregions |
| ElA (200I) | Country-level data on coal <br> production <br> Country-level data on number <br> of coal miners |  |
| ILO (I995b) | Data on noise exposure of <br> American workers | Applicable to A subregions, <br> extrapolated to B, C, D <br> and E subregions |
| NIOSH (I99I, I998); |  |  |

for their own household consumption. According to ILO (2002b), the majority of those who work in the informal sector are included in the "employed" category, and the remainder are in the "unemployed" category; thus, the informal sector workers are included in this analysis. At the same time, persons in precarious or contingent employment often face an increased risk of occupational health and safety hazards, which are not quantified here (Quinlan 2002). The use of EAR for persons aged $\geq 15$ years excludes children under 15 who work.

Estimates and projections of EAP were developed by ILO by applying estimates and projections of activity rates, by sex and age group, to the population estimates and projections assessed by the United Nations (ILO 1996). ILO estimates and projections of economic activity are taken primarily from population censuses and/or sample surveys carried out between 1975 and 1994. ILO also takes data from specific publications by national, interregional and/or international institutions.

Country-level data from the ILO electronic database were used to develop subregion-specific EARs for ages 15 years and above, for males and females (Table 21.3). EARs were estimated for $60-69$-year olds by using data for $60-64$-year olds. Data for people aged $\geq 65$ years were

Table 21.3 Economic activity rates by subregion, sex and age group

|  |  | Age group (years) |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :---: | :---: | :---: | :---: | :---: |
| Subregion $^{\text {a }}$ | Sex | $15-29$ | $30-44$ | $45-59$ | $60-69$ | $70-79$ | $\geq 80$ | Total $\geq 15$ |
| AFR-D | Male | 0.77 | 0.97 | 0.95 | 0.85 | 0.65 | 0.33 | 0.85 |
|  | Female | 0.50 | 0.61 | 0.62 | 0.48 | 0.28 | 0.14 | 0.53 |
| AFR-E | Male | 0.78 | 0.97 | 0.95 | 0.86 | 0.66 | 0.33 | 0.86 |
|  | Female | 0.64 | 0.72 | 0.69 | 0.54 | 0.36 | 0.18 | 0.65 |
| AMR-A (95\%) | Male | 0.70 | 0.93 | 0.87 | 0.50 | 0.13 | 0.07 | 0.73 |
|  | Female | 0.64 | 0.81 | 0.71 | 0.32 | 0.07 | 0.04 | 0.59 |
| AMR-B | Male | 0.78 | 0.97 | 0.89 | 0.66 | 0.33 | 0.17 | 0.82 |
|  | Female | 0.46 | 0.53 | 0.39 | 0.20 | 0.07 | 0.04 | 0.42 |
| AMR-D | Male | 0.71 | 0.98 | 0.96 | 0.86 | 0.61 | 0.31 | 0.82 |
|  | Female | 0.38 | 0.48 | 0.39 | 0.29 | 0.17 | 0.09 | 0.39 |
| EMR-B (90\%) | Male | 0.66 | 0.97 | 0.92 | 0.74 | 0.45 | 0.23 | 0.79 |
|  | Female | 0.33 | 0.37 | 0.26 | 0.18 | 0.09 | 0.05 | 0.31 |
| EMR-D (40\%) | Male | 0.73 | 0.97 | 0.94 | 0.76 | 0.44 | 0.22 | 0.82 |
|  | Female | 0.37 | 0.43 | 0.37 | 0.25 | 0.12 | 0.06 | 0.37 |
| EUR-A | Male | 0.66 | 0.96 | 0.84 | 0.35 | 0.05 | 0.03 | 0.68 |
|  | Female | 0.59 | 0.74 | 0.56 | 0.14 | 0.02 | 0.01 | 0.47 |
| EUR-B | Male | 0.72 | 0.96 | 0.80 | 0.41 | 0.22 | 0.11 | 0.74 |
|  | Female | 0.56 | 0.77 | 0.59 | 0.23 | 0.12 | 0.06 | 0.54 |
| EUR-C | Male | 0.72 | 0.97 | 0.89 | 0.30 | 0.11 | 0.06 | 0.74 |
| SEAR-B | Female | 0.61 | 0.94 | 0.74 | 0.17 | 0.05 | 0.03 | 0.58 |
|  | Male | 0.74 | 0.98 | 0.94 | 0.73 | 0.44 | 0.22 | 0.83 |
| SEAR-D (95\%) | Male | Female | 0.55 | 0.70 | 0.65 | 0.44 | 0.21 | 0.11 |
| WPR-A | Female | 0.45 | 0.98 | 0.95 | 0.72 | 0.53 | 0.27 | 0.85 |
|  | Male | 0.67 | 0.97 | 0.95 | 0.69 | 0.30 | 0.15 | 0.76 |
|  | Female | 0.57 | 0.70 | 0.67 | 0.36 | 0.13 | 0.07 | 0.52 |
|  | Male | 0.81 | 0.98 | 0.92 | 0.61 | 0.29 | 0.15 | 0.84 |
|  | Female | 0.77 | 0.89 | 0.67 | 0.29 | 0.09 | 0.05 | 0.71 |

[^78]applied to the $70-79$ age group. The $\geq 80$ age group was estimated at one half of the rate for the $\geq 65$ age group (by comparison with countrylevel data, which is reported by some countries for elderly workers) (ILO 2001).

## Proportion of the population working in an economic sector or occupational category

The distinction between "where people work", i.e. economic sector and "what they do", i.e. occupation, is important in exposure characterization. For example, within the economic subsector of manufacturing there

Table 21.4 Illustration of the ISIC classification system used in exposure assessment

| Economic sector | Economic subsectors | Occupational categories |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Professional | Administration | Clerical | Sales | Service | Agriculture | Production |

Agriculture

| Industry | Mining |
| :--- | :--- |
|  | Manufacturing |
|  | Electrical |
| Construction |  |
| Services | Trade |
|  | Transport |
|  | Finance |
|  | Services |

Source: ILO 1987.
are people who work as production workers, but also people who work as clerical or sales people (Table 21.4). EAP was used for injuries. EAP by economic sector and subsector was used for carcinogens and agents leading to COPD, because available data do not distinguish exposures by occupational category within economic sectors. For asthmagens, noise and ergonomic stressors, the analyses were conducted on the basis of exposure by occupational category within economic sectors.

The approach used here is based on the International Standard Industrial Classification of All Economic Activities (ISIC), an economic classification system of the United Nations, which organizes all economic activities by economic sectors and relevant subgroupings (ILO 1987; UN 2000). The ISIC system is used almost universally by national and international statistical services to categorize economic activity, and therefore allowed us to make global comparisons. Table 21.4 illustrates the ISIC classification scheme of economic sectors, economic subsectors and occupational categories that were used to estimate exposures to workers in this project. We did not subdivide agriculture into economic subsectors.

## Economic sector

For each subregion, a weighted proportion of working men and women (EAP) in each of the three economic sectors was constructed (Table 21.5) (World Bank 2001, data from 1990 and 1996-1998). Economic sector employment data were used to subdivide the number of workers in industry into the economic subsectors of mining, manufacturing, electricity (and other utilities) and construction. In a similar manner, the data for the service sector were subdivided into the economic subsectors of trade, transport, finance and services. The agriculture sector was not subdivided.
Table 21.5 EAP distribution in economic sectors and subsectors, by subregion and sex

| Subregion ${ }^{\text {a }}$ | Sex | Agriculture | Industry |  |  |  | Services |  |  |  | Sum |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Mining | Manufacturing | Electricity | Construction | Trade | Transport | Finance | Services |  |
| AFR-D (20\%) | Male | 0.55 | 0.01 | 0.09 | 0.01 | 0.04 | 0.06 | 0.04 | 0.03 | 0.16 | 1.00 |
|  | Female | 0.68 | 0.00 | 0.05 | 0.00 | 0.01 | 0.06 | 0.03 | 0.02 | 0.16 | 1.00 |
| AFR-E (20\%) | Male | 0.55 | 0.01 | 0.09 | 0.01 | 0.04 | 0.05 | 0.03 | 0.03 | 0.18 | 1.00 |
|  | Female | 0.65 | 0.00 | 0.03 | 0.00 | 0.00 | 0.05 | 0.02 | 0.02 | 0.22 | 1.00 |
| AMR-A (95\%) | Male | 0.05 | 0.01 | 0.21 | 0.01 | 0.09 | 0.21 | 0.06 | 0.10 | 0.26 | 1.00 |
|  | Female | 0.02 | 0.00 | 0.09 | 0.00 | 0.01 | 0.23 | 0.03 | 0.14 | 0.47 | 1.00 |
| AMR-B | Male | 0.20 | 0.01 | 0.15 | 0.04 | 0.08 | 0.10 | 0.05 | 0.25 | 0.12 | 1.00 |
|  | Female | 0.12 | 0.00 | 0.12 | 0.00 | 0.01 | 0.14 | 0.01 | 0.05 | 0.55 | 1.00 |
| AMR-D (70\%) | Male | 0.07 | 0.01 | 0.16 | 0.01 | 0.11 | 0.18 | 0.07 | 0.09 | 0.30 | 1.00 |
|  | Female | 0.03 | 0.00 | 0.12 | 0.00 | 0.00 | 0.31 | 0.01 | 0.06 | 0.46 | 1.00 |
| EMR-B (5\%) | Male | 0.15 | 0.01 | 0.16 | 0.03 | 0.11 | 0.15 | 0.08 | 0.05 | 0.27 | 1.00 |
|  | Female | 0.09 | 0.00 | 0.09 | 0.00 | 0.01 | 0.20 | 0.05 | 0.09 | 0.47 | 1.00 |
| EMR-D (20\%) | Male | 0.45 | 0.00 | 0.11 | 0.01 | 0.07 | 0.30 | 0.00 | 0.03 | 0.02 | 1.00 |
|  | Female | 0.68 | 0.00 | 0.10 | 0.01 | 0.01 | 0.20 | 0.00 | 0.00 | 0.01 | 1.00 |
| EUR-A | Male | 0.06 | 0.01 | 0.27 | 0.01 | 0.11 | 0.01 | 0.00 | 0.05 | 0.48 | 1.00 |
|  | Female | 0.05 | 0.00 | 0.15 | 0.00 | 0.02 | 0.01 | 0.01 | 0.12 | 0.64 | 1.00 |

Table 21.5 EAP distribution in economic sectors and subsectors, by subregion and sex (continued)

| Subregion ${ }^{\text {a }}$ | Sex | Agriculture | Industry |  |  |  | Services |  |  |  | Sum |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Mining | Manufacturing | Electricity | Construction | Trade | Transport | Finance | Services |  |
| EUR-B (70\%) | Male | 0.29 | 0.02 | 0.20 | 0.02 | 0.08 | 0.07 | 0.04 | 0.07 | 0.20 | 1.00 |
|  | Female | 0.44 | 0.00 | 0.16 | 0.01 | 0.01 | 0.03 | 0.01 | 0.04 | 0.30 | 1.00 |
| EUR-C (35\%) | Male | 0.21 | 0.04 | 0.14 | 0.03 | 0.15 | 0.05 | 0.20 | 0.12 | 0.06 | 1.00 |
|  | Female | 0.16 | 0.03 | 0.12 | 0.02 | 0.04 | 0.12 | 0.24 | 0.12 | 0.15 | 1.00 |
| SEAR-B (30\%) | Male | 0.46 | 0.01 | 0.12 | 0.01 | 0.07 | 0.14 | 0.06 | 0.01 | 0.13 | 1.00 |
|  | Female | 0.45 | 0.00 | 0.15 | 0.00 | 0.01 | 0.22 | 0.01 | 0.01 | 0.15 | 1.00 |
| SEAR-D (80\%) | Male | 0.53 | 0.02 | 0.13 | 0.02 | 0.03 | 0.01 | 0.06 | 0.04 | 0.17 | 1.00 |
|  | Female | 0.80 | 0.01 | 0.10 | 0.00 | 0.01 | 0.00 | 0.00 | 0.02 | 0.06 | 1.00 |
| WPR-A | Male | 0.05 | 0.00 | 0.24 | 0.01 | 0.13 | 0.17 | 0.10 | 0.09 | 0.21 | 1.00 |
|  | Female | 0.06 | 0.00 | 0.17 | 0.00 | 0.04 | 0.27 | 0.04 | 0.11 | 0.31 | 1.00 |
| WPR-B (95\%) | Male | 0.44 | 0.03 | 0.14 | 0.01 | 0.05 | 0.09 | 0.06 | 0.02 | 0.16 | 1.00 |
|  | Female | 0.40 | 0.01 | 0.12 | 0.01 | 0.01 | 0.17 | 0.06 | 0.06 | 0.17 | 1.00 |
| When data were not available for all countries, the percentage of the regional working age population ( $\geq 15$ years) represented by data is e.g. Grenada, were not included in these calculations. |  |  |  |  |  |  |  |  |  |  |  |
| Note: AFR-D and AFR-E (combined), EMR-B, EMR-D and SEAR-D data are based on 1990 employment data from the World Bank world develop limited. All others are taken from 1996-1998 World Bank EAP data. Subregional averages were calculated using country values weighted by |  |  |  |  |  |  |  |  |  |  |  |

## Occupational category

Regional tables of occupation within economic sector distributions were constructed using the number of employed people by occupation and economic sector. For comparison purposes, data were obtained from one source (ILO 1995a). For a subregion with only one country represented, the distribution of occupation within economic sector was assumed to represent the regional employment patterns. Where more than one country was represented, a weighted average was constructed. Where there were no data for the subregion, patterns for the most similar subregion were applied (EMR-B based on EMR-D, EUR-C based on EUR-B and WPR-A based on AMR-A). Because of limited data on occupational distribution by sex within an economic sector, the same distribution (i.e. proportional division) was applied within a subregion to ages 15 and above, and to males and females. The A subregions had higher proportions of EAP in the professional, managerial and administrative categories, while the $\mathrm{B}, \mathrm{C}, \mathrm{D}$ and E subregions had proportionally more workers in the production categories.

## Proportion of workers in an economic sector or occupational CATEGORY WITH EXPOSURE

Worldwide data on worker exposure are limited. Therefore, several assumptions were made, validated where possible, to establish the proportion of workers exposed to a specific risk factor within an economic sector (PEW). More detail is presented in the sections on specific risk factors.

## EXPOSURE PARTITIONING FACTOR (EPF)

In order to partition into high and low exposure groups those workers exposed to carcinogens, we chose the United States Occupational Health and Safety Administration (OSHA) Permissible Exposure Levels (PELs). For most carcinogens we were then able to estimate the risks for the low and high exposure groups from the literature.

The OSHA PELs state a level of the agent that can never be exceeded in the workplace (usually based on eight-hour time-weighted average exposures), and these have had the force of law in the United States as maximum limits of exposure since the creation of OSHA in 1971. Similar occupational exposure limits (OELs) have been promulgated as law by many countries, particularly in the A subregions, and as recommendations by professional expert groups. It is generally considered that a longterm mean exposure in a "minimally controlled" work environment will be in the range 0.3-0.5 times the PEL (Hewett 1996). For example, the American Industrial Hygiene Association suggests that a typical longterm average exposure may be one third of an eight-hour PEL (Roach 1992).

A different approach was used for asthmagens and agents leading to COPD. The actual disease-causing exposures themselves, within these occupations, are either generic (e.g. dust) or too numerous to be useful (e.g. there are over 200 known asthmagens). In both instances there were no international data on the number of workers exposed, which dictated the approach of using occupations or economic subsectors. For asthma, different relative risks were available for eight large occupational groups, while for COPD we partitioned the overall relative risk for the exposed population into high and low relative risks, and assigned these to different economic subsectors according to Korn et al. (1987).

## Occupational turnover (OT)

Cancers and lung diseases have long latency periods and once the disease process has begun the worker continues to be at risk, even after exposure ceases. This means that persons who were exposed in the past must be considered as ever-exposed, even if they are currently working in nonexposed jobs or have retired. Furthermore, OT increases the number of persons ever exposed to an occupational risk. This approach was consistent with cohorts represented in the epidemiological studies from which relative risks were taken. The OT factor was not utilized in estimating the numbers of workers exposed to noise, ergonomic stressors or risk factors leading to occupational injuries, as these risk factors do not have latent effects. No turnover was estimated for asthma and COPD owing to a lack of sufficient information on the applicability to studies in which relative risk was measured. Table 21.6 presents data from the literature on annual OT rates in various countries and industries throughout the world, organized by subregion. These reports did not indicate if employees were new to the job or to the industry, although several studies were at the company level, indicating that the worker was new to the company. Therefore, to account for previously unexposed workers entering jobs with carcinogen or dust exposures, an annual turnover rate (ATR) of $10 \%$ was selected for all subregions.

An adjustment factor (noted as OT) to account for annual turnover in jobs with exposure to occupational carcinogens was determined as follows:

Computation of adjustment factor to correct for occupational turnover (OT)

$$
\begin{align*}
& \text { Adjustment factor, } \mathrm{OT}=\mathrm{P}_{\mathrm{t}} / \mathrm{P}_{0} \\
& \quad=[\text { original workers }+ \text { new workers }- \text { deaths }] / \text { original workers }  \tag{3}\\
& =\left\{\mathrm{P}_{0}+\left[\mathrm{P}_{0} \times \text { ATR } \times \mathrm{t}\right]-\left[(\text { mortality rate })\left(\mathrm{P}_{0}+\left(\mathrm{P}_{0} \times \text { ATR } \times \mathrm{t}\right)\right]\right\} / \mathrm{P}_{0}\right.
\end{align*}
$$

where
$P_{t}=$ the proportion who have ever been occupationally exposed, during a period of 40 years, still living
Table 21.6 Turnover rates in various industries and countries

| Country or area | Basis of measurement | Annual turnover rate (ATR) | Comments | Source |
| :---: | :---: | :---: | :---: | :---: |
| A subregions Italy | Metal—mechanical engineering industry | 13.4\% | Industry level, based on 2729 observations | Lucifora (1998) |
| Italy |  | 26\% | Total worker turnover rate, including accession and separation | Lucifora (1998) |
| Spain Basque | Industrial production cooperative (manufacturing) | 3\% | 65 firms-cited as low rate | Johnson and Whyte (1977) |
| United States | Restaurant industry | 500\% | 8 southern restaurants | Butler and Twaddle (1979) |
| United States | Garment manufacturing | 140\% | I53 female workers at a single plant in the south-west | Koch and Rhodes (1981) |
| United States | One interstate trucking firm | 40\% | Expected rate for 1997-truck drivers | EIU (1997a) |
| United States California | Silicon Valley, one financial firm | 25\% | Software services group | EIU (1997f) |
| United States <br> New Mexico | State-wide survey by New Mexico Department of Labor, January-March 2001 | 25\% per quarter, ranging from $29 \%$ in agriculture to $15 \%$ in public administration | Agriculture rates show greatest seasonal variation | Moffett (2002) |
| B subregions Brazil | Brazilian labour market | 47\% | Cited as higher than most markets for which data are available | EIU (1997b) |

Table 21.6 Turnover rates in various industries and countries (continued)

| Country or area | Basis of measurement | Annual turnover <br> rate (ATR) | Comments |
| :--- | :--- | :--- | :--- | :--- | (1996a)

$\mathrm{P}_{0}=$ the proportion who are occupationally exposed at time $\mathrm{t}=0$
ATR $=$ turnover/year, taken as 0.10
$\mathrm{t}=$ time, taken as 40 years, a typical working lifetime
mortality $=20 \%$ of total cohort, based on published death rates of about 5 deaths per thousand over a period of 40 years (Minino and Smith 2001).

Equation 3 results in an adjustment factor of OT=4 to correct for occupational turnover over a 40 -year period with a median exposure duration of 10 years.

In addition to knowing the numbers of workers exposed to agents with latent health effects, in some cases it was also useful to know the duration of exposure to agents with latent effects for outcomes for which the risks were based on cumulative exposure. Cohort modelling was conducted to determine the typical duration of exposure (K. Steenland, personal communication, 2002). This modelling assumed that people worked for a maximum of 40 years, that $10 \%$ of the workers were replaced each year, and that $20 \%$ died over the 40 -year period. Exposure durations were randomly selected from a log-normal distribution. Persons were also randomly assigned a starting age at entry between 20 and 45 years, and were assumed to retire at age 65 years if they had not already left the cohort by that age. A steady-state working population was produced by using a log-normal distribution for exposure with a geometric mean of 9 years. Using this, the mean length of exposure (in years) at the end of 40 years could be estimated (by age) for all persons ever exposed in the cohort. The average estimated length of exposure, as shown in Table 21.7, was 9.8 years, which is consistent with data on a wide range of cohorts presented in the published literature (Steenland et al. 1991a, 1991b, 2001b).

### 1.5 RISK FACTOR-DISEASE RELATIONSHIP

Risk measures (relative risks or mortality rates) for the health outcomes resulting from exposure to the risk factors considered in this study were determined primarily from peer-reviewed, published studies. Adjustments were made, as appropriate, to account for differences in levels of exposure, exposure duration and/or age, sex and subregion.

- For carcinogens leading to cancer of the lung, trachea or bronchus, and for leukaemogens, composite values were taken from the literature and adapted to exposure patterns in the various subregions.
- For asthma, the relative risks for different occupations were taken from Karjalainen et al. (2002), with the exception of work in agriculture, for which the relative risk was taken from Kogevinas et al. (1999).

Table 21.7 Exposure duration after 40 years in model cohort

| Age group (years) | Number | Total exposure (years) | Average exposure (years) |
| :--- | :---: | :---: | :---: |
| $15-29$ | 12 | 50 | 4.2 |
| $30-44$ | 86 | 575 | 6.7 |
| $45-59$ | 117 | 1182 | 10.1 |
| $60-69$ | 105 | 1195 | 11.4 |
| $70-79$ | 53 | 618 | 11.7 |
| $\geq 80$ | 19 | 234 | 12.3 |
| All ages | 392 | 3854 | 9.8 |

- For COPD, the relative risks for different economic subsectors were taken from Korn et al. (1987).
- For noise, relative risks of noise-induced hearing loss were calculated from data on hearing loss in workers with different levels of noise exposure in the United States (NIOSH 1998).
- The relative risks of low back pain, given employment in different occupational categories, were taken from Leigh and Sheetz (1989).
- Owing to heterogeneity of factors leading to occupational injuries, relative risks could not be extrapolated from one setting to another. As a result, the mortality rates for workers exposed to risk factors leading to injuries were estimated for different subregions from various sources, including Laborsta (ILO 2001).


## 2. OcCupational Carcinogens

The International Agency for Research on Cancer (IARC 2002) has classified 150 chemical or biological agents or exposure situations as known or probable human carcinogens. IARC has classified 87 agents, mixtures or exposure circumstances as Group 1 (carcinogenic to humans), including various chemical compounds, pharmaceuticals and bacterial and viral infections. Many are encountered in occupational settings, e.g. asbestos and cadmium. An additional 63 agents, mixtures or exposure circumstances have been classified as Group 2A (probably carcinogenic to humans). Those with occupational significance include diesel fumes and benzidine-based dyes (IARC 2001). Although IARC classifies agents according to their overall carcinogenicity, specific sites are also considered.

Work-related malignant conditions can arise from a large variety of occupational exposures. However, the main groups of conditions are relatively few-lung cancer and leukaemia. The exposures selected for assessment in this study were based on how common they may be, the
risk arising from exposure, the strength of evidence and the availability of data. Table 21.8 shows the definition of each of the chemical and physical agents, along with the related cancer.

The analysis included relevant Group 1 and 2A carcinogens, with the following exceptions.

- Tetrachloroethylene and trichloroethylene, both classified in Group 2A, were not included as carcinogens because the evidence for cancer is weak.
- The aromatic amines and dyes, including 2-naphthylamine, benzidinebased dyes, benzidine and 4,4'-methylenebis(2-chloroaniline) (also known as MOCA) were excluded owing to lack of data for developing countries.
- Occupational carcinogens with extremely limited exposures (e.g. bischloromethyl ether, also known as BCME) were not included.
- Compounds for which exposure estimates were not available from the CAREX database (e.g. soot, xenylamine, 4-nitrobiphenyl and polycyclic aromatic hydrocarbons) were not included.
- Although radon is an IARC Group I carcinogen with large estimated exposures, it was excluded from consideration owing (i) to worldwide differences in naturally occurring radon emissions, (ii) to wide variations in climate and construction methods, which substantially affect the concentration of radon retained in buildings, and (iii) to difficulties in separating occupational and nonoccupational radon exposures.

Other conditions have insufficient relevant exposure data, insufficient risk data or insufficient number of cases worldwide to allow them to be usefully included. These conditions include:

- bladder cancer (aromatic amines, benzidine dyes, MOCA);
- liver (vinyl chloride);
- nasal cavity and middle ear (hardwood dust, chromium VI compounds, nickel compounds);
- bone and articular cartilage (ionizing radiation);

Table 21.8 Occupational carcinogens and health outcomes

| Occupational carcinogen | Outcome |
| :--- | :--- |
| Arsenic, asbestos, beryllium, cadmium, chromium, diesel exhaust, | Cancer of the trachea, <br> bronchus or lung <br> nickel, silica |
| Benzene, ethylene oxide, ionizing radiation | Leukaemia |

- skin (arsenic, by-products of distillation, ionizing radiation); and
- lung cancer due to passive smoking in the workplace.


### 2.1 EXposure variable and theoretical-minimum-risk EXPOSURE

Exposure was divided into three categories: background, low and high. The occupational risk factors for cancer involve workplace exposure, at concentrations higher than background level, to various chemical and physical agents that are known to cause malignant neoplasms. Thus, the theoretical minimum risk corresponds to "no occupational exposure to physical, chemical or biological agents or other factors above background levels".

### 2.2 Estimating risk factor levels

The general exposure assessment methodology was described earlier. This assessment was based on the distribution of the EAP by economic subsector, because the primary exposure data sources used in this analysis organized carcinogen exposure data by economic subsector (Equation 1). The regional distributions of workers into economic subsectors were adjusted by data on the carcinogens to which people in the various economic subsectors were exposed. As described earlier, an adjustment factor of 4 was used to account for turnover in jobs with exposure to occupational carcinogens.

The primary data source on work-related exposure to carcinogens for each economic subsector $(\operatorname{PEW}(\mathrm{es}(\mathrm{r}, \mathrm{g}) \mathrm{i})$ in Equation 1) is the CAREX database (FIOH 1999), which presents data on the number of workers in the European Union exposed to 139 carcinogens (IARC Group 1, 2A and selected 2B agents) at levels above background in 1990-1993. Table 21.9 lists the CAREX data for the carcinogens in our study. These estimates were based on national workforce data and exposure prevalence estimates from Finland and the United States, adjusted for the economic structure of each country, then refined by national experts.

It was assumed that the proportion of workers exposed to a particular carcinogen in a specific economic subsector was constant throughout the world. To check the validity of this assumption, the literature was searched for estimates of the number of workers exposed to silica. Silica was chosen as an indicator because there are more data on silica available for developing countries than on other carcinogens. This search yielded a range of study types, from rough estimates (Zou Changqi et al. 1997) to studies in which air concentrations were measured in workplaces (Yin et al. 1987). Estimates of the number of workers exposed to silica in China, Thailand and Viet Nam, and to benzene in China, were compared to the number of persons employed in that country, either in a specific economic sector or overall. The results obtained were compared with CAREX data. With a few exceptions, the estimated fraction
Table 21.9 Mean proportions of workers exposed to selected carcinogens, by economic sector and subsector, in the

| Carcinogen | Agriculture | Mining | Manufacturing | Electrical | Construction | Trade | Transport | Finance | Services |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Silica | 0.00372 | 0.23049 | 0.02327 | 0.01415 | 0.18860 | 0.00017 | 0.00476 | 0.00002 | 0.00061 |
| Cadmium | 0.00000 | 0.00000 | 0.00487 | 0.00287 | 0.00291 | 0.00002 | 0.00065 | 0.00000 | 0.00047 |
| Nickel | 0.00000 | 0.02025 | 0.01680 | 0.00352 | 0.00047 | 0.00007 | 0.00003 | 0.00000 | 0.00043 |
| Arsenic | 0.00054 | 0.00072 | 0.00400 | 0.00148 | 0.00134 | 0.00006 | 0.00000 | 0.00002 | 0.00011 |
| Chromium | 0.00000 | 0.00346 | 0.02079 | 0.00409 | 0.00237 | 0.00017 | 0.00370 | 0.00000 | 0.00225 |
| Diesel fumes | 0.00646 | 0.21970 | 0.01110 | 0.03358 | 0.05816 | 0.00485 | 0.13438 | 0.00000 | 0.00914 |
| Beryllium | 0.00000 | 0.00055 | 0.00207 | 0.00070 | 0.00004 | 0.00002 | 0.00011 | 0.00000 | 0.00003 |
| Asbestos | 0.01248 | 0.10248 | 0.00590 | 0.01702 | 0.05203 | 0.00292 | 0.00684 | 0.00016 | 0.00284 |
| Benzene | 0.00100 | 0.00200 | 0.00300 | 0.00100 | 0.00100 | 0.01000 | 0.00500 | 0.00000 | 0.02000 |
| lonizing radiation | 0.00000 | 0.01100 | 0.00000 | 0.03400 | 0.00000 | 0.00000 | 0.00400 | 0.00000 | 0.00000 |
| Ethylene oxide | 0.00012 | 0.00137 | 0.00060 | 0.00006 | 0.00027 | 0.00000 | 0.00002 | 0.00000 | 0.00057 |

[^79]of workers exposed to silica was equal to or higher in these countries than indicated by CAREX (Juengprasert 1997; T. Nguyen, personal communication, 2001; NIEHS 1999; Phan Hong Son et al. 1999; Yin et al. 1987; Zou Changqi et al. 1997). For example, the proportion of workers exposed to silica in manufacturing in Viet Nam is $3.7 \%$, compared to the CAREX estimate of $2.3 \%$.

It was assumed that, within a given economic subsector, both male and female workers and younger and older workers had the same probability of exposure. For example, if $2.3 \%$ of people working in manufacturing were exposed to silica, it was assumed that $2.3 \%$ of males and $2.3 \%$ of females working in manufacturing were exposed to silica, young and old alike. There were, however, fewer females working in manufacturing, so that at the population level the proportion of females with exposure to silica was lower than that of males.

There are few data on the distribution of exposure monitoring values, which are needed to accurately estimate the proportion of workers exposed to above or below a specific value ( $\mathrm{EPF}(\mathrm{r}$ ) in Equation 1). Therefore, the demarcation between low and high exposure was established as the PELs enforced by OSHA. Some reasons for selecting the PELs as partitioning values include the following.

- Exposure data for the United States are often reported based on "compliance with" or "exceeding" the PELs.
- The risks corresponding to low or high exposure have been linked to the PELs.
- As cancers have long latency periods, the exposures of concern have occurred several decades in the past. The OSHA PELs for many carcinogens have not changed since their adoption in 1971, allowing a stable benchmark for comparison (Table 21.10).

The peer-reviewed literature was searched for studies that included proportions of workers exposed above and below particular levels. There are many reports of exposures to contaminants in the literature, and even on the distribution of exposures at low and high levels in developed countries. However, there are few data on distribution of exposure values for developing countries. A summary of the major sources used to decide how to partition exposure values for carcinogens for the $\mathrm{B}, \mathrm{C}, \mathrm{D}$ and E subregions is presented in Tables 21.11 and 21.12 for benzene and metals, respectively.

The following data were used to partition exposure for A subregions:

- Finnish data (Partanen et al. 1995), indicating 11-94\% exposed above $0.2 \mathrm{mg} / \mathrm{m}^{3}$ respirable silica in a range of industries;
- NIOSH (1999) estimates of proportions of workers exposed above the PELs of $4 \%$ (asbestos) and $13.6 \%$ (silica); and

Table 21.10 OSHA permissible exposure levels (PELs) for carcinogens

| Chemical/ physical agent | PEL | Source | Comment |
| :---: | :---: | :---: | :---: |
| Arsenic | Inorganic: $10 \mu \mathrm{~g} / \mathrm{m}^{3}$ | OSHA, 29 CFR 1910.1018 | Effective 1978 |
|  | Organic: $0.5 \mathrm{mg} / \mathrm{m}^{3}$ | OSHA, 29 CFR 1910.1000, Table Z-I | Effective 1971 |
| Asbestos | Varied ${ }^{\text {a }}$ |  |  |
| Benzene | 10 ppm | OSHA, 29 CFR 1910.1000, Table Z-2 | Effective 1971 |
|  | I ppm | OSHA, 29 CFR 1910.1028 | Effective 1987 |
| Beryllium | $2 \mu \mathrm{~g} / \mathrm{m}^{3}$ | $\text { OSHA, } 29 \text { CFR I910.1000, }$ <br> Table Z-2 | Effective 1971 |
| Cadmium | Fume: $0.1 \mathrm{mg} / \mathrm{m}^{3}$ | $\begin{aligned} & \text { OSHA, } 29 \text { CFR 1910.1000, } \\ & \text { Table Z-2 } \end{aligned}$ | Effective 1971 |
|  | Dust: $0.2 \mathrm{mg} / \mathrm{m}^{3}$ | OSHA, 29 CFR 1910.1000, <br> Table Z-2 | Effective 1971 |
|  | $5 \mu \mathrm{~g} / \mathrm{m}^{3}$ | 29 CFR 1910.1027 | Effective 1992 |
| Chromium | Chromic acid and chromates: $0.1 \mathrm{mg} / \mathrm{m}^{3}$ Chromium metal: $1 \mathrm{mg} / \mathrm{m}^{3}$ | OSHA, 29 CFR 1910.1000, <br> Table Z-2 (ceiling) <br> OSHA, 29 CFR 1910.1000, <br> Table Z-I | Effective 1971 <br> Effective 1971 |
| Diesel exhaust | NA |  |  |
| Ethylene oxide | 1 ppm | OSHA, 29 CFR 1910.1047 | Effective 1984 |
| lonizing radiation | Rems/calendar quarter: whole body, I.25; hands, forearms, feet, ankles, 18.75; skin, 7.5 | OSHA, 29 CFR I910.1096, Table G-I8 | Effective 1974 |
| Nickel | Metal, insoluble and soluble compounds: $1 \mathrm{mg} / \mathrm{m}^{3}$ | OSHA, 29 CFR 1910.1000, Table Z-I | Effective 1971 |
| Silica | Respirable quartz: $\left(10 \mathrm{mg} / \mathrm{m}^{3}\right) /$ (per cent $\mathrm{SiO}_{2}+2$ ) | $\text { OSHA, } 29 \text { CFR I910.1000, }$ <br> Table Z-3 | Effective 1971. <br> For $100 \%$ silica dust, this is equivalent to $0.1 \mathrm{mg} / \mathrm{m}^{3}$. Halve this value for cristobalite and tridymite |

NA Not applicable.
a As shown in this table, most of the PELs have not changed since they were put in place. However, there were considerable changes in the United States PEL for asbestos during the years of interest to the current analysis, with a level before 1972 of 12 fibres $/ \mathrm{ml}$ before the first OSHAissued PEL in 1972 decreasing, through several steps, to 0.1 fibres $/ \mathrm{ml}$ in 1994 (Martonik et al. 2001; Nelson 1997).

Source: USDOL OSHA (2002a).

Table 2I.II Occupational exposure to benzene in developing countries

| Country | Industry | Concentration | Year (or year reported) |
| :--- | :--- | :--- | :---: |
| Egypt | Rubber coating | $0-74 \mathrm{mg} / \mathrm{m}^{3}$ | $(1986)$ |
| Turkey | Shoemaking | $48-96 \mathrm{mg} / \mathrm{m}^{3}$ <br> $672 \mathrm{mg} / \mathrm{m}^{3}($ maximum level) | 1970 |
| India | Petrol pump | $4.5 \mathrm{mg} / \mathrm{m}^{3}(\mathrm{mean})$ | 1991 |
| China | Various: paint, <br> chemical, varnish <br> works, shoemaking | $0.06-850 \mathrm{mg} / \mathrm{m}^{3}$ | $(1987)$ |
|  | Steel workers | $960-3200 \mathrm{mg} / \mathrm{m}^{3}$ per day <br> Brazil | $140 \mathrm{mg} / \mathrm{m}^{3}$, maximum of <br> Petrochemical |
|  |  |  | $(1993)$ |

Source: Pearce et al. (1994).

Table 21.12 Occupational exposures to metals in developing countries

| Country | Industry | Concentration | Year (or year reported) |
| :--- | :--- | :--- | :---: |
| China | Tin mine | Arsenic: $0.42 \mathrm{mg} / \mathrm{m}^{3}$, mean | 1952 |
|  |  | Arsenic: $0.01 \mathrm{mg} / \mathrm{m}^{3}$ | 1980 s |
| China | Cadmium refining | Cadmium: $0.04-0.074 \mathrm{mg} / \mathrm{m}^{3}$ | 1970 s |
| Singapore | Storage battery factory | Cadmium: $0.13-58.3 \mathrm{mg} / \mathrm{m}^{3}$, | 1980 |
|  |  | geometric means of three <br> sets of samples |  |
| China | Chromate production | $0.02-21.3 \mathrm{mg} / \mathrm{m}^{3}$ | $1960 \mathrm{~s}-1980$ |
|  |  | $0.55 \mathrm{mg} / \mathrm{m}^{3}$, mean | $(1989)$ |

Source: Pearce et al. (1994).

- NIOSH (2000b) data on miners, indicating silica exposures above the PEL for $8 \%$ of coal mine samples, $16 \%$ of metal mine samples, $9 \%$ of stone mine samples and $8 \%$ of sand and gravel facility samples.

For the $\mathrm{B}, \mathrm{C}, \mathrm{D}$ and E subregions, important evidence includes:

- Chinese data (Dosemeci et al. 1995), indicating roughly three quarters of samples above $0.1 \mathrm{mg} / \mathrm{m}^{3}$ respirable silica;
- a study of a South African brickworks (Myers et al. 1989), in which $45 \%$ of presented sample values were above $0.1 \mathrm{mg} / \mathrm{m}^{3}$ and roughly two thirds and four fifths of samples in medium and dusty areas, respectively, were above $0.1 \mathrm{mg} / \mathrm{m}^{3}$ respirable silica;
- a study of a South African pottery (Rees et al. 1992), where roughly three quarters of samples that included silica analysis were above the Threshold Limit Value (TLV); and

Table 21.13 Exposure partition factors for carcinogens for the A and for the $B, C, D$ and $E$ subregions

| Subregion | Proportion of exposed workers with <br> low exposures (at or below the PEL) | Proportion of exposed workers with <br> high exposures (above the PEL) |
| :--- | :---: | :---: |
| A | 0.90 | 0.10 |
| B, C, D and E | 0.50 | 0.50 |

- the Chinese benzene study (Yin et al. 1987), in which $35 \%$ of over 50000 workplaces had concentrations at or above $40 \mathrm{mg} / \mathrm{m}^{3}$, in comparison to the current OSHA PEL of $3.2 \mathrm{mg} / \mathrm{m}^{3}$ for benzene, and in which the benzene concentration in $86 \%$ of 141 shoe factories was above $25 \mathrm{mg} / \mathrm{m}^{3}$.

Based on these data, partition factors for carcinogen exposures were determined for the A and for the $\mathrm{B}, \mathrm{C}, \mathrm{D}$ and E subregions, as shown in Table 21.13.

## LUNG CARCINOGENS

The proportions of the population exposed to the occupational lung carcinogens included in the study (Table 21.8) are shown in Tables 21.14 and 21.16 by subregion, age, sex and level of exposure.

## LEUKAEMOGENS

The proportions of the population exposed to occupational leukaemogens (Table 21.8) are presented in Table 21.15 by subregion, age, sex and level of exposure.

### 2.3 RISK FACTOR-DISEASE RELATIONSHIPS

Relative risk estimates were used for lung carcinogens and leukaemogens. Table 21.17 summarizes the chemical or physical agent, the specific cancer and the key data sources that provided evidence of the link between the two. These review studies assessed risk measures for the main sites of occupational cancer, including the lung (which, for the purposes of this study, includes the trachea, bronchus and lung), the haematopoietic system (represented in this study by leukaemia) and malignant mesothelioma.

Relative risks for lung cancer and leukaemia were taken from studies of cohorts of workers with variable exposure durations and intensities, variable periods from the last exposure and variable lengths of followup. They therefore compare exposed with unexposed groups. In preparing relative risk estimates for exposure outcomes of interest, several assumptions were made:

Table 21.14 Proportions of the population exposed to lung carcinogens by subregion, age, sex and level of exposure

| Subregion | Sex | Exposure level | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | Background | 0.837 | 0.837 | 0.837 | 0.837 | 0.837 | 0.837 |
|  |  | Low | 0.082 | 0.082 | 0.082 | 0.082 | 0.082 | 0.082 |
|  |  | High | 0.082 | 0.082 | 0.082 | 0.082 | 0.082 | 0.082 |
|  | Female | Background | 0.934 | 0.934 | 0.934 | 0.934 | 0.934 | 0.934 |
|  |  | Low | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 |
|  |  | High | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 |
| AFR-E | Male | Background | 0.839 | 0.839 | 0.839 | 0.839 | 0.839 | 0.839 |
|  |  | Low | 0.080 | 0.080 | 0.080 | 0.080 | 0.080 | 0.080 |
|  |  | High | 0.080 | 0.080 | 0.080 | 0.080 | 0.080 | 0.080 |
|  | Female | Background | 0.929 | 0.929 | 0.929 | 0.929 | 0.929 | 0.929 |
|  |  | Low | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 |
|  |  | High | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 |
| AMR-A | Male | Background | 0.802 | 0.802 | 0.802 | 0.802 | 0.802 | 0.802 |
|  |  | Low | 0.178 | 0.178 | 0.178 | 0.178 | 0.178 | 0.178 |
|  |  | High | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 |
|  | Female | Background | 0.936 | 0.936 | 0.936 | 0.936 | 0.936 | 0.936 |
|  |  | Low | 0.058 | 0.058 | 0.058 | 0.058 | 0.058 | 0.058 |
|  |  | High | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 |
| AMR-B | Male | Background | 0.793 | 0.793 | 0.793 | 0.793 | 0.793 | 0.793 |
|  |  | Low | 0.103 | 0.103 | 0.103 | 0.103 | 0.103 | 0.103 |
|  |  | High | 0.103 | 0.103 | 0.103 | 0.103 | 0.103 | 0.103 |
|  | Female | Background | 0.951 | 0.951 | 0.951 | 0.951 | 0.951 | 0.951 |
|  |  | Low | 0.024 | 0.024 | 0.024 | 0.024 | 0.024 | 0.024 |
|  |  | High | 0.024 | 0.024 | 0.024 | 0.024 | 0.024 | 0.024 |
| AMR-D | Male | Background | 0.761 | 0.761 | 0.761 | 0.761 | 0.761 | 0.761 |
|  |  | Low | 0.119 | 0.119 | 0.119 | 0.119 | 0.119 | 0.119 |
|  |  | High | 0.119 | 0.119 | 0.119 | 0.119 | 0.119 | 0.119 |
|  | Female | Background | 0.961 | 0.961 | 0.961 | 0.961 | 0.961 | 0.961 |
|  |  | Low | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
|  |  | High | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
| EMR-B | Male | Background | 0.760 | 0.760 | 0.760 | 0.760 | 0.760 | 0.760 |
|  |  | Low | 0.120 | 0.120 | 0.120 | 0.120 | 0.120 | 0.120 |
|  |  | High | 0.120 | 0.120 | 0.120 | 0.120 | 0.120 | 0.120 |
|  | Female | Background | 0.963 | 0.963 | 0.963 | 0.963 | 0.963 | 0.963 |
|  |  | Low | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
|  |  | High | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
| EMR-D | Male | Background | 0.840 | 0.840 | 0.840 | 0.840 | 0.840 | 0.840 |
|  |  | Low | 0.080 | 0.080 | 0.080 | 0.080 | 0.080 | 0.080 |
|  |  | High | 0.080 | 0.080 | 0.080 | 0.080 | 0.080 | 0.080 |
|  | Female | Background | 0.955 | 0.955 | 0.955 | 0.955 | 0.955 | 0.955 |
|  |  | Low | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 |
|  |  | High | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 |

Table 21.14 Proportions of the population exposed to lung carcinogens by subregion, age, sex and level of exposure (continued)

| Subregion | Sex | Exposure level | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EUR-A | Male | Background | 0.802 | 0.802 | 0.802 | 0.802 | 0.802 | 0.802 |
|  |  | Low | 0.179 | 0.179 | 0.179 | 0.179 | 0.179 | 0.179 |
|  |  | High | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 |
|  | Female | Background | 0.937 | 0.937 | 0.937 | 0.937 | 0.937 | 0.937 |
|  |  | Low | 0.057 | 0.057 | 0.057 | 0.057 | 0.057 | 0.057 |
|  |  | High | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 |
| EUR-B | Male | Background | 0.779 | 0.779 | 0.779 | 0.779 | 0.779 | 0.779 |
|  |  | Low | 0.111 | 0.111 | 0.111 | 0.111 | 0.111 | 0.111 |
|  |  | High | 0.111 | 0.111 | 0.111 | 0.111 | 0.111 | 0.111 |
|  | Female | Background | 0.920 | 0.920 | 0.920 | 0.920 | 0.920 | 0.920 |
|  |  | Low | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 |
|  |  | High | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 |
| EUR-C | Male | Background | 0.654 | 0.654 | 0.654 | 0.654 | 0.654 | 0.654 |
|  |  | Low | 0.173 | 0.173 | 0.173 | 0.173 | 0.173 | 0.173 |
|  |  | High | 0.173 | 0.173 | 0.173 | 0.173 | 0.173 | 0.173 |
|  | Female | Background | 0.801 | 0.801 | 0.801 | 0.801 | 0.801 | 0.801 |
|  |  | Low | 0.099 | 0.099 | 0.099 | 0.099 | 0.099 | 0.099 |
|  |  | High | 0.099 | 0.099 | 0.099 | 0.099 | 0.099 | 0.099 |
| SEAR-B | Male | Background | 0.798 | 0.798 | 0.798 | 0.798 | 0.798 | 0.798 |
|  |  | Low | 0.101 | 0.101 | 0.101 | 0.101 | 0.101 | 0.101 |
|  |  | High | 0.101 | 0.101 | 0.101 | 0.101 | 0.101 | 0.101 |
|  | Female | Background | 0.922 | 0.922 | 0.922 | 0.922 | 0.922 | 0.922 |
|  |  | Low | 0.039 | 0.039 | 0.039 | 0.039 | 0.039 | 0.039 |
|  |  | High | 0.039 | 0.039 | 0.039 | 0.039 | 0.039 | 0.039 |
| SEAR-D | Male | Background | 0.805 | 0.805 | 0.805 | 0.805 | 0.805 | 0.805 |
|  |  | Low | 0.098 | 0.098 | 0.098 | 0.098 | 0.098 | 0.098 |
|  |  | High | 0.098 | 0.098 | 0.098 | 0.098 | 0.098 | 0.098 |
|  | Female | Background | 0.934 | 0.934 | 0.934 | 0.934 | 0.934 | 0.934 |
|  |  | Low | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 |
|  |  | High | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 |
| WPR-A | Male | Background | 0.745 | 0.745 | 0.745 | 0.745 | 0.745 | 0.745 |
|  |  | Low | 0.230 | 0.230 | 0.230 | 0.230 | 0.230 | 0.230 |
|  |  | High | 0.026 | 0.026 | 0.026 | 0.026 | 0.026 | 0.026 |
|  | Female | Background | 0.914 | 0.914 | 0.914 | 0.914 | 0.914 | 0.914 |
|  |  | Low | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |
|  |  | High | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
| WPR-B | Male | Background | 0.769 | 0.769 | 0.769 | 0.769 | 0.769 | 0.769 |
|  |  | Low | 0.115 | 0.115 | 0.115 | 0.115 | 0.115 | 0.115 |
|  |  | High | 0.115 | 0.115 | 0.115 | 0.115 | 0.115 | 0.115 |
|  | Female | Background | 0.875 | 0.875 | 0.875 | 0.875 | 0.875 | 0.875 |
|  |  | Low | 0.063 | 0.063 | 0.063 | 0.063 | 0.063 | 0.063 |
|  |  | High | 0.063 | 0.063 | 0.063 | 0.063 | 0.063 | 0.063 |

Table 21.15 Proportions of the population exposed to leukaemogens by subregion, age, sex and level of exposure

| Subregion | Sex | Exposure level | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | Background | 0.980 | 0.980 | 0.980 | 0.980 | 0.980 | 0.980 |
|  |  | Low | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
|  |  | High | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
|  | Female | Background | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
|  |  | Low | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
|  |  | High | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
| AFR-E | Male | Background | 0.979 | 0.979 | 0.979 | 0.979 | 0.979 | 0.979 |
|  |  | Low | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
|  |  | High | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
|  | Female | Background | 0.984 | 0.984 | 0.984 | 0.984 | 0.984 | 0.984 |
|  |  | Low | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
|  |  | High | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
| AMR-A | Male | Background | 0.973 | 0.973 | 0.973 | 0.973 | 0.973 | 0.973 |
|  |  | Low | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
|  | Female | Background | 0.970 | 0.970 | 0.970 | 0.970 | 0.970 | 0.970 |
|  |  | Low | 0.027 | 0.027 | 0.027 | 0.027 | 0.027 | 0.027 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
| AMR-B | Male | Background | 0.980 | 0.980 | 0.980 | 0.980 | 0.980 | 0.980 |
|  |  | Low | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
|  |  | High | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
|  | Female | Background | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 |
|  |  | Low | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
|  |  | High | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
| AMR-D | Male | Background | 0.967 | 0.967 | 0.967 | 0.967 | 0.967 | 0.967 |
|  |  | Low | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 |
|  |  | High | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 |
|  | Female | Background | 0.979 | 0.979 | 0.979 | 0.979 | 0.979 | 0.979 |
|  |  | Low | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
|  |  | High | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
| EMR-B | Male | Background | 0.969 | 0.969 | 0.969 | 0.969 | 0.969 | 0.969 |
|  |  | Low | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 |
|  |  | High | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 |
|  | Female | Background | 0.984 | 0.984 | 0.984 | 0.984 | 0.984 | 0.984 |
|  |  | Low | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
|  |  | High | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
| EMR-D | Male | Background | 0.984 | 0.984 | 0.984 | 0.984 | 0.984 | 0.984 |
|  |  | Low | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
|  |  | High | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
|  | Female | Background | 0.995 | 0.995 | 0.995 | 0.995 | 0.995 | 0.995 |
|  |  | Low | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |

Table 21.15 Proportions of the population exposed to leukaemogens by subregion, age, sex and level of exposure (continued)

| Subregion | Sex | Exposure level | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EUR-A | Male | Background | 0.968 | 0.968 | 0.968 | 0.968 | 0.968 | 0.968 |
|  |  | Low | 0.029 | 0.029 | 0.029 | 0.029 | 0.029 | 0.029 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
|  | Female | Background | 0.973 | 0.973 | 0.973 | 0.973 | 0.973 | 0.973 |
|  |  | Low | 0.024 | 0.024 | 0.024 | 0.024 | 0.024 | 0.024 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
| EUR-B | Male | Background | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
|  |  | Low | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
|  |  | High | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
|  | Female | Background | 0.983 | 0.983 | 0.983 | 0.983 | 0.983 | 0.983 |
|  |  | Low | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
|  |  | High | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
| EUR-C | Male | Background | 0.982 | 0.982 | 0.982 | 0.982 | 0.982 | 0.982 |
|  |  | Low | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
|  |  | High | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
|  | Female | Background | 0.981 | 0.981 | 0.981 | 0.981 | 0.981 | 0.981 |
|  |  | Low | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
|  |  | High | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
| SEAR-B | Male | Background | 0.980 | 0.980 | 0.980 | 0.980 | 0.980 | 0.980 |
|  |  | Low | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
|  |  | High | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
|  | Female | Background | 0.985 | 0.985 | 0.985 | 0.985 | 0.985 | 0.985 |
|  |  | Low | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
|  |  | High | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
| SEAR-D | Male | Background | 0.979 | 0.979 | 0.979 | 0.979 | 0.979 | 0.979 |
|  |  | Low | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
|  |  | High | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
|  | Female | Background | 0.995 | 0.995 | 0.995 | 0.995 | 0.995 | 0.995 |
|  |  | Low | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
| WPR-A | Male | Background | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 |
|  |  | Low | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
|  | Female | Background | 0.979 | 0.979 | 0.979 | 0.979 | 0.979 | 0.979 |
|  |  | Low | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
|  |  | High | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| WPR-B | Male | Background | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
|  |  | Low | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
|  |  | High | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
|  | Female | Background | 0.980 | 0.980 | 0.980 | 0.980 | 0.980 | 0.980 |
|  |  | Low | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
|  |  | High | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |

Table 21.16 Proportions of the population exposed to asbestos by subregion, age, sex and level of exposure

| Subregion | Sex | Exposure level | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | Background | 0.961 | 0.961 | 0.961 | 0.961 | 0.961 | 0.961 |
|  |  | Low | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
|  |  | High | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
|  | Female | Background | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
|  |  | Low | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
|  |  | High | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
| AFR-E | Male | Background | 0.961 | 0.961 | 0.961 | 0.961 | 0.961 | 0.961 |
|  |  | Low | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 |
|  |  | High | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 |
|  | Female | Background | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 |
|  |  | Low | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 |
|  |  | High | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 |
| AMR-A | Male | Background | 0.973 | 0.973 | 0.973 | 0.973 | 0.973 | 0.973 |
|  |  | Low | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
|  | Female | Background | 0.991 | 0.991 | 0.991 | 0.991 | 0.991 | 0.991 |
|  |  | Low | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
|  |  | High | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| AMR-B | Male | Background | 0.966 | 0.966 | 0.966 | 0.966 | 0.966 | 0.966 |
|  |  | Low | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 |
|  |  | High | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 |
|  | Female | Background | 0.992 | 0.992 | 0.992 | 0.992 | 0.992 | 0.992 |
|  |  | Low | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 |
|  |  | High | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 |
| AMR-D | Male | Background | 0.965 | 0.965 | 0.965 | 0.965 | 0.965 | 0.965 |
|  |  | Low | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 |
|  |  | High | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 |
|  | Female | Background | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 |
|  |  | Low | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
| EMR-B | Male | Background | 0.963 | 0.963 | 0.963 | 0.963 | 0.963 | 0.963 |
|  |  | Low | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 |
|  |  | High | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 |
|  | Female | Background | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 |
|  |  | Low | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
| EMR-D | Male | Background | 0.962 | 0.962 | 0.962 | 0.962 | 0.962 | 0.962 |
|  |  | Low | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
|  |  | High | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
|  | Female | Background | 0.985 | 0.985 | 0.985 | 0.985 | 0.985 | 0.985 |
|  |  | Low | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
|  |  | High | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |

Table 21.16 Proportions of the population exposed to asbestos by subregion, age, sex and level of exposure (continued)

| Subregion | Sex | Exposure level | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EUR-A | Male | Background | 0.971 | 0.971 | 0.971 | 0.971 | 0.971 | 0.971 |
|  |  | Low | 0.026 | 0.026 | 0.026 | 0.026 | 0.026 | 0.026 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
|  | Female | Background | 0.991 | 0.991 | 0.991 | 0.991 | 0.991 | 0.991 |
|  |  | Low | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
|  |  | High | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| EUR-B | Male | Background | 0.962 | 0.962 | 0.962 | 0.962 | 0.962 | 0.962 |
|  |  | Low | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
|  |  | High | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
|  | Female | Background | 0.982 | 0.982 | 0.982 | 0.982 | 0.982 | 0.982 |
|  |  | Low | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
|  |  | High | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
| EUR-C | Male | Background | 0.949 | 0.949 | 0.949 | 0.949 | 0.949 | 0.949 |
|  |  | Low | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 |
|  |  | High | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 |
|  | Female | Background | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 |
|  |  | Low | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 |
|  |  | High | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 |
| SEAR-B | Male | Background | 0.959 | 0.959 | 0.959 | 0.959 | 0.959 | 0.959 |
|  |  | Low | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 |
|  |  | High | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 |
|  | Female | Background | 0.981 | 0.981 | 0.981 | 0.981 | 0.981 | 0.981 |
|  |  | Low | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
|  |  | High | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
| SEAR-D | Male | Background | 0.959 | 0.959 | 0.959 | 0.959 | 0.959 | 0.959 |
|  |  | Low | 0.021 | 0.021 | 0.021 | 0.021 | 0.021 | 0.021 |
|  |  | High | 0.021 | 0.021 | 0.021 | 0.021 | 0.021 | 0.021 |
|  | Female | Background | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
|  |  | Low | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
|  |  | High | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
| WPR-A | Male | Background | 0.967 | 0.967 | 0.967 | 0.967 | 0.967 | 0.967 |
|  |  | Low | 0.030 | 0.030 | 0.030 | 0.030 | 0.030 | 0.030 |
|  |  | High | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
|  | Female | Background | 0.988 | 0.988 | 0.988 | 0.988 | 0.988 | 0.988 |
|  |  | Low | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
|  |  | High | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| WPR-B | Male | Background | 0.955 | 0.955 | 0.955 | 0.955 | 0.955 | 0.955 |
|  |  | Low | 0.022 | 0.022 | 0.022 | 0.022 | 0.022 | 0.022 |
|  |  | High | 0.022 | 0.022 | 0.022 | 0.022 | 0.022 | 0.022 |
|  | Female | Background | 0.974 | 0.974 | 0.974 | 0.974 | 0.974 | 0.974 |
|  |  | Low | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 |
|  |  | High | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 |

Table 21.17 Examples of sources used to assess the risk factor-disease relationship for selected occupational carcinogens

| Selected risk factor | Health outcome | Examples of key sources of evidence of causality |
| :--- | :--- | :--- |
| Lung carcinogens | Cancer of the trachea, <br> bronchus or lung | Nurminen and Karjalainen (200I); Steenland <br> et al. (1996, 2003) |
| Leukaemogens | Leukaemia | Lynge et al. (1997); BEIR V (I990); IARC <br> (1997) |

- that relative risks are the same for men and women;
- that relative risk values are constant with age; and
- that the relative risks apply equally to the risk of developing the malignant condition (incident cases) and to the risk of dying from the condition (fatal cases); where relative risk values were based on disease incidence studies, the incidence ratio was comparable to the corresponding mortality risk ratio.

Steenland et al. (1996) estimated for the United States the relative risk of exposure to nine lung carcinogens (arsenic, asbestos, beryllium, cadmium, chromium, diesel fumes, nickel, silica and radon). They did not consider agents to which relatively few workers were exposed (BCME, coke oven and coal gasification fumes and soot) and they did not consider smoking, beyond the selection where possible of relative risk factors that had been adjusted for smoking. Combined relative risk values (ranging from 1.31 to 3.69 ) were calculated for all but radon, using inverse variance and a random-effects model and relying on major cohort studies of the specific agents. The authors estimated that 9000-10000 men and 900-1900 women develop lung cancer annually in the United States owing to past exposure to occupational carcinogens (except radon). This would account for approximately $9 \%$ of lung cancer deaths in males and $2 \%$ in females, or $0.5 \%$ of all deaths annually in the United States.

Steenland et al. (2003) examined the population-attributable risk (PAR) from several studies (including Steenland et al. 1996). They applied the PAR to deaths occurring in 1997 in the United States to determine occupational deaths from lung cancer, among other outcomes. The authors determined a PAR for lung cancer in the range $6.1-17.3 \%$ for men and $2 \%$ for women. For overall cancer they determined a PAR of $7-19 \%$ for men and $11 \%$ for women. For leukaemia, a combined PAR for men and women of $0.8-2.8 \%$ was calculated.

Nurminen and Karjalainen (2001) estimated the proportion of fatalities related to occupational factors in Finland. The average number of exposed workers in Finland was estimated from census data by sex, age, occupation and industry, and the FINJEM national job-exposure matrix.

Relative risks were obtained from a review of epidemiological studies, focusing on risk estimates that were most valid for the Finnish exposure circumstances. The attributable fraction methodology was used to determine the proportion of deaths in the population attributable to occupational factors. The authors reported that $30 \%$ of deaths due to occupational disease in Finland in 1996 were caused by cancer. Occupational lung cancer accounted for $0.9 \%$ of all deaths. They attributed $24 \%$ of cancer of the bronchus and lung ( $29 \%$ for men and $5.3 \%$ for women) to occupational exposure to combined risk factors. The attributable fractions for urinary cancer were $10.3 \%$ overall- $14.2 \%$ for men and $0.7 \%$ for women. Combined occupational risk factors resulted in $10.9 \%$ ( $18.5 \%$ for males, $2.5 \%$ for females) of leukaemia deaths being attributed to occupational exposures, the majority ( $17.8 \%$ and $2.3 \%$, respectively) from electrical occupations, in contrast to $0.7 \%$ and $0.2 \%$, respectively, from benzene. An average of $71.3 \%$ ( $90 \%$ for males, $25 \%$ for females) of malignant mesothelioma was attributed to occupation.

The three review papers described above (Nurminen and Karjalainen 2001; Steenland et al. 1996, 2003) provided summary measures, or information that can be used to determine summary measures, of relative risk for one or more of the main agents and outcomes of interest. The study by Nurminen and Karjalainen (2001) focused on Finland, and preferentially used studies based in Scandinavia or thought to be most relevant to Finland. Most of its relative risk estimates relate to lung cancer, with attributable fractions presented for leukaemia. The paper by Steenland et al. (1996), although focused on the United States, was more inclusive of studies of suitable quality. The other paper by Steenland et al. (2003) provided information on relative mortality risks similar to the first (1996) paper. All papers provided similar summary measures of relative risk for lung cancers, but the Steenland et al. (1996) results were used preferentially because they are generally based on a broader range of studies. However, the Steenland paper provided information only on lung cancer. Table 21.18 gives a summary of the risk measures for each of the carcinogens and the relevant outcomes. The basis for these risk estimations is described in more detail below.

## LUNG CANCER

The evidence for substance-specific relative risk values, which were used to calculate the overall relative risk for the eight lung carcinogens, is briefly discussed below, relying heavily on the review paper by Steenland et al. (1996). The data in the paper provide a summary relative risk of 1.6 for occupational exposure to the set of lung carcinogens considered here.

Smoking is the main important potential confounder of lung cancer, and potentiates the effect of some exposures (notably with asbestos and lung cancer). In this analysis, where possible, studies were used that produced risk estimates for lung cancer after controlling for smoking.
Table 21.18 Summary of risk measures (relative risk and mortality rates) for occupational carcinogens

| Health outcome | Risk measure | Estimate and 95\% confidence interval (CI) | Comments | Primary data sources |
| :---: | :---: | :---: | :---: | :---: |
| Cancer of the trachea, bronchus and lung | Relative risk | Low exposure: 1.22 (I.09-1.35) to 1.32 (I.17-1.48) <br> High exposure: 1.79 (1.59-1.97) to 1.93 (1.71-2.16) | Composite relative risk based on individual relative risk of arsenic, asbestos, beryllium, cadmium, chromium, diesel exhaust, nickel and silica | Steenland et al. (1996) |
| Leukaemia | Relative risk | Low exposure: 1.67 (1.5I-2.00) to 1.93 (1.76-2.17) <br> High exposure: 3.06 (2.64-3.80) to 3.86 (3.48-4.32) | Composite relative risk based on individual relative risk of benzene, ionizing radiation and ethylene oxide | BEIR V (I990); IARC (2000); <br> Lynge et al. (I997); <br> Steenland et al. (2003) |

## Arsenic

Arsenic is accepted as a Group 1 carcinogen (IARC 1980, 1987a). The six principal epidemiological studies (covering nearly 18000 workers) reviewed by Steenland et al. (1996) indicated a combined relative risk of 3.69, with a range of $1.31-15.2$ reported for individual studies and a clear dose-response relationship. The lowest relative risk arose from a study in which exposures mostly ranged from 7 to $13 \mu \mathrm{~g} / \mathrm{m}^{3}$, compared to the OSHA level of $10 \mu \mathrm{~g} / \mathrm{m}^{3}$ (Enterline et al. 1987). Excess cancers in other studies were probably due to high exposures that occurred largely in the past. A combined relative risk of 3.69 ( $95 \%$ CI 3.06-4.46) was determined by the Steenland et al. (1996) review, whereas 3.2 was used by Nurminen and Karjalainen (2001).

## Asbestos

Both serpentine and amphibole asbestos have been shown to cause lung cancer in humans, with a clear dose-response relationship and a synergy between asbestos and tobacco (Lee 2001). Over 100 cohort studies and many case-referent studies, plus animal and cellular studies, provide ample evidence for causation. In six cohort studies of nearly 6000 asbestosis patients, the standardized mortality rate ranged from 3.5 to 9.1, with a combined relative risk of 5.9 . In 20 studies of over 100000 asbestos workers, the standardized mortality rate ranged from 1.04 for chrysotile workers to 4.97 for amosite workers, with a combined relative risk of 2.00. It is difficult to determine the exposures involved because few of the studies reported measurements, and because it is a problem to convert historical asbestos measurements in millions of dust particles per cubic foot to gravimetric units. Nevertheless, little excess lung cancer is expected from low exposure levels. These studies have been the subject of several reviews (IARC 1977; IPCS 1998; Nurminen and Karjalainen 2001; Steenland et al. 1996). The main papers provided a range of relative risks (1.04-7.4), with summary relative risks of 2.0 (Steenland et al. 1996) and 2.3 (Nurminen and Karjalainen 2001) cited in the two most recent reviews. The lower value $(2.0,95 \%$ CI 1.90-2.11), which is based on a wider range of studies, is accepted for this analysis.

## Beryllium

Beryllium is an IARC Group 1 carcinogen (IARC 1993), although epidemiological evidence is rather limited. A standardized mortality rate for lung cancer of 2.0 was determined from a registry cohort of 689 women and men (Steenland and Ward 1991), and an overall standardized mortality rate of 1.24 was found in a study of 9225 male workers from seven beryllium plants ( 1.49 at plants with higher exposure) (Ward et al. 1992). Steenland et al. (1996) utilized a smoking-adjusted relative risk of 1.49 (no $95 \%$ CI reported), based on a beryllium plant with high exposures.

## Cadmium

Cadmium is an IARC Group 1 carcinogen (IARC 1993). The best epidemiological evidence of its relationship to lung cancer comes from a cohort study by Stayner et al. (1992), although the evidence for carcinogenicity is stronger in animals and has recently been questioned in humans (Jarup and Nordberg 1998). The most recent follow-up study suggests a relative risk of 1.49 (95\% CI 0.96-2.22) (Steenland et al. 1996). Nurminen and Karjalainen (2001) used 1.2, based on a Scandinavian study.

## Chromium

Chromium is an IARC Group 1 carcinogen (IARC 1990a). There is ample epidemiological evidence of its causal association with lung cancer, with many cohort studies showing a dose-response relationship. Based on the largest and best designed studies of chromium production workers, producers of chromate paints and chromate plating workers, the overall relative risk is 2.78 ( $95 \%$ CI 2.47-3.52) (Steenland et al. 1996). Nurminen and Karjalainen (2001) used a lower relative risk of 1.4 from a hospital-based case-referent study.

## Diesel exhaust

Polycyclic aromatic hydrocarbons comprise the main components of diesel exhaust, which contains a mixture of substances. Diesel exhaust has been accepted as a Group 2A carcinogen (IARC 1989) and was scheduled for further review in 2001. Owing to limitations in exposure assessment to diesel exhaust, human epidemiology has been difficult to conduct. However, cohort studies and meta-analyses confirm a relationship between diesel exhaust exposure and lung cancer, with summary relative risks in the range 1.3-1.5 (Bhatia et al. 1998; Lipsett and Campleman 1999). Based on six relatively consistent recent studies with good documentation of exposure to diesel exhaust, in which the number of cases ranged from 50 to 1256, Steenland et al. (1996) determined a combined relative risk of 1.31 ( $95 \%$ CI 1.13-1.44), and Nurminen and Karjalainen (2001) used the same estimate.

## Nickel

Nickel is an IARC Group 1 carcinogen (IARC 1990a). Based on data from the 1990 report of the International Committee on Nickel Carcinogenesis in Man (ICNCM 1990), Steenland et al. (1996) calculated a combined relative risk of 1.56 ( $95 \%$ CI 1.41-1.73). Nurminen and Karjalainen (2001) used an estimate of 1.4 based on a Finnish study.

Silica
On the basis of detailed reviews, silica has been classified as an IARC Group 1 carcinogen (IARC 1987b, 1997). Several cohort studies in silica-
exposed and silicosis patients showed a dose-response relationship between silica exposure and lung cancer relative risk, and this was confirmed by meta-analyses and a pooled study (Steenland and Sanderson 2001). Animal and cellular studies provided supporting evidence. Controversy remains as to whether silicosis is a necessary precursor for the development of lung cancer, but this does not affect the underlying status of silica as a carcinogen (Checkoway 2000; Hnizdo and Sluis-Cremer 1991; Soutar et al. 2000). Steenland et al. (1996) based their combined relative risk of 1.33 ( $95 \%$ CI 1.21-1.45) on 13 large cohort and case-control studies of silica-exposed workers. These studies included granite workers, stone workers, pottery workers, brick workers, gold miners and diatomaceous earth miners, and covered a range of workers generally numbering from almost 1000 to over 5000. Half of the studies controlled for smoking. Nurminen and Karjalainen (2001) used a slightly higher estimate of 1.4.

## Combined estimates

A common methodology, similar to that used by Steenland et al. (1996) and Nurminen and Karjalainen (2001), was used in this analysis for all lung carcinogens, in that occupational exposure to carcinogens was estimated and applied to relative risk estimates to enable the determination of attributable fractions. A mean relative risk of 1.63 was determined for eight lung carcinogens (not including radon), using data reported by Steenland et al. (1996). This was done by calculating a weighted average of the substance-specific relative risks, and weighting the substance-specific relative risks by the proportion of workers exposed to each substance to determine a mean relative risk for workers exposed to the eight lung carcinogens. This was done separately for each subregion, using the proportion of workers in each subregion exposed to specific agents to weight the relative risk for each of the agents. However, the resulting average relative risks were not clearly different from each other (all were close to 1.6).

In addition, to estimate an uncertainty range for the initial mean relative risk, a weighted average was calculated of the lower and upper $95 \%$ CI values of the relative risk reported for each substance (except beryllium, for which there were no estimated CI). These values (not to be confused with the partitioned relative risk values for low- and highlevel exposure) were within $15 \%$ of the mean relative risk values. This is demonstrated for the AMR-A subregion (Table 21.19).

To produce relative risk estimates for low and high exposure, it was necessary to partition the mean relative risks into values that correspond to low- and high-level exposure. A mean relative risk (of 1.6) was determined for the United States. Based on the estimates of $90 \%$ of American workers exposed at or below about one fifth of the PEL values and $10 \%$ exposed at or above the PEL values, and an estimate of the American population-attributable fraction of lung cancer due to occupation

Table 21.19 Lung cancer relative risk, substance-specific and weighted average, for the AMR-A subregion

| Carcinogen | Combined relative risk ${ }^{\text {a }}$ (95\% Cl) | Proportion of workers exposed |
| :---: | :---: | :---: |
| Silica | 1.33 (1.21-1.45) | 0.0248 |
| Cadmium | 1.49 (0.96-2.22) | 0.0015 |
| Nickel | 1.56 (1.41-1.73) | 0.0039 |
| Arsenic | 3.69 (3.06-4.46) | 0.0011 |
| Chromium | 2.78 (2.47-3.52) | 0.0055 |
| Diesel fumes | 1.31 (1.13-1.44) | 0.0217 |
| Beryllium | 1.49 | 0.0005 |
| Asbestos | 2.00 (1.90-2.11) | 0.0094 |
| Total ${ }^{\text {b }}$ | 1.59 (1.41-1.77) |  |
| a Derived from major epidemiological studies. |  |  |
| Weighted summary relative risk, weighted by the proportion of workers exposed to each contributing carcinogen. |  |  |
| Source: Steenland et al. (1996). |  |  |

of $9 \%$ (Steenland et al. 1996), the mean relative risk of 1.6 was partitioned into a relative risk of 1.3 for low-level exposure to lung carcinogens, and 1.9 for high-level exposure. The United States ratios of the lower (1.3/1.6) and the higher (1.9/1.6) relative risks to the average relative risk were then applied to the average relative risks estimated for each subregion to produce estimated relative risks at low and high exposures for each subregion. In the same manner, upper and lower 95\% CI were produced for these relative risks, based on the limits estimated for the average relative risks. The results of this process are shown in Table 21.20.

## LEUKAEMIA

Leukaemia has been linked to exposure to benzene, ionizing radiation and ethylene oxide, all of which are IARC Group 1 carcinogens (IARC 2001; WHO 1999). There is also some evidence that exposure to low-frequency electric fields may be leukaemogenic (Nurminen and Karjalainen 2001; WHO 2001). However, as this physical agent has not been included in CAREX, it has been excluded from this study.

## Benzene

The causal relationship between leukaemia and benzene is well recognized, including data from cohort studies in China and the United States covering workers in chemical plants, refineries, machine production, and textile and cloth factories. Excesses of nonlymphocytic, myelogenous and acute myeloid leukaemias occurred. There is also limited evidence in

Table 21.20 Weighted summary relative risks for lung cancer for all subregions

| Subregion | Summary relative risk ${ }^{\text {a }}$ | Low exposure | High exposure |
| :---: | :---: | :---: | :---: |
|  |  | Combined relative risk ${ }^{\text {a }}$ (95\% CI) | Combined relative risk ${ }^{\text {a }}$ (95\% CI) |
| AFR-D | 1.61 | 1.31 (1.17-1.45) | 1.91 (1.72-2.11) |
| AFR-E | 1.62 | 1.32 (1.18-1.45) | 1.92 (1.72-2.12) |
| AMR-A | 1.59 | 1.29 (1.14-1.44) | 1.88 (1.67-2.1 1 ) |
| AMR-B | 1.58 | 1.28 (1.14-1.42) | 1.87 (1.67-2.08) |
| AMR-D | 1.56 | 1.26 (1.13-1.41) | 1.85 (1.64-2.05) |
| EMR-B | 1.56 | 1.26 (1.13-1.40) | 1.85 (1.64-2.05) |
| EMR-D | 1.61 | 1.31 (1.18-1.45) | 1.92 (1.72-2.10) |
| EUR-A | 1.62 | 1.32 (1.17-1.48) | 1.93 (1.71-2.16) |
| EUR-B | 1.59 | 1.29 (1.15-1.44) | 1.89 (1.69-2.10) |
| EUR-C | 1.50 | 1.22 (1.09-1.35) | 1.79 (1.59-1.97) |
| SEAR-B | 1.58 | 1.28 (1.15-1.42) | 1.88 (1.68-2.07) |
| SEAR-D | 1.61 | 1.31 (1.17-1.45) | 1.91 (1.70-2.09) |
| WPR-A | 1.57 | 1.27 (1.13-1.42) | 1.86 (1.65-2.08) |
| WPR-B | 1.58 | 1.28 (1.14-1.42) | 1.87 (1.67-2.07) |

a Weighted summary relative risk, weighted by the proportion of workers exposed to each contributing carcinogen in each subregion.
mammals (Hayes et al. 1997; IARC 1990b). A recent review (Lynge et al. 1997) provides a low-exposure relative risk of 2.0 ( $95 \%$ CI $1.8-2.2$ ) and a high-exposure relative risk of 4.0 (3.6-4.4).

## Ionizing radiation

The causal relationship between ionizing radiation and leukaemia is well recognized. There is consistency across numerous studies, strong association between exposure and outcome, and evidence of a dose-response gradient. Excess leukaemia has been observed in survivors of the atomic explosions at Hiroshima and Nagasaki, and also among patients medically treated with X-rays or $\gamma$-rays. The risk of leukaemia increases over fivefold at sufficiently high doses (BEIR V 1990; IARC 2000; ICRP 1991). Models describing risk have been proposed as: Linear RR model $1+5.5$ _ dose in Sv, quadratic $R R=1+0.24$ dose +0.27 dose $^{2}$ (dose in Sv) (BEIR V 1990). Relative risks of 1.22 (1.07-1.70) for low exposure and 1.57 (1.18-2.88) for high exposure are accepted as the best available estimates (BEIR V 1990; IARC 2000).

## Ethylene oxide

Workers have exposure to ethylene oxide either as a sterilant or as a chemical intermediary or final product. In a study in the United States,
ethylene oxide used as a sterilant was associated with lymphatic leukaemia and non-Hodgkin's lymphoma, with a rate ratio of 1.2 estimated for 45 -year exposure to 1 ppm . Other studies in Sweden and the United Kingdom of Great Britain and Northern Ireland showed nonsignificant excesses of these cancers. Of six studies of chemical plant workers (two in Sweden and one each in Germany, Italy, the United Kingdom and the United States), two found significant excesses, two found nonsignificant excesses and two found expected rates (IARC 1997). Relative risk was found to range from 1.1 to 3.5 (Steenland et al. 2003).

An approach similar to that used for lung carcinogens was applied to the leukaemogens. The separate relative risks for the development of leukaemia arising from exposures to the main relevant occupational carcinogens were combined into single summary relative risks, one for low exposure and one for high exposure. This was done separately for each subregion, using the exposure prevalence of the workforce in each subregion to weight the exposure-specific risks. However, the resulting average relative risks were not clearly different from each other. CI were estimated in the same manner, weighting the estimated CI for benzene and ionizing radiation (there were no estimated CI for ethylene oxide). Unlike lung cancer, the low- and high-exposure relative risks were available for each exposure, and these were directly incorporated into lowand high-exposure summary measures through the weighting process. An example of this approach is shown in Table 21.21, using the WPRB subregion. The results of this approach for each subregion are shown in Table 21.22.

Table 21.2I Leukaemia relative risk, substance-specific and weighted average, for the WPR-B subregion

| Carcinogen | Low exposure | High exposure |  |
| :---: | :---: | :---: | :---: |
|  | Combined relative risk ${ }^{\text {a }}$ (95\% CI) | Combined relative risk ${ }^{\text {a }}$ (95\% CI) | Proportion of workers exposed |
| Benzene | 2.0 (1.8-2.2) | 4.0 (3.6-4.4) | 0.0051 |
| lonizing radiation | 1.22 (1.07-1.7) | 1.57 (1.18-2.88) | 0.0010 |
| Ethylene oxide | 1.1 | 3.5 | 0.0003 |
| Total ${ }^{\text {b }}$ | 1.84 (1.68-2.12) | 3.60 (3.20-4.16) |  |

[^80]Table 21.22 Weighted summary relative risks for leukaemia for all subregions

| Subregion | Low exposure | High exposure |
| :---: | :---: | :---: |
|  | Combined relative risk ${ }^{\text {a }}$ (95\% Cl) | Combined relative risk ${ }^{\text {a }}$ (95\% CI) |
| AFR-D | 1.88 (1.72-2.15) | 3.73 (3.34-4.24) |
| AFR-E | 1.89 (1.73-2.15) | 3.75 (3.37-4.25) |
| AMR-A | 1.91 (1.74-2.16) | 3.80 (3.41-4.28) |
| AMR-B | 1.77 (1.61-2.07) | 3.38 (2.98-4.01) |
| AMR-D | 1.91 (1.74-2.16) | 3.78 (3.39-4.27) |
| EMR-B | 1.87 (1.70-2.13) | 3.66 (3.26-4.19) |
| EMR-D | 1.89 (1.72-2.15) | 3.73 (3.33-4.23) |
| EUR-A | 1.93 (1.76-2.17) | 3.86 (3.48-4.32) |
| EUR-B | 1.83 (1.67-2.11) | 3.57 (3.18-4.13) |
| EUR-C | 1.67 (1.51-2.00) | 3.06 (2.64-3.80) |
| SEAR-B | 1.89 (1.73-2.15) | 3.76 (3.37-4.26) |
| SEAR-D | 1.81 (1.65-2.10) | 3.51 (3.1I-4.09) |
| WPR-A | 1.90 (1.73-2.15) | 3.77 (3.38-4.26) |
| WPR-B | 1.84 (1.68-2.12) | 3.60 (3.21-4.16) |

a Weighted summary relative risk, weighted by the proportion of workers exposed to each contributing carcinogen in each subregion.

## Estimates of risk reversibility

There are limited data on risk reversibility from occupational exposure to carcinogens. The studies from which the estimated risks arise are based on cohorts of people exposed for different periods of time, followed up for various periods of time and with various periods of time between exposure cessation and follow-up, with follow-up periods varying between zero (still exposed) and many decades. Therefore, most of the absolute and relative risks produced by the studies already depend on whatever change in risk might occur once exposure ceases. However, some indication of the extent of risk reduction that might occur is given by a recent paper by Peto et al. (2000), which examined changes in the risk of developing lung cancer as a result of stopping smoking. The study estimated that, compared to the risk in persons who continued to smoke, the risk of lung cancer in males declined to about 0.66 within 10 years, to 0.44 between 10 and 20 years, to 0.2 between 20 and 30 years, and to 0.1 after 30 years.

## 3. Occupational airborne particulates

There are a vast number of respiratory conditions that can arise directly or indirectly from work. However, estimating exposures, risks and attrib-
utable proportions is not possible for many of these on an international (or even national) scale, because of lack of appropriate data sources. Therefore, only the more important of the work-related respiratory conditions, in terms of the total number of cases or the risks arising from exposure, are included here. All of these arise from exposure to particulates. Malignant respiratory disease is not included here because it is described in section 2.

Nonmalignant respiratory disease arises as a result of the exposure of workers to airborne agents, mostly in the form of particulates or dusts. ${ }^{2}$ The primary route of exposure is inhalation, whereby these agents gain access to the respiratory system and are either deposited (in the case of dusts) or enter the circulatory system. For some exposures, there is a very clear connection between the exposure and the disease (for example, silicosis is only caused by exposure to silica). Some exposures cause more than one type of disease, and even more than one type of respiratory disease. For example, asbestos can result in malignant conditions of the lung and the pleura (the inside lining of the chest), malignant conditions of the peritoneum (the inside lining of the abdomen) and nonmalignant conditions of the lungs (asbestosis and COPD). Other exposures have not been well characterized, but are believed to result in certain conditions (such as some forms of occupational asthma).

### 3.1 Exposure variable

## Causative agents of asthma

Asthma, which is a narrowing of the upper respiratory passages resulting in difficult breathing and wheezing, has both nonoccupational and occupational causes. Many hundreds of occupational agents, including some inorganic and organic dusts, have been associated with occupational asthma (Balmes et al. 2003; Chan-Yeung and Malo 1994; Venables and Chan-Yeung 1997). Biological agents include grains, flours, plants and gums, fur, feathers and other animal parts, insects and fungi, drugs and enzymes and various types of wood. Chemical agents include chlorofluorocarbons, alcohols, metals and their salts, and welding fumes (CCOHS 1997). These agents are found in a variety of workplaces, including food and natural products processing, animal handling facilities, manufacturing and construction.

It would not be possible to conduct exposure assessments and to obtain relative risk data for all the factors contributing to this important occupational disease, especially since they often occur in combination. We therefore used occupation as a proxy for exposure to agents that are associated with occupational asthma. The basis for this approach was the work of Karjalainen et al. (2001, 2002), who conducted extensive epidemiological studies of the entire Finnish workforce and developed relative risks for specific occupations. A similar but less extensive study based in 12 industrialized countries was also used (Kogevinas et al.
1999). Relative risks were applied to these occupational data to produce estimates of the number of deaths due to work-related asthma.

Causative agents of COPD
The causative agents of COPD are non-specific dust and fumes, with dusts showing a more consistent relationship than fumes (Becklake 1989). Because of a lack of worldwide data on the prevalence of occupational exposure to dusts and their combinations, work in specific economic subsectors was used as a surrogate for dust exposure. Relative risks were applied to these workforce data to produce estimates of the number of deaths from COPD arising from work-related exposures.

### 3.2 Estimating risk factor levels

The general exposure assessment methodology was described earlier. Occupation was used for asthma (Equation 2) and economic sector for COPD (Equation 1). The theoretical minimum risk corresponds to no occupational exposure above background levels to airborne particulates or other agents that cause nonmalignant respiratory disease.

## AgENTS CAUSING ASTHMA

The proportion of the total population with occupational exposure to asthmagens was estimated using Equation 2. Estimates were made for each occupational category by determining the proportion of the population working in occupations that matched as closely as possible to those identified by Karjalainen et al. $(2001,2002)$ and for which relative risk values were provided (Table 21.23). Those not working and those employed in administration were together considered to be the nonexposed reference category (relative risk=1). These calculations were done separately for men and women for each subregion of the world. Relative risks and the proportions exposed by occupational category were applied across all age groups from age 15 to $\geq 80$ years.

Table 21.24 summarizes the age-adjusted distribution of the labour force into occupations matching the categories for which relative risks were identified by Karjalainen et al. (2002).

## Agents causing COPD

It is not possible to estimate the proportion of the world's population exposed to the large number of agents identified in occupation-specific and agent-specific studies. Community-based studies have therefore been preferred. The most common exposure in these studies is exposure to dust and/or gas/fumes (e.g. Korn et al. 1987; Kryzanowski et al. 1986; Xu et al. 1992). Unfortunately, there are also no data to estimate the proportion of the world's workers exposed to dust and/or gas/fumes. The study by Korn et al. (1987) provides a link between self-reported exposure to dust (current and past exposure) and some categories of economic activity ${ }^{3}$ among the currently employed. Categories of economic activity
Table 21.23 Comparison of Finnish occupational categories with 1968 ISIC codes

| Finnish classification ${ }^{\text {a }}$ | Description | Examples ${ }^{\text {a }{ }^{\text {b }} \text { b }}$ | $\begin{aligned} & 1968 \\ & 1 S I C^{c} \\ & \hline \end{aligned}$ | Description | Examples | Comments |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| I | Administrative, managerial and clerical workers | No examples given | 2 3 | Administrative and managerial workers Clerical and related workers | Government officials, managers (upper level) Office managers, government workers, secretaries, bookkeepers, stock clerks | Combine Categories 2 and 3 (ISIC) for all economic subsectors |
| 0 | Technical, physical science, social science, humanistic and artistic workers | Engineers, medical personnel, child care givers, religious and social workers | 0/1 | Professional, technical and related workers | Scientists, technicians, engineers, medical and related workers, mathematicians, teachers, religious workers, artists | Use Category 0/I (ISIC) for all economic subsectors |
| 2 | Sales workers | Wholesale and retail dealers, other sales workers | 4 | Sales workers | Working proprietors, sales managers, sales workers, insurance agents | Use Category 4 (ISIC) for all economic subsectors |
| 3 | Agriculture, forestry, commercial fishing | Farmers and managerial workers in agriculture, forestry and horticulture, agricultural and horticultural workers, animal husbandry workers | 6 | Agricultural, animal husbandry and forestry workers, fishermen, hunters | Farm managers and supervisors, agriculture and animal husbandry workers, forestry workers, fishermen | Use Category 6 (ISIC) for all economic subsectors |
| 4 | Mine and quarry workers | Miners, quarrymen |  |  |  | Use Category 7/8/9 for mining economic subsector |

Table 21.23 Comparison of Finnish occupational categories with 1968 ISIC codes (continued)

| Finnish classification ${ }^{\text {a }}$ | Description | Examples ${ }^{\text {a, }}$ b | $\begin{aligned} & 1968 \\ & 1 S I C^{c} \end{aligned}$ | Description | Examples | Comments |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 8 | Service work | Firefighters and police, watch and security guards, cooks, housekeepers, domestic workers, building caretakers and cleaners, hygiene and beauty operators, launderers, dry cleaners and pressers | 5 | Service workers | Hotel managers, cooks, waiters, housekeepers, caretakers, beauty operators, firefighters, police | Use Category 5 (ISIC) for all economic subsectors |
| a Source: Karjalainen et al. (2001). <br> b Source: Karjalainen et al. (2002). <br> c Source: UN (2000). |  |  |  |  |  |  |

Table 21.24 Proportion of the population in occupational categories based on exposure to agents causing asthma, by subregion

| Subregion | Sex | Proportion exposed by occupation |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Background | Administration | Technical | Sales | Agricultural | Mining | Transport | Manufacturing | Services |
| AFR-D | Male | 0.1595 | 0.0498 | 0.0562 | 0.0513 | 0.4612 | 0.0081 | 0.0289 | 0.1342 | 0.0510 |
|  | Female | 0.4645 | 0.0249 | 0.0303 | 0.0324 | 0.3551 | 0.0012 | 0.0145 | 0.0482 | 0.0288 |
| AFR-E | Male | 0.1510 | 0.0524 | 0.0618 | 0.0447 | 0.4662 | 0.0082 | 0.0219 | 0.1377 | 0.0563 |
|  | Female | 0.3498 | 0.0360 | 0.0497 | 0.0328 | 0.4171 | 0.0010 | 0.0114 | 0.0547 | 0.0475 |
| AMR-A | Male | 0.2746 | 0.1971 | 0.1080 | 0.0875 | 0.0327 | 0.0035 | 0.0196 | 0.1940 | 0.0830 |
|  | Female | 0.4079 | 0.1772 | 0.1223 | 0.0789 | 0.0134 | 0.0005 | 0.0080 | 0.0928 | 0.0991 |
| AMR-B | Male | 0.1896 | 0.1124 | 0.0794 | 0.0665 | 0.1592 | 0.0077 | 0.0310 | 0.2225 | 0.1317 |
|  | Female | 0.5823 | 0.0662 | 0.0671 | 0.0410 | 0.0531 | 0.0013 | 0.0032 | 0.0878 | 0.0979 |
| AMR-D | Male | 0.1785 | 0.1894 | 0.0346 | 0.0912 | 0.0521 | 0.0020 | 0.0386 | 0.3115 | 0.1021 |
|  | Female | 0.6110 | 0.1033 | 0.0198 | 0.0466 | 0.0119 | 0.0001 | 0.0026 | 0.1328 | 0.0719 |
| EMR-B | Male | 0.2134 | 0.1042 | 0.1499 | 0.0854 | 0.1194 | 0.0046 | 0.0409 | 0.2072 | 0.0749 |
|  | Female | 0.6903 | 0.0523 | 0.0866 | 0.0443 | 0.0293 | 0.0001 | 0.0101 | 0.0463 | 0.0407 |
| EMR-D | Male | 0.1805 | 0.0419 | 0.0490 | 0.1742 | 0.3584 | 0.0014 | 0.0000 | 0.1465 | 0.0480 |
|  | Female | 0.6303 | 0.0110 | 0.0105 | 0.0509 | 0.2427 | 0.0002 | 0.0000 | 0.0401 | 0.0142 |

Table 21.24 Proportion of the population in occupational categories based on exposure to agents causing asthma, by subregion

| Subregion | Sex | Proportion exposed by occupation |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Background | Administration | Technical | Sales | Agricultural | Mining | Transport | Manufacturing | Services |
| EUR-A | Male | 0.3227 | 0.1176 | 0.2145 | 0.0252 | 0.0420 | 0.0040 | 0.0000 | 0.1934 | 0.0807 |
|  | Female | 0.5296 | 0.0942 | 0.1889 | 0.0131 | 0.0254 | 0.0008 | 0.0033 | 0.0733 | 0.0713 |
| EUR-B | Male | 0.2593 | 0.0747 | 0.0680 | 0.0400 | 0.2133 | 0.0136 | 0.0226 | 0.2490 | 0.0595 |
|  | Female | 0.4624 | 0.0453 | 0.0552 | 0.0131 | 0.2352 | 0.0020 | 0.0041 | 0.1379 | 0.0449 |
| EUR-C | Male | 0.2700 | 0.0946 | 0.0532 | 0.0317 | 0.1536 | 0.0212 | 0.1097 | 0.2239 | 0.0421 |
|  | Female | 0.4269 | 0.0818 | 0.0542 | 0.0512 | 0.0919 | 0.0138 | 0.1041 | 0.1239 | 0.0522 |
| SEAR-B | Male | 0.1756 | 0.0599 | 0.0482 | 0.0729 | 0.3664 | 0.0051 | 0.0336 | 0.1930 | 0.0452 |
|  | Female | 0.4240 | 0.0397 | 0.0368 | 0.0812 | 0.2487 | 0.0013 | 0.0032 | 0.1244 | 0.0407 |
| SEAR-D | Male | 0.1502 | 0.0645 | 0.0550 | 0.0150 | 0.4634 | 0.0149 | 0.0392 | 0.1535 | 0.0444 |
|  | Female | 0.5298 | 0.0134 | 0.0125 | 0.0028 | 0.3781 | 0.0026 | 0.0000 | 0.0509 | 0.0098 |
| WPR-A | Male | 0.2447 | 0.2058 | 0.1023 | 0.0787 | 0.0336 | 0.0014 | 0.0340 | 0.2270 | 0.0723 |
|  | Female | 0.4795 | 0.1410 | 0.0832 | 0.0755 | 0.0281 | 0.0002 | 0.0094 | 0.1118 | 0.0713 |
| WPR-B | Male | 0.1600 | 0.1023 | 0.0655 | 0.0454 | 0.3659 | 0.0194 | 0.0370 | 0.1399 | 0.0645 |
|  | Female | 0.2901 | 0.0928 | 0.0588 | 0.0753 | 0.2812 | 0.0066 | 0.0290 | 0.0925 | 0.0738 |
| Note: See also Table 21.26. |  |  |  |  |  |  |  |  |  |  |

among the currently employed are available worldwide, and can provide a broad approximation to the proportion of the world's population with current or past exposure to dust and/or gas/fumes. We based our estimates of exposed populations on data on employment in economic sectors of agriculture, industry and service from the World Bank (2001), supplemented by data from ILO (2000) on employment in economic activities. The proportions of the population with occupational exposure at medium and high levels to agents causing COPD were estimated using Equation 1.

Korn et al. (1987) defined as low-exposed those in finance, as mediumexposed those in the manufacture of non-durable goods, transport, utilities and the wholesale and retail trades, and as highly exposed those in the manufacture of durable goods, agriculture, mining and construction. Exposure was to "dusts" and to "gases", without these being further defined. We adopted these categories with some modification to account for our lack of data on the type of manufacturing industry and for the fact that agriculture in developed and developing countries probably involves different types of exposure to respirable dust. Lacking data that would have permitted us to divide manufacturing into medium and high potential for dust exposure, we have classified it as having potentially high dust exposure, given that in much of the world manufacturing involves more dust exposure than is typical in the United States where the Korn et al. study (1987) was done (Chien et al. 2002; Gomes et al. 2001). We have defined as nonexposed those not in the workforce and those in utility trade, finance and services. Those in agriculture, manufacturing and transportation were defined as having low exposure, while those in mining and construction were defined as having high exposure. Many workers in the medium and highly exposed economic activities are in fact not exposed to dusts, but on the whole the proportions in these industries are taken to represent the approximate proportion of those ever exposed to low and high levels of dusts in the general population. In Korn et al. (1987), the proportion of workers currently employed in the medium- and high-exposure industries listed above corresponded approximately to the proportion of those reporting ever having been occupationally exposed to dust in that study. This approach was followed in our study, in which it was assumed that the number of currently employed in specific industries corresponds roughly to the number ever occupationally exposed to dusts. The proportion exposed in different economic activities in each subregion was adjusted to account for an average labour force participation among the currently exposed in that subregion, which was applied across all ages. The results are presented in Table 21.25.

Table 21.25 Proportion of the population exposed to agents causing COPD, by subregion, sex and level of exposure

| Subregion | Exposure level | Proportion ever exposed |  |
| :---: | :---: | :---: | :---: |
|  |  | Male | Female |
| AFR-D | Background | 0.3722 | 0.5920 |
|  | Low | 0.5086 | 0.3776 |
|  | High | 0.1192 | 0.0305 |
| AFR-E | Background | 0.3744 | 0.5386 |
|  | Low | 0.5051 | 0.4365 |
|  | High | 0.1204 | 0.0249 |
| AMR-A | Background | 0.6879 | 0.9056 |
|  | Low | 0.0879 | 0.0314 |
|  | High | 0.2242 | 0.0630 |
| AMR-B | Background | 0.5653 | 0.8908 |
|  | Low | 0.2336 | 0.0553 |
|  | High | 0.2011 | 0.0539 |
| AMR-D | Background | 0.6465 | 0.9337 |
|  | Low | 0.1253 | 0.0169 |
|  | High | 0.2281 | 0.0494 |
| EMR-B | Background | 0.5829 | 0.9256 |
|  | Low | 0.2007 | 0.0441 |
|  | High | 0.2164 | 0.0303 |
| EMR-D | Background | 0.5818 | 0.7780 |
|  | Low | 0.2204 | 0.1776 |
|  | High | 0.1978 | 0.0444 |
| EUR-A | Background | 0.6819 | 0.8965 |
|  | Low | 0.0565 | 0.0253 |
|  | High | 0.2616 | 0.0781 |
| EUR-B | Background | 0.5096 | 0.6598 |
|  | Low | 0.2636 | 0.2469 |
|  | High | 0.2268 | 0.0933 |
| EUR-C | Background | 0.4312 | 0.6463 |
|  | Low | 0.3273 | 0.2409 |
|  | High | 0.2415 | 0.1128 |
| SEAR-B | Background | 0.4190 | 0.6694 |
|  | Low | 0.4112 | 0.2384 |
|  | High | 0.1698 | 0.0922 |
| SEAR-D | Background | 0.3965 | 0.5723 |
|  | Low | 0.4822 | 0.3869 |
|  | High | 0.1213 | 0.0408 |
| WPR-A | Background | 0.5994 | 0.8387 |
|  | Low | 0.1200 | 0.0531 |
|  | High | 0.2806 | 0.1082 |
| WPR-B | Background | 0.3700 | 0.5244 |
|  | Low | 0.4474 | 0.3807 |
|  | High | 0.1826 | 0.0949 |

### 3.3 RISK FACTOR-DISEASE RELATIONSHIPS

## ASTHMA

Occupational asthma is a condition characterized by variable airflow limitation or bronchial hyper-responsiveness related to workplace exposure. However, the precise definition of occupational asthma has been widely debated. The most controversial issue concerns whether only immunologically-mediated asthma should be considered to be occupational asthma or whether asthma arising as result of workplace exposure to irritants, or exacerbation of pre-existing asthma by workplace irritants, should also be considered in the definition (Lombardo and Balmes 2000; Malo and Chan-Yeung 2001; Wagner and Wegman 1998). Recently, consensus seems to have been reached in favour of a broad definition (American Thoracic Society review: Balmes et al. 2003). A broader approach has been supported by others (Blanc and Toren 1999; Karjalainen et al. 2001; Kogevinas et al. 1999; Milton et al. 1998; Toren et al. 1999), and recent studies of occupational asthma have tended to use a more inclusive approach (Karjalainen et al. 2001, 2002; Milton et al. 1998).

Occupational asthma is probably the most common work-related respiratory disorder in industrialized countries (Kogevinas et al. 1999), and is either stable (Singh and Davis 2002) or increasing in incidence (Sears 1997). Many hundreds of occupational agents, including some inorganic and organic dusts, have been associated with occupational asthma (Balmes et al. 2003; Chan-Yeung and Malo 1994; Venables and ChanYeung 1997).

Until recently, there has been limited information on the total risk of developing asthma from workplace exposure. The United States magnitude of mortality study (Steenland et al. 2003) estimated that about $5 \%$ of mortality from nonmalignant work-related respiratory disease was due to asthma. Studies of substance-specific risks have helped to identify or implicate particular substances as likely causative agents (e.g. Monso et al. 1998), but these studies have generally focused on agents thought to be sensitizers, and usually on only a limited number of these. They are therefore not useful for determining the true extent of asthma occurring as a result of work-related exposure. Several population-based studies have partially rectified this problem (Karjalainen et al. 2001, 2002; Kogevinas et al. 1996, 1999; Ng et al. 1994; Toren 1996; Toren et al. 1999), focusing on occupation-specific rather than substancespecific risks because of the plethora of potential causative exposures and the difficulty in characterizing them. These studies provided measures of relative risk and/or population-attributable fractions. Recent studies in Finland have estimated population-attributable fractions for occupational asthma of $18 \%$ (Nurminen and Karjalainen 2001) and of $17 \%$ (for women) and $29 \%$ (for men) (Karjalainen et al. 2002). A comprehensive review undertaken before these two Finnish studies found a
median value for population-attributable fraction of $9 \%$ for all relevant studies, and a median value of $15 \%$ for the highest-quality studies (Blanc and Toren 1999). The American Thoracic Society (Balmes et al. 2003) has recently reviewed the literature and estimated that approximately $15 \%$ of asthma is attributable to occupational exposure, based largely on studies in developed countries.

Of these studies, only that by Karjalainen et al. $(2001,2002)$ provides useable risk information to cover the whole workforce, while that by Kogevinas et al. provides useful information for agriculture. The study by Karjalainen et al. $(2001,2002)$ was a longitudinal study over 13 years covering the entire Finnish population, and provided relative risks for a large number of broad occupational categories. In that study, asthma was defined by the occurrence of clinically diagnosed asthma ( $\mathrm{n}=49$ 575) during the follow-up period; national medical records were linked to census data on an individual's occupation. The study population was composed of all those currently employed, aged 25-59 years at baseline, without prior history of asthma. Relative risks were calculated by comparing the occupation-specific incidence to the incidence of occupational asthma in administrative, managerial and clerical workers, whose risk was assumed to be similar to the background population risk. The relative risks were adjusted for age, and separate risks were available for males and females, although these were very close to each other and certainly within the limits of random variation. The study by Kogevinas et al. (1999) was a cross-sectional study of asthma involving 15000 people in 12 European countries. In both studies, relative risks of asthma morbidity were assumed to apply for asthma mortality. This assumption is likely to be reasonable in most circumstances, but may lead to some underestimation or overestimation of asthma mortality, depending on whether exposure results in asthma incidence or exacerbation.

The approach used here was based on the work of Karjalainen et al. (2001, 2002). The work by Kogevinas et al. (1999) was also used for the relative risk of asthma due to occupational exposure in agriculture. While the Finnish study was large, prospective and covered all occupations, there was concern that Finnish exposures within specific occupations might be atypical of the rest of the world. In particular, this was considered likely to be true for agriculture, since Finnish agriculture might involve more indoor work where the relative risks for asthma were relatively high. Therefore the Kogevinas et al. results were used for agriculture since they were believed to be more generalizable to agriculture in the rest of the world, especially the developing world.

We assumed that the relative risk of asthma morbidity owing to employment in occupational categories was approximately equal to the relative risks of asthma mortality. Those not working and those employed in administration were together considered to be the nonexposed reference category (relative risk $=1$ ). These calculations were done separately for men and women for each subregion of the world. Rela-

Table 21.26 Relative risks for occupational asthma by original occupation and economic subsector, and sex, age-adjusted

| Occupation | Relative risk <br> (males) | Relative risk <br> (females) | Source |
| :--- | :---: | :---: | :--- |
| Background | 1.00 | 1.00 | Non-working population, used as reference |
| Administration | 1.00 | 1.00 | Karjalainen et al. (2002), also used as reference |
| Technical | 1.05 | 1.06 | Karjalainen et al. (2002) |
| Sales | 1.14 | 1.13 | Karjalainen et al. (2002) |
| Agricultural | 1.41 | 1.41 | Kogevinas et al. (1999) |
| Mining | 1.95 | 1.00 | Karjalainen et al. (2002) |
| Transport | 1.31 | 1.22 | Karjalainen et al. (2002) |
| Manufacturing | 1.56 | 1.33 | Karjalainen et al. (2002) |
| Services | 1.53 | 1.41 | Karjalainen et al. (2002) |

tive risks and the proportions exposed by occupational category were applied across all age groups from age 15 to $\geq 80$ years. The relative risks by occupation are shown in Table 21.26.

## CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD)

Tobacco smoking is clearly the most important risk factor, but many work-related exposures have been demonstrated to cause COPD (Hendrick 1996). A recent United States study estimating the magnitude of mortality due to occupational exposure (Steenland et al. 2003) used an estimate of $14 \%$ for the population-attributable fraction for COPD due to occupational dust exposure (based on a community study of severe COPD) (Korn et al. 1987), and found that COPD represented $87 \%$ of all fatal work-related nonmalignant respiratory disease, although some of the other types of respiratory disease may have been underestimated. A review of Finnish data also used a population-attributable fraction of $14 \%$ for men (and $5 \%$ for women) (Nurminen and Karjalainen 2001), and a similar figure ( $15 \%$ ) was recently used in a review by the American Thoracic Society (Balmes et al. 2003).

As for asthma, difficulties arise from the vast array of definite, probable and possible causes of work-related COPD. The role of smoking, particularly in causing possible confounding effects, makes interpretation of studies difficult. Apparently significant individual differences in susceptibility, and uncertainty about pathological mechanisms, also cause problems. This area has been the subject of several reviews (Attfield and Wagner 1998; Balmes et al. 2003; Becklake 1989, 1994; Hendrick 1996; NIOSH 1996; Oxman et al. 1993), some covering all exposures and some concentrating on mineral dusts.

As a result of difficulties in characterizing all the likely causative occupational exposures, few published papers provide information that com-
prehensively describes the risk of developing COPD as a result of work. The paper by Korn et al. (1987) has been used in this analysis, as it provides relative risk information covering all workplace exposures. This study (the methods of which were described in more detail in an earlier study by Ferris et al. 1979) used data from a random sample of white adults aged 25-74 years from six United States cities and their surrounding areas ( 8515 people were included in the final sample). The definition of COPD was $\mathrm{FEV}_{1} / \mathrm{FVC}<0.6$, representing reasonably severe disease. Logistical regression analyses were undertaken, determining the odds ratios for various respiratory conditions and controlling for age, sex, current and lifetime smoking history and city of residence. These odds ratios for COPD morbidity from Korn et al. (1987) were assumed to apply to COPD mortality.

The study by Korn et al. (1987) provides a strict definition of COPD and relative risks for both men and women, and was based on a large number of participants. This study was therefore used as the basis of the relative risk and attributable fraction estimates presented here. Relative risks for COPD prevalence were used as an approximation of the relative risks for COPD mortality.

Korn et al. (1987) found relative risks of COPD of 1.62 for men and 1.24 for women for a history of exposure to dusts. We partitioned these relative risks into high- and low-exposure categories, and also used slightly different relative risks for low exposure in developed and developing countries. In developing countries the great majority of low-exposure employment is in agriculture, where much dust is nonrespirable. In developed countries much of the exposure in the low categories is in industries other than agriculture, where a higher percentage of dust exposure may be respirable and toxic. It was assumed that the relative risks applied across all age categories. The estimated relative risks are shown in Table 21.27.

RISK REVERSIBILITY
As for carcinogens, there are limited data on risk reversibility. The studies from which the estimated risks arise are based on cohorts of people

Table 21.27 Annual risks of COPD mortality

| Relative risk | Developing countries |  | Developed countries (AMR-A, EUR-A, WPR-A) |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Men | Women | Men | Women |
| Nonexposed | 1.0 | 1.0 | 1.0 | 1.0 |
| Low | 1.2 | 1.1 | 1.4 | 1.2 |
| High | 1.8 | 1.4 | 1.8 | 1.4 |

Source: Korn et al. (1987).
exposed for different periods of time, followed up for various periods of time and with various periods of time between exposure cessation and follow-up, with follow-up periods varying between zero (still exposed) and many decades. Therefore, most of the absolute and relative risks produced by the studies are already dependent on whatever change in risk might occur once exposure ceases. Indications of risk reversibility for COPD may be obtained from the literature on smoking.

## 4. Occupational noise

Noise is a common occupational hazard. The unit for sound (noise) level, whether measuring noise exposure or hearing loss, is the decibel (dB). Noise exposure levels as used in this document have the unit dBA. ${ }^{4}$ Noise-induced hearing loss is reported in dBHL. There is variability in the literature in the use of the terms to describe hearing ability. As used here, hearing loss refers to a decline in an individual's hearing ability. Hearing impairment refers to the effect of hearing loss on the individual's ability to function. The U.S. National Institute for Occupational Safety and Health (NIOSH) uses the term "material hearing impairment" ${ }^{5}$ to describe a hearing loss greater than 25 dB , and most occupational studies refer to 25 dBHL . The WHO definition used in this study is hearing loss greater than or equal to 41 dBHL . Therefore, extrapolations were made from the occupational studies to fit the requirements of the WHO study.

### 4.1 Exposure variable

The exposure variable used in this analysis is a direct measure of the risk factor, i.e. occupational exposure to noise, which is the causative agent of noise-induced hearing loss. As global data on the frequency of occurrence, duration and intensity of noise exposure do not exist, it was necessary to model this exposure for workers employed in various occupational categories. The theoretical minimum is based on expected background levels of noise, and consistency with national and international standards. Most experts agree that levels below 80 dBA would result in minimal risk of developing hearing loss.

For workers in various occupational categories, three levels of exposure were estimated:

- minimum exposure, less than 85 dBA ;
- moderately high noise, $\geq 85-90 \mathrm{dBA}$; and
- high noise, $>90$ dBA.

The choice of these levels was based on the recommended exposure limits (RELs) for occupational noise exposure around the world. In most developed countries the REL is 85 dBA as an eight-hour time-weighted average without hearing protection. In the United States the PEL is

90 dBA for an eight-hour day, although a hearing conservation programme is required for all employees exposed above 85 dBA for an eight-hour day. In developing countries, the REL is usually 90 dBA (Ahmed et al. 2001; Alidrisi et al. 1990; Hernandez-Gaytan et al. 2000; Hessel and Sluis-Cremer 1987; Osibogun at al 2000; Shaikh 1996; Sriwattanatamma and Breysse 2000).

Although the theoretical minimum exposure to noise was determined to be 80 dBA , it was not possible to estimate frequency of exposure by occupational category to occupational noise between 80 and 85 dBA . Therefore, persons with occupational exposure $<85 \mathrm{dBA}$ were included with the background population.

### 4.2 Estimating risk factor levels

## DATA SOURCES

Potentially useful studies were identified using the various approaches described in section 1. The key terms used were "occupational noise" and "occupational hearing impairment". Relevant studies were identified by critically appraising the references obtained. This included consideration of the approaches to selection, measurement, analysis and control of confounding. Potential confounders of noise-induced hearing loss include nonoccupational exposure to noise, undocumented occupational noise levels, use of personal protective equipment, use of some medicines, and outer- and middle-ear pathology. Recent review articles were used where available, and the main articles were obtained and appraised.

The main reason for excluding studies was that they did not contain data appropriate for determining risk of noise-induced hearing loss. Problems included an inappropriate (for this purpose) exposure measurement (such as reporting for only one or a few occupational groups or tasks); inappropriate (for this purpose) outcome measurement (such as dB per year loss with age or no data as to the number of cases vs total population); poorly characterized exposure or self-reported hearing loss; and inadequate control of confounding.

In the United States, about 9 million workers are exposed to timeweighted average sound levels of 85 dBA and above (Simpson and Bruce 1981, quoted in Suter 2000), and about 10 million have noise-induced hearing loss $>25 \mathrm{~dB}$ (USDOL OSHA 2002b). About 17\% of American "production workers" are exposed to average noise levels at or above 85 dBA (NIOSH 1998). In the European Union, $28 \%$ of workers surveyed reported that for at least $25 \%$ of the time they were occupationally exposed to noise loud enough to cause them to raise their voices during conversation (corresponding to approximately 85-90dBA) (EASHW 2000). The highest percentages of exposed workers were reported for mining, quarrying, manufacturing and construction. Australia compensates about 10000 people each year for noise-induced
hearing loss; evidence indicates that only one third of workers with noiseinduced hearing loss file compensation claims (NOHSC 1993). Summary statistics on noise exposure are not available for most industrializing and nonindustrialized countries. However, most published reports indicate that average noise exposure levels are well above the recommended occupational level in many industrialized countries, which is generally established at $85-90 \mathrm{~dB}$ for an eight-hour work day (Suter 2000; WHO/FIOSH 2001).

Information on noise exposures and noise-induced hearing loss in developing countries is given in Table 21.28. These studies are characterized by high occupational noise exposure levels, and many report hearing losses in exposed workers. The authors generally recommended engineering controls and hearing conservation programmes, including hearing protection, indicating that hearing protection is not widely used. Seventeen studies conducted in 12 countries in South America, Africa and Asia reported noise levels in a wide range of workplaces, including mining and the manufacture of food, fabrics, printed materials, metal products, drugs and watches. Most studies provided ranges of sound levels, with the lowest reported noise levels often below 80 dBA and the upper levels always above 90 dBA . All the studies that examined the hearing ability of workers revealed increased rates of hearing impairment in noise-exposed workers compared to nonexposed controls.

## Exposure estimation

Occupational exposure to elevated noise levels depends on a variety of factors, including (i) occupation and industry and (ii) workplace-specific factors such as type of facility and process, raw materials, machinery, tools, the existence of engineering and work practice controls, and the existence, condition and use of personal protective devices. Thus exposure assessment was conducted using the occupational category approach (Equation 2), modified to reflect different noise exposures in occupations in different economic subsectors.

Our estimation of the proportion of workers in each occupational category with exposure to noise at or above $85 \mathrm{dBA}(\operatorname{PEW}(\mathrm{oc}(\mathrm{r}, \mathrm{g}) \mathrm{i})$ ) was based on United States data on the prevalence of noise exposure at or above 85 dBA among production workers in nine economic subsectors (NIOSH 1998; USDHHS 1986) (see Table 21.29).

The prevalence values among production workers were calculated from the US National Occupational Exposure Survey conducted during 1981-1983 (NIOSH 1998), which estimated the number of production workers exposed to noise at or above 85 dBA , by economic subsector. All other prevalence values were estimated by us, based on the NIOSH values for production workers. The value of 0.20 calculated for production workers in agriculture was extrapolated to all agricultural workers in all economic subsectors. Similarly, the value of 0.12 for
Table 21.28 Studies of noise exposures and hearing impairment in selected developing countries

| Country or area | Facility/job | Sound levels | Hearing loss | Notes | Source |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Brazil | Rotogravure printing workers | Continuous noise levels from 71 to 93 dBA | Some $49 \%$ of 124 workers exposed to noise and organic solvents had hearing loss ( $>25 \mathrm{~dB}$ ) in the high frequencies, significantly associated with age |  | Morata et al. (1997) |
| Egypt | Road traffic policemen in Cairo | Average 97 dBA with horns, 85 without; 97 at railway crossings | About 20-dB loss at all frequencies compared to office policemen |  | Kamal et al. (1989) |
| Egypt | Textile factory | 78-91 dBA in wool sorting and combing units | Compared to nonexposed controls, workers exposed to $<85 \mathrm{dBA}$ had only $1 \%$ increase in hearing impairment after 12 years. In workers exposed to $>85 \mathrm{dBA}$ the increased risk was $9.6 \%$ | Hearing impairment was defined as average of left and right ear thresholds at $0.5, \mathrm{I}$ and 2 kHz , $>25 \mathrm{~dB}$ | Moselhi et al. (1979) |
| Hong Kong SAR | Five industries: weaving, bottling, metal working, spinning, airport | $\mathrm{L}_{\text {eq }}$ (8-hour time-weighted average, dBA): weaving 102; bottling 94; metal working 96; spinning 97; airport 80-90 | Compared to controls, noise-exposed workers had significantly higher thresholds in most age groups and in all five industries, closely matching predicted values | No evidence was found for any ethnic differences between western groups and Cantonese Chinese, either in general hearing ability or in response to long-term noise exposure | Evans and Ming (1982) |
| India | Heavy engineering industry: machine shop and press divisions | Ranged from 83-116dBA. At selected work sites: press 94-110; machine shop 83-92; foundry 86-116 | Mean hearing threshold: 40 controls $4-24 \mathrm{~dB}$; 53 machine shop employees $14-40 \mathrm{~dB}$; 60 press employees 19-70 dB | Hearing impairment was progressive with age for all groups. Use of hearing protection was recommended | Raja and Ganguly (1983) |

Bhattacharya et al.
(198।)
(198I)
Authors recommended
engineering controls and

hearing conservation $\quad$| Bhattacharya et al. |
| :--- |
| $(1990)$ |

I20 weavers, exposed I-I5 years.
In the age range $30-34$ years,
median threshold of audibility in
the right/left ear was $55 / 55$
compared to $15 / 15$ for controls;
for 35 - 39 -year olds the threshold
was $60 / 55$ compared to $15 / 15$ for
controls
Hearing thresholds of 165 workers were significantly higher than nonexposed controls, and correlated significantly with
employment duration

| Textile mill weavers | 102-104 dBA |
| :---: | :---: |
| Drug and pharmaceutical company | Noise levels in dBA: fermentation 100-105; air compressor 95-102; ammonia compressor 93-97; primary air filter 104-106. Night shift levels were I-3dBA higher |
| Watch factory in Bangalore | Maximum noise levels ranged from 74 in assembly to 99 dBA in the diesel generator room |
| Car assembly | 94-108dB |

## $\stackrel{\text { 즐 }}{\underline{C}}$

## India

Nigeria
Table 21.28 Studies of noise exposures and hearing impairment in selected developing countries (continued)

| Country or area | Facility/job | Sound levels | Hearing loss | Notes | Source |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Nigeria | Textile workers in five factories in Lagos | Continuous noise levels of 95-115dBA | Hearing thresholds of 61 noise-exposed workers were significantly higher than 90 nonexposed controls. After 7 years of employment, exposed workers lost $2-12 \mathrm{~dB}$ per year, compared to $0.6-1.8 \mathrm{~dB}$ per year in controls | No hearing protection worn. Exposed workers did not display $4000-\mathrm{Hz}$ notch, and the shape of the audiograms ${ }^{\text {a }}$ was convex upwards, indicating lower losses at the middle frequencies. (Typical audiograms with noise-induced hearing loss display a convex downwards shape, indicating higher losses at the middle frequencies) | Oleru (1980) |
| Pakistan | Polyester fibre plant | Average noise levels: filament take-up unit 93.2 dBA ; texturizing unit 94.8 dBA ; compressor house 99.5 dBA | - | Typical exposure is 48 hours per week in these areas. Author recommended engineering controls and hearing conservation, including use of hearing protection | Shaikh (1996) |
| Saudi Arabia | 78 factories producing food, chemicals, plastics, metals, paper and other products | $86 \%$ exceeded 85 dBA , at least in part of the factory. In 12\% all of the factory exceeded 85 dBA | - | None of the factories practised noise protection | Alidrisi et al. (1990) |
| Singapore | Audiometric testing of noise-exposed workers is mandatory in Singapore. Most cases of noise-induced | Noise dosimetry on 46 of these cases showed a mean time-weighted exposure of 90 dBA | 127 cases of NID identified from 1985-1994. On average, after 24 years of exposure, the mean hearing threshold at I, 2 and 3 kHz was 62 dB | Author stated that NID is the leading occupational disease in Singapore, with $>500$ new cases per year | Tay (1996) |

deafness (NID) are in those employed in shipping and metal manufacturing, the remainder in transport, quarrying and other manufacturing
Gold mining
(cross-sectional survey
 Johannesburg)

$$
99-107 \mathrm{~dB}
$$

$$
-
$$

92\% (60/73) of workers exposed to noise for $\geq 10$ years in weaving
departments had mean hearing impairment of 60 dB compared to

20 dB for control group
100 miners tested audiometrically.
Of those with over 20 years, $23 \%$ were completely deaf

| Use of hearing protection | $\begin{array}{l}\text { Hessel and Sluis- } \\ \text { increased from 13\% in } \\ \text { Cremer (1987) }\end{array}$ |
| :--- | :--- |
| 1979 to $17 \%$ in 1982 |  |

Khogali (1970)
Noweir et al. (1968)

Hearing impairment was defined as average hearing loss of $>25 \mathrm{~dB}$ for 500,1000 and 2000 Hz , with 5 times weighting of better ear. None of the miners <22 years
progressively to $22 \%$ of those $\geq 58$
progressively to $22 \%$ of those $\geq 58$
years old

$$
\begin{array}{ll}
\begin{array}{l}
\text { Newly mechanized facility } \\
\text { Audiometric test methods } \\
\text { not described; hearing } \\
\text { impairment not defined }
\end{array} & \text { Khogali (1970) } \\
\text { Noweir et al. (1968) }
\end{array}
$$

Table 21.29 Prevalence of noise exposure $\geq 85 \mathrm{dBA}$

| Economic <br> subsector | Professional | Administrative | Clerical | Sales | Service | Agriculture | Production $^{\text {a }}$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Agriculture | 0.05 | 0.05 | 0.05 | 0.12 | 0.12 | 0.20 |
| Mining | 0.05 | 0.05 | 0.05 | 0.12 | 0.12 | 0.20 | 0.20 |
| Manufacturing | 0.05 | 0.05 | 0.05 | 0.12 | 0.12 | 0.20 | 0.22 |
| Electricity | 0.05 | 0.05 | 0.05 | 0.12 | 0.12 | 0.20 | 0.15 |
| Construction | 0.05 | 0.05 | 0.05 | 0.12 | 0.12 | 0.20 | 0.18 |
| Trade | 0.02 | 0.02 | 0.02 | 0.12 | 0.12 | 0.20 | 0.13 |
| Transport | 0.02 | 0.02 | 0.02 | 0.12 | 0.12 | 0.20 | 0.12 |
| Finance | 0.02 | 0.02 | 0.02 | 0.12 | 0.12 | 0.20 | 0.02 |
| Services | 0.02 | 0.02 | 0.02 | 0.12 | 0.12 | 0.20 | 0.03 |

a Source: NIOSH (1998).
b Based on $1.5 \%$ of workers exposed to noise in "business services".
production workers in transportation was extrapolated to all sales and service workers. The value for professional, administrative and clerical workers was extrapolated from 0.02 indicated for production workers in business services. The remaining value, 0.05 for professional, administrative and clerical workers in agriculture, mining, manufacturing, electricity and construction, was based on expert judgement.

The prevalence values were then partitioned into moderately high and high noise exposures, i.e. $\geq 85-90 \mathrm{dBA}$ and $>90 \mathrm{dBA}$, to estimate the proportions of workers exposed to moderately high and high levels of noise (EPF(r), exposure partitioning factor). Data from the United States (USDHHS 1986), taken from the 1981 Occupational Safety and Health Administration Final Regulatory Analysis for the Hearing Conservation Amendment, provide the distribution of noise exposure of over nine million American production workers (see Table 21.30). Of the 6063000 production workers with exposure at or above 85 dBA , slightly over half ( 3407000 or $56 \%$ ) were exposed above 90 dBA . The distribution of noise exposure levels among workers exposed over 90 dBA was also used to determine that 95 dBA is a reasonable level of noise to estimate risks among the workers in the high-exposure group ( $>90 \mathrm{dBA}$ ).

The partitioning of workers by occupational category and noise level was assigned as follows, based on data in Table 21.30. Among production workers exposed at or above 85 dBA , half were considered to be exposed at $\geq 85-90 \mathrm{dBA}$ and half exposed at $>90 \mathrm{dBA}$. (Note that these partitioning values do not consider the use of personal protective equipment.) Of the agricultural workers and sales and service workers exposed at or above 85 dBA it was assumed, based on expert judgement, that approximately $70 \%$ are exposed at $\geq 85-90 \mathrm{dBA}$ and $30 \%$ at $>90 \mathrm{dBA}$.

Table 21.30 Distribution of 9368000 United States production workers who had noise exposure levels of 80 dBA or greater

| Noise-exposure level (dBA) | Number of workers |
| :--- | :---: |
| $80-85$ | 3305000 |
| $86-90$ | 2656000 |
| $91-95$ | 1936000 |
| $96-100$ | 965000 |
| $>100$ | 506000 |
| Total $>85$ | 6063000 |
| Total $>90$ | 3407000 |

Source: USDOL OSHA 198I, cited in NIOSH (1991).

Table 21.31 Prevalence of noise exposure $85-90 \mathrm{dBA}$ in A subregions

| Economic <br> subsector | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.14 | 0.10 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Professional | Administrative | Clerical | Sales | Service | Agriculture | Production |
| Agriculture | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.14 | 0.43 |
| Mining | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.14 | 0.11 |
| Manufacturing | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.14 | 0.08 |
| Electricity | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.14 | 0.09 |
| Construction | 0.05 | 0.02 | 0.02 | 0.09 | 0.09 | 0.14 | 0.07 |
| Trade | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.14 | 0.06 |
| Transport | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.14 | 0.01 |
| Finance | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.14 | 0.02 |
| Services | 0.02 |  |  |  |  |  |  |

All professional, administrative and clerical workers with noise exposure at or above 85 dBA were assumed to be at the $\geq 85-90-\mathrm{dBA}$ level. Tables 21.31 and 21.32 present the distribution of noise exposure levels among workers in the A subregions by occupational category within economic sectors.

In the absence of global data, it was assumed that the same proportion of workers in these occupational categories in the developing countries would be exposed to noise levels at or above $85 \mathrm{dBA}(\mathrm{B}, \mathrm{C}, \mathrm{D}$, and E subregions). Given the rarity of hearing conservation programmes in the developing subregions, it was assumed that $5 \%$ of production workers would be exposed in the $\geq 85-90 \mathrm{dBA}$ category and $95 \%$ in the $>90 \mathrm{dBA}$ category (as opposed to $50 / 50$ for the A subregions). Additionally, because mechanization is not widespread for D and E subregions, the majority ( $95 \%$ ) of the agricultural workers exposed at or above 85 dBA were assigned to the $\geq 85-90-\mathrm{dBA}$ level. Assignment of all

Table 21.32 Prevalence of noise exposure $>90 \mathrm{dBA}$ in A subregions

| Economic <br> subsector | Occupational category |  |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Professional | Administrative | Clerical | Sales | Service | Agriculture | Production |
|  | 0 | 0 | 0 | 0.03 | 0.03 | 0.06 | 0.10 |
|  | 0 | 0 | 0 | 0.03 | 0.03 | 0.06 | 0.43 |
|  | 0 | 0 | 0 | 0.03 | 0.03 | 0.06 | 0.11 |
|  | 0 | 0 | 0 | 0.03 | 0.03 | 0.06 | 0.08 |
|  | 0 | 0 | 0 | 0.03 | 0.03 | 0.06 | 0.09 |
|  | 0 | 0 | 0 | 0.03 | 0.03 | 0.06 | 0.07 |
|  | 0 | 0 | 0 | 0.03 | 0.03 | 0.06 | 0.06 |
| Finance | 0 | 0 | 0 | 0.03 | 0.03 | 0.06 | 0.01 |
| Services | 0 | 0 | 0 | 0.03 | 0.03 | 0.06 | 0.02 |
|  |  |  |  |  |  |  |  |

Table 21.33 Prevalence of noise exposure $85-90 \mathrm{dBA}$ in $B$ and $C$ subregions

| Economic <br> subsector | Professional | Administrative | Clerical | Sales | Service | Agriculture | Production |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.14 | 0.01 |
| Mining | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.14 | 0.04 |
| Manufacturing | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.14 | 0.01 |
| Electricity | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.14 | 0.04 |
| Construction | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.14 | 0.01 |
| Trade | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.14 | 0.01 |
| Transport | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.14 | 0.01 |
| Finance | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.14 | 0.00 |
| Services | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.14 | 0.00 |

other occupational categories was the same as for the A subregions. Tables 21.33-21.36 reflect the different partitioning for the $\mathrm{B}+\mathrm{C}$ and D + E subregions.

Table 21.37 presents the proportions of workers exposed to moderately high and to high noise levels by subregion, age and sex. The proportions of males exposed to these noise levels were consistently higher than those of females, owing both to higher rates of participation in the labour force and to higher rates of females working in the services sector.

### 4.3 RISK FACTOR-DISEASE RELATIONSHIPS

High noise levels in the workplace may cause elevated blood pressure, sleeping difficulties, annoyance and stress. Excessive noise can interfere

Table 21.34 Prevalence of noise exposure $>90 \mathrm{dBA}$ in B and C subregions

| Economic <br> subsector | Professional | Administrative | Clerical | Sales | Service | Agriculture | Production |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.06 | 0.19 |
| Mining | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.06 | 0.81 |
| Manufacturing | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.06 | 0.21 |
| Electricity | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.06 | 0.14 |
| Construction | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.06 | 0.17 |
| Trade | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.06 | 0.12 |
| Transport | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.06 | 0.11 |
| Finance | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.06 | 0.02 |
| Services | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.06 | 0.03 |
|  |  |  |  |  |  |  |  |

Table 21.35 Prevalence of noise exposure $85-90 \mathrm{dBA}$ in D and E subregions

| Economic <br> subsector | Professional | Administrative | Clerical | Sales | Service | Agriculture | Production |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Agriculture | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.19 |
| Mining | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.19 | 0.04 |
| Manufacturing | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.19 | 0.01 |
| Electricity | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.19 | 0.01 |
| Construction | 0.05 | 0.05 | 0.05 | 0.09 | 0.09 | 0.19 | 0.01 |
| Trade | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.19 | 0.01 |
| Transport | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.19 | 0.01 |
| Finance | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.19 | 0.00 |
| Services | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.19 | 0.00 |
|  |  |  |  |  |  |  |  |

with communications in the workplace, resulting in property damage or personal injury. Tinnitus ${ }^{6}$ and temporary threshold shift ${ }^{7}$ may also occur. However, the most serious effect is irreversible hearing impairment, resulting from damage to the delicate hearing mechanisms of the inner ear. Noise-induced hearing loss typically begins in the frequency range (pitch) of human voices and thus interferes with spoken communication.

Noise-induced hearing loss is caused by exposure to loud noises, such as those produced by woodworking equipment, chain saws, heavy machinery, gunfire, aircraft or amplified music. Permanent hearing loss from exposure to noise may happen quite early and an audiometric notch, or initial loss at or around 4000 Hz , may be noticeable within six

Table 21.36 Prevalence of noise exposure $>90 \mathrm{dBA}$ in D and E subregions

| Economic <br> subsector | Professional | Administrative | Clerical | Sales | Service | Agriculture | Production |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.01 | 0.19 |
| Mining | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.01 | 0.81 |
| Manufacturing | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.01 | 0.21 |
| Electricity | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.01 | 0.14 |
| Construction | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.01 | 0.17 |
| Trade | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.01 | 0.12 |
| Transport | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.01 | 0.11 |
| Finance | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.01 | 0.02 |
| Services | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | 0.01 | 0.03 |

months to one year from starting a job with a hazardous noise exposure. There is significant variation in the susceptibility to noise damage, so that two workers with the same exposure may not experience the same hearing impairment. With prolonged exposure to the same noise, hearing loss continues to worsen. For a given noise environment, most of the hearing loss occurs in the first few years, although there is a slower continuing progression as long as the noise exposure continues.

When a person is removed from the noise, hearing loss does not worsen but does remain permanent. Any additional hearing loss after termination of work in a noisy environment is due to other causes, most often presbycusis (age-related hearing loss). Most people are subject to presbycusis, which is the most common form of sensorineural hearing impairment. Data show that from as early as 30 years of age, and gradually increasing in later years, some hearing loss occurs in the general population. Individual variation is great, with around $50 \%$ of the population maintaining good hearing into old age. Other factors, such as ear infection secondary to airborne contaminants, mechanical injury or chemical substances can lead to or aggravate hearing impairment in the workplace.

## DESCRIPTION OF STUDIES

In the literature review, only three studies were found that indicated the frequency of hearing impairment at different thresholds of hearing (Davis 1989; Malchaire 2000; Waitzman and Smith 1999). Malchaire compared the expected percentage of subjects who, at age 55 , presented with hearing impairment at 25 and 50 dBHL with the personal level of exposure (Lpe) to noise in dBA , in the absence of noises $>140 \mathrm{~dB}$. The exposure time frame was 35 years (Table 21.38).

Table 21.37 Proportions of the working-age population occupationally exposed to different noise levels, by sex and subregion

| Subregion | Sex | Exposure level | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | $<85 \mathrm{dBA}$ | 0.87 | 0.84 | 0.84 | 0.86 | 0.89 | 0.95 |
|  |  | 85-90 dBA | 0.09 | 0.12 | 0.11 | 0.10 | 0.08 | 0.04 |
|  |  | $>90 \mathrm{dBA}$ | 0.04 | 0.04 | 0.04 | 0.04 | 0.03 | 0.01 |
|  | Female | $<85 \mathrm{dBA}$ | 0.92 | 0.90 | 0.90 | 0.92 | 0.96 | 0.98 |
|  |  | 85-90 dBA | 0.07 | 0.09 | 0.09 | 0.07 | 0.04 | 0.02 |
|  |  | $>90 \mathrm{dBA}$ | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 |
| AFR-E | Male | $<85 \mathrm{dBA}$ | 0.87 | 0.84 | 0.84 | 0.86 | 0.89 | 0.95 |
|  |  | 85-90 dBA | 0.09 | 0.12 | 0.11 | 0.10 | 0.08 | 0.04 |
|  |  | $>90 \mathrm{dBA}$ | 0.04 | 0.04 | 0.04 | 0.04 | 0.03 | 0.02 |
|  | Female | $<85 \mathrm{dBA}$ | 0.92 | 0.90 | 0.90 | 0.92 | 0.96 | 0.98 |
|  |  | 85-90 dBA | 0.07 | 0.08 | 0.08 | 0.07 | 0.04 | 0.02 |
|  |  | $>90 \mathrm{dBA}$ | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 |
| AMR-A | Male | $<85 \mathrm{dBA}$ | 0.92 | 0.90 | 0.90 | 0.91 | 0.93 | 0.97 |
|  |  | 85-90 dBA | 0.05 | 0.06 | 0.06 | 0.06 | 0.04 | 0.02 |
|  |  | $>90 \mathrm{dBA}$ | 0.03 | 0.04 | 0.04 | 0.03 | 0.02 | 0.01 |
|  | Female | $<85 \mathrm{dBA}$ | 0.96 | 0.95 | 0.95 | 0.96 | 0.98 | 0.99 |
|  |  | 85-90 dBA | 0.03 | 0.03 | 0.03 | 0.03 | 0.01 | 0.01 |
|  |  | $>90 \mathrm{dBA}$ | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 |
| AMR-B | Male | $<85 \mathrm{dBA}$ | 0.90 | 0.87 | 0.88 | 0.89 | 0.91 | 0.96 |
|  |  | 85-90 dBA | 0.05 | 0.06 | 0.06 | 0.05 | 0.04 | 0.02 |
|  |  | $>90 \mathrm{dBA}$ | 0.06 | 0.07 | 0.07 | 0.06 | 0.05 | 0.02 |
|  | Female | $<85 \mathrm{dBA}$ | 0.95 | 0.94 | 0.94 | 0.95 | 0.97 | 0.99 |
|  |  | 85-90 dBA | 0.03 | 0.03 | 0.04 | 0.03 | 0.02 | 0.01 |
|  |  | $>90 \mathrm{dBA}$ | 0.02 | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 |
| AMR-D | Male | $<85$ dBA | 0.91 | 0.89 | 0.89 | 0.90 | 0.93 | 0.96 |
|  |  | 85-90 dBA | 0.03 | 0.04 | 0.04 | 0.04 | 0.03 | 0.01 |
|  |  | $>90 \mathrm{dBA}$ | 0.05 | 0.07 | 0.07 | 0.06 | 0.05 | 0.02 |
|  | Female | $<85 \mathrm{dBA}$ | 0.95 | 0.94 | 0.94 | 0.96 | 0.97 | 0.99 |
|  |  | 85-90 dBA | 0.02 | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 |
|  |  | $>90 \mathrm{dBA}$ | 0.03 | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 |
| EMR-B | Male | $<85 \mathrm{dBA}$ | 0.91 | 0.88 | 0.89 | 0.90 | 0.92 | 0.96 |
|  |  | 85-90 dBA | 0.04 | 0.05 | 0.05 | 0.04 | 0.03 | 0.02 |
|  |  | $>90 \mathrm{dBA}$ | 0.05 | 0.07 | 0.07 | 0.06 | 0.05 | 0.02 |
|  | Female | $<85 \mathrm{dBA}$ | 0.96 | 0.95 | 0.95 | 0.96 | 0.98 | 0.99 |
|  |  | 85-90 dBA | 0.02 | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 |
|  |  | $>90 \mathrm{dBA}$ | 0.02 | 0.02 | 0.02 | 0.02 | 0.01 | 0.01 |
| EMR-D | Male | $<85 \mathrm{dBA}$ | 0.88 | 0.85 | 0.85 | 0.86 | 0.90 | 0.95 |
|  |  | 85-90 dBA | 0.09 | 0.11 | 0.11 | 0.10 | 0.07 | 0.04 |
|  |  | $>90 \mathrm{dBA}$ | 0.04 | 0.04 | 0.04 | 0.04 | 0.03 | 0.02 |
|  | Female | $<85 \mathrm{dBA}$ | 0.91 | 0.89 | 0.89 | 0.92 | 0.95 | 0.98 |
|  |  | 85-90 dBA | 0.07 | 0.09 | 0.09 | 0.07 | 0.04 | 0.02 |
|  |  | $>90 \mathrm{dBA}$ | 0.02 | 0.02 | 0.02 | 0.02 | 0.01 | 0.00 |
| EUR-A | Male | $<85 \mathrm{dBA}$ | 0.92 | 0.90 | 0.90 | 0.91 | 0.93 | 0.97 |
|  |  | 85-90 dBA | 0.05 | 0.06 | 0.06 | 0.06 | 0.04 | 0.02 |
|  |  | $>90 \mathrm{dBA}$ | 0.03 | 0.04 | 0.04 | 0.03 | 0.02 | 0.01 |

Table 21.37 Proportions of the working-age population occupationally exposed to different noise levels, by sex and subregion (continued)

| Subregion | Sex | Exposure level | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EUR-B | Female | $<85 \mathrm{dBA}$ | 0.96 | 0.96 | 0.95 | 0.97 | 0.98 | 0.99 |
|  |  | 85-90 dBA | 0.03 | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 |
|  |  | $>90 \mathrm{dBA}$ | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 |
|  | Male | $<85 \mathrm{dBA}$ | 0.88 | 0.85 | 0.85 | 0.87 | 0.90 | 0.95 |
|  |  | 85-90 dBA | 0.05 | 0.06 | 0.06 | 0.05 | 0.04 | 0.02 |
|  |  | $>90 \mathrm{dBA}$ | 0.07 | 0.09 | 0.09 | 0.08 | 0.06 | 0.03 |
|  | Female | $<85$ dBA | 0.93 | 0.91 | 0.91 | 0.93 | 0.96 | 0.98 |
|  |  | 85-90 dBA | 0.04 | 0.05 | 0.05 | 0.04 | 0.02 | 0.01 |
|  |  | $>90$ dBA | 0.03 | 0.04 | 0.04 | 0.03 | 0.02 | 0.01 |
| EUR-C | Male | $<85 \mathrm{dBA}$ | 0.88 | 0.85 | 0.85 | 0.87 | 0.90 | 0.95 |
|  |  | 85-90 dBA | 0.04 | 0.05 | 0.05 | 0.04 | 0.03 | 0.02 |
|  |  | $>90 \mathrm{dBA}$ | 0.08 | 0.10 | 0.10 | 0.09 | 0.07 | 0.04 |
|  | Female | $<85 \mathrm{dBA}$ | 0.93 | 0.92 | 0.92 | 0.94 | 0.96 | 0.98 |
|  |  | 85-90 dBA | 0.02 | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 |
|  |  | $>90 \mathrm{dBA}$ | 0.04 | 0.05 | 0.05 | 0.04 | 0.02 | 0.01 |
| SEAR-B | Male | $<85 \mathrm{dBA}$ | 0.88 | 0.84 | 0.84 | 0.87 | 0.91 | 0.95 |
|  |  | 85-90 dBA | 0.06 | 0.08 | 0.08 | 0.06 | 0.04 | 0.02 |
|  |  | $>90 \mathrm{dBA}$ | 0.06 | 0.08 | 0.08 | 0.07 | 0.05 | 0.03 |
|  | Female | $<85 \mathrm{dBA}$ | 0.92 | 0.88 | 0.88 | 0.91 | 0.96 | 0.98 |
|  |  | 85-90 dBA | 0.05 | 0.08 | 0.08 | 0.05 | 0.02 | 0.01 |
|  |  | $>90 \mathrm{dBA}$ | 0.04 | 0.04 | 0.04 | 0.03 | 0.02 | 0.01 |
| SEAR-D | Male | $<85 \mathrm{dBA}$ | 0.87 | 0.79 | 0.80 | 0.84 | 0.88 | 0.94 |
|  |  | 85-90 dBA | 0.09 | 0.15 | 0.15 | 0.11 | 0.08 | 0.04 |
|  |  | $>90 \mathrm{dBA}$ | 0.04 | 0.05 | 0.05 | 0.05 | 0.04 | 0.02 |
|  | Female | $<85 \mathrm{dBA}$ | 0.91 | 0.94 | 0.95 | 0.96 | 0.98 | 0.99 |
|  |  | 85-90 dBA | 0.07 | 0.04 | 0.03 | 0.02 | 0.01 | 0.01 |
|  |  | $>90$ dBA | 0.02 | 0.02 | 0.02 | 0.02 | 0.01 | 0.00 |
| WPR-A | Male | $<85 \mathrm{dBA}$ | 0.92 | 0.90 | 0.90 | 0.92 | 0.96 | 0.98 |
|  |  | 85-90 dBA | 0.04 | 0.06 | 0.06 | 0.04 | 0.02 | 0.01 |
|  |  | $>90 \mathrm{dBA}$ | 0.03 | 0.04 | 0.04 | 0.03 | 0.03 | 0.01 |
|  | Female | $<85 \mathrm{dBA}$ | 0.95 | 0.94 | 0.94 | 0.96 | 0.97 | 0.99 |
|  |  | 85-90 dBA | 0.03 | 0.04 | 0.04 | 0.03 | 0.02 | 0.01 |
|  |  | $>90 \mathrm{dBA}$ | 0.02 | 0.02 | 0.02 | 0.01 | 0.01 | 0.00 |
| WPR-B | Male | $<85 \mathrm{dBA}$ | 0.87 | 0.84 | 0.84 | 0.86 | 0.89 | 0.95 |
|  |  | 85-90 dBA | 0.06 | 0.08 | 0.08 | 0.07 | 0.05 | 0.03 |
|  |  | $>90 \mathrm{dBA}$ | 0.07 | 0.08 | 0.08 | 0.07 | 0.06 | 0.03 |
|  | Female | $<85 \mathrm{dBA}$ | 0.93 | 0.91 | 0.91 | 0.93 | 0.96 | 0.98 |
|  |  | 85-90 dBA | 0.04 | 0.05 | 0.05 | 0.04 | 0.02 | 0.01 |
|  |  | $>90 \mathrm{dBA}$ | 0.03 | 0.04 | 0.04 | 0.03 | 0.02 | 0.01 |

Table 21.38 Expected percentages of workers with hearing loss of $>25 \mathrm{~dB}$ or $>50 \mathrm{~dB}$ after 35 years of exposure, by personal level of exposure

|  | Personal exposure (dBA) |  |  |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
|  | Hearing loss | 85 | 90 | 92 | 94 | 97 | 98 | 99 |
| 50 dB | 4 | 5 | 7 | 9 | 15 | 18 | 21 | 26 |
| 25 dB | 29 | 36 | 40 | 46 | 59 | 65 | 70 | 75 |

Source: Malchaire (2000).

Waitzman and Smith (1999) analysed data from the United States Health Examination Survey (1960-1961) and the National Health and Nutrition Examination Survey (NHANES I, 1971-1975). Hearing loss was rated by a scheme developed by Klockhoff et al. (1974). The Klockhoff analysis used four hearing loss levels (Slight, Moderate, Severe1, Severe2), which were related in this analysis to the standard hearing impairment scales by using typical noise-induced hearing loss curves. The categories developed by Klockhoff et al. were based on the frequencies used in the National Health Examination Survey (500, 1000, $2000,3000,4000$ and 6000 Hz ), and the NHANES I study (500, 1000, 2000 and 4000 Hz ). As an example of this conversion procedure, the criteria of Klockhoff et al. for slight hearing loss were based on a $30-\mathrm{dB}$ loss at 4000 Hz . Using audiometric data presented by Klockhoff et al. (1974), we estimated the equivalent hearing losses for each category as follows:

- Slight: $>21 \mathrm{dBHL}$
- Moderate: $>38 \mathrm{dBHL}$
- Severe1: $>41 \mathrm{dBHL}$
- Severe2: $>50 \mathrm{dBHL}$

Waitzman and Smith reported odds ratios calculated by multivariate regressions for two age groups for workers in construction, manufacturing/mining and other subsectors (see Table 21.39). Blue-collar construction workers experienced between 2 and $>3.5$ times the risk experienced by white-collar workers in "other industries". The pattern of their hearing loss at normal speech frequencies significantly disrupted their ability to communicate.

Davis (1989) reported on the prevalence of hearing loss as a function of age in the adult population of Great Britain. Audiometric analyses on adults ranging in age from 17 to $\geq 80$ years were conducted in four cities. Hearing impairment was reported for $>25,>45$ and $>65 \mathrm{dBHL}$. Davis found a "significant" level of hearing loss ( $>25 \mathrm{dBHL}$ ) in $16 \%$ of the adult population (17- $\geq 80$ years).

Table 21.39 Odds ratios from logistic multivariate regressions on audiometric measures of hearing loss, by age group

| Type of worker/industry | Ages 25-44 years |  |  |  | Ages 45-65 years |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\begin{gathered} \text { Slight } \\ >21 \mathrm{~dB} \end{gathered}$ | Moderate $>38 \mathrm{~dB}$ | Severe I $>4 I \mathrm{~dB}$ | $\begin{aligned} & \text { Severe } 2 \\ & >50 \mathrm{~dB} \end{aligned}$ | $\begin{gathered} \text { Slight } \\ >21 \mathrm{~dB} \end{gathered}$ | Moderate $>38 \mathrm{~dB}$ | Severe I $>41 \mathrm{~dB}$ | Severe 2 $>50 \mathrm{~dB}$ |
| White collar |  |  |  |  |  |  |  |  |
| Construction | 0.85 | 0.00 | 0.00 | 0.00 | 2.50 | 2.17 | 1.68 | 2.18 |
| Manufacturing/ mining | 0.91 | 1.45 | 1.17 | 1.14 | 1.02 | 1.43 | 0.98 | 1.25 |
| Other industry groups | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Blue collar |  |  |  |  |  |  |  |  |
| Construction | 2.79 | 3.50 | 2.34 | 2.65 | 3.08 | 3.54 | 1.98 | 2.38 |
| Manufacturing/ mining | 2.01 | 3.03 | 1.94 | 2.40 | 2.33 | 1.86 | 1.88 | 1.90 |
| Other industry groups | 1.38 | 2.42 | 1.92 | 1.95 | 1.84 | 1.94 | 1.40 | 1.69 |

Source: Waitzman and Smith (1999).

Several studies were found that presented relative risk estimates by specific occupation. The risk estimates were higher than those determined in this analysis for exposed workers, as they focused on occupations with generally high noise exposures. These studies were based in Canada (Hessel 2000), Germany (Arndt et al. 1996) and Great Britain (Palmer et al. 2001) (see Table 21.40).

In Germany, a prevalence ratio for hearing loss of 1.5 was found in construction workers vs white-collar employees. Hearing impairment was defined as the sum of thresholds at 2000,3000 and 4000 Hz greater than 105 dB at least in one ear. Hessel (2000), in a similar study in Canada, found that construction workers (with the exception of boilermakers) had a lower prevalence of hearing loss than the Germans. Also, prevalence of hearing impairment in the comparison group in Canada was lower than in Germany. According to the author, the differences found may be due to the year of the study. The Canadian study was carried out 7-9 years later than the German study, and there may have been lower occupational noise levels and/or use of personal hearing protection. These potentially confounding factors were not described in the German study.

In contrast to the Canadian and German studies, prevalence ratios determined in Great Britain were based on self-reported hearing impairment. Prevalence of "ever employed in a noisy job" was compared against "never exposed in a noisy job". A noisy place was defined as one "where there was a need to shout to be heard". The questions used to define hearing impairment were modelled on those used in the Institute
Table 21.40 Prevalence ratio of occupational noise-induced hearing impairment in available studies

| Country | Reference | Definition of hearing impairment | Occupation | Prevalence ratio | 95\% Cl |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Canada (Edmonton) | Hessel (2000) | Greater than 105 dB hearing | Plumbers | 2.91 | - |
|  |  | loss at $2,3,4 \mathrm{kHz}$ (corresponds | Boilermakers | 3.88 | - |
|  |  | to > 35 dBHL ) | Electricians | 1.46 | - |
| Germany | Arndt et al. (1996) | Greater than 105 dB healing loss at $2,3,4 \mathrm{kHz}$ (corresponds to $>35 \mathrm{dBHL}$ ) | Carpenters | 1.77 | 1.48-2.12 |
|  |  |  | Unskilled workers | 1.75 | 1.47-2.09 |
|  |  |  | Plumbers | 1.49 | I.19-1.85 |
|  |  |  | Painters | 1.2 | 0.96-1.49 |
|  |  |  | Plasterers | 1.29 | 1.05-1.59 |
|  |  |  | Overall | 1.5 | 1.29-1.82 |
| Great Britain | Palmer et al. (2001) | Severe: wearing a hearing aid or having great difficulty in both ears on hearing conversation in a quiet room $>45 \mathrm{dBHL}$ | Male | 2.9 | - |
|  |  |  | Female | 1.8 | - |
|  |  | Moderate and worse: reported | Male | 3.6 | - |
|  |  | moderate difficulty in hearing conversation in a quiet room equivalent to 45 dBHL | Female | 2.9 | - |

[^81]of Hearing Research national survey of hearing, in which those who reported moderate or worse hearing impairment were found to have a mean hearing loss of 45 dBHL . The mean hearing loss of 45 dBHL is similar to the cut-off of 41 dBHL used by WHO and the Global Burden of Disease study, whereas the cut-off used in the German and Canadian studies corresponded to slight hearing impairment ( $>35 \mathrm{dBHL}$ ).

NIOSH, in a re-analysis of the data from its Occupational Noise and Hearing Survey (Prince et al. 1997) derived excess risk ${ }^{8}$ estimates with a model that used the average of $1000,2000,3000$ and 4000 Hz and a hearing loss $>25 \mathrm{dBHL}$. We used this information to develop excess risk estimates for workers exposed at $85-90 \mathrm{dBA}$ (defined by us as moderately high exposure) and $>90 \mathrm{dBA}$ (defined by us as high noise exposure, equivalent to 95 dBA ). To estimate excess risks for workers exposed to moderately high noise, we used the observed exposure-response relationships developed by NIOSH (1998) for workers of different ages who were exposed at 80,85 and 90 dBA for various numbers of years. The data show that at any noise level, hearing impairment increases with age and/or length of exposure. Also, the highest risk is found at the highest levels of exposure. Prince et al. (1997) found a small increase in excess risk in workers exposed to $80-84 \mathrm{dBA}$ vs a $<80 \mathrm{dBA}$ control group; however, these risk estimates are imprecise owing to the low numbers of workers in the study exposed to noise at these levels.

NIOSH (1998) also provides two data points of excess risk for workers exposed at 95 dBA for prolonged periods. Table 21.41 illustrates our estimation of excess risk of material hearing impairment at $>25 \mathrm{dBHL}$ for workers exposed at 95 dBA , based on these two data points. The excess risk value of $38.3 \%$ at 95 dBA for a 65 -year-old worker after 10 or more years of exposure was taken from NIOSH (1998), Appendix Table IV. In addition, an excess risk value of $19.5 \%$ was taken from the table for 30 -year-old workers exposed to noise at 95 dBA and for a duration of exposure of $5-10$ years. The values for $30-, 40-$ and 50 -year olds with $>10$ years of exposure were interpolated using the ratios of change of hearing loss with age at 90 dBA between each age group in Table 21.41. All calculations of NIOSH exposureresponse relationships were based on material hearing impairment at $>25 \mathrm{dBHL}$ and were adjusted by us to reflect noise-induced hearing loss at $\geq 41 \mathrm{dBHL}$.

The International Organization for Standardization (ISO) has also developed procedures for estimating hearing loss due to noise exposure. Their most recent model (referred to as the "1990-ISO model") and the 1997 NIOSH model (NIOSH 1998) are reasonably similar. Table 21.42 summarizes the excess risk estimates developed separately by NIOSH and ISO for material hearing impairment $>25 \mathrm{dBHL}$ caused by occupational noise exposure.

Table 21.4I Excess risk estimates for material hearing impairment $>25 \mathrm{dBHL}$, by age and duration of exposure

|  | Excess risk (\%) |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
| Average daily exposure (dBA) | Age 30 | Age 40 | Age 50 | Age 60 |
| $5-10$ years of exposure | 19.5 |  |  |  |
| $95^{\mathrm{a}}$ (estimated) | 5.4 | 9.7 | 14.3 | 15.9 |
| 90 | 1.4 | 2.6 | 4 | 4.9 |
| 85 | 0.2 | 0.4 | 0.6 | 0.8 |
| 80 |  |  |  |  |
| $>10$ years of exposure | 24 | 31 | 38 | 38.3 |
| $95^{\mathrm{a}}$ (estimated) | 10.3 | 17.5 | 24.1 | 24.7 |
| 90 | 2.3 | 4.3 | 6.7 | 7.9 |
| 85 | 0.3 | 0.6 | 1 | 1.3 |
| 80 |  |  |  |  |

a Estimates for 95 dBA were developed from NIOSH 1998 using methods described in the text. Source: NIOSH (1998).

Table 21.42 NIOSH and ISO estimated excess risk of incurring material hearing impairment ( $>25 \mathrm{dBHL}$ at I, 2, 3 and 4 kHz ) over a 40 -year working lifetime and at various average noise exposures

|  | Excess risk (\%) |  |  |
| :--- | :---: | :---: | :---: |
| Average daily noise exposure (dBA) | ISO | NIOSH |  |
| 90 | 17 | 25 |  |
| 85 | 6 | 8 |  |
| 80 | 1 | 1 |  |
| Source: | NIOSH (1998). |  |  |

## EXtrapolating from risks at >25dBHL TO RISKS at $\geq 41$ dBHL

Occupational hearing loss is usually denoted as $>25 \mathrm{dBHL}$ but WHO uses $\geq 41 \mathrm{dBHL}$ as a cut-off point to estimate prevalence of hearing loss. Therefore, extrapolations were made from studies of occupational risks at $>25 \mathrm{dBHL}$ to estimate risk to workers of hearing loss at 41 dBHL or greater. Data from the USDOL OSHA 1981 Final Regulatory Analysis for the Hearing Conservation Amendment (NIOSH 1991) provided a means of adjusting the various reports based on material hearing impairment $>25 \mathrm{dBHL}$ to $>40 \mathrm{dBHL}$, a level of hearing loss assumed for this project to be equivalent to the WHO definition of $\geq 41 \mathrm{dBHL}$. As presented in Table 21.43, OSHA estimated the number of workers with various levels of hearing loss or impairment. The number of expected

Table 21.43 Hearing levels (dBHL) of 9368000 United States production workers with noise exposure levels of $\geq 80 \mathrm{dBA}$

| Hearing threshold level <br> $(I, 2$ and 3 kHz$)$ | Cumulative cases | Expected cases | Excess cases |
| :--- | :---: | :---: | :---: |
| $>15 \mathrm{~dB}$ (mild hearing loss) | $3735000(40 \%)$ | $2111000(23 \%)$ | $1624000(17 \%)$ |
| $>25 \mathrm{~dB}$ (material hearing impairment) | $2025000(22 \%)$ | $965000(10 \%)$ | $1060000(11 \%)$ |
| $>40 \mathrm{~dB}$ (moderate to severe hearing | $718000(8 \%)$ | $245000(3 \%)$ | $473000(5 \%)$ |
| impairment) |  |  |  |

Source: USDOL OSHA, I98I, cited in NIOSH (1991).

Table 21.44 Estimated excess risk of incurring hearing impairment at 41 dBHL or greater over a 40-year working lifetime and at various average noise exposures

|  | Excess risk (\%) |  |
| :--- | :---: | ---: |
| Average daily noise exposure (dBA) | ISO | NIOSH |
| 90 | 7.6 | 11.2 |
| 85 | 2.7 | 3.6 |
| 80 | 0.4 | 0.4 |

cases (based on hearing levels of a nationwide sample of adults in U.S. Public Health Service hearing surveys) was subtracted to derive the number of excess cases at various levels of hearing loss or impairment in United States production workers (OSHA 1981, reported in NIOSH 1991).

Using the data in Table 21.43, a factor to correct excess risk data at $>25$ dBHL to WHO's excess risk at $\geq 41 \mathrm{dBHL}$ was determined as a ratio of the number of excess cases at $>40 \mathrm{~dB}$ divided by the number of excess cases at $>25 \mathrm{dBHL}$; i.e. 473000 divided by 1060000 yields a ratio of 0.446 . This correction factor of 0.446 was used to correct excess risk at $>25 \mathrm{dBHL}$ from the reported excess risk in Table 21.42 to the excess risk at $\geq 41 \mathrm{dBHL}$ as presented in Table 21.44 . As no additional data were available, the hearing impairment of production workers at $>40 \mathrm{dBHL}$ was assumed to be equivalent to the WHO definition of hearing loss of 41 dBHL or greater used in this project.

Excess risk estimates for hearing impairment at $\geq 41 \mathrm{dBHL}$ are presented in Table 21.45. They were generated by applying the same correction factor of 0.446 to Table 21.41.

## RELATIVE RISK ESTIMATION FOR NOISE INDUCED HEARING LOSS AT $\geq 41 \mathrm{dBHL}$

 The relative risk values were extrapolated using the following formula:Table 21.45 Estimated excess risk for hearing impairment at $\geq 41 \mathrm{dBHL}$, by age and duration of exposure

|  | Excess risk (\%) |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
| Average daily exposure (dBA) | Age 30 | Age 40 | Age 50 | Age 60 |
| $5-10$ years of exposure |  |  |  |  |
| 95 | 8.7 |  |  |  |
| 90 | 2.4 | 4.3 | 6.4 | 7.1 |
| 85 | 0.6 | 1.2 | 1.8 | 2.2 |
| 80 | 0.1 | 0.2 | 0.3 | 0.4 |
| $>10$ years of exposure |  |  |  |  |
| 95 | 10.7 | 13.8 | 16.9 | 17.0 |
| 90 | 4.6 | 7.8 | 10.7 | 11.0 |
| 85 | 1.0 | 1.9 | 3.0 | 3.5 |
| 80 | 0.1 | 0.3 | 0.4 | 0.6 |

## Relative risk $=1+($ excess risk $/$ expected risk $)$

Excess risk is defined in this study as "the percentage of workers with a hearing impairment in an occupationally noise-exposed population, minus the percentage who would normally incur such impairment from aging in an unexposed population". The expected risk is the prevalence for the general unexposed population. While the NIOSH document provides the excess risk of the exposed population, the expected risk is not reported by NIOSH. Data from Davis (1989) estimates prevalence as a function of age in the adult population of Great Britain. The average prevalence for both ears for a noise-induced hearing loss of $>45 \mathrm{dBHL}$ was calculated by us for the general population, using the data from Davis (1989) and the methods described above to adjust NIOSH data for noise-induced hearing loss at $>25 \mathrm{dBHL}$ to generate Table 21.45. The results are reported in Table 21.46.

In our study, relative risks for the age groups $0-4$ and $5-14$ years were not estimated, as occupational risks are not present and/or data are unavailable on levels or length of exposure. In the calculation of excess risk, the age group 15-29 years was assigned the lowest excess risk value of 8.7 in Table 21.45 for age 30 with $5-10$ years of exposure. For the category of workers with moderately high noise exposure ( $85-90 \mathrm{dBA}$ ), the excess risk estimate is the geometric mean of the excess risk estimates for 85 dBA and 90 dBA for each age group (see Table 21.45).

The older age groups ( $30-44$ years, $45-59$ years, etc.) did not neatly fit the age categories in Table 21.45, so worker-population-weighted averages were constructed for excess risk values at the required ages. For example, the excess risk estimate for the age group 30-44 years, at 85-90 dBA , was calculated by first taking the geometric mean of the excess risk

Table 21.46 Prevalence of hearing loss at $>45 \mathrm{dBHL}$ for the general population of Great Britain

| Age group (years) | Prevalence |
| :--- | :---: |
| $I 7-30$ | 1.25 |
| $3 I-40$ | 1.90 |
| $4 I-50$ | 4.75 |
| $5 I-6 I$ | 6.40 |
| $6 I-70$ | 9.35 |
| $7 I-80$ | 16.55 |
| $\geq 8 I$ | 25.35 |
| Source: Davis (I989). |  |

Table 21.47 Relative risks by age group and level of exposure

|  | Age group (years) |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| Level | $15-29$ | $30-44$ | $45-59$ | $60-69$ | $70-79$ | $\geq 80$ |
| $<85 \mathrm{dBA}$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $85-90 \mathrm{dBA}$ | 1.96 | 2.24 | 1.91 | 1.66 | 1.12 | 1.00 |
| $>90 \mathrm{dBA}$ | 7.96 | 5.62 | 3.83 | 2.82 | 1.62 | 1.00 |

estimates at 85 and 90 dBA , for people with $>10$ years of exposure in the 30 - and 40 -year-old categories. These values were then weighted by the worker population in the age groups 30-39 and 40-44 years. In a similar procedure, the prevalence values in Table 21.46 for the general population were adjusted to the same age groups. Relative risks for the age groups $70-79$ years and $\geq 80$ years were calculated from figures in Prince et al. (1997) and the prevalence data from Davis (1989). Table 21.47 presents the relative risks by age group and level of exposure.

## RISK REVERSIBILITY

It was assumed that risk reversibility following exposure removal was immediate. In other words, for those previously exposed who have not developed noise-induced hearing loss yet, the risk is removed after exposure removal (no new cases). Those with the condition, however, will continue to be affected by it.

## 5. OcCupational ergonomic stressors

The physical ergonomic features of work that are most frequently cited as risk factors for MSDs include a rapid pace of work and repetitive motion patterns; insufficient recovery time; heavy lifting and other forceful manual exertions; non-neutral body postures (either dynamic or
static); mechanical pressure concentrations; vibration (both segmental and whole-body); and low temperature. For the present analysis, the risk factor is exposure to the combination of occupational exposures that are implicated in the etiology of low back pain, including physical stressors and possibly psychosocial or work organization features as well.

### 5.1 Exposure variable

Assessing the fraction of back pain disorders that can be attributed to occupation requires that an indicator be identified that can be measured on a global scale and that can also be matched with data on known exposure-risk relationship(s). The various reviews of low back pain epidemiology have implicated an overlapping set of occupational exposures, such as lifting, forceful movements, awkward postures, whole-body vibration and perhaps psychosocial stressors. However, such exposures are rarely assessed in surveillance activities on a large scale, and thus data are not available for risk assessment calculations at the global level.

In contrast, low back pain (and other MSD morbidity) is commonly reported by broad industrial or occupational groupings. Thus, even though occupation is a less precise indicator of risk than a specific exposure, its widespread availability in administrative data sets and some epidemiological studies makes it useful in this context. Some epidemiological studies have also provided sufficient data to relate back pain to the same occupational categories. Occupation was therefore considered as a proxy for specific risk factors. The exposure variable is an occupational category with its assigned level of risk (low, medium or high rate of low back pain). This method thus required the assumption that the distribution of the combined individual risk factors is similar within each occupational group across geographical regions. This argument applies to psychosocial as well as physical exposures.

Given that differences can occur within occupations, the assumed homogeneity of occupational groups in their total exposure to ergonomic risk factors implies that differences in exposure among occupations are substantially larger than differences among workers within the occupation. Although this assertion appears self-evident, only a few investigators to date have examined it explicitly. Burdorf (1992) evaluated the homogeneity of the exposure to postural load on the back within and among four occupational groups, and reported that the exposure variability within occupational groups was small compared with differences among groups. The estimated contribution of the variance for postural load between occupational groups was the largest source of variance. Hollman et al. (1999) and Paquet et al. (1999) have similarly shown that, even within one subsector (health care and construction, respectively), differences in ergonomic exposures among jobs can be large relative to the variability within jobs. These studies provide strong evidence in support of the approach taken in this analysis. Although the literature is less conclusive regarding their effects, to the extent that any specific
ergonomic factor is etiologically important it is assumed to be internalized in the relative risks by occupation. At the same time, specific analyses at the national or local level could increase the precision by assessing specific physical risk factors at the workplace.

### 5.2 Theoretical minimum risk

For low back pain, "theoretical minimum risk" is considered to represent the level of disease that would occur in the population if all excessive physical workload were abated by effective implementation of ergonomic control measures. While interventions to reduce ergonomic stressors have not yet been widely implemented on a global scale, studies are available from selected settings demonstrating the great potential of exposure (and disease) reduction in this area. Certain interventions have shown that removal of ergonomic stressors can practically lead to the removal of back pain (or its reduction to negligible levels), which justifies the choice of theoretical minimum. For example, in jobs where the entire work activity consists of manually handling materials, lifting equipment can successfully reduce both the biomechanical exposure to the lower back and the risk of low back disorders (Marras et al. 2000).

Westgaard and Winkel (1997) reviewed 89 studies on ergonomic intervention studies, of which 20 were classified as mechanical exposure interventions and 32 as production system interventions. Most mechanical exposure interventions target the ergonomic exposure level directly, through redesign of the work station and work methods. The reviewers concluded that in work situations whose mechanical exposure level is initially high, a reduction in the level of mechanical exposure may be beneficial for musculoskeletal health. For comparison, several other risk factors considered in this chapter have not yet shown that interventions can have such highly effective results when applied to selected population groups.

Since occupation represents a proxy for the composite of etiological exposures, rather than being the exposure per se, it is not necessary that persons leave one occupation for another in order to achieve the theoretical minimum risk. Reduction in relative risk would occur through improved job design to reduce exposures within each occupational category. The number of individuals working in each category could remain constant, even though the nature of the risk in each category would change.

### 5.3 Estimating risk factor levels

Since data were not available worldwide on the prevalence of specific ergonomic exposures, occupations were grouped here by their risk of low back pain. Thus, for this outcome, estimation of risk factor levels is not independent of the levels of disease assigned to them in the next step (see section 5.4). The exposure assessment by occupation was utilized, as described in the Introduction. Using managers and professionals as a
baseline for comparison, epidemiological studies have indicated that clerical and sales workers have a slightly elevated risk, operators and service workers have a moderate risk, and farmers have the highest risk of low back pain (see section 5.4 for details).

The basic approach was to determine economic activity by subregion, age and sex, and then to determine the distribution of the working population in the various occupational categories. As each occupational category was assigned a single relative risk factor (based on methodology described below), it was not necessary to partition exposures into "high" or "low" levels.

The estimates of occupations with risk of low back pain were based on the published literature. The 1968 International Standard Classification Codes for Occupations utilizes the term "production workers", whereas the epidemiological studies refer to "operators". Based on the literature, we made the following assignments.

- Background exposure: professional and administrative workers
- Low exposure: clerical and sales workers
- Moderate exposure: operators (production workers) and service workers
- High exposure: farmers

The non-working population is not considered in this analysis, and is attributed the same relative risk as the background exposure category. It is likely that younger workers are represented more often in the production occupations, and that older workers have more opportunities to move into management and administration positions.

As seen from Table 21.48, the majority of the working-age population is employed in occupations with exposure to factors linked to low back pain. Males have higher exposure in general, owing to higher rates of participation in the labour force. Exposures are higher for men in the less developed subregions, owing to a higher proportion of workers in agriculture than in the more developed subregions.

### 5.4 RISK FACTOR-DISEASE RELATIONSHIPS

Pain in the soft tissues of the back is extremely common among adults. In the United States, the National Arthritis Data Workgroup reviewed national survey data showing that each year some $15 \%$ of adults report frequent back pain or pain lasting more than two weeks (Lawrence et al. 1998). Back pain is widespread in many countries, and is associated with substantial financial costs and loss of quality of life. In Canada, Finland and the United States, more people are disabled from working as a result of MSDs-especially back pain-than from any other group of diseases (Badley et al. 1994; Pope et al. 1991; Riihimäki 1995a). MSDs also constitute a major proportion of all registered and/or

Table 21.48 Proportions of the working-age population occupationally exposed to different levels of ergonomic stressor, by sex and subregion

| Subregion | Sex | Exposure category | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | Background | 0.29 | 0.11 | 0.13 | 0.22 | 0.40 | 0.70 |
|  |  | Low | 0.09 | 0.11 | 0.10 | 0.09 | 0.07 | 0.04 |
|  |  | Moderate | 0.20 | 0.25 | 0.25 | 0.22 | 0.17 | 0.09 |
|  |  | High | 0.42 | 0.53 | 0.52 | 0.46 | 0.35 | 0.18 |
|  | Female | Background | 0.53 | 0.42 | 0.41 | 0.55 | 0.73 | 0.87 |
|  |  | Low | 0.05 | 0.06 | 0.06 | 0.05 | 0.03 | 0.01 |
|  |  | Moderate | 0.09 | 0.11 | 0.11 | 0.08 | 0.05 | 0.02 |
|  |  | High | 0.34 | 0.41 | 0.42 | 0.32 | 0.19 | 0.09 |
| AFR-E | Male | Background | 0.29 | 0.12 | 0.14 | 0.22 | 0.40 | 0.70 |
|  |  | Low | 0.08 | 0.10 | 0.10 | 0.09 | 0.07 | 0.03 |
|  |  | Moderate | 0.20 | 0.25 | 0.25 | 0.22 | 0.17 | 0.09 |
|  |  | High | 0.42 | 0.53 | 0.52 | 0.47 | 0.36 | 0.18 |
|  | Female | Background | 0.41 | 0.34 | 0.37 | 0.50 | 0.67 | 0.83 |
|  |  | Low | 0.06 | 0.07 | 0.07 | 0.05 | 0.04 | 0.02 |
|  |  | Moderate | 0.11 | 0.13 | 0.12 | 0.10 | 0.06 | 0.03 |
|  |  | High | 0.41 | 0.46 | 0.44 | 0.35 | 0.23 | 0.12 |
| AMR-A | Male | Background | 0.49 | 0.33 | 0.37 | 0.64 | 0.91 | 0.95 |
|  |  | Low | 0.18 | 0.24 | 0.23 | 0.13 | 0.03 | 0.02 |
|  |  | Moderate | 0.29 | 0.39 | 0.36 | 0.21 | 0.05 | 0.03 |
|  |  | High | 0.03 | 0.04 | 0.04 | 0.02 | 0.01 | 0.00 |
|  | Female | Background | 0.58 | 0.47 | 0.53 | 0.79 | 0.95 | 0.98 |
|  |  | Low | 0.19 | 0.24 | 0.21 | 0.09 | 0.02 | 0.01 |
|  |  | Moderate | 0.22 | 0.28 | 0.24 | 0.11 | 0.02 | 0.01 |
|  |  | High | 0.01 | 0.02 | 0.02 | 0.01 | 0.00 | 0.00 |
| AMR-B | Male | Background | 0.32 | 0.16 | 0.23 | 0.43 | 0.71 | 0.86 |
|  |  | Low | 0.15 | 0.19 | 0.17 | 0.13 | 0.06 | 0.03 |
|  |  | Moderate | 0.37 | 0.46 | 0.43 | 0.32 | 0.16 | 0.08 |
|  |  | High | 0.15 | 0.19 | 0.17 | 0.13 | 0.06 | 0.03 |
|  | Female | Background | 0.63 | 0.57 | 0.69 | 0.84 | 0.94 | 0.97 |
|  |  | Low | 0.10 | 0.12 | 0.09 | 0.04 | 0.02 | 0.01 |
|  |  | Moderate | 0.21 | 0.24 | 0.18 | 0.09 | 0.03 | 0.02 |
|  |  | High | 0.06 | 0.07 | 0.05 | 0.03 | 0.01 | 0.00 |
| AMR-D | Male | Background | 0.41 | 0.19 | 0.20 | 0.29 | 0.49 | 0.75 |
|  |  | Low | 0.15 | 0.21 | 0.20 | 0.18 | 0.13 | 0.06 |
|  |  | Moderate | 0.39 | 0.55 | 0.53 | 0.48 | 0.34 | 0.17 |
|  |  | High | 0.04 | 0.06 | 0.06 | 0.05 | 0.04 | 0.02 |
|  | Female | Background | 0.70 | 0.62 | 0.69 | 0.77 | 0.86 | 0.93 |
|  |  | Low | 0.09 | 0.11 | 0.09 | 0.07 | 0.04 | 0.02 |
|  |  | Moderate | 0.20 | 0.26 | 0.21 | 0.16 | 0.09 | 0.05 |
|  |  | High | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 |
| EMR-B | Male | Background | 0.48 | 0.23 | 0.27 | 0.41 | 0.64 | 0.82 |
|  |  | Low | 0.15 | 0.22 | 0.21 | 0.17 | 0.10 | 0.05 |
|  |  | Moderate | 0.27 | 0.40 | 0.38 | 0.31 | 0.19 | 0.09 |
|  |  | High | 0.10 | 0.15 | 0.14 | 0.11 | 0.07 | 0.03 |

Table 21.48 Proportions of the working-age population occupationally exposed to different levels of ergonomic stressor, by sex and subregion (continued)

| Subregion | Sex | Exposure category | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EMR-D | Female | Background | 0.77 | 0.74 | 0.82 | 0.87 | 0.94 | 0.97 |
|  |  | Low | 0.10 | 0.11 | 0.08 | 0.05 | 0.03 | 0.01 |
|  |  | Moderate | 0.10 | 0.12 | 0.08 | 0.06 | 0.03 | 0.01 |
|  |  | High | 0.03 | 0.04 | 0.02 | 0.02 | 0.01 | 0.00 |
|  | Male | Background | 0.32 | 0.10 | 0.13 | 0.29 | 0.59 | 0.80 |
|  |  | Low | 0.19 | 0.25 | 0.24 | 0.20 | 0.11 | 0.06 |
|  |  | Moderate | 0.17 | 0.23 | 0.22 | 0.18 | 0.10 | 0.05 |
|  |  | High | 0.32 | 0.43 | 0.41 | 0.33 | 0.19 | 0.10 |
|  | Female | Background | 0.64 | 0.59 | 0.64 | 0.76 | 0.88 | 0.94 |
|  |  | Low | 0.06 | 0.07 | 0.06 | 0.04 | 0.02 | 0.01 |
|  |  | Moderate | 0.05 | 0.06 | 0.05 | 0.04 | 0.02 | 0.01 |
|  |  | High | 0.24 | 0.28 | 0.24 | 0.16 | 0.08 | 0.04 |
| EUR-A | Male | Background | 0.58 | 0.38 | 0.46 | 0.78 | 0.97 | 0.98 |
|  |  | Low | 0.11 | 0.16 | 0.14 | 0.06 | 0.01 | 0.00 |
|  |  | Moderate | 0.27 | 0.40 | 0.35 | 0.14 | 0.02 | 0.01 |
|  |  | High | 0.04 | 0.06 | 0.05 | 0.02 | 0.00 | 0.00 |
|  | Female | Background | 0.67 | 0.59 | 0.69 | 0.92 | 0.99 | 0.99 |
|  |  | Low | 0.11 | 0.14 | 0.11 | 0.03 | 0.00 | 0.00 |
|  |  | Moderate | 0.19 | 0.23 | 0.18 | 0.04 | 0.01 | 0.00 |
|  |  | High | 0.03 | 0.04 | 0.03 | 0.01 | 0.00 | 0.00 |
| EUR-B | Male | Background | 0.36 | 0.15 | 0.29 | 0.64 | 0.80 | 0.90 |
|  |  | Low | 0.10 | 0.13 | 0.11 | 0.06 | 0.03 | 0.01 |
|  |  | Moderate | 0.33 | 0.45 | 0.37 | 0.19 | 0.10 | 0.05 |
|  |  | High | 0.21 | 0.28 | 0.23 | 0.12 | 0.06 | 0.03 |
|  | Female | Background | 0.51 | 0.32 | 0.48 | 0.80 | 0.89 | 0.95 |
|  |  | Low | 0.05 | 0.07 | 0.05 | 0.02 | 0.01 | 0.01 |
|  |  | Moderate | 0.20 | 0.27 | 0.21 | 0.08 | 0.04 | 0.02 |
|  |  | High | 0.25 | 0.34 | 0.26 | 0.10 | 0.05 | 0.03 |
| EUR-C | Male | Background | 0.35 | 0.13 | 0.20 | 0.73 | 0.90 | 0.95 |
|  |  | Low | 0.11 | 0.14 | 0.13 | 0.04 | 0.02 | 0.01 |
|  |  | Moderate | 0.39 | 0.52 | 0.48 | 0.16 | 0.06 | 0.03 |
|  |  | High | 0.15 | 0.20 | 0.19 | 0.06 | 0.02 | 0.01 |
|  | Female | Background | 0.47 | 0.18 | 0.35 | 0.85 | 0.96 | 0.98 |
|  |  | Low | 0.13 | 0.20 | 0.15 | 0.04 | 0.01 | 0.01 |
|  |  | Moderate | 0.31 | 0.48 | 0.38 | 0.09 | 0.03 | 0.01 |
|  |  | High | 0.10 | 0.15 | 0.12 | 0.03 | 0.01 | 0.00 |
| SEAR-B | Male | Background | 0.32 | 0.09 | 0.13 | 0.33 | 0.59 | 0.80 |
|  |  | Low | 0.11 | 0.14 | 0.13 | 0.10 | 0.06 | 0.03 |
|  |  | Moderate | 0.25 | 0.33 | 0.32 | 0.25 | 0.15 | 0.07 |
|  |  | High | 0.33 | 0.44 | 0.42 | 0.32 | 0.20 | 0.10 |
|  | Female | Background | 0.49 | 0.36 | 0.40 | 0.60 | 0.81 | 0.90 |
|  |  | Low | 0.11 | 0.13 | 0.13 | 0.08 | 0.04 | 0.02 |
|  |  | Moderate | 0.16 | 0.21 | 0.19 | 0.13 | 0.06 | 0.03 |
|  |  | High | 0.24 | 0.30 | 0.28 | 0.19 | 0.09 | 0.05 |

Table 21.48 Proportions of the working-age population occupationally exposed to different levels of ergonomic stressor, by sex and subregion (continued)

| Subregion | Sex | Exposure category | Age group (years) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| SEAR-D | Male | Background | 0.29 | 0.10 | 0.13 | 0.34 | 0.51 | 0.76 |
|  |  | Low | 0.06 | 0.08 | 0.08 | 0.06 | 0.04 | 0.02 |
|  |  | Moderate | 0.23 | 0.29 | 0.28 | 0.21 | 0.16 | 0.08 |
|  |  | High | 0.42 | 0.53 | 0.52 | 0.39 | 0.29 | 0.14 |
|  | Female | Background | 0.57 | 0.45 | 0.52 | 0.69 | 0.85 | 0.92 |
|  |  | Low | 0.01 | 0.02 | 0.01 | 0.01 | 0.00 | 0.00 |
|  |  | Moderate | 0.06 | 0.08 | 0.07 | 0.04 | 0.02 | 0.01 |
|  |  | High | 0.36 | 0.46 | 0.40 | 0.26 | 0.13 | 0.06 |
| WPR-A | Male | Background | 0.51 | 0.29 | 0.30 | 0.49 | 0.78 | 0.89 |
|  |  | Low | 0.16 | 0.24 | 0.23 | 0.17 | 0.07 | 0.04 |
|  |  | Moderate | 0.30 | 0.43 | 0.42 | 0.31 | 0.13 | 0.07 |
|  |  | High | 0.03 | 0.04 | 0.04 | 0.03 | 0.01 | 0.01 |
|  | Female | Background | 0.59 | 0.50 | 0.52 | 0.74 | 0.91 | 0.95 |
|  |  | Low | 0.16 | 0.20 | 0.19 | 0.10 | 0.04 | 0.02 |
|  |  | Moderate | 0.21 | 0.26 | 0.25 | 0.13 | 0.05 | 0.02 |
|  |  | High | 0.03 | 0.04 | 0.04 | 0.02 | 0.01 | 0.00 |
| WPR-B | Male | Background | 0.27 | 0.11 | 0.17 | 0.45 | 0.74 | 0.87 |
|  |  | Low | 0.13 | 0.16 | 0.15 | 0.10 | 0.05 | 0.02 |
|  |  | Moderate | 0.25 | 0.30 | 0.28 | 0.19 | 0.09 | 0.04 |
|  |  | High | 0.35 | 0.43 | 0.40 | 0.27 | 0.13 | 0.06 |
|  | Female | Background | 0.30 | 0.20 | 0.39 | 0.74 | 0.92 | 0.96 |
|  |  | Low | 0.17 | 0.20 | 0.15 | 0.07 | 0.02 | 0.01 |
|  |  | Moderate | 0.22 | 0.25 | 0.19 | 0.08 | 0.03 | 0.01 |
|  |  | High | 0.30 | 0.35 | 0.27 | 0.11 | 0.04 | 0.02 |

compensable work-related diseases in many countries, representing a third or more of all registered occupational diseases in the United States, the Nordic countries and Japan (Bernard 1997; Pope et al. 1991; Vaaranen et al. 1994).

Guo et al. (1995) estimated that $65 \%$ of cases of low back pain in the United States are attributable to occupational activities. To date, there have been no other published estimates of the fraction of back pain (specifically) in the total population that is occupationally induced. However, low back pain was identified by the Pan American Health Organization as one of the top three occupational health problems to be targeted by surveillance within the WHO Region of the Americas (Choi et al. 2001).

Among MSDs caused by occupational ergonomic stressors, only low back pain is currently a separate category in the GBD database and could be assessed. For purposes of this chapter, low back pain is defined as all pain in the back that is not secondary to another disease or injury cause (such as cancer or a motor vehicle accident). This includes disk prob-
lems (displacement, rupture), sciatica and other sources of back pain. Cervical spine problems, such as neck pain or neck torsion problems, are excluded. This category of conditions is considered equivalent to what others have termed non-traumatic MSDs affecting the lower back.

In the epidemiological literature, MSDs of the back are often defined on the basis of pain reported on interview, usually with standardized study criteria referring to time of onset, frequency and/or severity of pain. Physical examinations have sometimes been used to supplement questionnaires, particularly to help localize the symptoms reported on interview and to rule out other causes of those symptoms. However, an important proportion of epidemiologically relevant (exposure-related) back disorders are negative on physical examination (e.g. Punnett et al. 1991) as well as on X-ray (e.g. Riihimäki et al. 1990). Most cases of low back pain cannot be diagnosed by objective criteria and are typically designated idiopathic or non-specific (Frank et al. 1995; Riihimäki 1991, 1995b), even if there are findings on examination or severe symptoms and loss of function.

It is difficult to measure directly the validity of questionnaire responses, since no consensus exists regarding a single "gold standard" against which all other measurements could be compared. The sensitivity and reliability of physical examination manoeuvres for identifying low back pain range from good to poor; not all pain results from known mechanisms for which there is a corresponding objective test (Deyo et al. 1992; Viikari-Juntura and Riihimäki 1999). Deyo et al. (1992) suggested that as many as $85 \%$ of cases of low back pain cannot be diagnosed because of the poor performance of examination and imaging tests.

A recent review by NIOSH (Bernard 1997) (see also below) emphasized that health outcomes defined subjectively should be included in any consideration of work-related back disorders. In 24 of the 42 epidemiological studies on low back pain reviewed, the health outcome was defined by reported symptoms on questionnaires or interview, ranging from any back pain to specific symptoms such as those consistent with sciatica. In several studies, MSD cases defined by symptoms alone and those defined by findings on physical examination have shown very similar associations with the ergonomic characteristics of subjects' jobs. Symptom-based case definitions generally appear to be both unbiased and more sensitive than those that require documented abnormalities on physical examination (e.g. Bernard et al. 1993; Punnett 1998; Punnett et al. 1991; Silverstein et al. 1986, 1987).

Other case definitions sometimes used epidemiologically include low back impairment or disability, typically indicated by reduced ability to perform activities of daily living or occupational tasks, work absenteeism and the seeking of medical care for back pain. Such behavioural measures, however, are less desirable than low back pain per se, because they are more distal from the direct morbidity and more likely to be affected by interpersonal variability (e.g. tolerance of pain before seeking medical
attention) or by differences in job demands (e.g. pain of the same severity may cause more low back disability in persons whose jobs are more demanding).

At the same time, there is a strong correlation between the frequency of musculoskeletal symptoms by occupation and the frequency of workers' compensation claims and recorded work-related repetitive trauma disorders in those same occupations (Fine et al. 1986; Silverstein et al. 1997). Outcomes such as days of restricted activity, long-term disability, health care utilization and use of medication are very common among people with back pain, indicating the public health importance and cost of these disorders even when self-reported pain is not confirmed objectively (Badley et al. 1994, 1995; Guo et al. 1999; Miedema et al. 1998; Punnett 1999; Westgaard and Jansen 1992).

Back pain has been defined operationally in various ways in epidemiological studies, including both prevalent and incident conditions. Variations in the definition are related to the recall period (e.g. pain now, or within the last week or the past year), the frequency or duration (e.g. at least three times in the past year, or lasting at least one week) and the severity (e.g. at least a "4" on a 7 -point pain intensity scale), among others. Even among studies that use similar definitions, prevalence estimates can vary substantially (Loney and Stratford 1999). However, for the purpose of evaluating the exposure-response relationship, as long as comparisons are made within a study population that has been evaluated with a consistent case definition, estimates of relative risk do not appear to be greatly affected (Ozguler et al. 2000).

## Evidence of causality

The evidence on low MSDs, including back pain, in relation to workplace factors has been thoroughly reviewed by NIOSH (Bernard 1997). The National Research Council, with the Institute of Medicine, has also published a comprehensive review of the evidence on MSDs in the workplace (National Research Council 2001). Strong or sufficient evidence was found for a number of risk factors at the workplace to be associated with back pain (Table 21.49). The National Research Council report (2001) summarized ranges of risk estimates for specific occupational stressors (Table 21.50).

In addition to these two comprehensive reviews from the United States, numerous other authors from Europe, Asia and Canada reviewed the same epidemiological literature or variously defined subsets, and most reached similar conclusions (e.g. Burdorf and Sorock 1997; Frank et al. 1996; Garg 1992; Hagberg et al. 1995; Hales and Bernard 1996; Hoogendoorn et al. 1999; Hulshof et al. 1987; Jensen 1988; Jin et al. 2000; Johanning et al. 1991; Lagerström et al. 1998; Nachemson and Jonsson 2000; Riihimäki 1991, 1995a; Viikari-Juntura 1997; Wikstrom et al. 1994). For example, in a systematic literature review that focused on 28 cohort and three case-control studies of highest methodological

# Table 21.49 Rating of evidence for causal relationships between specific occupational stressors and back disorders according to the NIOSH review 

| Strength of evidence | Specific stressor |
| :---: | :---: |
| Strong evidence ${ }^{\text {a }}$ | Lifting and forceful movements Whole-body vibration |
| Evidence ${ }^{\text {b }}$ | Awkward postures Heavy physical work |
| Insufficient evidence ${ }^{\text {c }}$ | Static work postures |
| Evidence of no effect ${ }^{\text {d }}$ | Other stressors |


|  | Strong evidence. A causal relationship is shown to be very likely between intense or long-duration exposure to the specific risk factor(s) and MSD of the back when the epidemiological criteria of causality are used. A positive relationship has been observed between exposure to the specific risk factor and MSD of the back where chance, bias and confounding factors could be ruled out with reasonable confidence in at least several studies. |
| :---: | :---: |
|  | Evidence. Some convincing epidemiological evidence shows a causal relationship when the epidemiological criteria for causality for intense or long-duration exposure to the specific risk factor(s) and MSD of the back are used. A positive relationship has been observed between exposure to the specific risk factor and MSD of the back in studies in which chance, bias and confounding factors are not the likely explanation. |
|  | Insufficient evidence. The available studies are of insufficient number, quality, consistency or statistical power to permit a conclusion regarding the presence or absence of a causal association. |
|  | Evidence of no effect. Adequate studies consistently show that the specific workplace risk factor is not related to development of MSD of the back. |
| Note: | In this review, 42 epidemiological studies were selected on the basis of the following criteria: (i) exposed and reference populations were well defined; (ii) MSDs of the back were measured by well defined, explicit criteria determined before the study; (iii) exposure was evaluated so that some inference could be drawn regarding repetition, force, extreme joint position, static loading or vibration and lifting tasks; (iv) study participation of more than $70 \%$. Thirty studies used a cross-sectional design and five a prospective cohort, four were case-control studies and two were retrospective cohorts. Full descriptions of these studies appear in Table 6-6 of the NIOSH review. Criteria for causality were based on strength of association, consistency, specificity of effect or association, temporality, exposure-response relationship and coherence of evidence. |
|  | Bernard |

quality, Hoogendoorn et al. (1999) found strong evidence for manual material handling, bending and twisting and whole-body vibration as risk factors for back pain. They found moderate evidence for patient handling and heavy physical work, and no evidence for standing or walking, sitting, sporting activities and total leisure-time physical activity. Specific psychological stressors, supported by weaker evidence, were reviewed by Burdorf and Sorock (1997), and are shown in Table 21.51.

Some of the results of these reviews on specific stressors were obtained from studies that were conducted within various occupational groups, such as operators, mine workers, dentists, office workers and nurses. The evidence presented above implies that preventive interventions reducing the exposure to these risk factors would decrease the occurrence of back disorders considerably, even within an occupation.

Table 21.50 Summary of epidemiological studies with risk estimates of null and positive associations of work-related risk factors and the occurrence of back disorders

| Work-related risk factor | Risk estimate expressed in relative risk or odds ratio |  |  |  | Attributable <br> fraction (\%) |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Null association ${ }^{\text {a }}$ |  | Positive association |  |  |  |
|  | $N^{\text {b }}$ | Range of odds ratio | $N^{\text {b }}$ | Range of odds ratio | $N^{\text {b }}$ | Range |
| Manual material handling | 4 | 0.90-1.45 | 24 | 1.12-3.54 | 17 | 11-66 |
| Frequent bending and twisting | 2 | 1.08-1.30 | 15 | 1.29-8.09 | 8 | 19-57 |
| Heavy physical load | 0 | NA | 8 | 1.54-3.71 | 5 | 31-58 |
| Static work posture | 3 | 0.80-0.97 | 3 | $1.30-3.29$ | 3 | 14-32 |
| Repetitive movements | 2 | 0.98-1.20 | 1 | 1.97 | 1 | 41 |
| Whole-body vibration | I | 1.10 | 16 | 1.26-9.00 | 11 | 18-80 |


| NA | Not applicable. |
| :--- | :--- |
| a | Confidence intervals of the risk estimates included the null estimate (I.0). In only 12 of 16 null <br> associations was the magnitude of risk estimate presented. |
| b | Number of associations presented in the epidemiological studies. |
| Source: | National Research Council (2001). |

Table 21.5I Summary of epidemiological studies on associations between specific occupational psychological risk factors and the occurrence of back disorders


Furthermore, there is sufficient evidence on biomechanical risk factors to conclude that many cases of low back pain could be prevented by workplace changes. For example, Marras et al. (2000) showed that lifting equipment and other engineering controls had demonstrable effects on lowering both biomechanical exposure to the lower back (compressive force, torque, etc.) and reported rates of low back disorders in 36 repetitive manual material handling jobs at 16 different companies.

The reviews by the National Research Council (2001), Westgaard and Winkel (1997) and Frank et al. (1996) each cited a number of well designed studies that identified opportunities to prevent risk of low back pain by reducing exposure to biophysical and psychosocial factors. To illustrate the improvements that can be obtained by reducing ergonomic stressors at work, selected interventions and their impact are shown in Table 21.52. Certain interventions have practically completely removed ergonomic stressors from the workplace.

Despite this extensive literature, some still dispute the importance of these factors, especially in relation to nonoccupational causes (e.g. Battie and Bigos 1991; Nachemson 1999; Waddell 1991). There are probably several interrelated reasons for the continuing controversy, many of which have been discussed by others (Frank et al. 1995, 1996; ViikariJuntura and Riihimäki 1999). The available epidemiological evidence still consists largely of cross-sectional and retrospective case-control investigations. With regard to assessment of morbidity, the use of selfreported symptoms for end-points has also generated discussion, as described above. Sparse longitudinal data leave important gaps in knowledge concerning latency of effect, natural history, prognosis and potential for selection bias in employed populations (e.g. the "healthy worker effect"). Few studies have attempted systematically to describe these features of back disorders in populations with defined levels of exposure to ergonomic stressors. Specific, quantitative comparisons of conclusions based on prevalence and incidence data within the same population are rare, and knowledge is still sparse as to the factors that predict recovery or persistence among workers who continue in their jobs after the onset of a disorder. No study has been identified that compares current and former workers with reference to both prior MSD morbidity and exposure status.

In addition, there are many known or suspected nonoccupational risk factors; some study populations have not provided enough statistical power to address potential confounding and effect modification of expo-sure-response relationships. Exposure assessment has often been limited, with too few exposures characterized to rule out confounding, or the use of crude exposure indicators leading to potential misclassification and unreliable conclusions. Attempts to partition risk between physical and psychosocial domains have obscured the overlap between the two and the distinction between preventable and nonpreventable risk factors (Bongers et al. 1993; MacDonald et al. 2001; Volinn and Punnett 2001; Volinn et al. 2001).

In the light of these issues, it is important to restate that the issue is not whether all back disorders are caused by work; there is a clear consensus that this is not the case. Nevertheless, most investigators and reviewers have concluded, equally clearly, that a large proportion of back disorders among persons with high exposure to ergonomic stressors at work could be prevented by reducing those exposures.
Table 21.52 Selected intervention studies on occupational ergonomic stressors and reduction of occurrence of low back pain

| Job title/activity | Intervention: engineering/administrative control | Outcome | Study ${ }^{\text {a }}$ |
| :---: | :---: | :---: | :---: |
| Nursing assistants (in two nursing home units) | Introduction of walking belts, mechanical hoists and shower chairs | Back injury incidence reduced from 83 to 42 per 100 full-time equivalents | Garg and Owen (1994) [Ex. 26-14I5] |
| Poultry processing employees | Introduction of workstation analysis and redesign, including altering heights of products, providing workstands, and installing tank tilters to reduce manual handling | Back injury rates declined from 4.4 to 3.0 per 200000 hours | Stuart-Buttle (1994) [Ex. 26-1045] |
| Office furniture manufacturing assembly workers | Scissor lifts installed to aid in packaging file cabinets of different sizes; small-assembly workstations altered to eliminate twisting and bending during lifting | Back injuries cut by 50\% | LaBar (1991) <br> [Ex. 26-1078] |
| Metal castings, unpacking operation | Crates modified by adding gates at each end and installing scissors lift to lift crates; changes made in the way castings were stacked in the crates to permit workers' arms to remain close to the body while unpacking | Back injuries associated with this operation eliminated | Oxenburgh (1994) [Ex. 26-104I] |
| Palletizing operation | Scissors lift table with turntable tops installed alongside each packing station | Five out of six back injuries eliminated | $\begin{aligned} & \text { Benson (1987) } \\ & \text { [Ex. 26-1062] } \end{aligned}$ |
| Lamp manufacturing | Addition of vacuum hoist; reduced equipment height; reduced box size and weight; introduction of back awareness programme for employees | Back and upper extremity disorders eliminated in the last four years | Carreau and Bessett (1991) [Ex. 26-107I] |
| Unpacking car parts | Plywood sheets modified to reduce weight and permit them to slide more easily in the grooves | Back injuries associated with this operation eliminated | Oxenburgh (1991) [Ex. 26-I04I] |

Manual handling of bulk paper

## Railway repairmen

## Nursing, hospital

## Video display terminal operators

Forestry workers
a All exhibit numbers refer to materials in USDOL OSHA (2000).
Source:

## OTHER OUTCOMES

Although the present analysis was limited to low back pain, the evidence on MSDs caused by occupational ergonomic stressors is broader. MSDs affecting the neck and the upper and lower limbs result from the same risk factors as are implicated in low back pain. For example, in a study of over 10000 manufacturing employees, the effect of "greater physical demands" of the job on low-back musculoskeletal injuries (relative risk of $1.6,95 \%$ CI 1.2-2.1) was only slightly higher than that for all other musculoskeletal injuries combined (relative risk of 1.4, $95 \%$ CI 1.1-1.7) (Tsai et al. 1992). De Zwart et al. (1997), studying over 7300 men in the Netherlands, found higher prevalences of shoulder disorders among employees in heavy physical work (e.g. heavy lifting and frequent stooping) and steeper increases over four years than among employees in less physically demanding jobs. The magnitude of these effects was very similar to those for low back injuries. The work-relatedness of upper and lower extremity MSDs has been discussed extensively, again by European as well as North American reviewers (e.g. Armstrong et al. 1993; Bernard 1997; Bongers et al. 1993; Buckle and Devereaux 1999; Hagberg et al. 1995; National Research Council 2001; Sluiter et al. 2000).

Also excluded here are other types of health effects related to ergonomic stressors, such as acute workplace injuries, cardiovascular disease, mental health and adverse reproductive effects (Punnett 2002). Thus, the total impact of excessively strenuous work activities on morbidity and related quality of life is greater than that estimated in this risk assessment.

## Hazard estimates

## Data sources

For the purposes of this analysis, studies were sought that compared the risk of low back pain among broad occupational groups (defined similarly to the economic subsectors explained above) and comprehensively enough to cover the range of paid occupations. Smaller, more specific studies limited to relatively narrow occupational groups (e.g. nurses, dock workers, drivers) were checked for consistency with the larger data sets. Studies where the reference groups were also engaged in substantial physical activity (e.g. house painters) were excluded. The most recent literature (1997-2001) was searched for exposure data and exposurerisk relationships, and published statistics of national occupational health and safety institutes were consulted.

In addition to this systematic search, a number of reviews and studies were identified to provide evidence supporting the selected approach. Search strategies were described in the Introduction. Medline was searched for articles more recent than 1985, using any of the keywords back pain, back disorder, back or musculoskeletal combined with any of
the following: occupation, occupational, workplace, work, workers, risk factors, risk.

## Description of literature

The studies specifically referred to in this section are summarized in Table 21.53.

The report of the National Research Council (2001) stated that the occupations with the highest risk among men were construction labourers, carpenters and industrial truck and tractor equipment operators, while among women they were nursing aides/orderlies/attendants, licensed practical nurses, maids and janitor/cleaners. Other high-risk occupations were hairdressers and automobile mechanics, often employed in small businesses or self-employed. No rates were listed against occupations in the report. The report stated that the highest-risk industries for men were lumber and building material retailing, crude petroleum and natural gas extraction and sawmills/millwork. Among women, the highest-risk industries were nursing and personal care facilities, beauty shops and motor vehicle equipment and manufacturing. No rates were listed for industries in the report.

Leigh and Sheetz (1989) measured low back pain on the basis of a national survey and a self-reported statement regarding "trouble with back or pain during the last year". They estimated relative risks by comparing the outcome frequency among occupational groups, using managers as a reference group (Table 21.54). This chapter places great weight on this study, because it was relatively large ( 1404 participants) compared to many others, it covered a comprehensive sample of occupations, and the results were adjusted for potential confounding variables. One important limitation, however, is that the multivariate analyses simultaneously included two ratings of physical exposure, socioeconomic status and occupational title. Since physical exposure is hypothesized to be the primary pathway through which occupational differences are manifested, these analyses would certainly lead to an underestimation of the effect of occupation on MSDs.

Although operators and service workers have very similar relative risks, it is common that intervention strategies differ among these occupational settings. For that reason, the relative risks and exposure assessments for those two occupational groups are presented separately throughout this analysis.

Within the limits of the available literature, the results of the Leigh and Sheetz analyses appear to be generally consistent with other reported relative risk values (Table 21.55). Since many other studies used office workers or other sedentary occupations as the reference group, it is appropriate to adjust the Leigh and Sheetz findings for comparative purposes. This can be done by dividing the relative risks for categories 3, 4, and 5 by 1.38 (the relative risk for clerical or sales work), in order to estimate a relative risk with clerical jobs as the reference group. The new
Table 21.53 Key studies and reviews on work-related back pain

| Study population (source) or literature reviewed | Population size | Outcome measured/reported | Magnitude or relative risk | Comments | Reference |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Reviews |  |  |  |  |  |
| Back pain in relation to heavy physical work, lifting and forceful movements, non-neutral postures, and/or whole-body vibration: Belgium, Finland, Italy, Japan, Netherlands, Russian Federation, Sweden, USA | 42 studies, evaluated according to 4 criteria for methodological quality | Back pain (multiple definitions among the studies) | Heavy physical work: ${ }^{\text {a }}$ I.0-12.1 <br> Lifting and forceful movements: ${ }^{\text {a }} 0.9-10.7$ <br> Non-neutral postures: 1.2-10.7 <br> Static work postures: 0.8-24.6 <br> Whole-body vibration: 1.0-39.5 <br> Conclusions: "strong evidence" for causal relationship with lifting and whole-body vibration; "evidence" for causal relationship with awkward posture and heavy physical work | Literature review: etiological studies | Bernard (1997) |
| Important risk factors for work-related back disorders, and strength of the association between the consistent risk factors and back disorder were identified | 35 epidemiological studies published from 1980 to 1996 | Back disorders (multiple definitions among the studies) | Manual materials handling: I.12-3.07 <br> Frequent bending and twisting: 1.29-8.09 <br> Heavy physical load: I.54-3.71 <br> Static work movement: 1.30-3.29 <br> Repetitive movement: 1.97 <br> Whole-body vibration: 1.47-9.00 <br> Mental stress: I.30-2.08 <br> Job satisfaction: 1.39-2.40 <br> Pace of work: 1.21 <br> Job decision latitude or monotonous work: I.25-2.34 <br> Conclusion: lifting or carrying loads, whole-body vibration and frequent bending and twisting consistently associated with work-related back disorder. Job dissatisfaction and low job decision latitude also important | Illustration on individual risk factors, such as age, smoking and education | Burdorf and <br> Sorock (1997) |


| Intensive and systematic review on physical load during work or leisure time as risk factors for back pain. Literature sources: Medline (1966-1997), Embase (1988-1997), Psyclit (1974-1997), NIOSHTIC, CISDOC and HSELINE (1977-1997) and Sportdiscus (1949-1997) | 31 publications in Dutch, English, French and German | Inclusion criteria (health outcome): back pain based on symptoms or signs of non-specific back pain, self-reported or measured (via sick leave, medical consultation, treatment and disability due to back pain, etc.) | Strong evidence for manual material handling, bending and twisting, and whole-body vibration as risk factors. Moderate evidence for patient handling and heavy physical work. No evidence for standing or walking, sitting, sports and total leisure-time physical activity | Cross-sectional studies were excluded | Hoogendoorn et al. (1999) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Review of studies of work-related low back pain in China among literature issued from 1983 to 1997 | 16 epidemiological studies selected; quality inclusion/ exclusion criteria were applied; all studies included were cross-sectional | Occurrence of work-related low back pain (multiple definitions among the studies) | Prevalence ratios: <br> Bending and twisting: 2.0-8.5 <br> Static posture: 1.5-14.3 <br> Whole-body vibration: 1.9-5.5 | Suggestion of potentially greater exposure in China's work environment and potential systematic bias | Jin et al. (2000) |
|  |  | Back pain (multiple definitions among the studies), muscle activity, tissue load and tissue damage | "linkages ... between external loads and biomechanical loading of the spine, biomechanical loading and internal tolerances of the spine, and internal tolerances and outcomes (from pain through disability) are well established...The literature relating to causal factors in work-related low back disorders is coherent and provides ample evidence on how adverse work situations can lead to them." (pp. 357-358) | Literature review: etiological, experimental and intervention studies | National Research Council (200I) |
| Systematic literature review of articles published in English, 1980-1995, 25 most populous countries | NA | Back pain (multiple definitions among the studies); point prevalence of low back pain/ annual or lifetime incidence | 2-4 times higher rates of back pain in high-income than in low-income countries. Within low-income countries, rates are higher among urban than among rural populations | Comparison of low-income and high-income countries | Volinn (1997) |

Table 21.53 Key studies and reviews on work-related back pain (continued)

| Study population (source) or literature reviewed | Population size | Outcome measured/reported | Magnitude or relative risk | Comments | Reference |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Original studies |  |  |  |  |  |
| Current full-time employees at one site of large chemical manufacturing company, USA, 1987-1989 | 5903 employees | Back or joint pain; back pain $>30$ days; visit to physician for back pain | $35.4 \%$ reported back or joint pain during the past year, $5.3 \%$ back pain lasting >30 days; managers and technicians had highest prevalences of back pain $>30$ days | Adjusted for age, sex and ethnicity | Burchfiel et al. (1992) |
| Random sample of retired workers living in the Paris area, members of a supplementary interprofessional retirement pension fund | 993 people in total ( 626 in first survey in 1982-1983, 464 in second survey in 1987-1988) | Osteoarticular disorders: presence of pain with/ without restricted joint movement for at least 6 months before interview | Increased frequency of osteoarticular pain during the 5 years between the two interviews; from $52 \%$ to $65 \%$ in men and $72 \%$ to $82 \%$ in women A significance increase in frequency of osteoarticular pain for those exposed to heavy weights and awkward postures (from 68\% to 77\% and $56 \%$ to $76 \%$, respectively) | Significantly higher frequencies of osteoarticular pain for women than men in both interviews | Derriennic et al. (1993) |
| Probability sample of total working population in USA, 1988 | 30074 workers | Back pain every day for a week or more during 12 months before interview | National estimate of back pain: 22.4 million cases and a prevalence of $17.6 \%$ in 1988 <br> $65 \%$ of cases were attributable to occupational activities <br> The risk was highest for construction labourers among males (prevalence 22.6\%) and nursing aides among females (18.8\%) |  | $\begin{aligned} & \text { Guo et al. (1995, } \\ & \text { 1999) } \end{aligned}$ |
| Random sample: all small and medium-sized factories listed with Labour Department in Delhi, India | 60 of 6076 factories; 631 workers selected randomly from 60 factories | Lumbar pain diagnosed by medical practitioner | Buffing workers, operators and assembly workers had highest pain prevalences (30.4\%, 28.7\% and $27.2 \%$, respectively) |  | Joshi et al. (2001) |


| Workers' compensation claims for back injuries in 26 states (among I 705674 workers' compensation claims), all workers employed in 26 states, USA, 1979 | 329474 claims for back injuries | Compensation claims filed for back injury or strains/sprains | Back injury constitutes 19.3\% of compensation claims in 26 states <br> Construction and mining industries have largest incidence: 1.6 and I.5 claims/ 100 workers, respectively Occupations with the largest incidence compared to total workers: miscellaneous labourers 12.3 claims/ 100 workers, garbage collectors II.I claims/ 100 workers | Percentage of workers employed in 26 states in 1979, obtained from 1970 census | Klein et al. (1984) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Diagnosed low back pain cases, among all employed persons in Japan, 1986 and 1988 | 13166 low back pain cases | Accidental low back pain caused by the action of a sudden force; required absence of $\geq 4$ days | Transport occupations: 8 cases per 10000 workers <br> Construction: 4 cases per 10000 workers <br> Sales/service: 0.5 cases per 10000 workers <br> Mining and cargo handling: 13.9 and 12.3 cases per 10000 workers | National survey data | Kuwashima et al. (1997) |
| Probability sample of United States workers employed $\geq 20$ hours a week, 1972-1973 | 959 working males and 455 working females | Low back pain: experience of trouble with back or spine in past year | I-year low back pain past prevalence: $19.4 \%$ males, $20.7 \%$ females Relative risk: farmers 5.17, operators 2.39, service 2.67 , clerical 1.38 compared to managers and professionals <br> Relative risk for high job demands: I. 68 | Health outcome did not distinguish between upper and lower back pain Over-adjustment by including socioeconomic status, occupation and physical risk factors in same multivariate analyses | Leigh and Sheetz (1989) |
| Three random samples of the Finnish population, 20-64 years of age, 1988-1990 | 7544 people | Any back pain in past 30 days/back pain in past 12 months | Odds ratios of back pain: 2.1 in farmers, 1.8 in manual workers, I. 7 in entrepreneurs, I. 4 in lower white-collar workers | Results adjusted for age, height, marital status, smoking, body mass index, physical activity, mental distress | Leino-Arjas et al. (1998) |
| Workers at 8 aluminium plants in Norway, 1998 | 5654 people | One-year prevalence of musculoskeletal symptoms by Standardized Nordic Questionnaire | Odds ratio for low back symptoms among operator vs office workers: I. 8 (range I.5-2.I) | Adjusted for sex, weight, height, smoking, and nonoccupational physical activity | Morken et al. (2000) |

Table 21.53 Key studies and reviews on work-related back pain (continued)

| Study population (source) or literature reviewed | Population size | Outcome measured/reported | Magnitude or relative risk | Comments | Reference |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Active workers from 4 occupational sectors (office, hospital, warehouse, airport registration), France, I99\| | 725 people | Six definitions of low back pain (pain or discomfort in a lumbar area in the previous 6 months/pain at least one day/pain > 30 days/intensity of pain above $3 /$ physician visit/ treatment for low back pain) | Prevalence of low back pain varied from $8 \%$ to $45 \%$ according to the case definition <br> Carrying heavy loads and bending postures showed consistently high odds ratios (1.88-2.14) for most low back pain definitions | Over-adjustment by including both occupation and physical risk factors in multivariate analyses; also adjusted for sex, age and body mass index | Ozguler et al. (2000) |
| All back disorders reported to clinic in one automobile assembly plant (referents from same production departments) | 95 assembly workers with back disorders (cases) and 124 without (referents) | Back cases: workers who sought medical attention for "new" back disorders during <br> a 10 -month period | Exposed workers: 84\% <br> Bending and twisting ( $100 \%$ vs $0 \%$ ): odds ratio 8.09 (range 1.5-44.0) <br> Lifting (>44.5 Newtons/min): odds ratio 2.2 (range I.0-4.7) | Adjusted for age, sex and sports activity | Punnett et al. (1991) |
| Employed persons in Washington State, USA, 1990-1998 | Approximately 1.23 million full-time equivalent workers per year | Workers' compensation claims for nontraumatic soft tissue disorders of the back | Rates by industry sector ranged from 43.5 to 280.0 per 10000 full-time equivalents | Surveillance data | Silverstein et al. (2002) |
| Population-based survey of approx 9.9 million adults ( 15 years or older), Belgium, 1991 | 3829 people | Reported symptoms: low back pain, history of low back pain, first low back pain and daily low back pain | Current low back pain: 33\% of population <br> Work dissatisfaction associated with low back pain history (odds ratio >2.4) | Adjusted for age, sex, language, residence, social class and job satisfaction | Skovron et al. (1994) |
| Full-time regular workers in Shell Oil Company's manufacturing facilities between 1987 and 1989, USA | 10350 people | Low-back injury (ICD-9 CM, $722,724)$ | Physically demanding jobs have relative risk of 1.57 for low back injury and 1.35 for non-low back musculoskeletal injury <br> Smoking and overweight showed high relative risks (1.54 and I.42, respectively) | Job title used to identify potential for increased physical demand at work | Tsai et al. (1992) |
| NA Not applicable. <br> a Range of effects for studies | met at least one criterion | epidemiological quality. |  |  |  |

Table 21.54 Relative risks of low back pain for occupational groups, with managers and professionals as the reference group

| Occupational activity | Relative risk (95\% Cl) |
| :--- | :---: |
| Managers and professionals | $1.0(\mathrm{NA})$ |
| Clerical or sales worker | $1.38(0.85-2.25)$ |
| Operators | $2.39(1.09-5.25)$ |
| Service workers | $2.67(1.26-5.69)$ |
| Farmers | $5.17(1.57-17.0)$ |

NA Not applicable.
Source: Based on data from Leigh and Sheetz (1989).
values would be $1.73,1.93$ and 3.75 , respectively. Keeping in mind that these estimates represent the average values for the entire occupational category, it can be seen that the other studies cited fall within the CIs, with very few exceptions, and in fact generally have similar point estimates (Table 21.55).

The only study that can directly and numerically be compared to that of Leigh and Sheetz (1989) is that by Leino-Arjas et al. (1998). However, the only value corresponding directly to one of the categories of Leigh and Sheetz is that for farmers. The relative risk is lower (2.13) than the one put forward by Leigh and Sheetz (5.17), which may reflect better working conditions for farmers in Finland. For this analysis we therefore used the average of these two results (see below).

Also available are administrative statistics from several countries on the number of cases of back conditions. These are generally compiled from employers' surveys or compensation statistics and typically report lower rates than those assessed by population surveys. Rates for certain occupations, as compared to managers and professionals, can be estimated on the basis of these statistics. Table 21.56 summarizes administrative workplace statistics on conditions involving the back, based on reports by employers in the United States of work-related injuries (Bureau of Labor Statistics 2001) and compensation statistics of the Australian workforce (National Occupational Health and Safety Commission 2001) and of the German national workforce (Bundesverband der Betriebskrankenkassen 2001).

All three of these data sets show higher risks for occupations other than managers and professionals, although the point estimates vary somewhat. None of these frequency estimates is adjusted for potential confounding variables. The incidents assessed in the first two data sets are limited to cases that have been recognized as work-related cases and involve behaviour such as absence from work or filing a claim against the employer. In contrast, the German study sought to assess the health status of the population more comprehensively and these data are

Table 21.55 Relative risks of occupational groups by exposure level

| Occupation (exposure category) | Source |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Leigh and Sheetz (1989) ${ }^{\text {a }}$ | $\begin{aligned} & \text { Astrand } \\ & (1987)^{\text {b }} \end{aligned}$ | Bongers et al. $(1990)^{\text {b }}$ | Bovenzi and Betta (1994) ${ }^{\text {b }}$ | $\begin{aligned} & \text { Burdorf } \\ & \text { et al. } \\ & (1993)^{\mathrm{b}} \end{aligned}$ | Hildebrandt $(1995)^{\mathrm{b}}$ | Johanning et al. $(1991)^{b}$ | Magnusson <br> et al. <br> $(1996)^{b}$ |
| Managers and professionals <br> Professionals Managers Teachers | 1.00/NA |  |  |  |  |  |  |  |
| Clerical or sales workers | 1.38/1.00 |  |  |  |  |  |  |  |
| Office workers (sedentary) Clerks |  | 1.00 |  | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Air force officers Civil servants Sales |  |  | 1.00 |  |  |  |  |  |
| Operators Construction labourers | 2.39/1.73 |  |  |  |  |  | 3.90 |  |
| Manual workers Pilots and aircrew |  | 2.28 | 9.00 |  |  |  |  |  |
| Drivers (bus, truck, tractor) |  |  |  | 1.83-5.49 | 2.51 | 1.32 |  | 1.55-2.10 |
| Crane operators |  |  |  |  | 3.29 |  |  |  |
| Dock workers |  |  |  |  |  |  |  |  |
| Plumbers <br> Carpenters <br> Technicians |  |  |  |  |  | 1.32 |  |  |
| Assembly, packing, food processing |  |  |  |  |  |  |  |  |
| Automobile mechanics Maintenance |  |  |  |  |  |  |  |  |
| Service workers Airport registration workers | 2.67/I. 93 |  |  |  |  |  |  |  |
| Hospital workers |  |  |  |  |  |  |  |  |
| Warehouse workers |  |  |  |  |  |  |  |  |
| Stock handlers, baggers |  |  |  |  |  |  |  |  |
| Janitors, cleaners |  |  |  |  |  |  |  |  |
| Waitresses <br> Nurses |  |  |  |  |  |  |  |  |
| Farmers | 5.17/3.75 |  |  |  |  |  |  |  |

NA Not applicable.
a Relative risks used in estimation of global burden of disease. The second set of relative risk values was estimated using clerical/sales jobs as the reference group, for the purpose of comparison with other studies.

| Source |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Partridge and Duthie $(1968)^{\mathrm{b}}$ | Riihimäki et al. $(1989)^{\mathrm{b}}$ | Riihimäki et al. (1994) ${ }^{\text {b }}$ | Videman et al. $(1990)^{b}$ | Burchfiel et al. (1992) | $\begin{aligned} & \text { Ozguler } \\ & \text { et al. } \\ & (2000) \end{aligned}$ | $\begin{aligned} & \text { Joshi } \\ & \text { et al. } \\ & (200 \mathrm{I}) \end{aligned}$ | Guo et al. (1995) ${ }^{\text {c }}$ [female] | Morken et al. (2000) | Leino-Arjas et al. (1998) (male) |
|  |  |  |  |  |  |  |  |  | 1.00 |
|  |  |  |  | 1.00 |  |  |  |  |  |
|  |  |  |  | 1.80 |  |  |  |  |  |
|  |  |  |  |  |  |  | [1.2] |  |  |
|  | 1.00 | 1.00 | 1.00 | 0.89 | 1.00 | 1.00 |  | 1.00 | 1.35 |
| 1.00 1.10 |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | $1.0-1.5$ | 1.40 |  | 1.10 |  | 1.83 |  | 1.80 |  |
|  |  |  |  |  |  |  | 2.10 |  |  |
|  |  |  | 3.60 | 1.49 |  |  |  |  | 1.84 |
|  |  |  | 2.90 |  |  |  | 2.00 |  |  |
| 1.27 |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | 1.70 |  |  |
|  |  | 1.50 |  |  |  |  | 2.10 |  |  |
|  |  |  |  | 1.20 |  | 1.59 |  |  |  |
|  |  |  |  |  |  | 1.73 |  |  |  |
|  |  |  |  |  |  |  | 1.80 |  |  |
|  |  |  |  |  |  | 1.59 | 1.70 |  |  |
|  |  |  |  | 1.03 |  |  |  |  |  |
|  |  |  |  |  | 0.86 |  |  |  |  |
|  |  |  |  |  | 1.13 |  |  |  |  |
|  |  |  |  |  | 0.54 |  |  |  |  |
|  |  |  |  |  |  |  | 1.70 |  |  |
|  |  |  |  |  |  |  | [2.0] |  |  |
|  |  |  |  |  |  |  | [1.6] |  |  |
|  |  |  |  |  |  |  | [1.5] |  |  |
|  |  |  |  |  |  |  | 1.80 |  | 2.13 |
| b Cited in Bernard (1997). |  |  |  |  |  |  |  |  |  |
| c Comp | d to all | le or fem | worker |  |  |  |  |  |  |

Table 21.56 Relative risks of occupational conditions involving the back by occupational title, compared to managers and professionals, from national surveillance data

|  | Relative risk for back conditions |  |  |
| :--- | :---: | :---: | :---: |
| Occupational activity $^{\mathrm{a}}$ | USA $^{\mathrm{b}}$ | Australia $^{\mathrm{c}}$ | Germany $^{\mathrm{d}}$ |
| Managers and professionals | 1.0 | 1.0 | 1.0 |
| Tradespeople | - | 5.5 | - |
| Clerks | - | 1.1 | 1.5 |
| Technical, sales and administrative support | 2.2 | - | - |
| Sales and service workers | - | 2.2 | 2.9 |
| Service workers | 7.4 | - | - |
| Operators | 9.1 | - | 2.4 |
| Farmers, fishermen and forestry workers | 4.3 | - | 3.6 |
| Operators and farmers | - | 8.8 | - |

- No data.
a Owing to the different reporting schemes, some rows (occupational activities) represent the sum of several other rows.
b Bureau of Labor Statistics (200I), nonfatal occupational injuries and illnesses involving days away from work, for injuries involving the back.
c National Occupational Health and Safety Commission (2001), conditions affecting the upper and lower back.
d Bundesverband der Betriebskrankenkassen (2001), musculoskeletal illnesses of the lower back.

Table 21.57 Exposure categories and relative risks of low back pain for occupational groups selected for this analysis, with managers and professionals as the reference group

| Exposure category | Relative risk | $95 \% ~ C l$ | Occupational activity |
| :--- | :---: | :---: | :--- |
| Background | 1.0 | NA | Managers and professionals |
| Low | 1.38 | $0.85-2.25$ | Clerical and sales workers |
| Moderate | 2.53 | $1.09-5.69$ | Operators and service workers |
| High | 3.65 | $1.57-17.0$ | Farmers |
| NA Not applicable. |  |  |  |
| Source: Exposure level adapted from Leigh and Sheetz (1989). |  |  |  |

therefore likely to be more comparable to those reported by Leigh and Sheetz. The values are, in fact, relatively close except for agricultural workers.

Given that the study by Leigh and Sheetz (1989) best fits the format required for this analysis, and the supporting evidence displaying very similar quantitative values, the proposed exposure categories and attributed relative risks are displayed in Table 21.57. The value for farmers is provided by an average of the relative risks for farmers in the Leigh and

Sheetz (5.17) and Leino-Arjas et al. (2.13) studies, resulting in a relative risk of 3.65. The CI, however, remains the same, because the CI from the Leigh and Sheetz study (1.57-17.0) includes the CI provided by Leino-Arjas et al. (1.6-2.9) and is wider, which is probably a truer representation of the statistical uncertainty of this estimate.

## Methodological quality of the literature

Many of the reviews cited above used systematic criteria to evaluate the potential for selection bias, information bias and confounding in the individual investigations. Several of them identified the methodologically stronger studies and relied primarily or exclusively on those to draw conclusions about the strength of the evidence.

Potential confounding by nonoccupational factors such as sex, age, anthropometry, smoking, heredity and general medical history was extensively investigated in the great majority of studies cited above. All of the studies on which NIOSH relied most heavily, as being rigorous and methodologically sound, controlled for multiple potential confounding variables, permitting the conclusion that physical job factors cause low back pain independently of other factors. Burdorf and Sorock's (1997) review also summarized the associations between low back pain and specific occupational exposures, and relied more heavily on data with adjustment for important covariates. For example, Smedley et al. (1995) adjusted for age, height and nonmusculoskeletal symptoms (the only nonoccupational factors associated with low back pain-see below) in their analysis of low back pain and patient handling demands among female nurses. Tsai et al. (1992) examined the effect of greater vs less physical demands in the job, adjusting for six nonoccupational covariates.

Leigh and Sheetz (1989) adjusted for sex, race, education, height and smoking. In addition, they included terms for occupation and for physical effort and repetitive work; this means that the effect of occupation is likely to be underestimated, since the primary intermediate variable (physical effort) was also included. There is also a great deal of discussion in the epidemiological literature about the mechanisms of the effect of socioeconomic status (see below). It could easily be argued that the inclusion of terms for education also results in overadjustment, since a lower level of education is strongly associated with employment in "unskilled" jobs with higher physical exposures and is likely to act at least in part through such limited job opportunities.

Ozguler et al. (2000) analysed multiple low back pain case definitions. The same set of covariates was examined for each one, and all those nonoccupational factors (sex, age, obesity, psychosomatic "well-being") that were associated with low back pain were kept in the model. Like that of Leigh and Sheetz, this study overadjusted the estimates for occupation, because exposure variables such as carrying heavy loads and bending posture were entered in addition to the occupation indicators.

Many investigators have treated socioeconomic status and sex as potential confounding variables that require adjustment in statistical analysis of MSD etiology. However, to the extent that these factors act through or are surrogates for working conditions, both physical and psychosocial, such analyses may in fact serve to obscure the role of those exposures. Both the incidence and the severity of low back pain show an inverse gradient with socioeconomic status (blue collar vs white collar jobs, income, education level) in both men and women (Bergenudd and Nilsson 1988; Broersen et al. 1996; Heistaro et al. 1998).

It seems highly plausible that a large part of the gradient of socioeconomic status in MSDs is due to differences in the work performed, since jobs with lower socioeconomic status consistently involve more physically strenuous and repetitive work (Behrens et al. 1994; Hollman et al. 1999). In a large study of metal working employees, psychosocial conditions at work and physical load were generally correlated with each other and were worse for blue-collar than for white-collar employees, as well as for women compared with men (Leino and Hänninen 1995). In each of these subgroups, adverse working conditions predicted the development or worsening of MSDs over a 10-year follow-up period. The effect of social class was not explained by "lifestyle" factors such as smoking, leisure-time physical activity, body mass index, alcohol consumption or marital status (Leino-Arjas et al. 1998).

## Sex differences

Sex is also often described as a "risk factor" for MSDs. In the great majority of studies relied on here, either the population was restricted to one sex or relative risks were adjusted for sex. However, in most studies of low back pain the prevalence was the same or only slightly higher in men than in women (e.g. Behrens et al. 1994; Guo et al. 1995; Morken et al. 2000; Tsai et al. 1992). Skovron et al. (1994) found a higher prevalence among men than among women aged 20-49 years, whereas from 50 years of age the prevalence in women gradually increased relative to that in men. Thus sex is not a strong risk factor for low back pain in any case, and confounding of these effect estimates is not of concern.

Women and men typically experience qualitatively and quantitatively different working conditions (Punnett and Herbert 2000). Women are overrepresented in "light", monotonous jobs that require precise, repetitive hand motions with less latitude for decision-making. Men are more often found in jobs with heavy whole-body workload, such as manual materials handling. In general, once job assignments and the consequent occupational exposures are taken into account, sex differences become negligible (Punnett and Herbert 2000).

Table 21.58 Prevalence and attributable risk of joint pain at 6 and II years after retirement, among workers with prior exposure to "heavy physical work"

|  | Interview I (after 6 years) | Interview 2 (after II years) |
| :--- | :--- | :---: |
| Men |  |  |
| Exposed | $58 \%$ | $75 \%$ |
| Unexposed | $49 \%$ | $57 \%$ |
| $\quad$ Attributable risk | $15.5 \%$ | $24.0 \%$ |
| Women |  |  |
| Exposed | $87 \%$ | $91 \%$ |
| Unexposed | $68 \%$ | $79 \%$ |
| Attributable risk | $21.8 \%$ | $13.2 \%$ |

Source: Derriennic et al. (1993).

## Estimates of risk reversibility

Although no explicit studies have been carried out on low back pain attributed to occupational factors before retirement, it has been assumed that leaving the job would reduce the risk of back pain. The burden of work-related back pain would diminish gradually once the theoretical minimum exposure was reached. Since the theoretical minimum is zero, no new cases would arise. However, morbidity from past exposure might persist or worsen after retirement (Derriennic et al. 1993; Holte et al. 2000; Sobti et al. 1997).

Derriennic et al. (1993) defined a closed cohort of retirees from mixed occupations in France, with an average elapsed period of six years from retirement (at age 63) to the baseline survey. A follow-up survey was conducted after five years. Joint pain was reported by $29 \%$ of men and $42 \%$ of women at baseline. At 11 years post-retirement, the attributable risk was even higher among men, although it decreased in women because of the high prevalence of joint pain among unexposed persons (Table 21.58).

In summary, there are few or no epidemiological data on whether or not new back disorders develop after leaving work that can be attributed to ergonomic stressors in previously high-exposed (vs low-exposed) workers. Thus, we have assumed the work-related incidence to be zero after retirement from paid employment. However, we do have the high impact of interventions on exposed workers, which supports a reversibility of $100 \%$. This means that the incidence of low back pain is zero for chronic and acute cases of low back pain after exposure ceases. However, chronic low back pain will continue (i.e. the incidence will be zero, but those who have already developed it will continue to experience it).

## Extrapolation of risk factor-disease relationships from one subregion to another

Because occupational group was used as an indicator of the average level of combined risk factors for low back pain within each occupation, differences in distribution of risk factors that might exist within occupations or between countries are an important consideration. Risk ratios among occupations vary somewhat from one country to another. This could be due to differences in distributions of risk factors for low back pain, or regional or cultural divergences in symptom reporting. These discrepancies become even more difficult to interpret when the comparisons are made between developing and developed countries. Unfortunately, the data are sparse regarding cross-national differences, both in exposure distributions (within similar types of job) and in reporting of low back pain.

One important element is the extent to which ergonomic interventions have been implemented in the various countries or regions. Although there are insufficient data to quantify the extent of effective ergonomic programmes in each region, it is generally true that occupational health and safety legislation, enforcement and adaptation of engineering controls (ergonomic changes) tend to be more widespread in developed countries, especially in northern Europe followed by North America. If this is correct, then application of occupation-specific relative risks from developed countries (e.g. Leigh and Sheetz 1989) to developing countries would produce conservative estimates.

Similarly, it would be easy to assume that, because of mechanization and other changes in production technology, more developed countries would typically have fewer ergonomic stressors in the same type of work than developing countries, even without intention to reduce ergonomic stressors. For example, Bao et al. (1997) compared shoulder-neck ergonomic exposures in a Chinese and a Swedish assembly line workplace. The Swedish workplace had a better ergonomic workstation design and was better balanced, as well as less sensitive to production irregularities, than the Chinese workplace. The Swedish operators were less exposed to awkward postures during work.

However, in contrast to the general assumption that low back pain rates should be higher in low-income than in high-income countries, a systematic review by Volinn (1997) showed 2-4 times higher rates among Belgian, German and Swedish general populations than among southern Chinese, Philippine, Indonesian and Nigerian farmers. Mentioning that the prevalence of low back pain is higher in the urban populations of low-income countries, and sharply higher in enclosed workshops in low-income countries compared to low-income rural populations, Volinn suggested that low back pain might be associated with urbanization and rapid industrialization, which imply more repetitive movements and loss of control over work pace and scheduling. The author noted that interpretation of the findings requires consideration of

Table 21.59 Comparison of ranges of effect estimates for selected risk factors for low back pain in some working populations of China, India and the Russian Federation

|  | China, India, Russian Federation |  |  | Developed countries $^{\mathrm{a}}$ |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Risk factor | Studies (n) | POR range |  | Studies (n) | POR range |
| Bending and twisting ${ }^{b}$ | 4 | $3.1-16.5$ |  | 9 | $1.3-8.1$ |
| Static posture $^{c}$ | 5 | $2.0-19.9$ |  | 3 | $1.3-3.3$ |
| Whole-body vibration $^{2}$ | 4 | $2.5-14.2$ |  | 14 | $1.5-9.0$ |
| Heavy manual lifting | $2^{c}$ | $1.4 \sim 3.5$ |  | 9 | $1.5 \sim 3.1$ |

[^82]the methodological quality of population surveys, such as sampling procedure, formulation of questions, procedures for administration of the survey, and nonresponse bias.

Little information on risk of low back pain by occupation is available from developing countries, in particular studies that would pass the quality criteria of the NIOSH or National Research Council reviews. One summary of the literature from China (Jin et al. 2000) reported risk factors for back pain similar to those reported in developed countries. However, in a comparison of the effect estimates for specific risk factors, the authors found slightly higher prevalence odds ratios (POR) in Chinese low back pain studies than other studies (Table 21.59). Alternative explanations would include unmeasured confounding or effect modification. Studies on occupational back pain performed in developing countries do generally report prevalences of back pain within specified occupations, but without comparing them to a reference group (Chiou and Wong 1992; Chiou et al. 1994; Joshi et al. 2001; Kumar et al. 1999; Muruka 1998; Omokhodion et al. 2000; Toroptsova et al. 1995; Yip 2001). Prevalences were generally high for the studied groups, but the lack of comparison to reference groups did not allow conversion into relative risk information, which was necessary for this analysis.

In summary, plausible arguments can be and have been advanced in favour of low back pain rates in specific occupational groups (farmers, factory workers, etc.) being both higher and lower in developing countries compared with developed countries, but the available data permit neither confirmation of this nor quantification of the differences in risk.

## 6. Occupational risk factors for injuries

Workplace injuries are a common hazard for workers. Deaths due to occupational injuries are defined as any potentially avoidable death due
to an external cause resulting from an exposure related to the person's work. The definition excludes death during commuting to or from the workplace. Workers travelling for work purposes are included.

Data in developed countries indicate that differential risks for injury exist by sector, being highest in agriculture and production, less in sales and service, and lowest in professional, administrative and clerical sectors. But similar data are unavailable for developing countries. At the same time, occupational registries provide some indication of injury out-comes-vs risk factor exposure-which can be used to assess the mortality associated with occupational factors. Applying the fatality rates due to occupational injuries per 100000 insured workers (Table 21.60) to the number of persons in the EAP, as defined earlier in the chapter, gives an indication of total deaths from injuries among workers. The rates reported here for fatal injuries were reported in most countries only for insured populations. Thus, we made the assumption that the same rates applied to all in the EAP, whether or not they were insured, despite some evidence that fatality rates are higher in uninsured populations (Dror 2001; Forastieri 1999; Loewenson 1998). Unfortunately, there is a lack of adequate data on work-related injuries in developing countries to make it possible to generate plausible rates for economic sectors by age, sex and subregion. Mortality outcomes were distributed in the same age pattern as reported in the United States for unintentional injuries.

Because no risk factor exposure is defined in this approach, the counterfactual risk (e.g. theoretical minimum risk level) was defined based on the outcome rather than risk factor exposure. To approximate the safest working conditions observed where all avoidable injury hazards are controlled by effective preventive measures, we chose the injury fatality rate of 0.1 (per 100000 workers) in the age group 16-17 years and in the occupation category "service" from the National Traumatic Occupational Fatalities surveillance system for the United States for the period 1980-1995 (Marsh and Layne 2001).

### 6.1 Outcomes considered

The outcomes considered were unintentional injuries, which include motor vehicle accidents, poisonings, falls, fires, drownings and the category "other unintentional injuries". Other unintentional injuries comprise exposure to inanimate mechanical forces, exposure to mechanical forces, other accidental threats to breathing, exposure to electric current, radiation and extreme ambient air temperature and pressure, contact with venomous animals and plants, exposure to forces of nature and accidental exposure to other and unspecified factors. Homicide at the workplace was not assessed owing to a complete lack of data from developing countries. To estimate the impact of the disability produced by nonfatal injuries, years lived with disability (YLD) were estimated using

Table 21.60 Fatality rates due to occupational injuries (per 100000 insured workers) by country and year

|  |  | Fatality rate <br> per 100000 | Source |
| :--- | :--- | :--- | :--- |
| Country | Year(s) | I982-1984 | 8.06 |
| Australia | $1998-1999$ | 4.0 | Harrison et al. (1989) <br>  <br>  <br> Wustria |
| Worker's Compensation Cases |  |  |  |
| (NOSHC 2002) |  |  |  |

the same attributable fractions as for mortality (age and sex) (i.e. it was assumed that an occupational injury had the same likelihood of being fatal as injuries caused by other factors).

### 6.2 UNDERREPORTING

Conventional sources of data on fatal injuries at work are compensation registries, insurance companies, death certificates and autopsy reports based on mortuary records. Data from compensation registries and insurance companies underestimate the magnitude of fatal injuries, either because they do not cover some sectors of the workforce or because they refer only to successful claims. To improve the accuracy of the reporting of fatal injuries at work, many countries gather data from different systems and data sources (death certificates, insurance companies, labour inspectorates, coroners' files, medical examiners' files) or develop specific projects. Wide disparities exist regarding the accuracy of these sources in identifying fatal injuries at work.

Despite the usefulness of the death records, data from the United States reveal that such records identify only between $67 \%$ and $90 \%$ of fatal injuries at work. A similar underreporting $(72.3 \%)$ has been found in the Mortality Registry of Tuscany in Italy (Chellini et al. 2002). The only study in a developing country that analysed underreporting showed that $28 \%$ of occupational fatalities in Cape Town, South Africa had not been reported in terms of statutory regulations (Lerer and Myers 1994). The level of underreporting increases to between $78 \%$ and $85 \%$ in rural areas (Schierhout et al. 1997). On the other hand, special registries also underreport; the National Fund for Occupational Diseases in Italy, for example, has a reporting rate of only $56.4 \%$ (Chellini et al. 2002).

To our knowledge, the most accurate system currently in place that uses multiple data sources to identify and classify work-related injuries is in the United States. Data are gathered from death certificates in two surveillance systems: the National Traumatic Occupational Fatality System (NTOF) of NIOSH and in the Bureau of Labor Statistics CFOI system. Thus it would appear that the United States has fairly complete records of occupational deaths due to injury (CDC 2001). Although, owing to paucity of data, we did not use the estimates of underreporting to calculate the rates of fatal injuries due to risks at work, this does indicate likely underestimation.

## 7. Results

Tables 21.61-21.63 present the overall attributable fractions, mortality and burden of disease for the selected occupational risk factors considered here.

In total, occupational risk factors considered here were responsible for 775000 deaths worldwide in 2000. There were five times as many deaths in males as in females: 647000 vs 128000 . The leading occupational

Table 21.6I Attributable fractions (\%) for the disease burden due to occupational exposure

| Risk factor | Outcome | Males | Females | Total |
| :---: | :---: | :---: | :---: | :---: |
| Ergonomic stressors | Low back pain | 41 | 32 | 37 |
| Noise | Hearing Loss | 22 | 11 | 16 |
| Agents leading to COPD | COPD | 18 | 6 | 13 |
| Asthmagens | Asthma | 14 | 7 | 11 |
| Risk factors for injuries | Unintentional injuries | 12 | 2 | 8 |
| Beryllium, cadmium, chromium, diesel exhaust, nickel, arsenic, asbestos, silica | Trachea, bronchus or lung cancer | 10 | 5 | 9 |
| Benzene, ethylene oxide, ionizing radiation | Leukaemia | 2 | 2 | 2 |

Table 21.62 Deaths (000s) due to occupational exposure ${ }^{\text {a }}$

|  |  |  | Total |  |  |
| :--- | :--- | :---: | :---: | :---: | :---: |
|  |  |  |  | \% total from <br> occupational <br> risk factors |  |
| Risk factor | Outcome | 240 | 78 | 318 | 41 |
| Agents leading to COPD | COPD | Females | Deaths | 19 | 310 |

[^83]Table 21.63 DALYs (000s) due to occupational exposure

| Risk factor | Outcome | Males | Females | Total |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | DALYs | \% total from occupational risk factors |
| Risk factors for injuries | Unintentional injuries | 9779 | 718 | 10496 | 48 |
| Noise | Hearing Loss | 2788 | 1362 | 4150 | 19 |
| Agents leading to COPD | COPD | 3020 | 713 | 3733 | 17 |
| Asthmagens | Asthma | 1110 | 511 | 1621 | 7 |
| Beryllium, cadmium, chromium, diesel exhaust, nickel, arsenic, asbestos, silica | Trachea, bronchus or lung cancer | 825 | 144 | 969 | 4 |
| Ergonomic stressors | Low back pain | 485 | 333 | 818 | 4 |
| Benzene, ethylene oxide, ionizing radiation | Leukaemia | 66 | 35 | 101 | 0 |
| Total |  | 18073 | 3816 | 21889 | 100.0 |

cause of death was COPD (41\%) followed by unintentional injuries $(40 \%)$ and trachea, bronchus or lung cancer ( $13 \%$ ). Workers who developed outcomes related to occupational risk factors lost about 22 million years of healthy life. By far the main cause of years of healthy life lost, within occupational diseases, was unintentional injuries (with $48 \%$ of the burden). This was followed by hearing loss due to occupational noise ( $19 \%$ ) and COPD due to occupational agents ( $17 \%$ ). Among the occupational factors analysed in this study, these three conditions accounted for $84 \%$ of years of healthy life lost. DALYs were almost five times greater in males than in females. Low back pain and hearing loss have in common the fact that they do not directly produce premature mortality, but substantial disability. This feature differentiates these conditions from the others analysed in the study. Results for specific risk factors are provided below.

### 7.1 Carcinogens

Tables 21.64-21.68 summarize the attributable fractions, mortality and burden of disease for the occupational carcinogens considered here.

For lung cancer, the attributable fraction varied from $5 \%$ in AMR-A to $14 \%$ in EUR-C, with overall attributable fractions for lung cancer estimated to be $10 \%$ for men and $5 \%$ for women ( $9 \%$ overall). For leukaemia, estimates of the attributable fraction varied from $1 \%$ in EMR-D to $3 \%$ in several subregions. There were estimated to be approximately 7000 deaths from leukaemia each year, with a much more even proportion between males and females than was seen for lung cancer, although approximately two thirds of the DALYs are due to male cases.

Table 21.64 Attributable fractions for lung cancer and leukaemia disease burden caused by workplace exposure

| Subregion | Lung cancer |  |  | Leukaemia |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Males | Females | Total | Males | Females | Total |
| AFR-D | 9 | 4 | 7 | 3 | 1 | 2 |
| AFR-E | 9 | 4 | 7 | 3 | 2 | 3 |
| AMR-A | 6 | 2 | 5 | 3 | 3 | 3 |
| AMR-B | 11 | 3 | 8 | 2 | 2 | 2 |
| AMR-D | 12 | 2 | 8 | 3 | 2 | 3 |
| EMR-B | 12 | 2 | 9 | 3 | 2 | 2 |
| EMR-D | 9 | 3 | 7 | 2 | 1 | 1 |
| EUR-A | 7 | 2 | 6 | 3 | 3 | 3 |
| EUR-B | 12 | 4 | 10 | 3 | 2 | 3 |
| EUR-C | 15 | 9 | 14 | 2 | 2 | 2 |
| SEAR-B | 10 | 4 | 9 | 2 | 2 | 2 |
| SEAR-D | 11 | 4 | 9 | 2 | 0 | 2 |
| WPR-A | 8 | 3 | 6 | 2 | 2 | 2 |
| WPR-B | 12 | 7 | 10 | 2 | 2 | 2 |
| World | 10 | 5 | 9 | 2 | 2 | 2 |

Table 21.65 Deaths (000s) from lung cancer and leukaemia caused by workplace exposure

| Subregion | Lung cancer |  |  | Leukaemia |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Males | Females | Total | Males | Females | Total |
| AFR-D | 1 | 0 | 1 | 0 | 0 | 0 |
| AFR-E | 1 | 0 | 1 | 0 | 0 | 0 |
| AMR-A | 7 | 2 | 8 | 0 | 0 | 1 |
| AMR-B | 4 | 0 | 4 | 0 | 0 | 0 |
| AMR-D | 0 | 0 | 0 | 0 | 0 | 0 |
| EMR-B | 1 | 0 | 1 | 0 | 0 | 0 |
| EMR-D | 1 | 0 | 1 | 0 | 0 | 0 |
| EUR-A | 11 | 1 | 12 | 1 | 0 | 1 |
| EUR-B | 6 | 0 | 6 | 0 | 0 | 0 |
| EUR-C | 12 | 1 | 14 | 0 | 0 | 0 |
| SEAR-B | 3 | 0 | 3 | 0 | 0 | 0 |
| SEAR-D | 11 | 1 | 12 | 0 | 0 | 1 |
| WPR-A | 3 | 0 | 4 | 0 | 0 | 0 |
| WPR-B | 27 | 7 | 34 | 1 | 1 | 2 |
| World | 88 | 14 | 102 | 4 | 3 | 7 |

Table 21.66 DALYs (000s) due to lung cancer and leukaemia caused by workplace exposure

| Subregion | Lung cancer |  |  | Leukaemia |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Males | Females | Total | Males | Females | Total |
| AFR-D | 6 | 1 | 7 | 2 | I | 3 |
| AFR-E | 9 | 2 | 11 | 4 | 2 | 6 |
| AMR-A | 53 | 13 | 65 | 4 | 3 | 7 |
| AMR-B | 34 | 4 | 38 | 4 | 4 | 8 |
| AMR-D | 2 | 0 | 2 | 2 | I | 2 |
| EMR-B | 10 | I | 11 | 2 | 1 | 3 |
| EMR-D | 14 | 2 | 16 | 3 | 1 | 4 |
| EUR-A | 89 | 9 | 99 | 6 | 4 | 10 |
| EUR-B | 60 | 5 | 65 | 3 | 2 | 5 |
| EUR-C | 127 | 14 | 140 | 2 | 2 | 4 |
| SEAR-B | 32 | 3 | 34 | 3 | 2 | 5 |
| SEAR-D | 109 | 11 | 120 | 10 | 1 | 11 |
| WPR-A | 23 | 3 | 26 | 1 | 1 | 2 |
| WPR-B | 257 | 76 | 333 | 19 | 11 | 30 |
| World | 825 | 144 | 969 | 66 | 35 | 101 |

Table 21.67 Age-specific attributable fractions, deaths and DALYs for lung cancer and leukaemia, males

|  | Age group (years) |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | 80-89 | All ages |
| Attributable fractions (\%) |  |  |  |  |  |  |  |
| Lung cancer | 11 | 11 | 10 | 10 | 10 | 9 | 10 |
| Leukaemia | 3 | 3 | 3 | 3 | 3 | 3 | 2 |
| Deaths (000s) |  |  |  |  |  |  |  |
| Lung cancer | 0 | 3 | 20 | 30 | 26 | 8 | 88 |
| Leukaemia | I | I | I | I | I | 0 | 4 |
| DALYs (000s) |  |  |  |  |  |  |  |
| Lung cancer | 10 | 76 | 306 | 279 | 136 | 18 | 825 |
| Leukaemia | 29 | 13 | 11 | 7 | 4 | I | 66 |

For each condition, deaths were predominantly among older persons up to 79 years, whereas DALYs tended to be highest in the younger age groups.

### 7.2 Nonmalignant respiratory diseases

Tables 21.69-21.74 summarize the attributable fractions, mortality and disease burden for asthma and COPD risk factors, each estimated as described earlier.

Table 21.68 Age-specific attributable fractions, deaths and DALYs for lung cancer and leukaemia, females

|  | Age group (years) |  |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
|  | $15-29$ | $30-44$ | $45-59$ | $60-69$ | $70-79$ | $80-89$ | All ages |
| Attributable fractions (\%) <br> $\quad$ Lung cancer | 5 | 5 | 5 | 5 | 4 | 4 | 5 |
| Leukaemia | 2 | 3 | 3 | 3 | 3 | 3 | 2 |
| Deaths (000s) <br> $\quad$ Lung cancer <br> Leukaemia | 0 | 1 | 3 | 4 | 4 | 2 | 14 |
| DALYs (000s) <br> Lung cancer <br> Leukaemia | 0 | 0 | 0 | 0 | 1 | 1 | 3 |
|  | 3 | 19 | 52 | 41 | 25 | 4 | 144 |

Table 21.69 Attributable fractions (\%) for mortality from asthma and COPD caused by workplace exposure

| Subregion | Asthma |  |  | COPD |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Males | Females | Total | Males | Females | Total |
| AFR-D | 21 | 15 | 18 | 16 | 5 | 11 |
| AFR-E | 23 | 18 | 20 | 16 | 5 | 11 |
| AMR-A | 15 | 9 | 11 | 18 | 3 | 11 |
| AMR-B | 20 | 8 | 13 | 17 | 3 | 11 |
| AMR-D | 19 | 7 | 13 | 15 | 2 | 9 |
| EMR-B | 18 | 5 | 12 | 17 | 2 | 11 |
| EMR-D | 20 | 10 | 16 | 17 | 3 | 11 |
| EUR-A | 16 | 7 | 11 | 19 | 4 | 13 |
| EUR-B | 22 | 14 | 18 | 19 | 6 | 14 |
| EUR-C | 21 | 12 | 18 | 21 | 6 | 16 |
| SEAR-B | 23 | 14 | 18 | 18 | 6 | 13 |
| SEAR-D | 23 | 14 | 18 | 16 | 5 | 11 |
| WPR-A | 17 | 9 | 13 | 21 | 5 | 16 |
| WPR-B | 22 | 16 | 19 | 19 | 7 | 12 |
| World | 21 | 13 | 17 | 18 | 6 | 12 |

It was estimated that 38000 deaths ( 23000 men and 15000 women) and 1.6 million DALYs result from occupational asthma each year. One quarter to one third of the asthma deaths and DALYs occurred in SEAR-D. The attributable fraction for mortality from asthma varied between subregions from $11 \%$ in AMR-A and EUR-A to 20\% in AFR-E, with worldwide attributable fractions estimated to be $21 \%$ for men and $13 \%$ for women ( $17 \%$ overall). The overall attributable

Table 21.70 Attributable fractions (\%) for burden of disease (DALYs) for asthma and COPD caused by workplace exposure

| Subregion | Asthma |  |  | COPD |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Males | Females | Total | Males | Females | Total |
| AFR-D | 11 | 7 | 10 | 16 | 5 | 11 |
| AFR-E | 13 | 9 | 11 | 16 | 5 | 12 |
| AMR-A | 9 | 4 | 7 | 18 | 3 | 11 |
| AMR-B | 12 | 4 | 8 | 17 | 3 | 10 |
| AMR-D | 11 | 3 | 7 | 13 | 1 | 7 |
| EMR-B | 11 | 2 | 7 | 17 | 2 | 12 |
| EMR-D | 14 | 6 | 10 | 17 | 3 | 11 |
| EUR-A | 11 | 4 | 8 | 19 | 4 | 12 |
| EUR-B | 15 | 8 | 12 | 19 | 6 | 13 |
| EUR-C | 18 | 8 | 14 | 21 | 6 | 14 |
| SEAR-B | 16 | 9 | 13 | 18 | 6 | 13 |
| SEAR-D | 17 | 10 | 13 | 16 | 5 | 11 |
| WPR-A | 12 | 5 | 9 | 21 | 5 | 14 |
| WPR-B | 15 | 9 | 12 | 19 | 7 | 14 |
| World | 14 | 7 | 11 | 18 | 6 | 13 |

Table 21.7I Numbers of deaths (000s) from asthma and COPD caused by workplace exposure

| Subregion | Asthma |  |  | COPD |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Males | Females | Total | Males | Females | Total |
| AFR-D | 1 | 1 | 2 | 4 | 1 | 6 |
| AFR-E | 2 | 1 | 3 | 5 | 1 | 7 |
| AMR-A | 0 | 0 | 1 | 12 | 2 | 14 |
| AMR-B | 1 | 0 | 1 | 8 | 1 | 9 |
| AMR-D | 0 | 0 | 0 | 0 | 0 | 0 |
| EMR-B | 0 | 0 | 0 | 1 | 0 | 1 |
| EMR-D | 2 | 1 | 2 | 7 | 1 | 8 |
| EUR-A | I | 1 | 1 | 16 | 2 | 18 |
| EUR-B | 1 | 1 | 2 | 5 | 1 | 7 |
| EUR-C | 2 | 1 | 3 | 12 | 2 | 15 |
| SEAR-B | 2 | 2 | 4 | 8 | 1 | 9 |
| SEAR-D | 7 | 5 | 12 | 47 | 13 | 60 |
| WPR-A | 1 | 0 | 1 | 3 | 0 | 4 |
| WPR-B | 3 | 3 | 6 | 109 | 52 | 161 |
| World | 23 | 15 | 38 | 240 | 78 | 318 |

Table 21.72 DALYs (000s) from asthma and COPD caused by workplace exposure

| Subregion | Asthma |  |  | COPD |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Males | Females | Total | Males | Females | Total |
| AFR-D | 63 | 27 | 90 | 43 | 10 | 53 |
| AFR-E | 84 | 56 | 141 | 57 | 12 | 69 |
| AMR-A | 37 | 15 | 51 | 147 | 21 | 168 |
| AMR-B | 98 | 27 | 125 | 115 | 17 | 132 |
| AMR-D | 16 | 4 | 19 | 6 | 0 | 6 |
| EMR-B | 18 | 3 | 21 | 20 | 1 | 20 |
| EMR-D | 74 | 27 | 100 | 75 | 13 | 87 |
| EUR-A | 41 | 14 | 55 | 176 | 29 | 205 |
| EUR-B | 30 | 13 | 43 | 75 | 19 | 94 |
| EUR-C | 32 | 9 | 41 | 135 | 34 | 169 |
| SEAR-B | 44 | 26 | 70 | 90 | 21 | 111 |
| SEAR-D | 310 | 166 | 476 | 552 | 149 | 701 |
| WPR-A | 23 | 9 | 33 | 44 | 9 | 53 |
| WPR-B | 241 | 115 | 356 | 1485 | 378 | 1862 |
| World | 1110 | 511 | 1621 | 3020 | 713 | 3733 |

Table 21.73 Age-specific mortality attributable fractions, deaths and DALYs for asthma and COPD, males

|  | Age group (years) |  |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
|  | $15-29$ | $30-44$ | $45-59$ | $60-69$ | $70-79$ | $80-89$ | All ages |
| Attributable fractions (\%) |  |  |  |  |  |  |  |
| Asthma | 23 | 23 | 23 | 22 | 22 | 21 | 21 |
| COPD | 17 | 18 | 18 | 18 | 18 | 19 | 18 |
| Deaths (000s) |  |  |  |  |  |  |  |
| Asthma | 3 | 4 | 6 | 4 | 4 | 2 | 23 |
| COPD | 0 | 3 | 29 | 56 | 91 | 62 | 240 |
| DALYs (000s) |  |  |  |  |  |  |  |
| Asthma | 670 | 228 | 144 | 43 | 20 | 5 | 1110 |
| COPD | 88 | 564 | 992 | 710 | 517 | 149 | 3020 |

fraction for asthma morbidity plus mortality was about two thirds of that for mortality, reflecting the fact that globally a great deal of asthma occurs at younger ages and is nonfatal and nonoccupational in origin.

For COPD mortality, the attributable fraction varied between subregions from 9\% in AMR-D to $16 \%$ in EUR-C and WPR-A (Table 21.69).

Table 21.74 Age-specific mortality attributable fractions, deaths and DALYs for asthma and COPD, females

|  | Age group (years) |  |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
|  | $15-29$ | $30-44$ | $45-59$ | $60-69$ | $70-79$ | $80-89$ | All ages |
| Attributable fractions (\%) | 13 | 14 | 14 | 13 | 13 | 12 | 13 |
| $\quad$ Asthma | 6 | 5 | 5 | 6 | 6 | 6 | 6 |
| COPD |  |  |  |  |  |  |  |
| Deaths (000s) | 2 | 3 | 4 | 2 | 2 | 2 | 15 |
| $\quad$ Asthma | 0 | 1 | 6 | 13 | 28 | 30 | 78 |
| COPD |  |  |  |  |  |  |  |
| DALYs (000s) | 228 | 95 | 81 | 28 | 15 | 5 | 511 |
| Asthma |  |  |  |  |  |  |  |
| COPD | 45 | 133 | 149 | 152 | 166 | 69 | 713 |

Worldwide attributable fractions for COPD were estimated to be $18 \%$ for men and $6 \%$ for women ( $12 \%$ overall). Overall attributable fractions (based on DALYs and reflecting mortality and morbidity) were very similar to the mortality-based fractions (see Tables 21.69 and 21.70). The estimated number of deaths is almost an order of magnitude higher for COPD than for asthma, with an estimated 318000 deaths (240000 men and 78000 women) and 3.7 million DALYs resulting from occupational COPD each year. Half of the COPD deaths and half of the DALYs occurred in WPR-B, owing in part to the large population of the subregion, high background COPD mortality rates and the relatively high employment in mining.

For both asthma and COPD, males predominated. Compared to females, males had nearly $50 \%$ higher attributable fraction for asthma mortality and three times that for COPD mortality. The ratio was about two for disease burden. Similar ratios were seen for the estimated numbers of deaths and DALYs due to these conditions. Asthma deaths were fairly evenly spread among all age groups from 30 to 79 years of age, whereas DALYs predominantly involved persons aged $30-59$ years. For COPD, the majority of deaths occurred in persons aged $\geq 60$ years, whereas DALYs were more evenly spread among all age groups from 30 to 79 years of age (see Tables 21.73 and 21.74).

### 7.3 Noise

Occupational noise-induced hearing loss accounted for more than four million DALYs, all of them produced by the disability associated with hearing loss (YLD). Worldwide, the burden of hearing loss attributed to occupational noise is $16 \%$, ranging between $7 \%$ in WPR-A and $21 \%$ in WPR-B. By sex, the effects of exposure to occupational noise are larger for males than for females in all subregions (Table 21.75). Attributable fractions are related to age group and sex in all subregions. Males usually

Table 21.75 Attributable fractions of occupational noise-induced hearing loss, by sex and subregion

| Subregion | Males | Females | All |
| :--- | :---: | :---: | ---: |
| AFR-D | 23 | 11 | 17 |
| AFR-E | 23 | 12 | 18 |
| AMR-A | 12 | 5 | 9 |
| AMR-B | 19 | 9 | 15 |
| AMR-D | 18 | 9 | 14 |
| EMR-B | 20 | 9 | 15 |
| EMR-D | 20 | 13 | 16 |
| EUR-A | 13 | 5 | 9 |
| EUR-B | 24 | 13 | 19 |
| EUR-C | 24 | 13 | 18 |
| SEAR-B | 23 | 16 | 19 |
| SEAR-D | 24 | 9 | 16 |
| WPR-A | 9 | 6 | 7 |
| WPR-B | 26 | 15 | 21 |
| World | 22 | 11 | 16 |

Table 21.76 Attributable fractions (\%) and DALYs (000s) for occupational noiseinduced hearing loss, by age group

|  | Age group (years) ${ }^{\text {a }}$ |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 15-29 |  | 30-44 |  | 45-59 |  | 60-69 |  | 70-79 |  | Total |  |  |
|  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | All |
| Attributable fraction | 29 | 16 | 29 | 16 | 21 | 11 | 13 | 6 | 3 | I | 22 | 11 | 16 |
| DALYs | 425 | 206 | 1144 | 530 | 925 | 482 | 271 | 136 | 23 | 9 | 2788 | 1362 | 4151 |

[^84]experience greater exposure to noise at work than females, owing to differences in occupational categories, economic sectors of employment and working lifetime. In this study, the attributable fraction decreased with age group after 30-44 years, indicating the heavy impact of occupational noise on the burden of hearing loss at younger ages (Table 21.76). The $30-44$-year age group accounted for the highest number of DALYs and the $70-79$-year age group for the lowest ( 1673000 vs 32000 ).

Table 21.77 provides estimates of the number of DALYs (in thousands) produced by occupational noise-induced hearing loss by subregion in the year 2000. Overall, four million DALYs were lost owing to noise-induced hearing loss. SEAR-D and WPR-B accounted for more

Table 21.77 DALYs (000s) due to occupational noise-induced hearing loss, by sex and subregion

| Subregion | Males | Females | All |
| :--- | ---: | :---: | ---: |
| AFR-D | 109 | 49 | 157 |
| AFR-E | 127 | 60 | 186 |
| AMR-A | 92 | 31 | 123 |
| AMR-B | 122 | 43 | 165 |
| AMR-D | 15 | 6 | 20 |
| EMR-B | 60 | 21 | 81 |
| EMR-D | 142 | 88 | 230 |
| EUR-A | 117 | 47 | 164 |
| EUR-B | 92 | 50 | 142 |
| EUR-C | 136 | 92 | 228 |
| SEAR-B | 219 | 185 | 404 |
| SEAR-D | 799 | 303 | 1101 |
| WPR-A | 26 | 22 | 48 |
| WPR-B | 735 | 365 | 1100 |
| World | 2788 | 1362 | 4151 |

Table 21.78 Attributable fraction and DALYs of low back pain due to occupational ergonomic stressors, by sex and subregion

| Subregion | AF (\%) |  |  | DALYs (000s) |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Males | Females | All | Males | Females | All |
| AFR-D | 36 | 29 | 33 | 21 | 16 | 37 |
| AFR-E | 36 | 31 | 33 | 25 | 20 | 45 |
| AMR-A | 35 | 25 | 30 | 17 | 10 | 27 |
| AMR-B | 41 | 23 | 33 | 32 | 15 | 47 |
| AMR-D | 34 | 18 | 27 | 4 | 2 | 6 |
| EMR-B | 31 | 12 | 22 | 9 | 3 | 12 |
| EMR-D | 36 | 25 | 31 | 25 | 16 | 41 |
| EUR-A | 34 | 22 | 29 | 21 | 11 | 32 |
| EUR-B | 43 | 37 | 40 | 18 | 12 | 30 |
| EUR-C | 45 | 36 | 41 | 21 | 14 | 34 |
| SEAR-B | 43 | 34 | 39 | 26 | 19 | 46 |
| SEAR-D | 43 | 34 | 38 | 111 | 78 | 189 |
| WPR-A | 38 | 27 | 33 | 9 | 5 | 14 |
| WPR-B | 44 | 38 | 41 | 146 | 110 | 256 |
| World | 41 | 32 | 37 | 485 | 333 | 818 |

Table 21.79 Summary results describing the global burden of occupational injuries

| Measure | Males | Females | Total |
| :--- | ---: | ---: | ---: |
| Attributable fraction for disease burden (\%) | 12 | 2 | 8 |
| Deaths (000s) | 291 | 19 | 310 |
| DALYs (000s) | 9779 | 718 | 10496 |

than half of the years of healthy life lost ( 1.1 million each in SEAR-D and WPR-B). Males lost twice the number lost by females (2788000 vs 1362000).

### 7.4 Ergonomic factors

The attributable fractions for low back pain ranged from $22 \%$ to $41 \%$ among the subregions (Table 21.78). Differences by age group were quite small, and the attributable fractions for the total working population (ages 15-65 years) were rather consistent. In most geographical regions, women have a lower attributable burden of low back pain than men, although the difference is most pronounced in the eastern Mediterranean region and the less developed countries in the Americas.

Occupational ergonomic stressors caused 818000 DALYs due to low back pain in 2000 (Table 21.78). Globally, $37 \%$ of low back pain was attributable to occupational causes. The occupational contribution to the burden of low back pain varied relatively little between subregions, with $22 \%$ being the lowest (EMR-B) and $41 \%$ the highest (EUR-C and WPRB). The attributable fraction in men ( $41 \%$ ) was slightly higher than that in women $(32 \%)$, which is mainly due to the type of work men perform, involving more vibration, heavy physical loads or handling of materials.

### 7.5 Risk factors for injuries

Work-related risk factors for unintentional injuries represent $8 \%$ of the burden of unintentional injuries. In all regions, the highest attributable fractions were found in males, reflecting the high number of males exposed to hazardous conditions in the workplace. Overall, the attributable fraction for males was $12 \%$ and for women $2 \%$. Occupational injuries were responsible for 310000 deaths ( 291000 males and 19000 females).

In 2000, there were 10496000 years of healthy life (DALYs) lost among exposed workers (Table 21.79). Overall, males lost about 90\% of healthy life years owing to unintentional injuries at work.

## 8. <br> Discussion

We have attempted to estimate the burden of disease due to selected occupational risk factors by considering exposure, rather than the
common actuarial approach. In this study, a methodology based on the EAP, economic sectors and subsectors and occupational categories was developed to quantify the exposure. Assignment of exposure (low/high) within these categories allowed us to make estimations about the amount of exposure to a given risk factor or groups of risk factors causing an outcome. The dominant source of uncertainty in this analysis was characterizing exposure, which was solely based on economic subsectors and/or occupations and involved a large number of extrapolations and assumptions. High-quality exposure data are lacking, especially in developing countries, and European and American exposure estimates were thus applied in many instances in developing regions (B, C, D and E subregions). This extrapolation could have substantial impact on the accuracy of analysis for the developing regions if exposures, as usually occur, vary from place to place and over time. Diseases with long latency (e.g. cancers) are those that are more susceptible to the assumptions and extrapolations. In addition to problems produced by the length of the latency period, the magnitude of the excess risk may vary depending on the age of the person when exposure began, the duration and strength of exposure and other concomitant exposures. The turnover of workers is another problem that affects both exposure and risk assessment.

The accuracy of the exposure data is fairly coarse because exposures vary greatly within an occupation. This indirect estimation of exposure may cause misclassification of the true exposure situation. The proportions of exposed workers with high exposure in the A and in the $\mathrm{B}, \mathrm{C}$, D and E subregions were less than the published data would indicate. This may be partly because the published literature often focuses on industries and/or occupations with high exposure, but may also indicate an underestimation of true exposure.

Sources of uncertainty in hazard estimates (relative risk and mortality rates) include variations determined from the literature (once again caused by the use of different exposure proxies), extrapolations to regions with different working conditions, the application to females of risk measures from male cohorts, and the application of the same relative risk values to all age groups (e.g. carcinogens).

Restricting the analysis to persons aged $\geq 15$ years excludes the quantification of child labour. The exclusion of children in the estimation was due to the wide variation in the youngest age group for which countries reported EARs. In addition to inconsistent data on EARs for children, there was virtually no data available on their exposure to occupational risk factors or the relative risks of such exposures. Specific, focused research on children is needed to quantify the global burden of disease due to child labour and the resulting implications.

Owing to lack of global data, we could not analyse occupational contributions to the global burden of infectious diseases, cardiovascular dis-
orders, MSDs of the upper extremities, skin disorders and other conditions with recognized occupational etiologies.

### 8.1 Occupational carcinogens

For each condition, deaths were predominantly in older persons up to 79 years, whereas DALYs tended to be highest in the younger age groups. The estimated overall attributable fractions for lung cancer of $10 \%$ for men and $5 \%$ for women ( $9 \%$ overall) are similar to those from a recent United States study, based on a review of relevant studies, in which the attributable fraction for lung cancer was estimated to be between $6 \%$ and $17 \%$ for men, and to be about $2 \%$ for women (Steenland et al. 2003). A similar Finnish study used estimates of $29 \%$ (men) and $5 \%$ (women) but these included a contribution from environmental tobacco smoke, which the study estimated to be about $2 \%$ or $3 \%$ (Nurminen and Karjalainen 2001). (The United States study did not include any contribution from environmental tobacco smoke, but separately estimated the contribution of workplace environmental tobacco smoke to be $5.7 \%$ [Steenland et al. 2003]).

The estimated $2 \%$ attributable fraction for leukaemia compares with $0.8-2.8 \%$ for the United States (Steenland et al. 2003) and $18 \%$ (men) and $2 \%$ (women) for Finland (Nurminen and Karjalainen 2001). The higher Finnish estimate seems to arise from the inclusion of occupational exposure to electromagnetic fields, from the reliance on different studies for relative risk estimates, and from the exposure patterns in the Finnish population.

### 8.2 NONMALIGNANT RESPIRATORY DISEASES

Many of the issues relevant to a discussion of the results for particulates are also relevant to carcinogens, and were discussed in detail under that rubric. The estimated attributable fractions for asthma mortality of $21 \%$ for men and $13 \%$ for women ( $17 \%$ overall) are similar to those from two recent reviews, both of which found an occupational attributable fraction of $15 \%$ (Balmes et al. 2003; Blanc and Toren 1999). The Finnish study on which most of the occupational relative risk estimates used in this study were based had higher estimates for men ( $29 \%$ ) and women ( $17 \%$ ) (Karjalainen et al. 2002), but these estimates are based on Finnish workforce patterns, which are likely to differ from those in most other countries.

Estimates of the attributable fraction for COPD mortality varied between subregions from $9 \%$ to $16 \%$. The overall value of $12 \%$ is very close to the few published estimates of occupational attributable fraction for COPD of $14 \%$ in the United States (Steenland et al. 2003, based on Korn et al. 1987), $14 \%$ for men and $5 \%$ for women in Finland (Nurminen and Karjalainen 2001) and $15 \%$ in a recent review by the American Thoracic Society (Balmes et al. 2003).

### 8.3 Noise

Occupational noise-induced hearing loss accounted for more than four million DALYs, all of them produced by the disability associated with hearing loss (YLD). Worldwide, the burden attributed to occupational noise is $16 \%$, ranging between $7 \%$ in WPR-A and $21 \%$ in WPR-B. By sex, the effects of exposure to occupational noise are larger for males than for females in all subregions. The attributable fraction decreased with age group after 30-44 years, indicating the heavy impact of occupational noise on the burden of hearing loss at younger ages.

In addition to causing irreversible hearing loss, high noise levels in the workplace cause elevated blood pressure, sleeping difficulties, annoyance and stress. Our findings indicate that occupational noise has multiple consequences, both for the individual and for society, and particularly for those suffering hearing loss at young ages. Most occupational noise exposure can be minimized by the use of engineering controls to reduce the generation of noise at its source, within complete hearing loss prevention programmes that include noise assessment, audiometric monitoring of workers' hearing, appropriate use of hearing protectors and worker education.

### 8.4 ERgonomic factors

Human capacity for work depends on many functions and attributes: body size, muscle strength, aerobic fitness, sensory perception and cognitive capacity. Features of the work environment that do not accommodate these needs may produce physical or psychosocial stressors on the human system. Work features that have received attention because of their adverse health effects include heavy manual handling and other types of strenuous work, and awkward body postures.

For this analysis, the exposure variable was work in an occupational category with its assigned level of risk (low, medium or high rate of low back pain). This exposure variable is the "proxy" for the combination of occupational exposures found in the specified occupation that are implicated in the etiology of low back pain.

Occupational ergonomic stressors caused 818000 DALYs from low back pain in 2000. The attributable fractions of low back pain ranged from $22 \%$ to $41 \%$ among subregions, with the global fraction amounting to $37 \%$. The African countries had the highest attributable fraction of low back pain for all age groups analysed. Fractions of $40 \%$ or above were reached in EUR-B and EUR-C and in SEAR-B. The attributable fraction in men ( $41 \%$ ) was slightly higher than that in women $(32 \%)$, which is mainly due to the type of work men perform, involving more vibration, heavy physical load or material handling. Subregional variations reflect differences in occupational types and exposure. Over half of the working population in AFR-D and AFR-E was employed in agriculture. In contrast, about one third of the working populations in the AMR
and EUR subregions were employed in production occupations ("operators") and another large fraction ( $40 \%$ or more) in professional, sales and clerical jobs. In general, males are more exposed than females because they constitute a higher proportion of the labour force. In the less developed subregions, males are generally more exposed because of the higher proportions of workers in formal agriculture than in the developed subregions. The proportion of females in the labour force was particularly low in EMR-B and EMR-D.

The available literature demonstrates the feasibility and benefits of workplace ergonomic interventions (training and engineering controls) that have been implemented by employers in numerous economic sectors. Effective abatement measures include redesigning workstations to eliminate the need for bending and twisting; installing material or patient hoists and other lifting devices; a greater variety of work tasks, to avoid repetitively loading the same body tissues; and improving the mechanical isolation of seating to reduce transmission of whole-body vibration. Training programmes are most effective when they address job design, target supervisory and management personnel along with the manual labour force, and take place in a setting where workers are empowered to utilize the knowledge imparted. In general, the coordination of multiple activities-workstation improvements, training, enhanced medical surveillance and management-within an intervention programme appears to be the most effective. This is consistent with the conclusions of Shannon et al. $(1996,1997)$ that lower injury rates are associated with workplace characteristics such as general workforce empowerment and top management's active leadership, together with delegation of decisionmaking authority regarding occupational safety.

### 8.5 Risk factors for injuries

To our knowledge, this is the first study to estimate attributable fractions of work-related risk factors for unintentional injuries within the overall burden of DALYs. Lack of data on exposure did not allow a risk based approach and the estimates were based on occupational injury registries. This will limit the applicability of these estimates to preventive purposes which are based on exposure. Our findings show that the overall attributable fraction of $9 \%$ reported in this study is above the upper range of values reported by Chen et al. (2001) in the United States. Chen et al. reported an overall attributable fraction of $3.8 \%$, varying between $1.5 \%$ in Arizona and $9.8 \%$ in Alaska. The difference in the findings between the two studies is explained by the heavy burden of mortality in the DALY estimation in developing countries, especially when deaths occur in younger populations.

Our findings understate the importance of the impact of occupational risk factors leading to injuries in the overall burden of disease due to injuries. A major factor in the underestimation was our use of data from an insured population from one country. There is some evidence that
mortality can be greater in uninsured populations, but in the absence of consistent evidence, a similar mortality in the insured and uninsured populations was assumed (Dror 2001; Forastieri 1999; Loewenson 1998). Lerer and Myers (1994) found that $28 \%$ of occupational fatalities in Cape Town, South Africa, were not reported despite a statutory requirement to do so. Using this fraction, we may have missed about 100000 occupational injury deaths due to underreporting. Also, we did not estimate the injury mortality due to intentional injuries such as homicides in the workplace, owing to the lack of data from developing countries. However, current evidence shows that intentional injuries must be present in such countries; thus the lack of an estimation of deaths due to this cause increases the degree of underestimation of the number of deaths due to injuries (e.g. by approximately $4 \%$ in Australia and New Zealand).

Analysis of the full contribution of injuries at work within the overall burden of injuries requires indicators that measure not only mortality but also morbidity. In some countries and regions, with constant or slightly decreasing mortality patterns, it has been observed that the decline in mortality is balanced by an increase in the severity of injuries and morbidity, especially long-lasting or permanent disabilities (CDC 2001; Guerrero et al. 1999). In these cases, evaluation of the effectiveness of preventive measures is also hampered.

Injuries are largely preventable by improvements to make work safer and healthier. Engineering controls, administrative policies, health and safety information and education to promote safety-conscious attitudes and behaviour are needed. Surveillance data must be developed to provide the basis for targeting preventive measures towards high-risk groups of workers. The distribution of burden by type of external cause of mortality has allowed the developed countries to focus on preventive actions at work, resulting in a reduction in injury rates over time. Similar analysis and preventive actions in other countries could greatly reduce injuries at the workplace.

### 8.6 Conclusion

The aim of this study was to estimate the attributable fractions of selected occupational exposures. The risk factors were selected according to the availability of data, the strength of evidence linking the occupational exposure and the outcome, and the amount of risk arising from the exposure. An important feature of these risk factors and the resulting disease burden is their concentration among the working population, especially those in high-risk occupations and sectors. Hazards at workplaces and the resulting illness and injury are understood most accurately in the formal sector, and even there much undercounting occurs. The burden in the informal sector in developing countries, where large proportions of the population work, is high and largely lacks description. Neither household and family agricultural work by women nor child

Table 21.80 Expected rate of growth of the economically active population between 2000 and 2010, by sex and subregion

|  | Growth rate |  |
| :--- | :---: | :---: |
| Subregion | Males | Females |
| AFR-D | 0.33 | 0.38 |
| AFR-E | 0.27 | 0.25 |
| AMR-A | 0.08 | 0.12 |
| AMR-B | 0.17 | 0.27 |
| AMR-D | 0.27 | 0.45 |
| EMR-B | 0.32 | 0.66 |
| EMR-D | 0.31 | 0.53 |
| EUR-A | -0.35 | 0.03 |
| EUR-B | -0.26 | 0.16 |
| EUR-C | 0.00 | 0.00 |
| SEAR-B | 0.17 | 0.24 |
| SEAR-D | 0.21 | 0.27 |
| WPR-A | -0.03 | 0.04 |
| WPR-B | 0.11 | 0.11 |

Source: ILO (2002a).
labour were addressed in our study. Due primarily to lack of data in developing countries, we were unable to include important occupational risks for infectious diseases, dermatitis, reproductive disorders, some cancers, ischaemic heart disease, musculoskeletal disorders of the upper extremities, and other conditions such as workplace stress.

The estimated burden of occupational risk factors can be diminished by improving working conditions, as many examples from different countries have shown. Work-related diseases are largely preventable. For example, many dusty activities can be made safer by using wet methods, thus reducing workers' exposures to silica. Work surfaces can be adjusted to a worker's height, thereby reducing suffering from low back pain. Substituting safe chemicals for known carcinogens can prevent many cancers. A change of process can reduce noise levels, thus protecting workers' hearing. Attention to electrical safety or machine guarding can eliminate tragic injuries at the workplace.

## 9. Projections of future exposure

In the next 50 years, the population of the developing regions will steadily rise, whereas that of more developed regions is expected to change little because fertility levels will remain below replacement level (UN 2001). There will also be differences in growth rates between the
sexes. A negative growth rate among economically active males is expected to occur between 2000 and 2010 in developed regions such as Europe, while comparable female rates will continue increasing in most of the regions, including the developed ones (ILO 2002a) (Table 21.80).

The expected changes in the world population will affect the EAP as well as the median age of workers (Fullerton and Toosi 2001). These changes in the characteristics of the working population will be accompanied by a different distribution of employment in the economic sectors (agriculture, industry and services). Currently, the service sector of many economies is growing at a fast rate, while the agricultural sector is rapidly declining in developing countries and remains at a stable low level in developed countries. It is expected that these different patterns of growth within the economic sectors will continue in the coming years. Moreover, the expected changes will affect the distribution of occupations within an economic sector. In developed countries in which a change in the structure of the economy has been observed, there has been a shift in the proportion of workers from the "production" category in favour of professional, managerial, clerical and sales occupations.

### 9.1 Exposure estimation for the years 2010, 2020 AND 2030

As mentioned above, the EAP by economic sector was used to estimate the working population exposed to some risk factors, including car-

Table 21.81 Projected EARs for the year 2010 by sex and subregion

| Subregion | Males | Females |
| :--- | :--- | :---: |
| AFR-D | 0.84 | 0.55 |
| AFR-E | 0.85 | 0.64 |
| AMR-A | 0.7 | 0.59 |
| AMR-B | 0.8 | 0.45 |
| AMR-D | 0.81 | 0.44 |
| EMR-B | 0.78 | 0.39 |
| EMR-D | 0.81 | 0.42 |
| EUR-A | 0.57 | 0.47 |
| EUR-B | 0.69 | 0.56 |
| EUR-C | 0.75 | 0.59 |
| SEAR-B | 0.82 | 0.62 |
| SEAR-D | 0.84 | 0.48 |
| WPR-A | 0.71 | 0.52 |
| WPR-B | 0.81 | 0.68 |

Source: ILO (2002a).

Table 21.82 Projected distribution of EAP by subregion in 2010, 2020 and 2030

| Subregion | 2010 |  | 2020 |  | 2030 |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | EAP | \% total | EAP | \% total | EAP | \% total |
| AFR-D | 151300284 | 4.6 | 199904691 | 5.4 | 260126023 | 6.3 |
| AFR-E | 181011306 | 5.5 | 232504879 | 6.3 | 300771980 | 7.3 |
| AMR-A | 176129373 | 5.4 | 191817362 | 5.2 | 201632376 | 4.9 |
| AMR-B | 218574298 | 6.7 | 251138963 | 6.8 | 278348992 | 6.8 |
| AMR-D | 35233802 | 1.1 | 43360108 | 1.2 | 51164255 | 1.2 |
| EMR-B | 67730185 | 2.1 | 82896189 | 2.2 | 98826530 | 2.4 |
| EMR-D | 165776470 | 5.1 | 214302617 | 5.8 | 269408795 | 6.5 |
| EUR-A | 172528633 | 5.3 | 171619225 | 4.6 | 165811266 | 4.0 |
| EUR-B | 110565142 | 3.4 | 118115432 | 3.2 | 123452320 | 3.0 |
| EUR-C | 125923283 | 3.8 | 118343637 | 3.2 | 112001074 | 2.7 |
| SEAR-B | 173799078 | 5.3 | 196214683 | 5.3 | 213861114 | 5.2 |
| SEAR-D | 663743911 | 20.3 | 784531784 | 21.1 | 885891293 | 21.5 |
| WPR-A | 77452109 | 2.4 | 76367545 | 2.1 | 72515258 | 1.8 |
| WPR-B | 952086321 | 29.1 | 1030847264 | 27.8 | I 086544 \| 12 | 26.4 |
| Total | 3271854196 | 100.0 | 3711964378 | 100.0 | 4120355388 | 100.0 |

cinogens, while occupational category within a sector was used for others, including noise and ergonomic stressors. Therefore, to project the exposed population for the years 2010, 2020 and 2030 a three-step procedure was followed: (i) the EAP was estimated; (ii) the EAP was distributed among economic sectors; and, where needed, (iii) occupational categories were distributed within the economic sectors.

## EAP Estimation

To obtain the EAP for the year 2010, we multiplied the overall population (2010) by the EARs by subregion for the year 2010 as estimated by ILO (See Table 21.81). Then, in the absence of other data, the same EAR by subregion was used for the years 2020 and 2030 to generate the EAP (see Equation 4). Calculations were restricted to persons aged $\geq 15$ years by sex and subregion, thus allowing regional patterns to be preserved.

$$
\begin{align*}
& E A P_{15+j}=\sum\left[E A R_{2010}(\text { for each age group } \geq 15\right. \\
& \left.\left.\quad \times \text { Population }_{i}(\text { for each age group } \geq 15)\right)\right] \tag{4}
\end{align*}
$$

where
$\mathrm{EAP}_{15+\mathrm{j}}=$ economically active population $\geq 15$ years, $\varphi=$ year (2010, 2020, 2030)

Population $_{\mathrm{i}}=$ population year 2010, 2020, 2030
EAR $\quad=$ economic activity rate, year 2010
The EAP will increase steadily towards 2010, 2020 and 2030, but the amount of the increase and the patterns are somewhat different between developed and developing countries, as well as among countries having a similar degree of development. The percentage distribution of the EAP by subregion reflects the growth of the overall population, with greater growth in developing countries. WPR-B and SEAR-D will contribute $49.4 \%$ of global EAP in the year 2010, whereas developed subregions will contribute only $13.1 \%$ (Table 21.82).

## EAP DISTRIBUTION AMONG ECONOMIC SECTORS FOR 2010, 2020 AND 2030

The basic approach to estimating the EAP among economic sectors was to use regression analysis to identify the relationship between the distribution of the economic sectors and the projected years of interest. The dependent variables (proportion of EAP employed in agriculture, industry or services) were separately compared to the independent variable time, ${ }^{10}$ using the following model:

$$
\begin{equation*}
\text { PEAPA }=\ln \left(a Y_{T}\right)+\ln b \tag{5}
\end{equation*}
$$

where

Figure 21.3 Projected distribution of the agricultural sector by year and subregion, 2010, 2020 and 2030


PEAPA $=$ proportion of EAP in agriculture (similarly, PEAPI and PEAPS for industry, or service) in Year T
$\mathrm{Y}_{\mathrm{T}} \quad=$ Time (Year T)
The slope factors and intercepts obtained by regression analysis, using the EAP proportion by economic subsector for the years 1990-2000, were then used to estimate the proportion of the EAP for the years 2010, 2020 and 2030, separately for each economic subsector. We did not include economic development (e.g. measured as GDP per capita) as an additional variable in the analysis, assuming that previous trends capture the effects of trends in GDP. Given the economic and social factors that determine occupational distributions, the changes in the EAP in the future are subject to behavioural decisions by individuals, policy decisions in home countries and abroad, and developments in education. The project distribution of EAP among economic sectors showed different patterns among different subregions. As an example, Figure 21.3 presents the distribution of EAP in agriculture.

## Occupational categories adjustment

No data were available to develop trends for employment in occupational categories in 2010, 2020 and 2030. Therefore, proportions of exposed workers within occupational categories were adjusted according to the distribution pattern of the year 2000, adjusted only for the new proportions employed within economic sectors in the year of interest.

## Notes

1 See preface for an explanation of this term.
2 Dusts are technically defined as dry particle aerosols produced by mechanical processes such as breaking, grinding and pulverizing (Johnson and Swift 1997). Particle sizes range from less than $1 \mu \mathrm{~m}$ to over $100 \mu \mathrm{~m}$. The smaller particles present a greater hazard, as they remain airborne longer and are more likely to enter the respiratory tract. Dusts may be organic (e.g. grain dust) or inorganic (e.g. silica, asbestos and coal dust).
3 Economic activities comprise agriculture, mining, manufacturing, utilities, construction, trade, transport, finance and services.

4 dBA is the unit of sound pressure level in decibels that has been A-weighted, i.e. measured with an A-weighted sound level meter. Sound levels measured in dBA have been widely used to evaluate occupational and environmental exposures because of the good correlations between the "A" scale and human hearing ability at different frequencies, hearing damage and environmental annoyance.
5 The average of the hearing threshold levels for both ears that exceeds 25 dB at $1000,2000,3000$ and 4000 Hz .

6 Tinnitus is noise originating in the ear rather than in the environment. The noise may be a buzzing, ringing, roaring, whistling, humming or hissing in the ears. Ringing in the ears is an extremely common phenomenon experienced by up to a third of the adult population at one time or another.
7 A temporary increase in the threshold of hearing for an ear caused by exposure to high-intensity noise.
8 The percentage of workers with a hearing impairment in an occupationally noise-exposed population, after subtracting the percentage in an unexposed population who would normally incur such impairment owing to ageing.
9 Year was the predictor of the data.

## References

Ahmed HO, Dennis JH, Badran O et al. (2001) Occupational noise exposure and hearing loss of workers in two plants in eastern Saudi Arabia. Annals of Occupational Hygiene, 45:371-380.
Alidrisi M, Jamil ATM, Jiffry MSA, Jefri MA, Erturk F (1990) Evaluation of noise stresses in Jeddah Industrial State. Journal of Environment Science and Health, A25:873-896.
Amweelo M (2000) Accident prevention in Namibia. African Newsletter on Occupational Health and Safety 10(1).

Armstrong TJ, Buckle P, Fine LJ et al. (1993) A conceptual model for workrelated neck and upper-limb musculoskeletal disorders. Scandinavian Journal of Work, Environment and Health, 19:73-84.
Arndt V, Rothenbacher D, Brenner H et al. (1996) Older workers in the construction industry: results of a routine health examination and a five-year follow-up. Occupational and Environmental Medicine, 53:686-691.
Astrand NE (1987) Medical, psychological, and social factors associated with back abnormalities and self reported back pain: a cross sectional study of male employees in a Swedish pulp and paper industry. British Journal of Industrial Medicine, 44:327-336.
Atallah ZR, Jamous LW, AbuDhaise BA, Alwash RH (1998) Fatal occupational injuries in Jordan during the period 1980 through 1993. Safety Science, 28:177-187.

Attfield M, Wagner G (1998) Respiratory disease in coal miners. In: Environmental and occupational medicine, 3rd edn. Rom W, ed. Little, Brown and Co., Boston, MA.
Badley EM, Rasooly I, Webster GK (1994) Relative importance of musculoskeletal disorders as a cause of chronic health problems, disability, and health care utilization: findings from the 1990 Ontario Health Survey. The Journal of Rheumatology, 21:505-514.
Badley EM, Webster GK, Rasooly I (1995) The impact of musculoskeletal disorders in the population: are they just aches and pain? Findings from the 1990 Ontario Health Survey. The Journal of Rheumatology, 22:733-739.

Balmes J, Becklake M, Blanc P et al. (2003) American Thoracic Society statement on occupational contribution to the burden of airway disease. American Journal of Respiratory and Critical Care Medicine, 167:787-797.

Bao S, Winkel J, Mathiassen SE, Shahnavan H (1997) Interactive effect of ergonomics and production engineering on shoulder-neck exposure: a case study of assembly work in China and Sweden. International Journal of Industrial Ergonomics, 20:75-85.
Battie MC, Bigos SJ (1991) Industrial back pain complaints: a broader perspective. Orthopedics Clinics of North America, 22:273-282.
Becklake M (1989) Occupational exposures: evidence for a causal association with chronic obstructive pulmonary disease. American Review of Respiratory Disease, 140:S85-91.

Becklake MR (1994) Symptoms and pulmonary functions as measures of morbidity. Annals Occupational Hygiene, 38:569-580, 418.

Behrens V, Seligman P, Cameron L et al. (1994) The prevalence of back pain, hand discomfort, and dermatitis in the U.S. working population. American Journal of Public Health, 84:1780-1785.
BEIR V (1990) Health effects of exposure to low levels of ionizing radiation: BEIR V. Committee on the biological effects of ionizing radiation. National Academy Press, Washington, DC.
Bergenudd H, Nilsson B (1988) Back pain in middle age: occupational work load and psychologic factors: an epidemiologic study. Spine, 13:58-60.
Bernard BP, ed. (1997) Musculoskeletal disorders and workplace factors: a critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back. CDC/NIOSH, Cincinnati, OH.

Bernard BP, Sauter SL, Petersen M et al. (1993) Health hazard evaluation report: Los Angeles Times. (NIOSH Report No. 90-013-2277). Cincinnati, OH.

Bhatia R, Lopipero P, Smith A (1998) Diesel exhaust exposure and lung cancer. Epidemiology, 9:84-91.
Bhattacharya SK, Saiyed HN, Roy A, Chatterjee SK (1981) Hearing acuity in weavers of a textile mill. Indian Journal of Medical Research, 74:779-785.
Bhattacharya SK, Tripathi SR, Kashyap S (1990) Heat and noise problems in a firm in a drug and pharmaceutical firm in India. Industrial Health, 28:203-207.

Blanc P, Toren K (1999) How much adult asthma can be attributed to occupational factors? American Journal of Medicine, 107:580-587.

Bongers PM, de Winter CR, Kompier MAJ et al. (1993) Psychosocial factors at work and musculoskeletal disease. Scandinavian Journal of Work, Environment and Health, 19:297-312.

Bongers PM, Hulshof CTJ, Dijkstra L, Boshuizen HC, Groenhout HJM, Valken E (1990) Back pain and exposure to whole body vibration in helicopter pilots. Ergonomics, 33:1007-1026.

Bovenzi M, Betta A (1994) Low-back disorders in agricultural tractor drivers exposed to whole-body vibration and postural stress. Applied Ergonomics, 25:231-241.

Broersen JPJ, de Zwart BCH, Dijk F et al. (1996) Health complaints and working conditions experienced in relation to work and age. Occupational and Environmental Medicine, 53:51-57.
Buckle PW, Devereaux J (1999) Work-related neck and upper limb musculoskeletal disorders. European Agency for Safety and Health at Work, Luxembourg.
Bundesverband der Betriebskrankenkassen (2001) Krankheitsartenstatistik 1997. (Internet communication of 20 September 2001 at: http://de.osha.eu.int/statistics/bkk/index.stm).

Burchfiel CM, Boice JA, Stafford BA, Bond GG (1992) Prevalence of back pain and joint problems in a manufacturing company. Journal of Occupational Medicine, 34:129-134.

Burdorf A (1992) Sources of variance in exposure to postural load on the back in occupational groups. Scandinavian Journal of Work, Environment and Health, 18:361-367.

Burdorf A, Naaktgeboren B, deGroot HC (1993) Occupational risk factors for low backpain among sedentary workers. Journal Occupational Medicine, 35:1213-1220.
Burdorf A, Sorock G (1997) Positive and negative evidence of risk factors for back disorders. Scandinavian Journal of Work, Environment and Health, 23:243-256.

Bureau of Labor Statistics (2001) Safety and health statistics. Office of Safety, Health and Working Conditions. U.S. Department of Labor. (Internet Communication of 10 September 2001 at: http://stats.bls.gov/oshhome.htm).

Burton, DJ (1997) Chapter 31. General methods for the control of airborne hazards. In: The industrial environment-its evaluation and control. Di Nardi S, ed. AIHA, Fairfax, VA.
Butler S, Twaddle S (1979) Turnover in the restaurant industry: a structural explanation. Mid-South Sociological Association. Sociological Abstracts, 27(4).
CCOHS (1997) OSH answers. Canadian Centre for Occupational Health and Safety. Available at http://www.ccohn.ca/oshanswers/diseases.
CDC (2001) Non-fatal occupational injuries and illnesses treated in hospital emergency departments-Unites States 1998. Centers for Disease Control and Prevention. Journal of the American Medical Association, 285:2443-2444.

Chan-Yeung M, Malo J (1994) Aetiological agents in occupational asthma. European Respiratory Journal, 7:346-371.
Checkoway (2000) Epidemiological evidence on the carcinogenicity of silica: factors in scientific judgement: letter. Annals of Occupational Hygiene, 44:483-487.

Chellini E, Baldasseroni A, Giovannetti L, Zoppi O (2002) A survey on fatal work accidents based on mortality registry data: results of the Tuscany study
on INAIL and RMR cases in the period 1992-1996. Epidemiologia e Prevenzione, 26:11-17.

Chen W, Zhuang Z, Attfield MD et al. (2001) Exposure to silica and silicosis among tin miners in China: exposure-response analyses and risk assessment. Occupational and Environmental Medicine, 58:31-37.

Chien V, Chai S, Hai D et al. (2002) Pneumoconiosis among workers in a Vietnamese refractory brick facility. American Journal of Industrial Medicine, 42:397-402.

Chiou, WK, Wong MK (1992) Epidemiology of low back pain in the nurses of Chang Gung Memorial Hospital. Chang Gung Medical Journal, 15: 64-71.
Chiou WK, Wong MK, Lee YH (1994) Epidemiology of low back pain in Chinese nurses. International Journal of Nursing Studies, 31:361-368.
Choi BCK, Tennassee LM, Eijkemans GJM (2001) Developing regional workplace health and hazard surveillance in the Americas. Pan American Journal of Public Health, 10: 376-381.

Davis AC (1989) The prevalence of hearing impairment and reported hearing disability among adults in Great Britain. International Journal of Epidemiology, 18:911-917.
de Zwart BCH, Broersen JP, Frings-Dresen MHW, van Dijk FJH (1997) Repeated survey on changes in musculoskeletal complaints relative to age and work demands. Occupational and Environmental Medicine, 54:793-799.
Derriennic F, Iwatsubo Y, Monfort C et al. (1993) Evolution of osteoarticular disorders as a function of past heavy physical work factors: longitudinal analysis of 627 retired subjects living in the Paris area. British Journal of Industrial Medicine, 50:851-860.

Deyo RA, Rainville J, Kent D (1992) What can the history and physical examination tell us about low back pain? Journal of the American Medical Association, 268:760-765.

Dosemeci M, McLaughlin JF, Chen J-Q et al. (1995) Historical total and respirable silica dust exposure levels in mines and pottery factories in China. Scandinavian Journal of Work, Environment and Health, 2:S39-43.
Dror DM (2001) Reinsurance of health insurance for the informal sector. Bulletin of the World Health Organization, 79:672-678.
Dupre D (2001) Accidents at work in the EU 1998-1999. EUROSTAT (Theme 3-16/2001). Statistics in Focus, Population and Social Conditions, Luxembourg.

EASHW (2000) Monitoring the state of occupational safety and health in the European Union-pilot study. European Agency for Safety and Health at Work. Luxembourg.

EIA (2001) Energy Information Administration, U.S. Department of Energy, International Energy Database, January 2001, available at www.eia.doe.gov/emeu/iea/coal.html.
EIU (1995) Business operations report, India, 4th quarter, 1995. Shortage of managers. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1996a) China hand. Ch. 13. Human resources: knowing why staff leave, 1 Sept. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1996b) China labour: unemployed reluctant to trade welfare for work. Country view, 27 Nov. Economist Intelligence Unit. Available at http://www.eiu.com.
EIU (1997a) 28 February. Economist Intelligence Unit. Available at http://www.eiu.com.
EIU (1997b) Brazil regulations: rules eased on loans for exporters. Country alert, 6 May. Economist Intelligence Unit. Available at http://www.eiu.com.
EIU (1997c) Business operations report, Laos. Importance of training: training breeds staff loyalty, 26 Jun. Economist Intelligence Unit. Available at http://www.eiu.com.
EIU (1997d) Caribbean industry: region seeks to lure informatics firms. Country alert, 2 Jul. Economist Intelligence Unit. Available at http://www.eiu.com.
EIU (1997e) Country commerce South Korea. Human resources. Overview, 1 Jul. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1997f) USA industry: Silicon Valley survey—of mice \& men. Country view, 1 Apr. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1998) Business operations report Eastern Europe. Business services: background checks on security personnel essential, 16 March.
EIU (2001a) Country commerce China. Human resources. Overview, 28 Feb. Economist Intelligence Unit. Available at http://www.eiu.com.
EIU (2001b) Country forecast Africa. AIDS/Africa: labour costs are set to increase substantially, 14 May. Economist Intelligence Unit. Available at http://www.eiu.com.
EIU (2001c) Country profile Hong Kong. Resources and infrastructure: education, 1 Jul. Economist Intelligence Unit. Available at http://www.eiu.com.
EIU (2002) Country commerce China. Human resources. Overview, 28 Feb. Economist Intelligence Unit. Available at http://www.eiu.com.
Emmett EA (2002) Occupational contact dermatitis I: incidence and return to work pressures. American Journal of Contact Dermatitis, 13:30-34.
Enterline P, Marsh G, Esmen N, Henderson V, Callahan C, Paik M (1987) Some effects of cigarette smoking arsenic, and $\mathrm{SO}_{2}$ on mortality among US copper smelter workers. Journal of Occupational Medicine, 29:831-838.

Evans WA, Ming HY (1982) Industrial noise-induced hearing loss in Hong Kong: a comparative study. Annals of Occupational Hygiene, 25:63-80.
Ezenwa A (2001) A study of fatal injuries in Nigerian factories. Occupational Medicine, 51:485-489.
Ferris BG Jr, Speizer FE, Spengler JD et al. (1979) Effects of sulfur and respirable particles on human health: methodology and demography of populations in study. American Review of Respiratory Disease, 120:767-779.
Feyer AM, Williamson AM, Stout N, Driscoll T, Usher H, Langley JD (2001) Comparison of work related fatal injuries in the United States, Australia, and New Zealand: method and overall findings. Injury Prevention, 7:22-28.

Fine LJ, Silverstein BA, Armstrong TJ, Anderson CA, Sugano DS (1986) Detection of cumulative trauma disorders of upper extremities in the workplace. Journal of Occupational Medicine, 28:674-678.

FIOH (1999) CAREX database. Finnish Institute of Occupational Health. Available at http://www.ocuphealth.fi/internet/english.
FIOH (2001) CAREX database. Finnish Institute of Occupational Health. Available at http://www.occuphealth.fi/list/data/CAREX.
Forastieri V (1999) Improvement of working conditions and environment in the informal sector through safety and health measures. International Labour Organization, Geneva.
Frank JW, Kerr MS, Brooker A-S et al. (1996) Disability resulting from occupational low back pain, Part I: What do we know about primary prevention? A review of the scientific evidence on prevention before disability begins. Spine, 21:2908-2917.

Frank JW, Pulcins IR, Kerr MS et al. (1995) Occupational back pain-an unhelpful polemic. Scandinavian Journal of Work, Environment and Health, 21:3-14.

Fullerton H, Toosi M (2001) Labor force projections to 2010:steady growth and changing composition. Monthly Labor Review, 124:21-38.
Furuya S, Natori Y, Ikeda R (2003) Asbestos in Japan. International Journal Occupational Environmental Health, 9:260-265.
Garg A (1992) Occupational biomechanics and low-back pain. Occupational Medicine: State of the Art Review, 7:609-628.
Gomes J, Lloyd O, Norman N, Pahwa P (2001) Dust exposure and impairment of lung function at a small iron foundry in a rapidly developing country. Occupational and Environmental Medicine, 58:656-662.

Guerrero JL, Sniezek JE, Sehgal (1999) The prevalence of disability from chronic conditions due to injury among adults ages 18-69 years: United States, 1994. Disability and Rehabilitation, 21:187-192.

Guo H-R, Tanaka S, Cameron LL et al (1995) Back pain among workers in the United States: national estimates and workers at high risk. American Journal of Industrial Medicine, 28:591-602.
Guo H-R, Tanaka S, Halperin WE, Cameron LL (1999) Back pain prevalence in U.S. industry and estimates of lost workdays. American Journal of Public Health, 89:1029-1035.
Hagberg M, Silverstein BA, Wells RP et al., eds (1995) Work-related musculoskeletal disorders (WMSD): a handbook for prevention. Taylor and Francis. London.

Hales TR, Bernard BP (1996) Epidemiology of work-related musculoskeletal disorders. Orthopedic Clinics of North America, 27:679-709.

Harrison JE, Frommer MS, Ruck EA, Blyth FM (1989) Deaths as a result of work-related injury in Australia, 1982-1984. Medical Journal of Australia, 50:118-125.

Hayes R, Yin S, Dosemeci M et al. (1997) Benzene and the dose-related incidence of hematologic neoplasms in China. Chinese Academy of Preventive Medicine-National Cancer Institute Benzene Study Group. Journal of the National Cancer Institute, 89:1065-1071.
Heistaro S, Vartiainen E, Heliövaara M et al. (1998) Trends of back pain in eastern Finland, 1972-1992, in relation to socioeconomic status and behavioral risk factors. American Journal of Epidemiology, 148:671-682.
Hendrick DJ (1996) Occupational and chronic obstructive pulmonary disease (COPD). Thorax 51:947-955.
Hernandez-Gaytan SI, Santos-Burgo AC, Becker-Meyer JP, Macias-Carrillo C, Lopez-Cervantes M (2000) Prevalence of hearing loss and correlated factors in a cement plant. Salud Publica de Mexico, 42:106-111.

Hessel P (2000) Hearing loss among construction workers in Edmonton, Alberta, Canada. Journal of Occupational and Environmental Medicine, 42:57.

Hessel PA, Sluis-Cremer GK (1987) Hearing loss in white South African goldminers. South African Medical Journal, 71:364-367.

Hewett P (1996) Interpretation and use of occupational exposure limits for chronic disease agents. Occupational Medicine: State of the Art Reviews, 11:561-590.
Hildebrandt VH (1995) Back pain in the working population: prevalence rates in Dutch trades and professions. Ergonomics, 38:1283-1298.
Hnizdo E, Sluis-Cremer G (1991) Silica exposure, silicosis, and lung cancer: a mortality study of South African gold miners. British Journal Industrial Medicine, 48:53-60.

Hollman S, Klimmer F, Schmidt K-H et al. (1999) Validation of a questionnaire for assessing physical work load. Scandinavian Journal of Work, Environment and Health, 25:105-114.

Holte HH, Tambs K, Bjerkedal T (2000) Manual work as predictor for disability pensioning with osteoarthritis among the employed in Norway 1971-1990. International Journal of Epidemiology, 29:487-494.
Hoogendoorn WE, van Poppel MNM, Bongers PM, Koes BW, Bouter LM (1999) Physical load during work and leisure time as risk factors for back pain. Scandinavian Journal of Work, Environment and Health, 25:387403.

Hulshof CTJ, Veldhuijzen van Zanten B (1987) Whole-body vibration and low-back pain: a review of epidemiologic studies. International Archives of Occupational and Environmental Health, 59:205-220.

Human Resources Development Canada (2000) Surveys until 2000. Available at http://www.hrdc-drhc.gc.ca/hrib/hrp-prh/ssd-des/english/glossary/ survey.shtml.
IARC (1977) Asbestos. (Monograph 14.) International Association of Cancer Registries, Lyon.
IARC (1980) Some metals and metallic compounds. (Monograph 23.) International Association of Cancer Registries, Lyon.

IARC (1987a) Overall evaluations of carcinogenicity: an updating of IARC. (Monographs 1-42, Supplement 7.) International Association of Cancer Registries, Lyon.

IARC (1987b) Silica and some silicates. (Monograph 42.) International Association of Cancer Registries, Lyon.
IARC (1989) Diesel and gasoline engine exhausts and some nitroarenes. (Monograph 46.) International Association of Cancer Registries, Lyon.
IARC (1990a) Chromium, nickel and welding. (Monograph 49.) International Association of Cancer Registries, Lyon.
IARC (1990b) Cancer. IARC Scientific Publications. International Association of Cancer Registries, Lyon.
IARC (1993) Beryllium, cadmium, mercury, and exposures in the glass manufacturing industry. (Monograph 58.) International Association of Cancer Registries, Lyon.

IARC (1997) Silica, some silicates, coal dust and para-aramid fibrils. (Monograph 68.) International Association of Cancer Registries, Lyon.

IARC (2000) Ionizing radiation, part 1: X- and gamma (g)-radiation, and neutrons. (Monograph 75.) International Association of Cancer Registries, Lyon.

IARC (2001) Overall evaluation of carcinogenicity to bumans. International Association of Cancer Registries, Lyon. Available at http://193.51.164.11/monoeval/crthall.html.
IARC (2002) IARC Monographs Programme on the Evaluation of Carcinogenic Risks to Humans. Available at http://193.51.164.11.
ICNCM (1990) Report of the International Committee on Nickel Carcinogenesis in Man. Scandinavian Journal of Work, Environment and Health, 16:1-84.

ICRP (1991) Recommendations of the International Commission on Radiological Protection. (ICRP Publication No. 60. Vol 21:1-3.) International Commission on Radiological Protection, Oxford.

ILO (1987) Resolutions adopted by the International Conference of Labour Statisticians. Annex 1. Major, sub-major, minor and unit group titles. International Labour Organization. Available at http://www.ilo.org/public/ english/bureau/stat/download/res/isco.pdf.
ILO (1995a) Yearbook of labour statistics 1995, 54th issue. International Labour Organization, Geneva.
ILO (1995b) Sectoral activities programme, recent developments in the coalmining industry, coal mines committee, 13th session, Report I. International Labour Organization, Geneva.
ILO (1996) ILO Yearbook of labour statistics 1996, 55th issue. International Labour Organization, Geneva.

ILO (2000) Yearbook of labour statistics, 59th issue. International Labour Organization, Geneva.

ILO (2001) Laborsta, the labour statistics database. International Labour Organization. Available at http://laborsta.ilo.org/.

ILO (2002a) Economically active population 1950-2010, 4th edn, rev. 2. International Labour Organization, Geneva.
ILO (2002b) The International Labor Conference, Report VI. Decent work and the informal economy. International Labour Organization. Geneva.

IPCS (1998) Chrysotile asbestos. Environmental health criteria 203. International Programme on Chemical Safety, Geneva.
Jarup L, Nordberg G (1998) Cancer. In: Health effects of cadmium exposure: a review of the literature and a risk estimate. Jarup L, Berglund M, Elinder CG, Nordberg G, Vahter M, eds. Scandinavian Journal Work, Environment and Health, 24:34-36.
Jensen RC (1988) Epidemiology of work-related back pain. Topics Acute Care Trauma Rehabilitation, 23:1-15.
Jin K, Sorock GS, Courtney T et al (2000) Risk factors for work-related low back pain in the People's Republic of China. International Journal of Occupational and Environmental Health, 6:26-33.

Johanning E (1991) Back disorders and health problems among subway train operators exposed to whole-body vibration. Scandinavian Journal of Work Environment and Health, 17:414-419.
Johanning E, Wilder D, Landrigan P (1991) Whole-body vibration exposure in subway cars and review of adverse health effects. Journal of Occupational Medicine, 33:605-612.
Johnson AG, Whyte WF (1977) The Mondragon system of worker production cooperatives. Industrial and Labor Relations Review, 31:18-30.
Johnson D, Swift D (1997) Sampling and sizing particles. In: The occupational environment-its evaluation and control. Di Nardi SR, ed. The American Industrial Hygiene Association, Fairfax, VA.

Joshi TK, Menon KK, Kishore J (2001) Musculoskeletal disorders in industrial workers of Delhi. International Journal of Occupational and Environmental Health, 7:217-222.

Juengprasert W (1997) Towards elimination of silicosis in Thailand. Mineral dusts and prevention of silicosis. Asian-Pacific Regional Network on Occupational Safety and Health Information (ASIA-OSH), 4(2).
Kam Lam TF (2000) Occupational safety and health in China. Asian labour update. Available at http://www.amrc.org.hk/alu/Alu39/013906.htlm.
Kamal AA, Eldamati SE, Faris F (1989) Hearing threshold of Cairo traffic policemen. International Archives of Occupational and Environmental Health, 61:543-545.

Karjalainen A, Kurppa K, Martikanen R, Karjalainen J, Klaukka T (2002) Exploration of asthma risk by occupation -extended analysis of an incidence study of the Finnish population. Scandinavian Journal of Work, Environment and Health, 28:49-57.

Karjalainen A, Kurppa K, Martikanen R, Klaukka T, Karjalainen J (2001) Work is related to a substantial portion of adult-onset asthma incidence in the Finnish population. American Journal of Respiratory Critical Care Medicine, 164:565-568.

Kauppinen T, Toikkanen J, Pedersen D et al. (2000) Occupational exposure to carcinogens in the European Union. Occupational and Environmental Medicine, 57:10-18

Khogali M (1970) Industrial health problems of developing countries. 1.1. Types of risk and their prevention. Journal of Tropical Medicine and Hygiene, 73:269-274.

Klein BP, Jensen RC, Sanderson LM (1984). Assessment of workers' compensation claims for back strains/sprains. Journal of Occupational Medicine, 26:443-448.

Klockhoff I, Drettner B, Svedberg A (1974) Computerized classification of the results of screening audiometry in groups of persons exposed to noise. Audiology, 13:326-334.

Koch JL, Rhodes, SR (1981) Predictors of turnover of female factory workers. Journal of Vocational Behavior, 18:145-161.
Kogevinas M, Antó JM, Soriano JB, Tobias A (1996) The risk of asthma attributable to occupational exposures: a population-based study in Spain. American Journal of Respiratory Critical Care Medicine, 154:137-143.

Kogevinas M, Anto JM, Sunyer J, Tobias A, Kromhout H, Burney P (1999) Occupational asthma in Europe. The Lancet, 353:1750-1754.
Korn RJ, Dockery DW, Speizer FE, Ware JH, Ferris BG Jr (1987) Occupational exposures and chronic respiratory symptoms: a population based study. American Review of Respiratory Disease, 136:298-304.
Krzyzanowski M, Jedrychowski W, Wysocki M (1986) Factors associated with the change in ventilatory function and the development of chronic obstructive pulmonary disease in a 13 -years follow-up of the Cracow study. American Review of Respiratory Disease, 134:1011-1019.

Kumar A, Varghese M, Mohan D, Mahajan P, Gulati P, Kale S (1999) Effect of whole-body vibration on the low back. Spine, 24:2506-2515.

Kuwashima A, Aizawa Y, Nakamura K, Taniguchi S, Watanabe M (1997) National survey on accidental low back pain in workplace. Industrial Health, 35:187-193.

Lagerström M, Hansson T, Hagberg M (1998) Work-related low back problems in nursing. Scandinavian Journal of Work, Environment and Health, 24: 449-464.

Lahiri S (2002) Analysis of cost-effectiveness of interventions to reduce low back pain as a consequence of occupational ergonomic stressors. World Health Organization, Geneva.

Lawrence RC, Helmick CG, Arnett FC et al. (1998) Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis and Rheumatism, 41:778-799.

Lee P (2001) Relation between exposure to asbestos and smoking jointly and the risk of lung cancer. Occupational and Environmental Medicine, 58:145-153.
Leigh J, Driscoll T (2003) Malignant mesothelioma in Australia, 1945-2002. International Journal Occupational Environmental Health, 9:206-217.

Leigh JP, Sheetz RM (1989) Prevalence of back pain among full-time United States workers. British Journal of Industrial Medicine, 46:651-657.

Leino PI, Hänninen V (1995) Psychosocial factors at work in relation to back and limb disorders. Scandinavian Journal of Work, Environment and Health, 21:134-142.

Leino-Arjas P, Hanninen K, Puska P (1998) Socioeconomic variation in back and joint pain in Finland. European Journal of Epidemiology, 14:79-87.
Lerer LB, Myers JE (1994) Application of two secondary documentary sources to identify the underreporting of fatal occupational injuries in Cape Town, South Africa. American Journal of Industrial Medicine, 26:521-527.
Lipsett M, Campleman S (1999) Occupational exposure to diesel exhaust and lung cancer: a meta-analysis. American Journal of Public Health, 89:1009-1017.

Loewenson R (1998) Health impact of occupational risks in the informal sector in Zimbabwe. International Journal of Occupational and Environmental Health, 4:264-274.

Lombardo LJ, Balmes JR (2000) Occupational asthma: a review. Environmental Health Perspectives, 8:S697-704.
Loney P, Stratford P (1999) The prevalence of low back pain in adults: a methodological review of the literature. Physical Therapy, 79:384-396.
Lucifora C (1998) The impact of unions on labour turnover in Italy: evidence from establishment level data. International Journal of Industrial Organization, 16:353-376.
Lynge E, Anttila A, Hemminki K (1997) Organic solvents and cancer. Cancer Causes and Control, 8:406-419.
MacDonald LA, Karasek RA, Punnett L, Scharf T (2001) Covariation between workplace physical and psychosocial stressors: evidence and implications for occupational health research and prevention. Ergonomics, 44:696-718.

Magnusson ML, Pope MH, Wilder DG,Areskoug B (1996) Are occupational drivers at an increased risk for developing musculoskeletal disorders? Spine, 6:710-717.
Malchaire J (2000) Strategy for prevention and control of the risks due to noise. Occupational and Environmental Medicine, 57:361-369.
Malo J, Chan-Yeung (2001) Occupational asthma. Journal of Allergy and Clinical Immunology, 108:317-328.
Marras WS, Allread DL, Burr DL, Fathallah FA (2000). Prospective validation of low-back disorder risk model and assessment of ergonomic interventions associated with manual materials handling tasks. Ergonomics, 43: 1866-1886.

Marsh SM, Layne LA (2001) Fatal injuries to civilian workers in the United States, 1980-1995 (National profile). Department of Health and Human Services. National Institute for Occupational Safety and Health (DHHS/NIOSH) 2001-129. Available at http://www.cdc.gov/niosh/NTOF2000/2001129.htlm.

Martonik JF, Nash E, Grossman E (2001) The history of OSHA's asbestos rulemakings and some distinctive approaches that they introduced for regulating occupational exposure to toxic substances. American Industrial Hygiene Association Journal, 62:208-217.
Miedema HS, Chorus AMJ, Wevers CWJ, van der Linden S (1998) Chronicity of back problems during working life. Spine, 23:2021-2029.
Milton DK, Solomon DM, Roseillo RA, Herrick RF (1998) Risk and incidence of asthma attributable to occupational exposure among HMO members. American Journal of Industrial Medicine, 33:1-10.
Minino AM, Smith BL (2001) Deaths: preliminary data for 2000. National Vital Statistics Reports, 49:1-40.

Ministerio de Trabajo y Asuntos Sociales (2002) [Ministry of Labour and Social Affairs (2002)] Gabinete de Comunicación. Información Estadística. [Communications Bureau. Statistical Information.] Available at http://www.mtas.es/. Informe de Resultados España [Report of results for Spain].
Moffett M (2002) New Mexico labor market dynamics. Labor market review, special article. New Mexico Department of Labor, February. Available at http://www3.state.nm.us/dol/dol_lmrsa6.htm.
Monso E, Munoz-Rino F, Izquierdo J et al. (1998) Occupational asthma in the community: risk factors in a western Mediterranean population. Archives of Environmental Health, 53:93-98.

Morata TC, Fiorini AC, Fischer FM et al. (1997) Toluene-induced hearing loss among rotogravure printing workers. Scandinavian Journal of Work, Environment and Health, 23:289-298.

Morken T, Moen B, Riise T et al (2000) Prevalence of musculoskeletal symptoms among aluminum workers. Occupational Medicine, 50:141-421.
Moselhi M, El-Sadik YM, El-Dakhakhny A (1979) A six-year follow up study for evaluation of the 85 dBA safe criterion for noise exposure. American Industrial Hygiene Association Journal, 40:424-426.
Mukherjee A K, Nag DP, Kakde Y, Prakash MN, Rao SR (1995) Noise level monitoring in a watch factory in Bangalore. Indian Journal of Industrial Medicine, 41:42-44.
Muruka AO (1998) Age, height and duration in of service in relation to back pain among tea pickers in Kenya. African Newsletter for Occupational Health and Safety, 8(3). Accessed at http://www.occuphealth.fi/e/info/anl/398/ in May 2002.
Myers JE, Lewis P, Hofmeyr W (1989) Respiratory health of brickworkers in Cape Town, South Africa: background, aims, and dust exposure determinations. Scandinavian Journal of Work, Environment and Health, 15:180-187.
Nachemson A (1999) Back pain: delimiting the problem in the next millennium. International Journal of Law and Psychiatry, 22:473-490.
Nachemson A, Jonsson E, eds (2000) Neck and back pain: the scientific evidence of causes, diagnosis, and treatment. Lippincott, Williams and Wilkins, Philadelphia, PA.

National Research Council (2001) Musculoskeletal disorders and the workplace: low back and upper extremities. Panel on musculoskeletal disorders and the workplace. Commission on Behavioral and Social Sciences and Education, National Research Council and Institute of Medicine. National Academy Press, Washington, DC. (also accessed at http://www.nap.edu/openbook).
National Statistics Office Philippines (2000) Proceedings of the seminar on the ILO project on occupational injuries, 6 December. National Statistics Office, Manila.
Nelson DI (1997) Risk assessment in the workplace. In: The occupational envi-ronment-its evaluation and control. Di Nardi SR, ed. The American Industrial Hygiene Association, Fairfax, VA.

Ng TP, Hong CY, Goh LG, Wong ML, Koh KT, Ling SL (1994) Risks of asthma associated with occupations in a community-based case-control study. American Journal of Industrial Medicine, 25:709-718.
NIEHS (1999) United States. A world-class program. National Institute on Environmental Health Sciences. Environmental Health Perspectives, 107(7). Available at http://ehpnet1.niehs.nih.gov/docs/1999/107-7/niehsnews.html.
NIOSH (1991) NIOSH publications on noise and hearing. Perspectives in disease prevention and health promotion, leading work-related diseases and injuries-United States. Noise-induced Loss of Hearing, July. National Institute for Occupational Safety and Health, Washington, DC.

NIOSH (1996) Violence in the workplace: risk factors and prevention strategies. National Institute for Occupational Safety and Health (NIOSH-CDC), Cincinnati, OH. Available at http://www.cdc.gov/niosh/violpr.htlm.
NIOSH (1998) Criteria for a recommended standard: occupational noise exposure. Revised criteria 1998. National Institute for Occupational Safety and Health, Cincinnati, OH. Available at http://www.cdc.gov/niosh/98-126.html.
NIOSH (1999) Work-related lung disease surveillance report 1999. U.S. Department of Health and Human Services, National Institute for Occupational Safety and Health, Washington, DC.
NIOSH (2000a) Worker health chartbook, 2000. NIOSH-CDC. US Department of Health and Human Services, National Institute for Occupational Safety and Health, Washington, DC.

NIOSH (2000b) Injuries, illnesses, and hazardous exposures in the mining industry, 1986-1995: a surveillance report. (No. 2000-117.) National Institute for Occupational Safety and Health, Washington, DC.
NOHSC (1993) National code of practice for management and protection of hearing at work (NOHSC:2009.) National Occupational Health and Safety Commission, Australia. Available at http://www.nohsc.gov.au/OHSInforma tion/NOHSCPublications/fulltext/docs/h3/19.htm.
NOHSC (2001) National Occupational Health and Safety Commission, Australia. Available at http://www.nohsc.gov.au.

NOHSC (2002) Statistics. National Occupational Health and Safety Commission, Australia. Available at http://www.nohsc.gov.au/statistics/default.htm.

Noweir MH, El-Dakhakhny AA, Valic F (1968) Exposure to noise in the textile industry of the U.A.R. American Industrial Hygiene Association Journal, 29:541-546.

Nurminen M, Karjalainen A (2001) Epidemiologic estimate of the proportion of fatalities related to occupational factors in Finland. Scandinavian Journal Work, Environment and Health, 27:161-213.

Obiako MN (1979) Deafness and the mining industry in Zambia. East African Medical Journal, 56:445-449.
Oleru UG (1980) Comparison of the hearing levels of Nigerian textile workers and a control group. American Industrial Hygiene Association Journal, 41:283-287.

Oleru UG, Ijaduola GTA, Sowho EE (1990) Hearing thresholds in an auto assembly plant: prospects for hearing conservation in an [sic] Nigerian factory. International Archives of Occupational and Environmental Health, 62:199-202.

Omokhodion FO, Umar US, Ogunnowo BE (2000) Prevalence of low back pain among staff in a rural hospital in Nigeria. Occupational Medicine, 50:107-110.

Ory FG, Rahman FU, Katagade V, Shukla A, Burdorf A (1997) Respiratory disorders, skin complaints, and low-back trouble among tannery workers in Kanpur, India. American Industrial Hygiene Association Journal, 58: 740-746.

Osibogun A, Igweze IA, Adeniran LO (2000). Noise-induced hearing loss among textile workers in Lagos Metropolis. The Nigerian Postgraduate Medical Journal, 7:104-111.

Oxman A, Muir D, Shannon H, Stock S, Hnizdo E, Lange H (1993) Occupational dust exposure and chronic obstructive pulmonary disease. A systematic overview of the evidence. American Review of Respiratory Disease, 148:38-48.

Ozguler A, Leclerc A, Landre M, Pietri-Taleb F, Niedhammer I (2000) Individual and occupational determinants of low back pain according to various definitions of low back pain. Journal of Epidemiology and Community Health, 54:215-220.

PAHO/WHO (1998) Informe del proyecto sistematización de datos básicos sobre salud de los trabajadores en países de las Américas-Agosto [Report of the project: systematisation of basic data of health of workers in LAC countriesAugust]. División del Ambiente y Salud [Division for Environment and Health]. Pan American Health Organization/World Health Organization.
Palmer K, Pannett B, Griffin M (2001) Occupational exposure to noise and hearing difficulties in Great Britain. (Contract Report 361/2001.) University of Southampton for the Health and Safety Executive, Southhampton.

Paquet VL, Punnett L, Buchholz BO (1999) An evaluation of manual materials handling in highway construction work. International Journal of Industrial Ergonomics, 24:431-444.

Partanen T, Jaakkola J, Tossavainen A (1995) Silica, silicosis and cancer in Finland. Scandinavian Journal of Work, Environment and Health, 2:S84-86.

Partridge REH, Duthie JJR (1968) Rheumatism in dockers and civil servants: a comparison of heavy manual and sedentary workers. Annals of Rheumatic Diseases, 27:559-568.

Pearce N, Matos E, Vainio H, Boffetta P, Kogevinas M (1994) Occupational cancer in developing countries. (IARC Scientific Publications No. 129) Lyon.
Perkio-Makela M, Riihimäki H (1997) Intervention on seat adjustment among drivers of forest tractors. International Journal of Industrial Ergonomics, 19:231-237.

Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R (2000). Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. British Medical Journal, 321: 323-329.

Peto J, Decarli A, La Vecchia C, Levi F, Negri E (1999) The European mesothelioma epidemic. British Journal of Cancer, 79:666-672.

Phan Hong Son, Tran Ngoc Lan, Le Van Trung et al. (1999) Use geographical information system to identify risk for silicosis and planning strategies for prevention, July, 1999. Funded by the Fogarty International Center, NIH. Washington, DC.
Pope MH, Andersson GBJ, Frymoyer JW, Chaffin DB, eds (1991) Occupational low back pain: assessment, treatment and prevention. Mosby-Year Book, Inc., St. Louis, MO.

Price B, Ware A (2004) Mesothelioma trends in the United States: an update based on surveillance, epidemiology, and End Results program. Data for 1973 through 2003. American Journal of Epidemiology, 159:107-112.

Prince MM, Stayner LT, Smith RJ, Gilbert S J (1997) A re-examination of risk estimates from the NIOSH Occupational Noise and Hearing Survey (ONHS) . Journal of the Acoustic Society of America, 101:950-963.
Punnett L (1998) Ergonomic stressors and upper extremity disorders in vehicle manufacturing: cross-sectional exposure-response trends. Occupational and Environmental Medicine, 55:414-420.
Punnett L (1999) The costs of work-related musculoskeletal disorders in automotive manufacturing. New Solutions: A Journal of Environmental and Occupational Health Policy, 9:403-426.
Punnett L (2002) Ergonomics and public health. Oxford textbook of public health. 4th edn. Oxford University Press, Oxford.

Punnett L, Fine LJ, Keyserling WM, Herrin GD, Chaffin DB (1991) Back disorders and non-neutral trunk postures of automobile assembly workers. Scandinavian Journal of Work, Environment and Health, 17:337-346.
Punnett L, Herbert R (2000) Work-related musculoskeletal disorders: is there a gender differential, and if so, what does it mean? In: Women and health. Goldman MB, Hatch MC, eds. Academic Press. San Diego, CA.

Quinlan M (2002) Workplace health and safety effects of precarious employment. In the Global Occupational Health Network Newsletter, Issue No. 2. World Health Organization, Geneva.

Raja S, Ganguly G (1983) Impact of exposure to noise on the hearing acuity of employees in a heavy engineering industry. Indian Journal of Medical Research, 78:100-113.
Rees D, Cronje R, du Toit RSJ (1992) Dust exposure and pneumoconiosis in a South African pottery. 1. Study objectives and dust exposure. British Journal of Industrial Medicine, 49:459-464.

Riihimäki H (1991) Low-back pain, its origin and risk indicators. Scandinavian Journal of Work, Environment and Health, 17:81-90.

Riihimäki H (1995a) Back and limb disorders. In: Epidemiology of work related diseases. McDonald C, ed. BMJ Publishing Group, London.

Riihimäki H (1995b) Hands up or back to work- future challenges in epidemiologic research on musculoskeletal diseases. Scandinavian Journal of Work, Environment and Health, 21:401-403.

Riihimäki H, Mattsson T, Zitting AJ et al. (1990) Radiographically detectable degenerative changes of the lumbar spine among concrete reinforcement workers and house painters. Spine, 15:114-119.
Riihimäki H, Tola S, Videman T, Hänninen K (1989) Low-back pain and occupation: a cross-sectional questionnaire study of men in machine operating, dynamic physical work, and sedentary work. Spine, 14:204-209.
Riihimäki H, Viikari-Juntura E, Moneta G, Kuha J, Videman T, Tola S (1994) Incidence of sciatic pain among men in machine operating, dynamic physical work and sedentary work. Spine, 19:138-142.

Roach S (1992) Health risks from hazardous substances at work—assessment, evaluation and control. Pergamon Press, New York.
Schierhout GH, Midgley A, Myers JE (1997) Occupational fatality underreporting in rural areas of the Western Cape Province, South Africa. Safety Science, 25:113-122.
Sears MR (1997) Descriptive epidemiology of asthma. The Lancet, 350:S1-4.
Shaikh GH (1996) Noise problem in a polyester fiber plant in Pakistan. Industrial Health, 34:427-431.
Shannon HS, Mayr J, Haines T (1997) Overview of the relationship between organizational and workplace factors and injury rates. Safety Science, 26:201-17.

Shannon HS, Walters V, Lewchuk W et al. (1996) Workplace organizational correlates of lost-time accident rates in manufacturing. American Journal of Industrial Medicine, 29:258-268.

Silverstein BA, Fine LJ, Armstrong TJ (1986) Hand wrist cumulative trauma disorders in industry. British Journal of Industrial Medicine, 43:779-784.

Silverstein BA, Fine LJ, Stetson DS (1987) Hand-wrist disorders among investment casting plant workers. The Journal of Hand Surgery, 12A:838-844.

Silverstein BA, Stetson DS, Keyserling WM, Fine LJ (1997) Work-related musculoskeletal disorders: comparison of data sources for surveillance. American Journal of Industrial Medicine, 31:600-608.
Silverstein BA, Viikari-Juntura E, Kalat J (2002) Use of a prevention index to identify industries at high risk for work-related musculoskeletal disorders of the neck, back, and upper extremity in Washington state, 1990-1998. American Journal of Industrial Medicine, 41:149-169.
Simpson M, Bruce R (1981) Noise in America: the extent of the noise problem. (EPA Report No. 550/9-81-101). Environmental Protection Agency, Washington, DC.

Singapore Government (2000) Statistics Singapore-statistical resources-statistical classifications-Singapore standard classification (SSOC), online portal available at http://www.gov.sg.

Singh N, Davis GS (2002) Review: occupational and environmental lung disease. Current Opinions on Pulmonary Medicine, 8:117-125.

Skovron ML, Szpalski M, Nordin M, Melot C, Cukier D (1994) Sociocultural factors and back pain: A population-based study in Belgian adults. Spine, 19:129-137.

Sluiter JK, Rest KM, Frings-Dresen MHW (2000) Criteria document for evaluation of the work-relatedness of upper extremity musculoskeletal disorders. SALTSA Joint Programme for Working Life Research in Europe and Academic Medical Center, University of Amsterdam, Amsterdam.
Smedley J, Egger P, Cooper C, Coggon D (1995) Manual handling activities and risk of low back pain in nurses. Occupational and Environmental Medicine, 52:160-3

Sobti A, Cooper C, Inskip H et al. (1997) Occupational physical activity and long-term risk of musculoskeletal symptoms: a national survey of post office pensioners. American Journal of Industrial Medicine, 32:76-83.
Soutar C, Robertson A, Miller B, Searl A, Bignon J (2000) Epidemiological evidence on the carcinogenicity of silica: factors in scientific judgement. Annals of Occupational Hygiene, 44:3-14.
Sriwattanatamma P, Breysse P (2000) Comparison of NIOSH noise criteria and OSHA hearing conservation criteria. American Journal of Industrial Medicine, 37:334-338.

Stayner L, Smith R, Thun M, Schnorr T, Lemen R (1992) A dose response analysis and quantitative assessment of lung cancer risk and occupational cadmium exposure. Annals of Epidemiology, 2:177-194.
Steenland K, Beaumont J, Elliot L (1991a) Lung cancer in mild steel welders. American Journal of Epidemiology, 133:220-229.
Steenland K, Burnett C, Ward E, Lalich N, Hurrell J (2003) Dying for work: the magnitude of US mortality from selected causes of death associated with occupation. American Journal of Industrial Medicine, 43:461-482.
Steenland K, Loomis D, Shy C, Simonsen N (1996) Review of occupational lung carcinogens. American Journal of Industrial Medicine, 29:474-490.

Steenland K, Sanderson W (2001) Lung cancer among industrial sand workers exposed to crystalline silica. American Journal of Epidemiology, 53:695-703.
Steenland K, Sanderson W, Calvert GM (2001b) Kidney disease and arthritis in a cohort study of workers exposed to silica. Epidemiology, 12:405-412.

Steenland K, Stayner L, Greife A et al. (1991b) Mortality among workers exposed to ethylene oxide. New England Journal of Medicine, 324: 1402-1407.

Steenland K, Ward E (1991) Lung cancer incidence among patients with beryllium disease: A cohort mortality study. Journal of the National Cancer Institute, 83:1380-1385.
Suter A (2000) Standards and regulations. In: The noise manual, 5th edn, Berger EH, Royster LH, Rozster JD, Driscoll DP, Layne M, eds. American Industrial Hygiene Association, Fairfax, VA.

Tay P (1996) Severe noise-induced deafness: a 10 -year review of cases. Singapore Medical Journal, 37:362-364.

Toren K (1996) Self-reported rate of occupational asthma in Sweden 1990-1992. Occupational and Environmental Medicine, 53:757-761.

Toren K, Balder B, Brisman J et al. (1999) This risk of asthma. European Respiratory Journal, 13:496-501.
Toroptsova NV, Benevolenskaya LI, Karyakin AN, Sergeev IL, Erdesz S (1995) "Cross/sectional" study of low backpain among workers at an industrial enterprise in Russia. Spine, 20:328-332.
Tsai SP, Gilstrap EL, Cowles SR, Waddell LC, Ross CE (1992) Personal and job characteristics of musculoskeletal injuries in an industrial population. Journal of Occupational Medicine, 34:606-612.
UN (2000) International standard industrial classification of all economic activities (ISIC). Third revision. United Nations publication (St/ESA/STAT/SER.M/4/Rev.3). United Nations, New York.

UN (2001)World population prospects. The 2000 revision highlights. Population Division. Department of Economic and Social Affairs, United Nations, New York.

USDHHS (1986) Perspectives in disease prevention and health promotion, leading work-related diseases and injuries-United States. Morbidity and Mortality Weekly Report, 35(12). (Noise-induced Loss of Hearing. Reprinted in USDHHS 1991, NIOSH Publications on Noise and Hearing.)
USDOL OSHA (2000) Docket for the Federal Register. (Vol. 65, No. 220.) U.S. Department of Labor, Occupational Safety and Health Administration, Washington, DC.

USDOL OSHA (2002a) Permissible exposure limits, codified at 29 CFR 1910.1000. U.S. Department of Labor, Occupational Safety and Health Administration. Available at http://www.osha.gov/SLTC/pel/index.html.
USDOL OSHA (2002b) Noise and hearing conservation. U.S. Department of Labor, Occupational Safety and Health Administration. Available at http://www.osha-slc.gov/SLTC/noisehearingconservation/index.html

USEIA (2001) U.S. Energy Information Administration, U.S. Department of Energy, International Energy Database. Accessed January 2001 at www.eia.doe.gov/emeu/iea/coal.html.
Vaaranen V, Vasama M, Toikkanen J et al. (1994) Occupational diseases in Finland 1993. Institute of Occupational Health, Helsinki.

Venables KM, Chang-Yeung M (1997) Occupational asthma. The Lancet, 349:1465-1469.

Videman T, Nurminen M, Troup JDG (1990) Lumbar spine pathology in cadaveric material in relation to history of back pain, occupation, and physical loading. Spine, 15:728-740.
Viikari-Juntura E (1997) The scientific basis for making guidelines and standards to prevent work-related musculoskeletal disorders. Ergonomics, 40:1097-1117.

Viikari-Juntura E, Riihimäki H (1999) New avenues in research on musculoskeletal disorders. Scandinavian Journal of Work, Environment and Health, 25:564-568.

Volinn E (1997) The epidemiology of low back pain in the rest of the world. A review of surveys in low-and middle-income countries. Spine, 22:1747-1754.

Volinn E, Punnett L (2001) Point of view. Spine, 26:1902-1903.
Volinn E, Spratt KF, Magnusson M, Pope MH (2001) The Boeing prospective study and beyond. Spine, 26:1613-1622.

Waddell G (1991) Low back disability: a syndrome of Western civilization. Neurosurgery Clinics of North America, 2:719-738.
Wagner GR, Wegman DH (1998) Occupational asthma: prevention by definition. American Journal of Industrial Medicine, 33:427-429.
Waitzman N, Smith K (1999) Unsound conditions: work-related hearing loss in construction, 1960-75. The Center to Protect Worker's Rights. CPWR Publications, Washington, DC.

Ward E, Okan A, Ruder A, Fingerhut M, Steenland K (1992) A mortality study of workers at seven beryllium processing plants. American Journal of Industrial Medicine, 22:885-904.

Westgaard RH, Jansen T (1992) Individual and work related factors associated with symptoms of musculoskeletal complaints: I. A quantitative registration system. British Journal of Industrial Medicine, 49:147-153.
Westgaard RH, Winkel J (1997) Ergonomic intervention research for improved musculoskeletal health: a critical review. International Journal of Industrial Ergonomics, 20:463-500.
WHO (1999) International statistical classification of diseases and related health problems (ICD-10) in occupational health. Karjalainen A, ed. Protection of the Human Environment, Occupational and Environmental Health Series, (WHO/SDE/OEH/99.11). World Health Organization, Geneva.

WHO (2001) Electromagnetic fields and public health: extremely low frequency fields and cancer. (Fact Sheet No. 263). World Health Organization, Geneva. Available at http://www.who.int/inf-fs/en/fact263.html.

WHO/FIOSH (2001) Occupational exposure to noise: evaluation, prevention and control. Goelzer B, Hansen CH, Sehrndt GA, eds. On behalf of WHO by the Federal Institute for Occupational Safety and Health (FIOSH), Dortmund.

Wikstrom B-O, Kjellberg A, Landstrom U (1994) Health effects of long-term occupational exposure to whole-body vibration: a review. International Journal of Industrial Ergonomics, 14:273-292.
World Bank (2001) World development indicators 2001. Available at http://worldbank.com.
Xia Z, Courtney TK, Sorock GS, Zhu J, Fu H, Liang Y, Christiani DC (2000) Fatal occupational injuries in a new development area in the People's Republic of China. Occupational and Environmental Medicine, 42: 917-922.

Xu X, Christiani DC, Dockery DW et al. (1992) Exposure-response relationships between occupational exposures and chronic respiratory illness: a commu-nity-based study. American Review of Respiratory Disease, 146:413-418.

Yin SN, Li Q, Liu Y, Tian F, Du C, Jin C (1987) Occupational exposure to benzene in China. British Journal of Industrial Medicine, 44:192-195.
Yip YB (2001) A study of work stress, patient handling activities and the risk of low back pain among nurses in Hong Kong. Journal of Advanced Nursing, 36:794-804.

Zou Changqi, Gao Yun, Ma Qingyan (1997) Pneumoconiosis in China: current situation and countermeasures. Mineral dusts and prevention of silicosis, 4(2) September 1997. Asian-Pacific Regional Network on Occupational Safety and Health Information (ASIA-OSH). Available at http://www.ilo.org/public/english/region/asro/.

## Chapter 22

# Contaminated injections in 

## HEALTH CARE SETTINGS

Anja M. Hauri, Gregory L. Armstrong and Yvan J.F. Hutin

## Summary

Injections given in health care settings with injection equipment reused in the absence of sterilization have been associated with infection with hepatitis B virus (HBV), hepatitis C virus (HCV) and HIV.

Input parameters included the annual number of injections per person, the proportion of injections administered with equipment reused in the absence of sterilization, the probability of transmission following percutaneous exposure, the age-specific prevalence of active infection, the prevalence of immunity (i.e. antibody to the hepatitis B core antigen or HbcAg [anti-HBc], anti-HCV and anti-HIV) and the incidence of HBV, HCV and HIV infections. We used mathematical models to transform diverse sources of data available into the prevalence of contaminated injections and the relative risk associated with these practices.

Four subregions ${ }^{1}$ (AMR-A, EMR-B, EUR-A and WPR-A) where reuse of injection equipment in the absence of sterilization was negligible were assumed to have zero risk. In the remaining 10 subregions, the annual number of injections per person ranged from 1.9 to 11.3 and the proportion of injections administered with reused equipment ranged from $1.2 \%$ to $75 \%$.

In 10 subregions, in 2000, injections caused an estimated 21 million HBV infections, two million HCV infections and 260000 HIV infections, accounting for $32 \%, 40 \%$ and $5 \%$ of new infections, respectively. Thus, the burden in 2000 due to past and present exposure accounted for 501000 deaths and 10461000 disability-adjusted life years (DALYs).

Injection overuse and unsafe practices are common worldwide and account for a high burden of infections with bloodborne pathogens. There is a need for policies and programmes for the safe and appropriate use of injections in countries where poor practices occur.

## 1. Introduction

During the twentieth century, injection use increased tremendously and today injections are probably the most common health care procedure (Drucker et al. 2001). Poor injection practices, including injection overuse and unsafe practices, have been reported in many developing and transitional countries (Simonsen et al. 1999). Many injections given for curative purposes in developing and transitional countries are unnecessary as they are prescribed for the treatment of conditions that could be treated with oral drugs or for which medications are not needed (Reeler 1990; Simonsen et al. 1999). In addition to being unnecessary, many injections are unsafe. Of particular concern is the reuse of injection equipment in the absence of sterilization. A common practice consists of rinsing injection equipment between injections in a pot of tepid water (Figure 22.1).

Unsafe injection practices constitute an important route of infection for bloodborne pathogens. Recently, a study suggested that the spread of HCV through unsafe injections in Egypt may represent the largest nosocomial outbreak ever reported (Frank et al. 2000). Epidemiological studies have reported an association between contaminated injections and infection with bloodborne pathogens, including HBV, HCV and HIV

Figure 22.I Injection equipment soaked in tepid water before reuse in the absence of sterilization, Africa, 2000


[^85](Simonsen et al. 1999). The causal nature of this association is supported by many criteria. First, transmission through unsafe injection practices is biologically plausible because all three viruses are present in blood and body fluids of infected individuals (Choo et al. 1989; Molina et al. 1994; Shikata et al. 1977). They can be transmitted by transfusion (Aach et al. 1991; Busch et al. 1996; Senior et al. 1974) and other percutaneous routes, including needle-stick injuries among health care workers (Cardo et al. 1997; CDC 1997; Seeff et al. 1978). Second, several studies in developing countries have demonstrated an association between receiving injections and infection with bloodborne pathogens. Third, the measures of association (e.g. odds ratios) often exceed 2 (Luby et al. 1997; Narendranathan and Philip 1993; Quigley et al. 2000) and show a dose-response relationship (Khan et al. 2000; Ko et al. 1991a; Quigley et al. 2000). Fourth, studies have also reported an association between recent, incident cases of infection with HBV (Hutin et al. 1999), HCV (El-Sakka 1997) and HIV (Quigley et al. 2000), and exposure to injections during the time period that patients were likely to have been infected, indicating that the exposure preceded the outcome.

The proportion of new infections with HBV, HCV and HIV attributable to unsafe injection practices in specific populations can be estimated from case-control and cohort studies. A total of 12 studies (Table 22.1) were identified to examine the association between HBV infection and injections, with population attributable fractions ranging between $21 \%$ and $61 \%$ (Anonymous 1998; Hsu et al. 1993; Hussain 2001; Hutin et al. 1999; Ko and Chung 1991; Ko et al. 1991a; Luby et al. 1997; Narendranathan and Philip 1993; Simard et al. 2000; Singh et al. 2000; Thuring et al. 1993; Val Mayans et al. 1990). Of these eight ( $67 \%$ ) were based upon recent, incident cases. A total of 10 studies (Table 22.2) were identified to examine the association between HCV infection and injections, with population attributable fractions ranging between $20 \%$ and 84\% (Chang et al. 1996; Chen et al. 1995; El-Sakka 1997; Ho et al. 1997; Khan et al. 2000; Luby et al. 1997; Mohamed et al. 1996; Sun et al. 1999, 2001; Thuring et al. 1993). Of these, three were based upon recent, incident cases. A total of four studies (Table 22.3) based upon recent, incident cases were identified to examine the association between HIV infection and injections, with population attributable fractions ranging between $8 \%$ and $45 \%$ (Bultreys et al. 1994; N'Galy et al. 1988; Quigley et al. 2000; Wawer et al. 1994). (Studies based upon prevalent cases of HIV infection are not included in this report as the high frequency of HIV transmission through sexual exposure raises the possibility of reverse causation.)

Two limitations were common among the studies of the association between injections and infections. A first limitation was that studies of persons with prevalent, chronic infections are generally unable to distinguish the direction of the causal relationship between injections and infection. While study subjects could have acquired infections because
Table 22.I Studies examining the association between health care injections and HBV infection

| Country or area | Author(s) | Year of study | Study design | Types of cases | Attributable fraction (\%) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Cambodia | Thuring et al. (1993) | 1990-1991 | Survey | Prevalent | 2-13 |
| China (Province of Taiwan) | Hsu et al. (1993) | 1994 | Case-control | Prevalent | 24.6 |
| China (Province of Taiwan) | Ko and Chung (1991) | 1984-1989 | Cohort | Incident | 43.1 |
| China (Province of Taiwan) | Ko et al. (1991a) | $1991{ }^{\text {a }}$ | Cohort | Incident | 73.9 |
| Gambia | Val Mayans et al. (1990) | 1988 | Cohort | Incident | * |
| Egypt | Anonymous (1998) | 1994 | Case-control | Incident | 27.7 |
| India | Narendranathan and Philip (1993) | $1993{ }^{\text {a }}$ | Case-control | Incident | 53.3 |
| India | Singh et al. (2000) | 1998 | Case-control | Incident | 49.7 |
| Pakistan | Hussain (2001) | 2000-2001 | Case-control | Prevalent | 52 |
| Pakistan | Luby et al. (1997) | 1994 | Case-control | Prevalent | 35-41 |
| Romania | Hutin et al. (1999) | 1998 | Case-control | Incident | 40 |
| Republic of Moldova | Hutin et al. (1999) | 1994-1995 | Case-control | Incident | $21,52^{\text {b }}$ |
| * No association found. However, only immunization injections were consid <br> a Year of publication. <br> b $21 \%$ among children, $52 \%$ among adults. |  |  |  |  |  |

Table 22.2 Studies examining the association between health care injections and HCV infection

| Country or area | Author(s) | Year of study | Study design | Types of cases | Attributable fraction (\%) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Cambodia | Thuring et al. (1993) | 1990-1991 | Survey | Prevalent | 90.6 |
| China (Province of Taiwan) | Chang et al. (1996) | 1991 | Survey | Prevalent | 50.4 |
| China (Province of Taiwan) | Chen et al. (1995) | 1990-1994 | Case-control | Incident | 20.1 |
| China (Province of Taiwan) | Ho et al. (1997) | 1993 | Case-control | Prevalent | 51-88 |
| China (Province of Taiwan) | Sun et al. (1999) | 1992 | Case-control | Prevalent | 44 |
| China (Province of Taiwan) | Sun et al. (2001) | 1994 | Case-control | Incident | 36.4 |
| Egypt | El-Sakka (1997) | 1996-1997 | Case-control | Incident | 87.9 |
| Egypt | Mohamed et al. (1996) | 1996 | Survey | Prevalent | 9.9 |
| Pakistan | Khan et al. (2000) | 1995 | Case-control | Prevalent | 24.4-78.5 |
| Pakistan | Luby et al. (1997) | 1994 | Case-control | Prevalent | $1.4,62.9^{\text {a }}$ |
| a $1.4 \%$ for injections received during past year; $62.9 \%$ for injections received during the past 10 years. |  |  |  |  |  |

Table 22.3 Studies examining the association between health care injections and HIV infection ${ }^{\text {a }}$

| Country | Author(s) | Year of study | Study design | Types of <br> cases | Attributable <br> fraction (\%) |
| :--- | :--- | :---: | :--- | :--- | :---: |
| Democratic <br> Republic of <br> the Congo | N'Galy et al. (1988) | $1984-1986$ | Cohort | Incident | 28 |
| Rwanda | Bultreys et al. (1994) | $1989-1993$ | Cohort | Incident | 45 |
| Uganda | Quigley et al. (2000) | $1990-1997$ | Case-control | Incident | $16,41^{\text {b }}$ |
| Uganda | Wawer et al. (1994) | $1989-1990$ | Cohort | Incident | 8 |

${ }^{\text {a }}$ Restricted to studies recruiting recent, incident cases of HIV infection.
b $16 \%$ among women, $41 \%$ among men.
they received injections, they could also have received injections as a result of complications of their infection. Studies examining risk factors for HCV and HIV infections are more often affected by this bias because recent, acute cases of infection with these two pathogens are difficult to identify. However, three elements suggest that reverse causation is unlikely. First, most case patients in these studies were asymptomatic and therefore unlikely to seek injections for treatment of their infection. Second, a study that included prevalent cases of infection and examined the association between injections received during different time periods reported that injections received in a distant past were more strongly associated with infection than those received in a recent past, precisely the opposite of what could be expected if the hypothesis of reverse causation were true (Luby et al. 1997). Third, studies that included incident, recent cases of infections have reported similar associations (Chen et al. 1995; El-Sakka 1997; Sun et al. 2001). This includes a prospective cohort study examining the risk factors for HCV infection that validated the results of a cross-sectional survey conducted in the same population (Sun et al. 2001).

A second limitation was that the association between injections and infections with bloodborne pathogens may have been confounded by a number of other exposures, including sexually transmitted infections (STIs). In some cases, the apparent association between injections and infection may be secondary to two hidden associations-between STIs and injections on the one hand, and between STIs and infection on the other. However, confounding is unlikely to explain the associations observed because most studies also examined risk factors other than injections, including STIs, and a number of studies used stratification and multivariate analysis to control for these potential confounders. Nonetheless, there is still a need for research to determine the degree to
which STIs and injections confound each other's relationship with bloodborne pathogens, particularly in the case of HIV infection.

Because information from epidemiological studies was too limited to permit estimation of the global burden of disease attributable to unsafe injections, a global mass action mathematical model was generated in 1995 (Aylward et al. 1995) and further developed to formulate regional estimates in 1999 (Kane et al. 1999). This model included input parameters reflecting injection frequency, injection safety, the percutaneous transmission potential of bloodborne pathogens and the epidemiology of infection with HBV, HCV and HIV. Results of this analysis suggested that each year, in the world, reuse of injection equipment in the absence of sterilization accounts for 8 to 16 million HBV infections, 2.3 to 4.7 million HCV infections and 80000 to 160000 HIV infections (Kane et al. 1999).

This mass action model had three main limitations. First, it did not address variations of input parameters (i.e. injection frequency, prevalence of immunity and incidence of HIV infection) across age and sex groups within subregions. Second, no systematic procedure was used to review the literature and generate subregional estimates for injection frequency and injection safety. In this work, we used a new mathematical model to estimate the global burden of disease from unsafe injection practices, which although based on the same general approach as Kane et al. (1999) improves on some of the data limitations. In our analysis, we considered only HBV, HCV and HIV infections because of the substantial information on their association with unsafe injections and because these pathogens probably account for the majority of injectionassociated infections. Other complications of unsafe injections not included in this model include abscesses (Fontaine et al. 1984; Soeters and Aus 1989), septicaemia (Archibald et al. 1998), malaria (Abulrahi et al. 1997) and infection with viral haemorrhagic fever viruses (FisherHoch et al. 1995; WHO 1976).

## 2. Methods

### 2.1 Definitions

## Health care injection

We defined a health care injection as a procedure that introduces a substance into the body through a piercing of the skin or of a mucosal membrane, including intradermal, subcutaneous, intramuscular and intravenous injections, for curative or preventive health care purposes, whether administered in a formal health care setting (e.g. clinic, hospital) or other settings (e.g. homes, pharmacies). Injections of illicit drugs were not considered in this work (see chapter 13).

## Reuse of injection equipment in the absence of sterilization

We defined reuse of injection equipment as the administration of an injection to a recipient with a syringe or needle that had been previously used on another person and that was reused in the absence of sterilization. In this chapter, reuse of injection equipment in the absence of sterilization will simply be referred to as "reuse of injection equipment".

## Choice of exposure variable contaminated injections

Reuse of injection equipment in itself would not be a risk factor in the absence of source patients infected with bloodborne pathogens. Thus, contaminated injections were the risk factor of interest. An injection contaminated with a bloodborne pathogen was defined as an injection given with a needle or a syringe used on an infected patient and reused on a second patient. The exposure under consideration for this study was defined as receiving at least one injection contaminated with HBV, HCV or HIV in one year. Exposure status would therefore depend on reuse of equipment, injection frequency and prevalence of active infection with HBV, HCV and HIV in the population. Persons receiving no contaminated injection in one year were considered unexposed. Four subregions (AMR-A, EMR-B, EUR-A and WPR-A) where reuse of injection equipment in the absence of sterilization was negligible were assumed to have zero risk.

## Theoretical minimum level of EXposure

The theoretical minimum level of exposure was zero contaminated injections per person and per year. This theoretical minimum is also an achievable goal as there are no reports of reuse of injection equipment in many industrialized countries.

### 2.2 Transmission model

Data on the risk associated with contaminated injections are generally not available as relative risks, especially since these can change from one place or time to another due to changes in background prevalence. Instead, information from diverse sources such as case-control studies, cross-sectional studies and observational studies of injection practices were brought together and integrated by means of mathematical models to develop internally consistent estimates of prevalence and hazard. The hazard estimates were based on the mass action principle, which states that

$$
I_{u}=p_{s}\left[1-\left(1-p_{t} p_{r} p_{v}\right)^{n}\right]
$$

where $p_{s}$ is the proportion of the population susceptible to infection (in most cases, 1 minus prevalence of antibody to the virus), $p_{t}$ is the probability of transmission after percutaneous exposure to a particular
pathogen, $p_{r}$ is the probability that injection equipment will have been reused, $p_{v}$ is the prevalence of active infection and $n$ is the annual number of injections per person. This model implicitly assumes that the whole population is equally likely to be currently infected or receive an injection. For HBV, HCV and HIV, the three pathogens under consideration, this incidence is small enough that the equation can be simplified to

$$
I_{u}=p_{s} \times p_{t} \times p_{r} \times p_{v} \times n
$$

which can be further reduced to

$$
I_{u}=p_{s} \times p_{t} \times n_{c}
$$

in which $n_{c}$ is the average annual number of contaminated injections and

$$
n_{c}=p_{r} \times p_{v} \times n
$$

All parameters were assumed to be different for each of the three pathogens except the annual number of injections per person ( $n$ ), which was assumed to be constant within a particular age, sex and subregional stratum and the probability of reuse of injection equipment $\left(p_{r}\right)$, which was assumed to be constant within a particular subregion. The probability of transmission $\left(p_{t}\right)$ was based upon studies estimating the risk of infection with HBV, HCV and HIV following a needle-stick exposure from an infected patient. For HBV, $p_{t}$ was assumed to vary according to the proportion of the infected population that was negative for hepatitis B e-antigen ( HBeAg ), ( $p_{t}=0.06$ ), or HbeAg positive ( $p_{t}=0.3$ ), (Seeff et al. 1978). For HCV, $p_{t}$ was assumed to be 0.018 (CDC 1997). For HIV, the generally accepted value of $p_{t}$ of 0.003 (Cardo et al. 1997) for needle-stick injuries was too low, since most injuries on which this estimate was based were superficial and did not involve hollow-bore needles. At the same time, the estimated risk from a deep needle-stick injury that can be estimated from the same study, 0.021 , was too high (it is higher than the estimated $p_{t}$ for HCV) because time can elapse during which HIV can be inactivated between the initial use and the reuse of a syringe on a second patient. As a compromise, the mean of the estimates for superficial and deep injuries, 0.012 , was used in the model as $p_{t}$ for HIV.

### 2.3 Estimates of the proportion of the population exposed to contaminated injections from THE MASS ACTION MODEL

If $n_{c}<1$ and each person in the population could receive only one injection, then the probability of receiving a contaminated injection, $p_{c}$, would equal $n_{c}$. However, it is possible for someone to receive two, three or more contaminated injections in any given year. Because contaminated
injections are small probability events (Table 22.6), it can be assumed that the number of contaminated injections per individual follows a Poisson distribution in the population with a mean of $n_{c}$ per individual, then the probability of receiving zero injections would be exp $\left(-n_{c}\right)$, and the probability of receiving at least one injection would be

$$
p_{c}=1-\exp \left(-n_{c}\right)
$$

Thus when $n_{c}$ is very small, $p_{c}$ is approximately equal to $n_{c}$ and each exposed person will receive on average only one contaminated injection per year, as noted above. In most other situations, $p_{c}$ will be slightly smaller than $n_{c}$ and each exposed person will receive on average $n_{c} /\left(1-\exp \left(-n_{c}\right)\right)$ contaminated injections per year.

## $2.4 \quad$ Estimates of the relative risk from the mass ACTION MODEL

For the purposes of the model, we considered the total incidence of infection in the population, $I_{t}$, to be composed of two components: the incidence due to contaminated injections, $I_{u}$, and the baseline incidence, $I_{b}$, which can also be thought of as the incidence in the population if contaminated injections could be eliminated. $I_{t}$ can be estimated from incidence or prevalence surveys and $I_{b}$ can be estimated if $I_{t}$ and $I_{u}$ are known:

$$
I_{b}=I_{t}-I_{u}
$$

and the proportion of infections attributable to unsafe injections is

$$
A F=I_{u} / I_{t}
$$

As this proportion of infections would have occurred only among the exposed proportion of the population $\left(p_{c}\right)$, the risk among the exposed relative to the unexposed, by back-calculation from attributable fraction (AF) relationship, would be:

$$
R R_{c}=1+A F /\left(p_{c} \times(1-A F)\right)
$$

In most situations where the necessary variables are available or can be estimated from existing data, this equation can estimate the relative risk. However, in situations where a substantial proportion of infections are attributable to contaminated injections (i.e. situations where $I_{u}$ approaches $I_{t}$ ), this equation produces unstable estimates of the relative risk, and other methods were used, as below.

### 2.5 Estimates of relative risks from analytical EPIDEMIOLOGICAL STUDIES

Cohort and case-control studies that examined the association between injections and infection defined exposure as receiving at least one injection, contaminated or not, and the absence of exposure as receiving no injections. If $R R_{i}$ is the estimate of relative risk from such a study and $n_{i}$ is the average number of injections received by the cases, then $\left[1+\left(R R_{i}-1\right) / n_{i}\right]$ is the relative risk attributable to one injection. Only a portion ( $p_{r} \times p_{v}$ ) of the injections received are contaminated and persons who do not receive contaminated injections are at no increased risk. Therefore, the relative risk of infection in a person who receives only one contaminated injection is $\left[1+\left(R R_{i}-1\right) /\left(n_{i} \times p_{r} \times p_{v}\right)\right]$. In practice, this often underestimates the relative risk because persons who receive injections are more likely to have been infected in the past and are therefore less likely to be susceptible to infection. Because of this phenomenon, the calculated $R R_{i}$ will be an underestimate of the true $R R_{i}$ if nonsusceptible controls are not excluded from the relative risk calculation. To account for this phenomenon, we assumed that the number of injections received in the prior year was approximately proportional to the probability of having been previously infected, such that the relative risk from receiving a single contaminated injection is $\left[1+\left(R R_{i}\right.\right.$ $\left.-1) /\left(n_{i} \times p_{r} \times p_{v} \times p_{s}\right)\right]$. This method was used to estimate hazard in cases where a substantial proportion of infections was attributable to contaminated injections, as described above.

### 2.6 Data sources

## InJECTION PRACTICE PARAMETERS

Sources of information available to estimate the annual number of injections per person ( $n$ ) included, by decreasing order of data quality, pop-ulation-based injection frequency surveys and other population-based data providing injection frequency estimates. Sources of information for estimating the proportion of reuse ( $p_{r e}$ ) included, by decreasing order of data quality, observational studies of injection practices using the World Health Organization (WHO) standardized injection safety assessment survey tool (WHO 2002), studies of injection practices conducted using other, non-standardized methods, and back-calculations in published analytical epidemiological studies using the mass action equation and the relative risks of infection with bloodborne pathogens associated with receiving injections.

Sources of information were obtained through Medline searches, searches in WHO unpublished documents, including evaluations of the Expanded Programme on Immunization (EPI) and unpublished reports made available through the electronic mail list server of the Safe Injection Global Network (SIGN) (Bass 2000; Hutin and Chen 1999). All studies were reviewed using a standardized study abstraction instrument
and entered in a database. Estimates were generated for each subregion for proportion of reuse ( $p_{r e}$ ) or number of injections per person for each age, sex and subregional stratum ( $n$ ) using a standardized decisionmaking algorithm to use the best source of data available.

The frequency distribution of the annual number of injections per person that was available from two studies conducted in EUR-B (CDC 1999) and in EUR-C (WHO 1999) indicated that a small proportion of the population above the 90th percentile received more than 20 injections per year. To avoid overestimating the attributable fraction, we made the conservative assumption that those receiving such a high number of injections had already been infected and were already immune. Thus, for these two subregions (EUR-B and EUR-C) where the injection frequency distribution was available, we excluded those who had received more than 20 injections per year (approximately above the 90th percentile), thereby reducing the annual number of injections per year. For the other subregions, data were available in tabulated form in published reports. This format already eliminated the upper $10 \%$ of the frequency distribution and no adjustment was necessary (e.g. persons reporting more than seven injections per year were all considered to have received eight injections per year). When more than one source of information regarding injection frequency or reuse of equipment was available for one age, sex and subregion stratum, all were used to compute an estimate.

## Prevalence and incidence of HBV, HCV and HIV infection

We used the prevalence of active infection in the general population to estimate the proportion of patients representing a source of contamination for reused syringes and/or needles $\left(p_{v}\right)$. Therefore, we did not assume the prevalence of active infection to be higher in a health care setting, nor considered different strata according to selected settings (e.g. immunization vs clinic for the management of STIs) (see discussion). Estimates for the proportion of the population chronically infected with HBV, HCV and HIV were obtained from the WHO programmes on HBV (C. Nelson, personal communication, 2000) and HCV (D. Lavanchy, personal communication, 2000), and from the Joint United Nations Programme on HIV/AIDS (UNAIDS 2000). In the case of HBV and HCV, catalytic models in which the annual risk of infection was constant over time and over age groups were generated. These models were fitted so that annual risk of infection led to region-specific estimates of the prevalence of active infection. Once the annual risk of infection was obtained, it was used to estimate the age-specific prevalence of susceptibility and the total incidence of infection among susceptible individuals. In the case of HIV, incidence estimates were obtained from UNAIDS (2000).

### 2.7 Estimates of the proportion of the population exposed TO CONTAMINATED INJECTIONS

## Proportion of reuse

Sources of information used to generate the estimates (Table 22.4) included observational studies of injection practices using the WHO standardized injection safety assessment survey tool (AFR-D, AFR-E and EUR-B), observational studies of injection practices conducted using non-standardized methods (SEAR-B, SEAR-D and WPR-D), backcalculations using the mass action equation and the relative risks of infection with bloodborne pathogens associated with receiving injections (EUR-C), and a combination of the second and the third methods (EMRD). No quantitative data were available for six subregions. For two of them, AMR-B and AMR-D, there were qualitative reports of reuse. For AMR-B, these reports suggested that reuse was uncommon (Flaskerud and Nyamathi 1996; Ugalde and Homedes 1988; Villanueva et al. 1997). Thus, estimates from the other subregion with the lowest frequency of reuse (EUR-B) were extrapolated. For AMR-D, as qualitative reports suggested that reuse was more common than in AMR-B (Janszen and Laning 1993), estimates from EUR-C, with the second lowest frequency of reuse, were extrapolated. For EUR-A, EMR-B, AMR-A and WPR-A, representing mostly countries with high per capita gross national product, the proportion of reuse was considered negligible. Among subregions for which quantified estimates were available, SEAR-D had the highest proportion of reuse ( $75 \%$ ), followed by EMR-D ( $70 \%$ ) and WPR-B ( $30 \%$ ). EUR-B had the lowest proportion of reuse (1.2\%). (See Figure 22.2.)

## ANNUAL NUMBER OF INJECTIONS PER PERSON

Sources of information used to generate subregional input parameters (Table 22.5) included population-based injection frequency surveys, and other population-based studies that provided information about injection frequency. No injection frequency estimates were generated for those subregions for which reuse was considered negligible as no risk applied. Among subregions with quantified information available, EUR-C was the subregion with the highest injection frequency ( 11.3 injections per person and per year), followed by EUR-B ( 5.2 injections per person and per year) (CDC 1999; WHO 1999). However, when the top 10th percentile of injection frequency was removed, EMR-D was the subregion with the highest injection frequency ( 4.3 injections per person and per year), followed by SEAR-D (4 injections per person and per year). The subregions with the lowest annual number of injections per person were AMR-B (1.7 injections per person and per year) and AMR-D (1.9 injections per person and per year).
Table 22.4 Subregional estimates of the proportion of injections administered with reused equipment and data sources used,


[^86]Figure 22.2 Number of injections per person and per year, and proportion of these administered with injection equipment reused in the absence of sterilization, by subregion, $2000^{\text {a }}$


[^87]
## Proportion of the population exposed to contaminated injections

The proportion of the population exposed to contaminated injections reflected the frequency of injections received, the frequency of reuse of injection equipment and the prevalence of active infection with HBV, HCV and HIV (Table 22.6). This estimate varied from $0.03 \%$ (AMR-B) to $13.33 \%$ (EMR-D) in the case of HBV, from less than $0.03 \%$ (AMRB) to $16.73 \%$ (EMR-D) in the case of HCV, and from $0.00 \%$ (EUR-B) to $2.05 \%$ (AFR-E) in the case of HIV.

### 2.8 Estimates for the relative risks of infection for RECEIVING CONTAMINATED INJECTIONS

In the case of HIV, contaminated injections did not account for most new infections (i.e. $I_{u}$ was not close to $I_{t}$ ). Thus, model-based estimates of relative risk were used for all subregions (Table 22.7[c]). In the case
Table 22.5 Subregional injection frequency estimates and data sources used, 2000

|  | AFR-D | AFR-E | AMR-B | AMR-D | EMR-D | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-B |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Annual Crude | 2.2 | 2.0 | 1.7 | 1.9 | 4.3 | 5.2 | 11.3 | 2.1 | 4.0 | 2.4 |
| ```number of Truncated }\mp@subsup{}{}{b injections per person (n)a``` | NA | NA | NA | NA | NA | 2.5 | 3.5 | NA | NA | NA |
| Countries from which injection frequency surveys were used | Guinea-Bissau (Ferry 1995) | Central African Republic, Côte d'Ivoire, United Republic of Tanzania, Zambia, Burundi (Ferry 1995), Uganda (Priotto et al. 2001) | Brazil <br> (Ferry 1995) | NA | Egypt <br> (Talaat et al. 2001) | Romania (CDC 1999) | Republic of Moldova (WHO 1999) | Thailand (Ferry 1995; Reeler and Hematorn 1994) Indonesia (Ferry 1995; van Staa and Hardon 1996) | India (Anand et al. 2001; Ferry 1995) | NA |


Table 22.6(a) Proportion of the population (\%) receiving injections contaminated with HBV annually, 2000 ${ }^{\text {a }}$

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 5.51 | 4.47 | 4.47 | 4.47 | 4.47 | 4.47 | 4.47 | 4.47 |
|  | Female | 5.51 | 4.47 | 4.47 | 4.47 | 4.47 | 4.47 | 4.47 | 4.47 |
| AFR-E | Male | 4.72 | 3.31 | 3.31 | 3.31 | 3.31 | 3.31 | 3.31 | 3.31 |
|  | Female | 4.72 | 4.20 | 4.20 | 4.20 | 4.20 | 4.20 | 4.20 | 4.20 |
| AMR-B | Male | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
|  | Female | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| AMR-D | Male | 0.52 | 0.41 | 0.41 | 0.41 | 0.41 | 0.41 | 0.41 | 0.41 |
|  | Female | 0.52 | 0.41 | 0.41 | 0.41 | 0.41 | 0.41 | 0.41 | 0.41 |
| EMR-D | Male | 13.33 | 12.01 | 12.01 | 12.01 | 12.01 | 12.01 | 12.01 | 12.01 |
|  | Female | 13.33 | 12.01 | 12.01 | 12.01 | 12.01 | 12.01 | 12.01 | 12.01 |
| EUR-B | Male | 0.28 | 0.12 | 0.09 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 |
|  | Female | 0.28 | 0.13 | 0.16 | 0.19 | 0.19 | 0.19 | 0.19 | 0.19 |
| EUR-C | Male | 1.86 | 1.09 | 1.29 | 1.42 | 1.42 | 1.42 | 1.42 | 1.42 |
|  | Female | 1.78 | 0.92 | 1.26 | 1.78 | 1.78 | 1.78 | 1.78 | 1.78 |
| SEAR-B | Male | 9.95 | 5.15 | 5.15 | 5.15 | 5.15 | 5.15 | 5.15 | 5.15 |
|  | Female | 9.95 | 5.15 | 5.15 | 5.15 | 5.15 | 5.15 | 5.15 | 5.15 |
| SEAR-D | Male | 11.22 | 10.02 | 10.02 | 10.02 | 10.02 | 10.02 | 10.02 | 10.02 |
|  | Female | 11.22 | 10.02 | 10.02 | 10.02 | 10.02 | 10.02 | 10.02 | 10.02 |
| WPR-D | Male | 11.74 | 7.84 | 7.84 | 7.84 | 7.84 | 7.84 | 7.84 | 7.84 |
|  | Female | 11.74 | 7.84 | 7.84 | 7.84 | 7.84 | 7.84 | 7.84 | 7.84 |
| a AMR-A, EMR-B, EUR-A and WPR-A excluded as reuse of injection equipment is negligible in these subregions. |  |  |  |  |  |  |  |  |  |

Table 22.6(b) Proportion of the population (\%) receiving injections contaminated with HCV annually, 2000 ${ }^{\text {a }}$

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 1.29 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 |
|  | Female | 1.29 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 |
| AFR-E | Male | 1.12 | 0.78 | 0.78 | 0.78 | 0.78 | 0.78 | 0.78 | 0.78 |
|  | Female | 1.12 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| AMR-B | Male | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
|  | Female | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| AMR-D | Male | 0.62 | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 |
|  | Female | 0.62 | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 |
| EMR-D | Male | 16.73 | 15.10 | 15.10 | 15.10 | 15.10 | 15.10 | 15.10 | 15.10 |
|  | Female | 16.73 | 15.10 | 15.10 | 15.10 | 15.10 | 15.10 | 15.10 | 15.10 |
| EUR-B | Male | 0.09 | 0.04 | 0.03 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
|  | Female | 0.10 | 0.05 | 0.06 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 |
| EUR-C | Male | 1.19 | 0.70 | 0.82 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 |
|  | Female | 1.14 | 0.59 | 0.81 | 1.14 | 1.14 | 1.14 | 1.14 | 1.14 |
| SEAR-B | Male | 3.31 | 1.68 | 1.68 | 1.68 | 1.68 | 1.68 | 1.68 | 1.68 |
|  | Female | 3.31 | 1.68 | 1.68 | 1.68 | 1.68 | 1.68 | 1.68 | 1.68 |
| SEAR-D | Male | 5.92 | 5.27 | 5.27 | 5.27 | 5.27 | 5.27 | 5.27 | 5.27 |
|  | Female | 5.92 | 5.27 | 5.27 | 5.27 | 5.27 | 5.27 | 5.27 | 5.27 |
| WPR-D | Male | 3.28 | 2.16 | 2.16 | 2.16 | 2.16 | 2.16 | 2.16 | 2.16 |
|  | Female | 3.28 | 2.16 | 2.16 | 2.16 | 2.16 | 2.16 | 2.16 | 2.16 |

[^88]Table 22.6(c) Proportion of the population (\%) receiving injections contaminated with HIV annually, 2000 ${ }^{\text {a }}$

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 0.64 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 |
|  | Female | 0.64 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 |
| AFR-E | Male | 2.05 | 1.43 | 1.43 | 1.43 | 1.43 | 1.43 | 1.43 | 1.43 |
|  | Female | 2.05 | 1.82 | 1.82 | 1.82 | 1.82 | 1.82 | 1.82 | 1.82 |
| AMR-B | Male | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
|  | Female | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| AMR-D | Male | 0.13 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 |
|  | Female | 0.13 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 |
| EMR-D | Male | 0.10 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 |
|  | Female | 0.10 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 |
| EUR-B | Male | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
|  | Female | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| EUR-C | Male | 0.08 | 0.05 | 0.05 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 |
|  | Female | 0.07 | 0.04 | 0.05 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 |
| SEAR-B | Male | 0.33 | 0.16 | 0.16 | 0.16 | 0.16 | 0.16 | 0.16 | 0.16 |
|  | Female | 0.33 | 0.16 | 0.16 | 0.16 | 0.16 | 0.16 | 0.16 | 0.16 |
| SEAR-D | Male | 1.19 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 |
|  | Female | 1.19 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 |
| WPR-D | Male | 0.06 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |
|  | Female | 0.06 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |
| a AMR-A, EMR-B, EUR-A and WPR-A excluded as reuse of injection equipment is negligible in these subregions. |  |  |  |  |  |  |  |  |  |

Table 22.7(a) Relative risks associated with receiving injections contaminated with HBV, 2000a

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 3.53 | 3.45 | 3.45 | 3.45 | 3.45 | 3.45 | 3.45 | 3.45 |
|  | Female | 3.53 | 3.45 | 3.45 | 3.45 | 3.45 | 3.45 | 3.45 | 3.45 |
| AFR-E | Male | 3.47 | 3.37 | 3.37 | 3.37 | 3.37 | 3.37 | 3.37 | 3.37 |
|  | Female | 3.47 | 3.43 | 3.43 | 3.43 | 3.43 | 3.43 | 3.43 | 3.43 |
| AMR-B | Male | 75.21 | 74.68 | 74.68 | 74.68 | 74.68 | 74.68 | 74.68 | 74.68 |
|  | Female | 75.21 | 74.68 | 74.68 | 74.68 | 74.68 | 74.68 | 74.68 | 74.68 |
| AMR-D | Male | 25.43 | 24.78 | 24.78 | 24.78 | 24.78 | 24.78 | 24.78 | 24.78 |
|  | Female | 25.43 | 24.78 | 24.78 | 24.78 | 24.78 | 24.78 | 24.78 | 24.78 |
| EMR-D | Male | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 |
|  | Female | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 |
| EUR-B | Male | 6.49 | 6.44 | 6.43 | 6.45 | 6.45 | 6.45 | 6.45 | 6.45 |
|  | Female | 6.49 | 6.45 | 6.46 | 6.47 | 6.47 | 6.47 | 6.47 | 6.47 |
| EUR-C | Male | 7.06 | 6.77 | 6.85 | 6.9 | 6.9 | 6.9 | 6.9 | 6.9 |
|  | Female | 7.04 | 6.71 | 6.84 | 7.03 | 7.03 | 7.03 | 7.03 | 7.03 |
| SEAR-B | Male | 7.02 | 5.56 | 5.56 | 5.56 | 5.56 | 5.56 | 5.56 | 5.56 |
|  | Female | 7.02 | 5.56 | 5.56 | 5.56 | 5.56 | 5.56 | 5.56 | 5.56 |
| SEAR-D | Male | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 |
|  | Female | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 | 11.71 |
| WPR-D | Male | 7.86 | 6.28 | 6.28 | 6.28 | 6.28 | 6.28 | 6.28 | 6.28 |
|  | Female | 7.86 | 6.28 | 6.28 | 6.28 | 6.28 | 6.28 | 6.28 | 6.28 |

[^89]Table 22.7(b) Relative risks associated with receiving injections contaminated with HCV, $2000^{\text {a }}$

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 19.74 | 18.88 | 18.88 | 18.88 | 18.88 | 18.88 | 18.88 | 18.88 |
|  | Female | 19.74 | 18.88 | 18.88 | 18.88 | 18.88 | 18.88 | 18.88 | 18.88 |
| AFR-E | Male | 17.5 | 16.6 | 16.6 | 16.6 | 16.6 | 16.6 | 16.6 | 16.6 |
|  | Female | 17.5 | 17.15 | 17.15 | 17.15 | 17.15 | 17.15 | 17.15 | 17.15 |
| AMR-B | Male | 31.36 | 31.28 | 31.28 | 31.28 | 31.28 | 31.28 | 31.28 | 31.28 |
|  | Female | 31.36 | 31.28 | 31.28 | 31.28 | 31.28 | 31.28 | 31.28 | 31.28 |
| AMR-D | Male | 21.35 | 20.81 | 20.81 | 20.81 | 20.81 | 20.81 | 20.81 | 20.81 |
|  | Female | 21.35 | 20.81 | 20.81 | 20.81 | 20.81 | 20.81 | 20.81 | 20.81 |
| EMR-D | Male | 27.77 | 27.77 | 27.77 | 27.77 | 27.77 | 27.77 | 27.77 | 27.77 |
|  | Female | 27.77 | 27.77 | 27.77 | 27.77 | 27.77 | 27.77 | 27.77 | 27.77 |
| EUR-B | Male | 31.9 | 31.4 | 31.3 | 31.49 | 31.49 | 31.49 | 31.49 | 31.49 |
|  | Female | 31.9 | 4.01 | 31.52 | 31.62 | 31.62 | 31.62 | 31.62 | 31.62 |
| EUR-C | Male | 31.9 | 31.4 | 31.3 | 31.49 | 31.49 | 31.49 | 31.49 | 31.49 |
|  | Female | 31.9 | 4.01 | 31.52 | 31.62 | 31.62 | 31.62 | 31.62 | 31.62 |
| SEAR-B | Male | 38.0 | 23.86 | 23.86 | 23.86 | 23.86 | 23.86 | 23.86 | 23.86 |
|  | Female | 38.0 | 23.86 | 23.86 | 23.86 | 23.86 | 23.86 | 23.86 | 23.86 |
| SEAR-D | Male | 37.64 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 |
|  | Female | 37.64 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 |
| WPR-D | Male | 37.64 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 |
|  | Female | 37.64 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 | 26.72 |

Table 22.7(c) Relative risks associated with receiving injections contaminated with HIV, 2000 ${ }^{\text {a }}$

| Subregion | Sex | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | 14.05 | 13.84 | 5.57 | 5.57 | 5.57 | 5.57 | 5.57 | 5.57 |
|  | Female | 14.05 | 13.84 | 3.39 | 3.39 | 3.39 | 3.39 | 3.39 | 3.39 |
| AFR-E | Male | 5.95 | 5.79 | 2.3 | 2.3 | 2.3 | 2.3 | 2.3 | 2.3 |
|  | Female | 5.95 | 5.89 | 1.72 | 1.72 | 1.72 | 1.72 | 1.72 | 1.72 |
| AMR-B | Male | 122.12 | 121.85 | 21.03 | 21.03 | 21.03 | 21.03 | 21.03 | 21.03 |
|  | Female | 122.12 | 121.85 | 41.1 | 41.1 | 41.1 | 41.1 | 41.1 | 41.1 |
| AMR-D | Male | 66.27 | 65.09 | 8.12 | 8.12 | 8.12 | 8.12 | 8.12 | 8.12 |
|  | Female | 66.27 | 65.09 | 16.25 | 16.25 | 16.25 | 16.25 | 16.25 | 16.25 |
| EMR-D | Male | 1 | I | 64.41 | 64.41 | 64.41 | 64.41 | 64.41 | 64.41 |
|  | Female | 1 | 1 | 135.38 | 135.38 | 135.38 | 135.38 | 135.38 | 135.38 |
| EUR-B | Male | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|  | Female | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| EUR-C | Male | 1 | 1 | 6.48 | 6.48 | 6.48 | 6.48 | 6.48 | 6.48 |
|  | Female | 1 | 1 | 31.5 | 31.7 | 31.7 | 31.7 | 31.7 | 31.7 |
| SEAR-B | Male | 198.4 | 150.67 | 26.01 | 26.01 | 26.01 | 26.01 | 26.01 | 26.01 |
|  | Female | 198.4 | 150.67 | 43.86 | 43.86 | 43.86 | 43.86 | 43.86 | 43.86 |
| SEAR-D | Male | 213.52 | 166.26 | 16.61 | 16.61 | 16.61 | 16.61 | 16.61 | 16.61 |
|  | Female | 213.52 | 166.26 | 33.35 | 33.35 | 33.35 | 33.35 | 33.35 | 33.35 |
| WPR-D | Male | 1 | , | 41.69 | 41.69 | 41.69 | 41.69 | 41.69 | 41.69 |
|  | Female | 1 | 1 | 127.3 | 127.3 | 127.3 | 127.3 | 127.3 | 127.3 |

of HCV, contaminated injections did not account for most of new infections in all subregions apart from EUR-C, SEAR-D and EMR-D. For EUR-C and SEAR-D, model-based relative risks for EUR-B and WPRB, which had similar prevalence patterns, were used. In EMR-D, the available study-based relative risk was used (Table 22.7[b]) (El-Sakka 1997). In the case of HBV, contaminated injections did not account for most of new infections in all subregions apart from SEAR-D and EMRD. In EMR-D, the available study-based relative risk was used (Anonymous 1998). For SEAR-D, the study-based relative risk for EMR-D was used as this subregion had HBV prevalence patterns and injection practices close to those of EMR-D (Table 22.7[a]).

### 2.9 Progression of HBV, HCV and HIV infection to disability and death

## TIME INTERVAL BETWEEN INFECTION AND THE OCCURRENCE OF DISABILITY AND DEATH

The majority of the burden of disease associated with infections with HBV, HCV and HIV is delayed in time. For HBV infection, a small proportion of acute infections lead to death through fulminant liver failure. However, most of the burden of disease is secondary to the long-term consequences of chronic HBV infection, including end-stage liver disease and hepatocellular carcinoma. For HCV infection, the death to case ratio for acute infections is negligible and all the burden of disease is secondary to the long-term consequences of chronic infection, including end-stage liver disease and hepatocellular carcinoma. For HIV infection, the burden of disease is secondary to the progression to AIDS and to death following AIDS.

To take into account the time interval between infection and the progression towards death and disability, two measures need to be distinguished. First, the attributable burden in 2000 includes the current burden in the year 2000 that is secondary to past and present unsafe injection practices. Second, the future burden due to current unsafe injection practices in 2000 includes the future long-term consequences in terms of end-stage liver disease, hepatocellular carcinoma and AIDS of the HBV, HCV and HIV infections acquired in 2000 because of contaminated injections.

## CUrrent burden due to past and present unsafe injection practices

In the absence of information regarding injection practices in the past, we were unable to model the fraction of HBV, HCV and HIV infection that was attributable to contaminated injections before the year 2000. Thus, we made the assumption that the fraction of HBV, HCV and HIV infections attributable to contaminated injections in the past was identical to the one modelled for the year 2000. We then applied these attributable fractions to the current burden in 2000 in terms of DALYs and
deaths that were associated with the consequences of HBV, HCV and infection (i.e. acute infections, hepatocellular carcinoma, end-stage liver diseases and HIV infection/AIDS).

## Future burden due to current unsafe injection practices

To reflect the delay between infection and disease outcomes, the fraction of new infections with HBV, HCV and HIV attributable to contaminated injections was converted into years of life lost (YLL) using synthetic cohorts of infected individuals followed for mortality associated with HBV, HCV or HIV infection (AIDS or chronic liver disease) and background mortality. Background mortality was taken into account using age, sex and subregion-specific Global Burden of Disease (GBD) life tables. ${ }^{2}$ To estimate the years of life lost secondary to HBV infections, the model parameters included:

- a rate of progression to chronic infection of $30 \%$ among persons infected under the age of 5 years and of $6 \%$ for persons infected at the age of 5 years or older (McMahon et al. 1985);
- an annual rate of clearance of infection (i.e. sero-reversion) of $1 \%$ following chronic infection (Alward et al. 1985); and
- an age-dependent yearly mortality rate associated with chronic liver disease among persons chronically infected with HBV (Figure 22.3) that was modelled on the basis of African and Asian studies (Gay et al. 2001).

To estimate the years of life lost from HCV infections, two sets of assumptions were used according to the age of the individual at infection (Figure 22.4). For persons infected before the age of 40, the model parameters included:

- a rate of progression to chronic infection of $63 \%$, the average of rates observed in two large studies conducted in this age group (Alter and Seeff 2000);
- a cumulated incidence rate of cirrhosis of $5 \%$ at 20 years among patients with chronic infection (Alter and Seeff 2000; Freeman et al. 2001); and
- a yearly mortality rate associated with hepatocellular carcinoma and chronic liver disease of $3.7 \%$ after the onset of cirrhosis, the average of two large studies (Alter and Seeff 2000).

For persons infected at the age of 40 years or older, the model parameters included:

- a rate of progression to chronic infection of $80 \%$ (Alter and Seeff 2000);
Figure 22.3 Decision tree for the theoretical cohort used for the calculation of the years of life lost (YLL) secondary to hepatitis $B$ virus infection

| Year 0 |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 100000 initial HBV infections at age N |  |  |  |  |  |
| Year 1 | $\square$ |  |  |  |  |  |
|  | Clearance of viral infection (70\% among those aged <5 years; $94 \%$ among those aged $\geq 5$ years) | Mortality from other causes at age N (GBD life tables)* |  | Survival with chronic HBV infection | chBV infection |  |
| Year 2 |  |  |  |  |  |  |
| Year 2 | Clearance of viral infection (sero-reversion) (1\%) |  | Mortality from other causes at age $\mathrm{N}+1$ (GBD life tables)* |  | Mortality from chronic HBV infection (Age- and sex-specific modelled rate) | Survival with chronic HBV infection |
| Year 3 and subsequent years |  |  |  |  |  |  |
|  |  |  |  |  |  | Follow-up over time using an annual scenario similar to year 2 (...) |

* GBD life tables prepared for World health report 2001 (WHO 2001)
Figure 22.4 Decision tree for the theoretical cohort used for the calculation of the years of life lost (YLL) secondary to hepatitis C virus infection

* GBD life tables prepared for World health report 2001 (WHO 2001).
- a cumulated incidence rate of cirrhosis of $20 \%$ at 20 years (Alter and Seeff 2000; Freeman et al. 2001); and
- a yearly mortality rate associated with chronic liver disease of $3.7 \%$ after the onset of cirrhosis, as in the younger age group.

Disability and death attributable to acute viral hepatitis were considered negligible for HBV and HCV in comparison with the disability and death secondary to chronic infection. In the case of HIV, parameters of progression of HIV infection to AIDS and death developed by WHO and UNAIDS were used (N. Walker, personal communication, 2002).

## Uncertainty analysis

Standard errors were calculated for selected key input parameters, including the annual number of injections per person and the proportion of reuse. Standard formulae for the calculation of confidence intervals for means and proportions were used. In the specific case of the proportion of reuse estimated on the basis of measures of association, total sample size was assumed to be the total number of study participants included in the study.

For subregions for which good quality data were available on injection frequency (injection frequency surveys) or injection safety (standardized or non-standardized injection safety surveys), a $95 \%$ confidence interval was calculated for the input parameter on the basis of the standard error ( $\pm 2 \mathrm{SE}$ ). For subregions for which only lower quality data were available for injection frequency (other population-based injection frequency data) or injection safety (back-calculated estimates), an arbitrary larger interval was used to account for added uncertainty ( $\pm 4 \mathrm{SE}$ ). For subregions for which no data were available and inferences were made using other subregions, an even larger interval was arbitrarily used to account for added uncertainty ( $\pm 6$ SE).

Lower and upper bounds of the $95 \%$ confidence intervals for the proportion of the population exposed and for relative risk estimates were obtained by including the values for lower and upper bounds of the input parameters in the model equations. Confidence intervals for the relative risks that were study-based rather than model-based were calculated on the basis of the confidence interval of the relative risk in the original epidemiological studies. See section 4 for further discussion of sources of uncertainty.

## 3. Results

### 3.1 Fraction of infections attributable to contaminated INJECTIONS IN 2000

Globally, the fractions of incident HBV, HCV and HIV infections attributable to contaminated injections in the subregions where reuse of injec-
tions was reported were $31.9 \%$, $39.9 \%$ and $5.4 \%$, respectively (Table 22.8). For HBV, this proportion was highest in EMR-D (58.3\%) and lowest in EUR-B ( $0.9 \%$ ). For HCV, this proportion was highest in EMRD $(81.7 \%)$ and lowest in EUR-B ( $0.9 \%)$. For HIV, this proportion was highest in SEAR-D ( $24.3 \%$ ) and lowest in AMR-B $(0.00 \%)$. In absolute numbers of infections, our analysis indicated that globally, in 2000, contaminated injections may have caused 20.6 million cases of new HBV infections, 2.0 million cases of HCV infections and 260000 cases of HIV infections.

### 3.2 Current burden due to past and present unsafe INJECTION PRACTICES

The current burden in 2000 due to past and present unsafe injection practices reached 501000 deaths, with the majority of deaths occurring in Asia ( $39 \%$ and $31 \%$ in WPR-B and SEAR-D, respectively) (Table 22.9 ) and among persons aged $\geq 15$ years ( $n=444000,88 \%)$. When death and disability were combined, the burden reached 10461000 DALYs, with a similar predominance in Asia ( $27 \%$ and $39 \%$ in WPRB and SEAR-D, respectively) and adults ( $n=8419000,81 \%$ of DALYs among persons aged $\geq 15$ years). Taken together, viral hepatitis $B$ and $C$ and their chronic consequences accounted for $74 \%$ and $61 \%$ of the deaths and DALYs, respectively, and HIV accounted for the remainder. There were no substantial differences in the distribution of death and disability by sex.

### 3.3 Future burden due to current unsafe INJECTION PRACTICES

Models of natural history and background mortality allowed estimation of the burden of disease attributable to contaminated injections received in 2000. This analysis suggested that the 20.6 million HBV infections in the year 2000 would lead to 26492 deaths in 2000 from fulminant hepatitis and an additional 49000 early deaths from the consequences of chronic infection between 2000 and 2030. With respect to HCV infection, we estimated that the two million infections in 2000 would lead to 24000 early deaths between 2000 and 2030. Finally, 210000 of the 260000 persons infected with HIV through contaminated injections in 2000 are expected to die prematurely from AIDS between 2000 and 2030. While our analytic horizon did not go beyond year 2030, it is anticipated that persons infected with HBV and HCV because of contaminated injections in 2000 would continue to develop long-term complications leading to death beyond this date.

## 4. Discussion

While the consequences of poor injection practices have been recognized for many decades (Anonymous 1945; Wyatt 1984), the safe and
Table 22.8 HBV, HCV and HIV infections attributable to contaminated injections (attributable fraction and absolute numbers,

| Subregion | HBV |  | HCV |  | HIV |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Attributable fraction (\%) | Number of infections | Attributable fraction (\%) | Number of infections | Attributable fraction (\%) | Number of infections |
| AFR-D | $\begin{gathered} 10.9 \\ (8.2-\mid 3.9) \end{gathered}$ | $\begin{gathered} 639498 \\ (478834-81435 I) \end{gathered}$ | $\begin{gathered} 16.4 \\ (12.3-20.8) . \end{gathered}$ | $\begin{gathered} 5468 । \\ (41078-69402) \end{gathered}$ | $\begin{gathered} 2.5 \\ (1.9-3.1) \end{gathered}$ | $\begin{gathered} 18317 \\ (13765-23243) \end{gathered}$ |
| AFR-E | $\begin{gathered} 9.2 \\ (6.9-11.5) \end{gathered}$ | $\begin{gathered} 630976 \\ (474379-792536) \end{gathered}$ | $\begin{gathered} 13.0 \\ (9.8-16.2) \end{gathered}$ | $\begin{gathered} 54131 \\ (40819-67794) \end{gathered}$ | $\begin{gathered} 2.5 \\ (1.9-3.1) \end{gathered}$ | $\begin{gathered} 64412 \\ (48520-80759) \end{gathered}$ |
| AMR-B | $\begin{gathered} 2.3 \\ (0.0-16.3) \end{gathered}$ | $\begin{gathered} 14118 \\ (112-98872) \end{gathered}$ | $\begin{gathered} 0.9 \\ (0.0-6.4) \end{gathered}$ | $\begin{gathered} 2282 \\ (18-15985) \end{gathered}$ | $\begin{gathered} 0.2 \\ (0.0-1.5) \end{gathered}$ | $\begin{gathered} 305 \\ (2-2 \mid 32) \end{gathered}$ |
| AMR-D | $\begin{gathered} 9.3 \\ (0.0-26.9) \end{gathered}$ | $\begin{gathered} 28570 \\ (16-82490) \end{gathered}$ | $\begin{gathered} 9.2 \\ (0.0-26.7) \end{gathered}$ | $\begin{gathered} 6304 \\ (4-18215) \end{gathered}$ | $\begin{gathered} 1.5 \\ (0.0-4.5) \end{gathered}$ | $\begin{gathered} 911 \\ (1-2626) \end{gathered}$ |
| EMR-D | $\begin{gathered} 58.3 \\ (26.2-82.4) \end{gathered}$ | $\begin{gathered} 2533443 \\ (\mathrm{I} \mid 40352-35806 \mathrm{I}) \end{gathered}$ | $\begin{gathered} 81.7 \\ (52.1-95.0) \end{gathered}$ | $\begin{gathered} 645486 \\ (412078-750452) \end{gathered}$ | $\begin{gathered} 7.1 \\ (5.7-8.5) \end{gathered}$ | $\begin{gathered} 2210 \\ (1775-2668) \end{gathered}$ |
| EUR-B | $\begin{gathered} 0.9 \\ (0.0-3.3) \end{gathered}$ | $\begin{gathered} 21122 \\ (156-78639) \end{gathered}$ | $\begin{gathered} 0.9 \\ (0.0-3.4) \end{gathered}$ | $\begin{gathered} 2110 \\ (16-7729) \end{gathered}$ | $\begin{gathered} 0.0 \\ (0.0-0.0) \end{gathered}$ | $(0-0)^{0}$ |
| EUR-C | $\begin{gathered} 7.7 \\ (1.8-15.0) \end{gathered}$ | $\begin{gathered} 193636 \\ (46035-378229) \end{gathered}$ | $\begin{gathered} 21.2 \\ (6.1-34.7) \end{gathered}$ | $\begin{gathered} 35668 \\ (10287-58378) \end{gathered}$ | $\begin{gathered} 0.6 \\ (0.2-1.2) \end{gathered}$ | $\begin{gathered} 1526 \\ (374-2903) \end{gathered}$ |
| SEAR-B | $\begin{gathered} 22.4 \\ (16.5-28.7) \end{gathered}$ | $\begin{gathered} 942038 \\ (694606-1205102) \end{gathered}$ | $\begin{gathered} 30.8 \\ (22.8-39.2) \end{gathered}$ | $\begin{gathered} 94873 \\ (70235-120979) \end{gathered}$ | $\begin{gathered} 7.0 \\ (5.2-8.9) \end{gathered}$ | $\begin{gathered} 6260 \\ (4638-7980) \end{gathered}$ |
| SEAR-D | $\begin{gathered} 53.6 \\ (21.6-79.9) \end{gathered}$ | $\begin{gathered} 8019210 \\ (3237944-11954579) \end{gathered}$ | $\begin{gathered} 59.5 \\ (40.4-93.6) \end{gathered}$ | $\begin{gathered} 498166 \\ (338548-784474) \end{gathered}$ | $\begin{gathered} 24.3 \\ (18.3-0.1) \end{gathered}$ | $\begin{gathered} 156663 \\ (1\|8235-194\| 87) \end{gathered}$ |
| WPR-B | $\begin{gathered} 33.6 \\ (0.0-79.0) \end{gathered}$ | $\begin{gathered} 7610161 \\ (2126-17868925) \end{gathered}$ | $\begin{gathered} 37.6 \\ (0.0-89.8) \end{gathered}$ | $\begin{aligned} & 608200 \\ & (172-1454478) \end{aligned}$ | $\begin{gathered} 2.5 \\ (0.0-5.9) \end{gathered}$ | $\begin{aligned} & 5549 \\ & (2-13378) \end{aligned}$ |
| World | $\begin{gathered} 31.9 \\ (9.4-56.9) \end{gathered}$ | $\begin{gathered} 20632772 \\ (6074558-36854335) \end{gathered}$ | $\begin{gathered} 39.9 \\ (18.2-66.7) \end{gathered}$ | $\begin{gathered} 2001901 \\ (913254-3347885) \end{gathered}$ | $\begin{gathered} 5.4 \\ (3.9-7.0) \end{gathered}$ | $\begin{gathered} 256152 \\ (187312-329877) \end{gathered}$ |

Table 22.9 Current burden in 2000 due to past and present contaminated injections

|  |  | AFR-D | AFR-E | AMR-B | AMR-D | EMR-D | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-B | World |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\begin{aligned} & \text { Deaths } \\ & (000 \mathrm{~s}) \end{aligned}$ | HBV and HCV | 6 | 6 | 1 | 1 | 41 | 1 | 10 | 25 | 88 | 193 | 372 |
|  | HIV/AIDS | 10 | 44 | 0 | 0 | 4 | 0 | 0 | 3 | 66 | 1 | 129 |
|  | Total | 16 | 51 | 1 | 2 | 44 | 1 | 10 | 28 | 154 | 194 | 501 |
| $\begin{aligned} & \text { DALYs } \\ & (000 \mathrm{~s}) \end{aligned}$ | HBV and HCV | 106 | 111 | 15 | 22 | 721 | 13 | 162 | 413 | 2007 | 2781 | 6349 |
|  | HIV/AIDS | 325 | 1436 | 2 | 10 | 106 | 0 | 3 | 99 | 2094 | 38 | 4112 |
|  | Total | 430 | 1547 | 17 | 31 | 826 | 13 | 165 | 512 | 4100 | 2819 | 10461 |

appropriate use of injections remains a target that has not been reached in most developing and transitional countries. Since the early 1990s, epidemiological studies of new HBV, HCV and HIV infections have indicated that unsafe injections are a risk factor for each disease (Simonsen et al. 1999). In most transitional and developing countries where HBV and HCV lead to a high burden of chronic liver disease, unsafe injections account for a high proportion of these infections (Hutin et al. 1999; Khan et al. 2000; Luby et al. 1997; Narendranathan and Philip 1993).

This chapter made use of the best available evidence regarding the rates of injection use, the frequency of reuse and the association between injections and infections. Use of a mass action model was needed for the communicable nature of outcomes and lack of transferability of hazard from across populations. The results indicate that in 2000, four decades after the widespread availability of disposable injection equipment and two decades into the HIV pandemic, contaminated injections accounted for close to a third of new HBV infections, $40 \%$ of new HCV infections and $5 \%$ of new HIV infections globally. The burden of disease in 2000 due to past and current infections reached 501000 deaths and 10461000 DALYs, with the majority of deaths occurring among adults, mostly in Asia.

Using available studies, we described injection practices worldwide in terms of safety and frequency. Our analysis indicated high rates of injection worldwide, with marked subregional variations, for a total exceeding 16 thousand million injections in the 10 (of 14) subregions that were included in our study. This order of magnitude is validated by the market analysis suggesting that in Japan, the United States of America and western Europe, 17.5 thousand million syringes are sold annually (Kaninda 2001).
Four subregions stood out with particularly high estimates. The crude annual number of injections per person was the highest in the former socialist economies of Europe and central Asia, reaching 11.3 and 5.2 in EUR-B and EUR-C, respectively. Most injections in these countries are administered in public health care facilities by physicians or nurses, with a high number of injections per prescription (CDC 1999; WHO 1999). While patients' demand is commonly quoted by health care providers as a major driver of injection overuse, knowledge, attitude and practice (KAP) surveys find that health care providers have a tendency to overestimate patients' preference for injections and that the population is open to alternatives to injected medications (CDC 1999; Vong et al. 2002). In fact, KAP surveys conducted among physicians indicate that prescribers have false preconceptions about the effectiveness of injectable medications and that these preconceptions are sometimes supported by non-evidence-based official treatment protocols. Thus, prescribers' attitudes also contribute to injection overuse (Stoica et al. 1999).

Injection use was also high in the Middle East and in south Asia where the annual number of injections per person reached 4.3 and 4.0 in EMR-

D and SEAR-D, respectively. In these countries, a high proportion of injections are administered by private providers who, in some cases, are not medically qualified (Khan et al. 2000; Kosen 1999; Talaat et al. 2001). In such settings, health care providers' attitudes also drive injection overuse (Khan et al. 2000; Luby et al. 1997). However, practices are different. The reference to any guideline is uncommon. Injections are frequently used on an ad hoc basis to administer mixtures of antibiotics, analgesics, vitamins or antihistamines in the desire to meet what is believed to be the demand of the user (Khan et al. 2000).

Reducing injection overuse would only be a matter of promoting rational drug use if injections were administered safely. Unfortunately, our analysis indicated that injections are given in a way that may harm the injection recipient. Determinants of these unsafe injection practices include the lack of supplies of new, sterile, single-use, disposable injection equipment (Dicko et al. 2000), the lack of awareness among patients and providers regarding the risks associated with unsafe practices (Anand et al. 2001; Khan et al. 2000), and the absence of an efficient sharps waste management system to prevent recycling of contaminated equipment (Hofmann 2001). Of interest, our analysis suggests that injection practices are safer in sub-Saharan Africa ( $19 \%$ and $17 \%$ of reuse in AFR-D and AFR-E, respectively) than in the Middle East and south Asia ( $70 \%$ and $75 \%$ reuse in EMR-D and SEAR-D, respectively). The proportion of the population aware of the potential risk of HIV infection through unsafe injections was $24 \%$ in Pakistan in 1998 (Luby et al. forthcoming), $19 \%$ in India in 1999 (Anand et al. 2001) and $52 \%$ in Burkina Faso in 2001 (Logez 2001). In addition, the social and economic consequences of the HIV pandemic have been perceived more acutely on the African continent than in Asia. Thus, a higher awareness regarding the risks of HIV infection associated with unsafe injections in subSaharan Africa (Birungi 1998) may partly explain this difference observed in the proportion of reuse.

HBV infection was the most common consequence of unsafe injection practices in the world, with more than 20 million cases of infection annually. Among the three pathogens that we examined, HBV is the most prevalent globally (Maynard et al. 1989) and the one most easily transmitted through unsafe injections (Seeff et al. 1978). The subregionspecific fractions of new HBV infections attributable to contaminated injections were compatible with those reported in analytical studies, including $2 \%$ (Thuring et al. 1993) to $73.9 \%$ (Ko et al. 1991a) in WPRB (compared with $33.6 \%$ in our model), $49.7 \%$ (Singh et al. 2000) to $53.3 \%$ (Narendranathan and Philip 1993) in SEAR-D (compared with $53.6 \%$ in our model) and $27.7 \%$ (Anonymous 1998) to $52 \%$ (Hussain 2001) in EMR-D (compared with $58.3 \%$ in our model). Because attributable fractions were also high among children aged $<5$ years, a substantial proportion of unexplained transmission of HBV among preschool children may thus be attributed to unsafe injections
(Davis et al. 1989). Such infections would entail a substantial burden of disease and death in the future since the long-term consequences of HBV infections are most severe among persons infected during childhood (McMahon et al. 1985). The natural history of the infection is well described for HBV and there was relatively little uncertainty around the disease progression parameters that we used.

The burden of disability and death secondary to injection-associated HBV infections was estimated to be low in comparison to the number of infections because of the low proportion of progression to chronic infections and the delay between infection and death. This delay between initial infection and death reduced the burden because of the 30 -year horizon of this work and because background mortality would lead infected persons to die from other causes during this time interval. In addition, the burden of disease avoidable through the control of contaminated injections as a risk factor is limited because universal childhood immunization against hepatitis B is being increasingly introduced in resource-poor countries with the support of the Global Alliance for Vaccines and Immunization (GAVI). If high vaccination rates are indeed reached in the future, as in our assumptions for 2030, herd immunity will progressively reduce the incidence of injection-associated infections through a high prevalence of immunity and a low prevalence of active infection (Wittet 2001).

HCV infection was estimated to be the second most common consequence of contaminated injections worldwide, with more than two million infections each year. Injection-associated HCV infection was less common than injection-associated HBV infection because of the lower prevalence of HCV infection (WHO 2000a) and the lower percutaneous transmission potential of HCV (CDC 1997). However, the fraction of new HCV infection attributable to contaminated injections was high among all age groups, including adults, and was higher than for HBV infection. These high attributable fractions are compatible with those reported in analytical studies, including $20.1 \%$ (Chen et al. 1995) to $90.6 \%$ (Thuring et al. 1993) in WPR-B (compared with $37.6 \%$ in our model), and 9.9\% (Mohamed et al. 1996) to 87.9\% (El-Sakka 1997) in EMR-D (compared with $81.7 \%$ in our model).

HCV is primarily transmitted through percutaneous or permucosal exposure to blood (Alter 1995). Transmission among sexual partners occurs but is not efficient (Alter 1995), occurring mostly among individuals engaging in high-risk sexual behaviour that may expose them to blood, or body fluids contaminated with blood (Williams et al. 1999). Sexual transmission may account for a higher proportion of HCV infections in industrialized countries, where contaminated health care injections and other unsafe percutaneous procedures are uncommon (Alter et al. 1999). However, in developing and transitional countries, the exposures most likely to transmit HCV include unsafe injections, transfusion of blood, blood components and blood products, and other unsafe
percutaneous exposures conducted in medically-related and other settings (e.g. dental care, surgery, circumcision, shaving).

The risk of HCV infection following transfusion of contaminated blood—about $92 \%$ (Aach et al. 1991)—greatly exceeds the risk of HCV infection following a contaminated injection. However, transfusion of blood, blood components and blood products is an infrequent exposure in comparison with injections. Annually, worldwide, it is estimated that over 75 million blood donations occur (WHO 2000b). Our analysis suggested that for developing and transitional countries alone, more than 16 thousand million injections might occur annually. Thus, despite a lower risk associated with each unsafe event, the higher frequency of injections explains why, globally, our analysis suggested that a high proportion of HCV infection was attributable to contaminated injections. While percutaneous exposures other than injections have been associated with HCV infection in epidemiological studies, they rarely explain a high proportion of infections (Wasley and Alter 2000).

HCV infection is currently not preventable through immunization and, in contrast to HBV infection, its long-term consequences may be more severe among persons infected during adulthood than among persons infected during childhood (Alter and Seeff 2000; Vogt et al. 1999). To estimate the burden of disease secondary to current injectionassociated HCV infections, we used conservative estimates for the parameters describing the progression of HCV infection towards chronic liver disease and its consequences. However, there is substantial uncertainty as to whether these estimates obtained from studies conducted in industrialized countries are representative of the natural history of HCV infection in developing and transitional countries as environmental factors could influence the risk of progression to chronic liver diseases and its consequences. In addition, because the parameters that we selected assume that infected patients only die after 20 years, the 30 -year analytic horizon of this work only captured a small proportion of the future early deaths. Nevertheless, if the parameters used in our model were indeed representative, our analysis would suggest that injection-associated HCV infections would not constitute a major avoidable burden of disease between 2000 and 2030. If we underestimated the severity of the natural history of HCV infection, then the burden of chronic liver disease and death secondary to injection-associated HCV infection that could be avoided in the future, particularly in countries highly endemic for HCV, would be estimated as substantial.

Historically, health care injections have not been viewed as a major vehicle of HIV infection and the promotion of safe injection practices has not held a high priority in HIV prevention programmes. However, most nosocomial outbreaks of HIV infection have been reported from countries with low prevalence of HIV infection, including Colombia (Shaldon 1995), the Libyan Arab Jamahiriya (Yerly et al. 2001), Romania (Hersh et al. 1993) and Ukraine (Simonsen et al. 1999). In
other countries where HIV infection and poor injection practices are more common and where sexual transmission accounts for the majority of infections, injection-associated HIV infections are likely to occur but they have rarely been reported or suspected.

Our analysis suggests that contaminated injections may cause $5.4 \%$ of new cases of HIV infection worldwide, representing the largest burden of disease that could be avoided through safe and appropriate use of injection policies. Few epidemiological studies are available with which to validate our estimates, either because transmission through injections was not examined or because these studies were not based upon recent, incident HIV infections. This lack of information represents a substantial source of uncertainty. In AFR-E where several studies were available (Bultreys et al. 1994; N'Galy et al. 1988; Quigley et al. 2000; Wawer et al. 1994), the lowest attributable fraction (8\%) calculated on the basis of the data provided by Wawer et al. (1994) largely exceeded our estimate of $2.5 \%$. In EMR-D and SEAR-D, the model suggests that the attributable fraction could reach $7.1 \%$ and $24.3 \%$, respectively. These estimates are not validated by epidemiological studies and may be overestimated because the epidemic is still concentrated, violating the assumptions made in the mass action model about the distribution of contaminated injections. Despite large uncertainty in attributable fractions in these two subregions, the high frequency of unsafe injection practices coinciding with emerging HIV epidemics must lead to urgent preventive measures.

In the future, studies of the risk factors for HIV infection should ensure that data are collected in a way that allows examination of the association between HIV infection and injections. In the meantime, HIV prevention programmes should communicate the risk of HIV infection associated with health care injections since safe and appropriate use of injection policies constitute effective interventions against HIV infection (CDC 2001; Logez 2001; Prawitasari Hadiyono et al. 1996).

While much emphasis was put on gathering the best available data from published and unpublished sources, this analysis has a number of limitations due to data scarcity.

- Our model was constrained by the limited number of studies with adequately described injection practices. Moreover, some of these studies employed non-standardized methodologies, which could not be used. The high frequency of injections reported in developing and transitional countries contrasts with the paucity of data available to describe practices. Until recently, few standardized tools for assessment or evaluation were available to routinely collect information on injection frequency or safety. However, WHO has recently developed new assessment tools that utilize standardized methods, which will prospectively generate information of appropriate quality. This should
allow future revisions of these burden of disease estimates to be based upon data of better quality.
- The transmission potential of HBV, HCV and HIV through percutaneous exposure was obtained on the basis of epidemiological studies that estimated the risk of infection with bloodborne pathogens among health care workers following a needle-stick injury. Although these figures are based upon many well-conducted studies that included a large number of study participants, they estimated a different risk: infection associated with a needle-stick injury. Factors that could cause the risk from contaminated injections to be higher than the risk from needle-stick injuries include the liquid flow rinsing the needle that occurs during an injection and the potential survival of HBV and HCV in the pots of tepid water often used to rinse injection equipment between injections. Factors that could cause the risk from contaminated injections to be lower than the risk from needle-stick injuries include a longer time interval between injections that could cause inactivation of some virus particles and a dilution effect in the pots of tepid water when injection equipment is reused.
- Our model only estimated the incidence of infections with HBV, HCV and HIV secondary to the reuse of injection equipment on one patient. It did not take into account the transmission secondary to the reuse of equipment on multiple patients, the transmission associated with unhygienic use of multi-dose medication vials and the transmission that may occur through cross-contamination while preparing injections. Failure to address these specific unsafe practices may have led to an underestimation in our results.
- Our analysis did not take into account any transmission networks by which injection frequencies, background prevalence of infection or probability of exposure to unsafe practices, were assessed. As a result, exposure to contaminated injections would not be distributed independently, thus creating various population subgroups with different bloodborne pathogen transmission dynamics. This would be particularly important in settings with concentrated groups of infected persons (e.g. persons with HIV in SEAR-D). However, we excluded persons presenting with very high injection frequencies to calculate the subregional injection frequency input parameters and adjusted the model for the possibility that persons receiving a high number of injections could already have immunity against infection with bloodborne pathogens. Further a potential network effect could involve percutaneous and sexual transmission (e.g. use of injected antibiotics among commercial sex workers), thereby transforming a dendritic transmission network into a more effective cyclic one (Potterat et al. 1999; Rothenberg et al. 1998).

Safe and appropriate use of injection policies aims to eliminate unnecessary injections and achieve safe injection practices. Such initiatives should not constitute separate programmes but should be integrated with other activities (WHO 2000c) to provide more effective interventions including:

- communication of risks associated with unsafe injections to patients and health care workers through disease prevention programmes such as HIV prevention;
- ensuring access to sufficient quantities of single-use, disposable injection equipment in health care facilities; and
- management of sharps waste to prevent reuse of dirty equipment and needle-stick injuries.

This study generated initial estimates of the burden of disease that could be avoided through the implementation of such policies. Further studies will address sources of uncertainty that remain in the natural history of HCV infection and in the proportion of HIV infections attributable to contaminated injections.

## 5. Projections of future Levels of exposure

The future prevalence of exposure to contaminated injections and the future relative risk of HBV, HCV or HIV infection associated with contaminated injections were calculated by including different input parameters into the same model. Assumptions regarding the injection practice parameters in 2030, including the annual number of injections per person and proportion of reuse, were generated through a survey of nine experts from four of the six WHO regions. These projections took into account the high effectiveness of planned or implemented interventions aimed at improving the safety of injections, the moderate effectiveness of interventions aimed at decreasing injection overuse, the prospective for future increased access to health care in sub-Saharan Africa (which could lead to an increase in injection use), and the potential for health care reform in the former socialist countries of eastern Europe and central Asia (which could lead to a decrease in injection overuse). Assumptions regarding the expected number of injections per person and the proportion of reuse were compatible with a slight decrease of injection frequency and a marked improvement of injection safety, although subregions were expected to remain heterogeneous (Table 22.10).

Epidemiological parameters of HBV infection were modified to account for the expected increased use of hepatitis B vaccine secondary to the accelerated introduction of this vaccine into immunization programmes supported by GAVI: the three-dose vaccine coverage was assumed to be $90 \%$ among persons aged $<15$ years and $50 \%$ among persons aged 15-29 years. The prevalence of active infection in the pop-

Table 22.10 Assumptions regarding projected injection practices in 2030

|  | AFR-D | AFR-E | AMR-B | AMR-D | EMR-D | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-B |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Proportion <br> of reuse <br> (\%) | 8 | 7 | 0 | 4 | 25 | 0 | 4 | 8 | 17 | 9 |
| Annual <br> no. of <br> injections <br> per <br> person | 2.3 | 2.2 | 1.2 | 1.9 | 3.0 | 1.6 | 2.2 | 1.7 | 2.4 | 1.8 |

ulation was assumed to be $2 \%$ in all subregions, except for AMR-B and AMR-D, where it was assumed to be $0.5 \%$. The annual incidence among susceptible individuals was assumed to be $0.1 \%$ in all subregions, except for AMR-B and AMR-D where it was assumed to be $0.01 \%$. In the absence of available epidemiological projections, the incidence and the prevalence of HCV and HIV were assumed to remain constant. Changes in parameters were assumed to be linear between 2000 and 2030.

Our projections into the future of the risk of infection with bloodborne pathogens associated with exposure to contaminated injections did not take into account the dynamic effect of new injection-associated infections on the prevalence of infection with bloodborne pathogens. Our model, a Bernoulli risk projection model, is more adapted to the estimation of the current and past incidence of injection-associated infections. In the case of HIV, where contaminated injections account for only a small proportion of new infections and the pandemic continues to be largely driven by sexual transmission, this limitation is unlikely to substantially affect our results. In the case of HBV, contaminated injections account for a substantial proportion of new infections. The impact of prevention policies on the future burden of disease is likely to be underestimated because prevented cases of HBV infection will reduce the pool of chronically infected persons who constitute sources of infection. However, in countries of intermediate or high HBV endemicity, agespecific prevalence of infection and historical data suggest that the endemicity level has not substantially changed over the past decades, and there is no evidence of injection practices playing a major role in the introduction of HBV in a community.

In contrast, in the case of HCV, where the proportion of infections attributable to injections is high, the effect of this limitation is likely to be considerable. In addition, there is evidence that in some countries, including China (Province of Taiwan) (Sun et al. 1999), Egypt (Frank et al. 2000) and Pakistan (Luby et al. 1997), HCV was recently introduced,
largely through contaminated injections, and rapidly reached high prevalence levels. In fact, in some of these countries, the prevalence is heterogeneous and areas persist where the virus has not been widely introduced (Mujeeb et al. 2000; Sun et al. 1999). In subregions where reuse of injection equipment is common but the prevalence of HCV infection is not yet high (e.g. SEAR-D), there is an opportunity at present to prevent future community-wide outbreaks of HCV infections. Our model does not reflect this opportunity.

## Acknowledgements

We are grateful to Steve Luby for critical comments on the methodology and on the manuscript; to David Gisselquist for assistance with the review of epidemiological studies that examined the association between injections and HIV infection; and to John Potterat for assistance in addressing the network effect issue.

## Notes

1 See preface for an explanation of this term.
2 Prepared for World health report 2001 (WHO 2001).

## References

Aach RD, Stevens CE, Hollinger FB et al. (1991) Hepatitis C virus infection in post-transfusion hepatitis. New England Journal of Medicine, 325:13251329.

Abulrahi HA, Bohlega EA, Fontaine RE, Al-Seghayer SM, Al-Ruweis AA (1997) Plasmodium falciparum malaria transmitted in hospital through heparin locks. The Lancet, 349:23-25.
Alter MJ (1995) Epidemiology of hepatitis C in the West. Seminars in Liver Disease, 15:5-14.
Alter MJ, Kruszon-Moran D, Nainan OV et al. (1999) The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. New England Journal of Medicine, 341:556-562.
Alter HJ, Seeff LB (2000) Recovery, persistence and sequelae in hepatitis C virus infection: a perspective on long-term outcome. Seminars in Liver Disease, 20:17-35.
Alward WL, McMahon BJ, Hall DB et al. (1985) The long-term serological course of asymptomatic hepatitis B virus carriers and the development of primary hepatocellular carcinoma. Journal of Infectious Diseases, 151: 604-609.
Anand K, Pandav CS, Kapoor SK, Undergraduate study team (2001) Injection use in a village in north India. National Medical Journal of India, 14: 143-144.

Anonymous (1945) Role of syringes in the transmission of jaundice. A memorandum by medical officers of the Ministry of Health. The Lancet, 2:116119.

Anonymous (1998) Possible risk factor for acute hepatitis B infection, case control study, Qena and Luxor city fever hospitals, 1994. Epidemiology Bulletin-Egypt, 1:6-8.
Archibald LK, Ramos M, Arduino MJ et al. (1998) Enterobacter cloacae and Pseudomonas aeruginosa polymicrobial bloodstream infections traced to extrinsic contamination of a dextrose multidose vial. Journal of Pediatrics, 133:640-644.

Aylward B, Kane M, McNair-Scott R, Hu DJ (1995) Model-based estimates of the risk of human immunodeficiency virus and hepatitis B virus transmission through unsafe injections. International Journal of Epidemiology, 24: 446-452.

Bass A (2000) SIGNPOST: the safe injection global network internet forum listserve. Bulletin of the World Health Organization, 18:277.
Birungi H (1998) Injections and self-help: risk and trust in Ugandan health care. Social Science and Medicine, 47:1455-1462.
Bultreys M, Chao A, Habimana P et al. (1994) Incident HIV-1 infection in a cohort of young women in Butare, Rwanda. AIDS, 8:1585-1591.
Busch MP, Operskalski EA, Mosley JW et al. (1996) Factors influencing human immunodeficiency virus type 1 transmission by blood transfusion. Journal of Infectious Diseases, 174:26-33.
Cardo DM, Culver DH, Ciesielski CA et al. (1997) A case-control study of HIV seroconversion in health care workers after percutaneous exposure. New England Journal of Medicine, 337:1542-1543.

CDC (Centers for Disease Control and Prevention) (2001) Injection practices among nurses-Valcea, Romania, 1998. Morbidity and Mortality Weekly Report, 50:59-61.
CDC (1997) Recommendations for follow-up of health care workers after occupational exposure to hepatitis C virus. Morbidity and Mortality Weekly Report, 46:603-606.
CDC (1999) Frequency of vaccine-related and therapeutic injection-Romania 1998. Morbidity and Mortality Weekly Report, 48:271-274.

Chang SJ, Chen HC, Ying J, Lu CF, Ko YC (1996) Risk factors of hepatitis C virus infection in a Taiwanese aboriginal community. Kao Hsiung I Hsueh Ko Hsueh Tsa Chih, 12:241-247.
Chen TZ, Wu JC, Yen FS et al. (1995) Injection with nondisposable needles as an important route for transmission for acute community-acquired hepatitis C virus infection in Taiwan. Journal of Medical Virology, 46: 247-251.
Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M (1989) Isolation of a cDna clone derived from a bloodborne non-A, non-B viral hepatitis genome. Science, 244:359-362.

Davis LG, Weber DJ, Lemon SM (1989) Horizontal transmission of hepatitis B virus. The Lancet, 1:889-893.

Deivanayagam N, Nedunchelian K, Ahamed SS, Ashok TP, Mala N, Ratnam SR (1993) Intramuscular injection as a provoking factor for paralysis in acute poliomyelitis. A case control study. Indian Pediatrics, 30:335-340.
Dicko M, Oni AQ, Ganivet S, Kone S, Pierre L, Jacquet B (2000) Safety of immunization injections in Africa: not simply a problem of logistics. Bulletin of the World Health Organization, 78:163-169.
Drucker E, Alcabes PG, Marx PA (2001) The injection century: massive unsterile injections and the emergence of human pathogenes. The Lancet, 358: 1989-1992.

El-Sakka H (1997) Risk factors for HCV infection. Gharbia Governorate, Egypt, 1996-1997. (Presentation given at the EIS conference, April 1997.) Atlanta, GA.

Ferry B (1995) Risk factors related to HIV transmission. In: Sexual behaviour and AIDS in the developing world. Cleland J, Ferry B, eds. World Health Organization, Geneva.
Fisher-Hoch SP, Tomori O, Nasidi A et al. (1995) Review of cases of nosocomial Lassa fever in Nigeria: the high price of poor medical practice. British Medical Journal, 311:857-859.
Flaskerud JH, Nyamathi AM (1996) Home medication among Latina women in Los Angeles: implications for health education and prevention. AIDS Care, 8:95-102.

Fontaine O, Diop B, Beau JP (1984) La diarrhee infantile au Senegal [Childhood diarrhoea in Senegal]. Medicina Tropical, 44:27-31.

Frank C, Mohamed MK, Strickland GT (2000) The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. The Lancet, 355:887-891.

Freeman AJ, Dore GJ, Law MG et al. (2001) The natural history of chronic hepatitis C: a systematic review. (Unpublished manuscript communicated by Dr Gregory Dore, April 2001.) University of New South Wales, Sydney.
Gay NJ, Edmunds WJ, Bah E (2001) Estimating the global burden of hepatitis B. (Unpublished manuscript prepared for the Global Burden of Disease study, April 2001.) World Health Organization, Department of Vaccines and Biologicals, Geneva.
Grosskurth H, Mosha F, Todd J et al. (1995) A community trial of the impact of improved sexually transmitted disease treatment on the HIV epidemic in rural Tanzania. 2. Baseline survey results. AIDS, 9:927-934.
Guyer B, Bisong AAE, Gould J, Brigoud M, Aymard M (1980) Injections and paralytic poliomyelitis in tropical Africa. Bulletin of the World Health Organization, 58:285-291.
Hersh BS, Popovici F, Jezek Z et al. (1993) Risk factors for HIV infection among abandoned Romanian children. AIDS, 7:1617-1624.

Ho M-S, Hsu C-P, Yuh Y et al. (1997) High rate of hepatitis C virus infection in an isolated community: persistent hyperendemicity or period-related phenomena? Journal of Medical Virology, 52:370-376.
Hofmann CA (2001) Presentation made at the annual meeting of the Safe Injection Global Network. World Health Organization, Geneva. (Unpublished document WHO/BCT/DCT/ 01.04; available on request from the Department of Essential Health and Technologies, World Health Organization, 1211 Geneva 27, Switzerland.)
Hsu SC, Chang MH, Ni YH, Hsu HY, Lee CY (1993) Horizontal transmission of hepatitis B virus in children. Journal of Pediatric Gastroenterology and Nutrition, 16:66-69.

Hussain R (2001) A case-control study of risk factors for acute hepatitis B virus infection in Karachi, Pakistan [Dissertation]. Aga Khan University, Karachi.

Hutin YJF, Chen RT (1999) Injection safety: a global challenge. Bulletin of the World Health Organization, 77:787-788.

Hutin YJF, Harpaz R, Drobeniuc J et al. (1999) Injections given in healthcare settings as a major source of acute hepatitis B in Moldova. International Journal of Epidemiology, 28:782-786.
Janszen E, Laning W (1993) Injections: more expensive, more effective and faster. A report from Ecuador. In: The impact of injections on daily medical practice. Bloem M, Wolfers I, eds. VU University Press, Amsterdam.
Kane A, Lloyd J, Zaffran M, Simonsen L, Kane M (1999) Transmission of hepatitis B , hepatitis C and human immunodeficiency viruses through unsafe injections in the developing world: model-based regional estimates. Bulletin of the World Health Organization, 77:801-807.

Kaninda AV (2001) Frequency of injections worldwide: market data analysis and injections given for selected purposes. (Unpublished WHO report, Department of Blood Safety and Clinical Technology.) World Health Organization, Geneva.
Khan AJ, Luby SP, Fikree FF et al. (2000) Unsafe injections and the transmission of hepatitis B and C in a periurban community in Pakistan. Bulletin of the World Health Organization, 78:956-963.

Ko YC, Chung DC (1991) Transmission of hepatitis B virus infection by iatrogenic intramuscular injections in an endemic area. Kao Hsiung I Hsueh Ko Hsueh Tsa Chih, 7:313-317.

Ko YC, Chung DC, Pai HH (1991b) Intramuscular-injection-associated gluteal fibriotic contracture and hepatitis B virus infection among school children. Kao Hsiung I Hsueh Ko Hsueb Tsa Chih, 7 :358-362.
Ko YC, Li SC, Yen YY, Yeh SM, Hsieh CC (1991a) Horizontal transmission of hepatitis B virus from siblings and intramuscular injection among preschool children in a familial cohort. American Journal of Epidemiology, 133:1015-1023.

Kosen S (1999) Assessment on the safety of EPI injection practices in four provinces of Indonesia. (Unpublished document.) Health Services and Development Center, Ministry of Health, Indonesia.

Lakshman M, Nichter M (2000) Contamination of medicine injection paraphernalia used by registered medical practitioners in south India: an ethnographic study. Social Science and Medicine, 51:11-28.
Logez S (2001) Accessibility to injection equipment: impact on injection safety. (WHO Mission Report.) Pharmaciens Sans Frontières Comité International, Clermont-Ferrand.

Luby SB, Hoodboy F, Aziz M, Shah A (forthcoming) Long term improvement of unsafe injection practices following community intervention.
Luby SP, Qamruddin K, Shah AA et al. (1997) The relationship between therapeutic injections and high prevalence of hepatitis C infection in Hafizabad, Pakistan. Epidemiology and Infection, 119:349-356.

Maynard JE, Kane MA, Hadler SC (1989) Global control of hepatitis B through vaccination: role of hepatitis B vaccine in the Expanded Programme on Immunization. Reviews of Infectious Diseases, 11:S574-578.

McMahon BJ, Alward WL, Hall DB, Heyward WL, Bender TR, Francis DP (1985) Acute hepatitis B virus infection: relation of age to the clinical expression of disease and subsequent development of the carrier state. Journal of Infectious Diseases, 151:599-603.
Mohamed MK, Hussein MH, Massoud AA et al. (1996) Study of the risk factors for viral hepatitis C infection among Egyptians applying for work abroad. Journal of the Egyptian Public Health Association, 71:113-142.
Molina JM, Ferchal F, Chevret S et al. (1994) Quantification of HIV-1 virus load under ziovudine therapy in patients with symptomatic HIV infection: relation to disease progression. AIDS, 8:27-33.

Mujeeb SA, Shahad S, Hyder AA (2000) Geographical display of health information: study of hepatitis C infection in Karachi, Pakistan. Public Health, 114:413-415.

N'Galy B, Ryder RW, Bila K et al. (1988) Human immunodeficiency virus infection among employees in an African hospital. New England Journal of Medicine, 319:1123-1127.
Narendranathan M, Philip M (1993) Reusable needles—a major risk factor for acute virus B hepatitis. Tropical Doctor, 23:64-66.
Otu AA (1987) Hepatocellular carcinoma, hepatic cirrhosis and hepatitis B virus infection in Nigeria. Cancer, 60:2581-2585.
Pape JW, Liautaud B, Thomas F et al. (1985) The acquired immunodeficiency syndrome in Haiti. Annals of Internal Medicine, 103:674-678.

Potterat JJ, Rothenberg RB, Muth SQ (1999) Network structural dynamics and infectious disease propagation. International Journal of STD and AIDS, 10:182-185.

Prawitasari Hadiyono JE, Suryawati S, Danu SS, Sunartono, Santoso B (1996) Interactional group discussion: results of a controlled trial using a behavioural intervention to reduce the use of injections in public health facilities. Social Science and Medicine, 42:1177-1183.

Priotto G, Ruiz A, Kyobutungi C (2001) Injection use in the population of Mbarara district, Uganda, cross sectional population survey. Epicentre, Paris.

Quigley MA, Morgan D, Malamba SS (2000) Case-control study of risk factors for incident HIV infections in rural Uganda. Journal of Acquired Immune Deficiency Syndromes, 23:418-425.

Quigley MA, Munguti K, Grosskurth H et al. (1997) Sexual behaviour patterns and other risk factors for HIV infection in rural Tanzania: a case-control study. AIDS, 11:237-248.
Reeler AV (1990) Injections: a fatal attraction. Social Science and Medicine, 31:1119-1125.
Reeler AV, Hematorn C (1994) Injection practices in the third world. World Health Organization, Geneva. (Unpublished document WHO/DAP/94.8; available on request from the Drug Action Programme, World Health Organization, 1211 Geneva, Switzerland.)

Rothenberg RB, Sterk C, Toomey KE et al. (1998) Using social network and ethnographic tools to evaluate syphilis transmission. Sexually Transmitted Diseases, 25:154-160.
Schnurr A, Xiaojun W, Jingjin Y (1999) Sterilizable syringes and needles use in China. (Presentation made at the TechNet meeting, December 1999 in Harare, Zimbabwe.)
Seeff LB, Wright EC, Zimmerman HJ, Alter HJ, Dietz AA, Felsher BF (1978) Type B hepatitis after needle-stick exposure: prevention with hepatitis B immune globulin. Annals of Internal Medicine, 88:285-293.

Senior JR, Sutnick AI, Goeser E, London WT, Dahlke MB, Blumberg BS (1974) Reduction of post-transfusion hepatitis by exclusion of Australia antigen from donor blood in an urban public hospital. American Journal of the Medical Sciences, 267:171-177.

Shaldon S (1995) Transmission of HIV in dialysis centres. The Lancet, 346:451.
Shikata T, Karasaw T, Abe K et al. (1977) Hepatitis B antigens and infectivity of hepatitis B virus. Journal of Infectious Diseases, 136:571.
Simard E, Wasley A, Pistol A, Popa M, Hutin Y, Mast E (2000) Surveillance for acute hepatitis B among children, Romania, 1997-1999. (Presentation given at the American Public Health Association meeting, December 2000.) Boston, MA.

Simonsen L, Kane A, Lloyd J, Zaffran M (1999) Unsafe injections in the developing world and transmission of bloodborne pathogens. Bulletin of the World Health Organization, 77:789-800.

Singh J, Bhatia R, Patnaik SK et al. (2000) Community studies on hepatitis B in Rajahmundry town of Andra Pradesh, India, 1997-1998: unnecessary therapeutic injections are a major risk factor. Epidemiology and Infection, 125:367-375.

Soeters R, Aus C (1989) Hazards of injectable therapy. Tropical Doctor, 19: 124-126.

Stoica A, Hutin YJF, Paun M, Mast EE, Margolis HS (1999) Attitudes of physicians regarding the use of therapeutic injections, Arges district, Romania. (Abstract presented at the annual meeting of the Society for Healthcare Epidemiology of America, April 1999.) San Francisco, CA.
Sun CA, Chen HC, Lu CF (1999) Transmission of hepatitis C virus in Taiwan: prevalence and risk factors based on a nationwide survey. Journal of Medical Virology, 59:290-296.
Sun CA, Chen HC, Lu SN et al. (2001) Persistent hyperendemicity of hepatitis C virus infection in Taiwan: the important role of iatrogenic risk factors. Journal of Medical Virology, 65:30-34.
Talaat M, Oun S, El-Halawani A, Bodenschatz C, Mahoney F (2001) A рори-lation-based survey of injection practices among Egyptians. (Proceeding of the regional FETP meeting), Amman.

Thuring EG, Joller-Jemelka HI, Sareth H, Sokhan U, Reth C, Grob P (1993) Prevalence of markers of hepatitis viruses A, B, C and of HIV in healthy individuals and patients of a Cambodian province. Southeast Asian Journal of Tropical Medicine and Public Health, 24:239-249.
Ugalde A, Homedes N (1988) Medicines and rural health services: an experiment in the Dominican Republic. In: The context of medicines in developing countries. van der Geest S, Whyte SR, eds. Kulwer Academic Publishers, Dordrecht.

UNAIDS (2000) Global HIV epidemic update: December 2000. UNAIDS, Geneva.

Val Mayans M, Hall AJ, Inskip HM et al. (1990) Risk factors for transmission of hepatitis B virus to Gambian children. The Lancet, 336:1107-1109.
van Staa A, Hardon A (1996) Injection practices in the developing world. World Health Organization, Geneva. (Unpublished document WHO/DAP/96.4; available on request from the Drug Action Programme, World Health Organization, 1211 Geneva, Switzerland.)
Villanueva A, Calderon RV, Vargas BA et al. (1997) Report on an outbreak of postinjection abscesses due to Mycobacterium abscessus, including management with surgery and clarithromycin therapy and comparison of strains by random amplified polymorphic DNA polymerase chain reaction. Clinical Infections Diseases, 24:1147-1153.
Vogt M, Lang T, Frosner G et al. (1999) Prevalence and clinical outcome of hepatitis C infection in children who underwent cardiac surgery before the implementation of blood-donor screening. New England Journal of Medicine, 341:866-870.
Vong S, Perz J, Hutin Y, Drobeniuc J, Bell BP, 1998 International Field Epidemiology Course Participants (2002) Determinants of high frequency of therapeutic injections, Chisinau, Republic of Moldova. (Abstract presented at the 51st Epidemic Intelligence Service conference, April 2002) Centers for Disease Control and Prevention, Atlanta, GA.
Wasley A, Alter MJ (2000) Epidemiology of hepatitis C: geographic differences and temporal trends. Seminars in Liver Disease, 20:1-16.

Wawer MJ, Sewankambo NK, Berkley S et al. (1994) Incidence of HIV-1 infection in a rural region of Uganda. British Medical Journal, 308:171-173.

WHO (1976) Ebola haemorrhagic fever in Zaire. Bulletin of the World Health Organization, 56:271-293.

WHO (1999) High frequency of therapeutic injections, Republic of Moldova. Weekly Epidemiological Record, 11:84-86.

WHO (2000c) "Aide Mémoire" for a national strategy for the safe and appropriate use of injections. World Health Organization, Department for Blood Safety and Clinical Technology, Geneva.
WHO (2000b) Blood safety for too few. (Press release 2000/25, April 7.) World Health Organization, Geneva.
WHO (2000a) Hepatitis C: global prevalence (Update). Weekly Epidemiological Record, 75:18-19.
WHO (2001) World health report 2001. World Health Organization, Geneva.
WHO (2002) Tool for the assessment of injection safety. World Health Organization, Geneva. (Unpublished document WHO/V\&B/01.30; available on request from the Department of Immunization, Vaccines and Biologicals, World Health Organization, 1211 Geneva 27, Switzerland.)

Williams CL, Mast EE, Coleman PJ et al. (1999) Sexual practices associated with hepatitis $C$ virus infection among non injecting-drug-using female prostitutes in the United States. (Program and abstracts of the 6th International Symposium on Hepatitis and Related Viruses: Molecular Virology and Pathogenesis, June 6-9.) National Institutes of Health, Bethesda, MD.
Wittet S (2001) Introducing hepatitis B vaccine. (Immunization focus, March 6-7.) Available at http://www.vaccinealliance.org.
Wyatt HV (1984) The popularity of injections in the third world: origins and consequences for poliomyelitis. Social Science and Medicine, 19:911-915.

Yerly S, Quadri R, Negro F et al. (2001) Nosocomial outbreak of multiple bloodborne viral infections. Journal of Infectious Diseases, 184:369-372.

## Chapter 23

# Child sexual abuse 

Gavin Andrews, Justine Corry, Tim Slade, Cathy Issakidis and Heather Swanston

## Summary

This chapter summarizes the evidence of a relationship between child sexual abuse and subsequent mental disorder. Child sexual abuse (CSA) typically includes unwanted and inappropriate sexual solicitation of, or exposure to, a child by an older person; genital touching or fondling; or penetration in terms of oral, anal or vaginal intercourse or attempted intercourse. CSA can vary along a number of dimensions including frequency, duration, age at onset, and relationship of victim to perpetrator. However, the most common dimension used to define CSA is type of abuse. Three categories are commonly reported in the literature. Noncontact abuse encompasses a range of acts and includes inappropriate sexual solicitation or indecent exposure. The two other categories are contact abuse, which includes touching or fondling, and intercourse, which includes oral, anal or vaginal intercourse. In this chapter the upper age limit used to define childhood was 18 years. The theoretical minimum exposure was defined as no abuse.

The disease outcomes chosen for the current analysis were depression, panic disorder, post-traumatic stress disorder (PTSD), alcohol abuse/ dependence, drug abuse/dependence and suicide attempts. Evidence for causality came from twin studies, prospective studies and representative community studies. In particular, the three twin studies available provided strong evidence of a causal relationship as they inherently controlled for the genetic and family environment factors that are also associated with mental disorders (Dinwiddie et al. 2000; Kendler et al. 2000; Nelson et al. 2002). These studies provide evidence of significant associations between CSA and depression, panic disorder, alcohol abuse/dependence, drug abuse/dependence and suicide attempts. While PTSD was not considered as an outcome in the twin studies, data from a prospective study (Silverman et al. 1996) and three representative community studies (Davidson et al. 1991; Molnar et al. 2001; Saunders
et al. 1999) provided consistent evidence that a strong association exists. There was insufficient evidence to support the relationship with obsessive-compulsive disorder (OCD) and so it was excluded. A number of other mental disorder outcomes that were not considered here have been linked to child sexual abuse. Eating disorders have long been conceptualized as a response to a dysfunctional family environment, of which child sexual abuse can be a part. A number of studies have also looked at the association between child sexual abuse and personality disorders, particularly antisocial and borderline personality disorder. Additionally, child sexual abuse does not only produce an increased risk of mental disorder. There has been anecdotal and experimental evidence suggesting that CSA increases the probability of negative psychological outcomes such as poor self-esteem, lack of a sense of control or agency, difficulties with intimacy and continuing sexual difficulties.

Review articles on the prevalence of child sexual abuse have commonly reported a range of prevalence anywhere from $2 \%$ to $62 \%$. Previous studies have demonstrated a significant difference in the prevalence of CSA, depending on a number of methodological factors, including method of data collection, number of questions used to assess CSA, definition of childhood and the type of sample assessed. Therefore, multiple linear regression analyses were employed separately for males and females to identify the methodological characteristics that significantly contribute to the variation in prevalence. Any unwanted variation was removed by adjusting individual prevalence estimates accordingly. These "adjusted" prevalence estimates were then broken down into the three levels of abuse (non-contact, contact and intercourse) and the eight age groups. Meta-analysis was then used to combine the estimates within each country. These country estimates were weighted by the population of the country and combined to provide the final subregional ${ }^{1}$ estimates. Prevalence estimates were higher in females than in males and varied across subregions with the highest prevalence estimates found in AFR-E and SEAR-D.

Relative risk estimates were derived from studies examining the relationship between CSA and psychiatric outcomes. Where relative risks were reported only for overall exposure to CSA, these were extrapolated into the three levels of abuse. Few studies adjusted for the confounding effects of family dysfunction. An external method of adjustment was applied to the relative risks from those studies that did not control for family dysfunction. The small number of studies available for analysis meant that separate relative risks for different age, sex and subregional groups could not be obtained. Relative risks were stratified by psychiatric outcome before being combined using meta-analysis. Results showed that the relative risks are not significantly different across types of mental disorder suggesting that CSA is not particularly associated with any one disorder. Additionally, across types of abuse there was a general trend for increased risk to be associated with "increased" exposure to CSA.

That is, as more severe forms of CSA were experienced the risks for developing a mental disorder increased. After external adjustment the contact and intercourse categories of abuse remained significant across most disorders but non-contact abuse became non-significant.

Across the world CSA contributed to between $4 \%$ and $5 \%$ of the burden of disease in males and between $7 \%$ and $8 \%$ of the burden of disease in females, for each of the conditions depression, alcohol abuse/dependence and drug abuse/dependence. The attributable fractions were higher for panic disorder ( $7 \%$ for males and $13 \%$ for females) and higher still for PTSD ( $21 \%$ for males and $33 \%$ for females). For suicide attempts attributable fractions were $6 \%$ for males and $11 \%$ for females. There were slight regional variations in the amount of burden that could be attributed to CSA, with AFR-E and SEAR-D having higher attributable fractions. Prevalence was estimated to be higher in these subregions. However, data for these subregions came from a few studies that were methodologically poor. The burden of disease attributed to CSA was greater in the younger age groups and declined in the older age groups. Since risk was assumed to be constant across age, this merely reflected the age distribution of the mental disorder disease burden, which impacts largely on the younger age groups due to its early onset and chronic nature. Avoidable burden would be the same as attributable since it was assumed that the prevalence of CSA does not change over time under a "business-as-usual" scenario.

## 1. Introduction

### 1.1 DEFINITIONS OF RISK FACTOR

Definitions of exposure to child sexual abuse vary. Patterns of interpersonal behaviour are being described, not the results of measuring a physical attribute like body mass index, hypertension or blood lead levels. In physical inactivity we accept the report of a person's habitual behaviour and calculate the health consequences of that behaviour to that person. In CSA we quantify one person's behaviour with a child, and usually have to rely on the retrospective report of that child when adult. We then estimate the health consequences of that occurrence. It is important to point out that, even in prospective studies, data on CSA are gathered retrospectively. It is unethical, and in many countries illegal, to prospectively identify CSA and not intervene. The problems of measurement have proven to be difficult but not insurmountable.

In its broadest sense CSA includes unwanted and inappropriate sexual solicitation of, or exposure to, a child by an older person (non-contact abuse), genital touching or fondling (contact abuse), and penetration in terms of oral, anal or vaginal intercourse or attempted intercourse (intercourse). Many studies have used a narrow definition of CSA to include contact abuse and intercourse only. Definitions of CSA also differ
depending on the cut-off age used to define childhood. While in most countries 18 years of age is the legal cut-off used to define childhood, in many countries the age of consent, especially for sexual activity, is lower. However, the most widely-reported definition of childhood in large population surveys of CSA is 18 years or less. Very few studies provided estimates of prevalence by different age groups and none provided estimates of mental disorder risk by different age groups. Furthermore, as described below, it has been shown that the first onset of CSA is less likely to occur between the ages of 15 and 18 years than in younger children. For these reasons the cut-off age used to define childhood in this chapter was set at 18 years.

### 1.2 Choice of exposure variable, reasons and implications

The causes of adult mental disorders have proven difficult to define. There is considerable evidence from longitudinal and twin studies that both genetic and environmental factors are implicated in different proportions in different disorders. Because we are dealing with human behaviour there is also evidence of substantial gene-environment interaction (Kendler et al. 2000; Rutter 1999). Advances in medicine generally have been simplified by the availability of animal models of the condition, which allow the biology to be explicated. Animal models of mental disorders just do not exist, and progress in understanding has to be made by association and inference and does not come from experimental paradigms.

It is simplistic to assume that genetic contributions are immutable. Few would claim that the genetic factors are, except in the case of Huntington chorea and some dementias, a full and sufficient explanation. Genetic factors act to enhance vulnerability. For example, the indicated prevention programmes in anxiety disorders almost certainly work by inhibiting the expression of that genetic vulnerability. Nevertheless, it is inherently more plausible to identify a risk factor that is purely environmental, as CSA is, when looking for a risk factor that might be avertible.

In broad terms, the risk factors for mental disorders can be grouped into vulnerability produced by temperament, by adversity and by deprivation. Trauma and CSA are examples of adversity; family dysfunction and neglect are examples of deprivation (Bryant and Range 1995; Kessler et al. 1997). Temperament, adversity and deprivation co-occur more often than is predicted by chance and, as these impact throughout childhood and adolescence in complex ways, their influence in adult functioning is not likely to be simple or linear (Mullen et al. 1996, 2000; Rutter 1999).

CSA is no exception to this rule. It is more frequent in situations in which the other factors are present. Fortunately, it is possible to estimate the independent contribution of CSA to adult mental disorders from both
twin studies on twin pairs discordant for CSA (where both genetics and family environment like deprivation and other adversity are controlled by virtue of the twin method), and longitudinal studies of young people growing up (where data were prospectively gathered on temperament, non-CSA adversity and deprivation) (Borowsky et al. 1999; Kendler et al. 2000; Molnar et al. 2001; Mullen et al. 1993, 1996; Yama et al. 1995; Zuravin and Fontanella 1999). The link between CSA and adult mental disorders has been established, even after controlling for these other determinants of adult mental disorders.

Two other issues make CSA an appropriate exposure variable to choose. First, it is not rare-most reviews have concluded that close to one child in six experiences an episode of CSA as defined using the broad definition above (Fergusson and Mullen 1999). Second, there is an extensive literature on CSA, much of it recent and much using the definitions presented above, which made a systematic review possible. The literature on deprivation or other adversity is nowhere near as extensive, coherent or accessible.

### 1.3 Choice of theoretical minimum

Given the nature of CSA and the way it is defined in this chapter, the only acceptable theoretical minimum is zero.

## 2. Estimating risk factor levels

### 2.1 Methods

The methods used to identify sources and studies for estimation of both risk factor levels and risk factor-disease relationships are presented in the following section.

### 2.2 Criteria for considering sources and studies

The following inclusion criteria were used:

- any study which determined the prevalence of childhood sexual abuse in any sample;
- any study which determined both the presence and absence of CSA and the subsequent presence and absence of our chosen outcomes;
- any review chapters or reports published in the last 10 years where the topic was CSA (books where the central topic was CSA were included regardless of year of publication);
- methodological papers to assist with the interpretation of the results; and
- meta-analyses of original research results.

The following exclusion criteria were then applied before collection of the articles:

- prevalence studies with total sample sizes of less than 100 (unless data was from an underrepresented country); and
- studies where the population was sampled on the basis of the presence of one of our chosen outcomes.

The first exclusion criterion was applied on the basis of the recommendation of Fergusson and Mullen (1999) and because samples of less than 100 may not provide accurate prevalence estimates for CSA. The second exclusion criterion was applied because studies that sampled on the basis of the presence or absence of an outcome could not be used to calculate a relative risk for that outcome (Streiner 1998). In epidemiological terms this means we could not determine the number of CSAexposed individuals who developed the outcome of interest. Instead, we determined the number of people with outcome A who were exposed to CSA, thus answering a different question. Case-control studies were therefore only included where "cases" were individuals exposed to CSA and "controls" were those who were not exposed.

### 2.3 Search strategy for identification of studies

Several strategies were used to locate studies for this chapter. First, computer searches of 16 databases were conducted. Databases searched were the following: Medline; Embase; Psychinfo; E-psych; Healthstar; Cinahl; Cochrane; Social Work Abstracts; Health \& Society; Family \& Society; General Science Abstracts; Cambridge Life Sciences; Family Studies; Dissertation Abstracts; Child Abuse; Child Welfare \& Adoption; Social Sciences Citation Index.

Subject heading (SH) and key word (KW) searches were carried out in two stages and were defined as follows:

Stage 1
SH: Child Sexual Abuse
OR
KW: "child*" (child, childhood, children) AND KW: ["sexual abuse", "sexual assault", "molestation", "incest"]

AND
Stage 2 (RISK FACTOR LEVELS)
SH: Epidemiology OR KW: ["epidemiology", "prevalence", "incidence"]

STAGE 2 (RISK FACTOR-DISEASE RELATIONSHIPS)
SH: Depression, OR KW: "depress"" (depression, depressive)
SH: Anxiety Disorders, OR SH: [Panic Disorder, Agoraphobia, Obsessive Compulsive Disorder, post-traumatic stress disorder] OR KW: ["panic disorder", "agoraphobia", "obsessive compulsive disorder", "OCD", "post-traumatic stress disorder", "PTSD"]

SH: [Substance Related Disorders, Alcoholism] OR KW: ["alcohol abuse", "alcohol dependence", "alcohol use", "substance abuse", "substance dependence", "substance use", "drug abuse", "drug dependence", "drug use"]

SH: [Suicide, Suicide Attempted], OR KW: ["suicid" " (suicide, suicidal)]
These searches generated a list of over 12000 studies, of which approximately 4000 were duplicates. The abstracts of the remaining 8000 were examined to isolate potentially appropriate studies. The inclusion criteria were applied and the number of potentially relevant articles was reduced to approximately 1000. Many of the articles excluded at this step were those that either did not measure CSA separately from other types of abuse, or did not include a non-exposed control group.

After the exclusion criteria were applied to the 1000 articles, an initial sample of 460 articles was identified for collection. As the articles were reviewed the reference lists were examined in an attempt to uncover additional studies. The tables of contents of Child Abuse and Neglect, the leading journal in the area of child sexual abuse, were searched to locate any articles missed in searches. Experts in the area were contacted for information on unpublished data or for data from countries underrepresented in the usual databases and journals. These experts were located through the membership directory of the International Society for Prevention of Child Abuse and Neglect (ISPCAN), an organization sponsored by UNICEF-NY, Child Protection Division. The Society has over 300 members. All members were contacted via email or fax, with the exception of those from countries where we had adequate published data. A request for data was also put out on a "child maltreatment researchers" list on the Internet which yielded further information. The researchers had access to translation services so articles were not excluded on the basis of language.

The final database consisted of 604 articles. Of these, a further 91 were excluded after collection on the basis of the exclusion criteria defined above (CSA not reported separately from other types of abuse or from abuse as an adult $n=33$; subjects sampled on the basis of outcome $n=33$; no non-exposed control group $n=11$; sample size too small $n=11$; individual case study data only $n=3$ ). In these cases it was not possible to determine from the abstract alone whether the study
should be included. This produced a final set of 513 articles or reports. They were as follows:

- 103 reviews or meta-analyses;
- 55 methodological papers;
- 179 coded studies;
- 48 studies not coded as data had been collected from another paper or report;
- 52 prevalence studies not coded as data on CSA were derived from a secondary source such as official reports or records;
- 13 studies not coded as subjects were gathered from a special sample;
- 48 studies not coded as the outcome data were either not adequately measured or outcome was measured using a continuous as opposed to a categorical measure; and
- 15 studies not coded, as we were unable to obtain copies of the relevant papers or reports.

The 179 coded studies were coded for risk factor levels, risk factor-disease relationships or both.

### 2.4 Methods for obtaining estimates where more than one data source exists

More than one estimate was available for each subregion. Therefore a meta-analysis was conducted to combine estimates. However, before this was carried out the prevalence estimates were adjusted to remove any differences that can be explained by variations in methodology employed in the studies.

## Coding of studies

Prevalence of CSA was obtained from each study included in this analysis. Where possible, an overall prevalence of any CSA was coded as well as a breakdown into the mutually exclusive groups of non-contact, contact and intercourse. A number of methodological characteristics were also coded from each study. These included the type of sample, the representativeness of the sample, age and sex distributions, the method of data collection employed, how childhood was defined, survey response rate, how many questions were used to elicit the presence of CSA and whether restrictions were placed on the definition of abuse.

## Correcting for methodological variability within studies

Previous studies have noted that the prevalence of CSA differs depending on a number of methodological factors, including method of data collection, number of questions used to assess CSA, definition of child-
hood and the type of sample assessed (Bolen and Scannapieco 1999; Gorey and Leslie 1997; Haugaard and Emery 1989; Wynkoop et al. 1995). In order to examine the methodological characteristics that influence the variability in prevalence, multiple linear regression analyses were conducted. As it was assumed that the impact of these factors may differ for males and females, separate regression models were fitted for all estimates of prevalence for males ( $N=93$ estimates) and all estimates of prevalence for females ( $N=143$ estimates). Too few data points existed to carry out the regression analyses in each of the levels of exposure. Thus, the independent variable was prevalence of any CSA (coded as either broad or narrow definition).

## Model-building steps

1. Each independent variable was screened for adequate cell sizes as well as its relationship with all other independent variables. Cells were collapsed where necessary and choices were made between two variables when they were significantly related (Table 23.1 displays the variables chosen for the regression analysis and their categories).
2. The distribution of the dependent variable was tested for skewness, kurtosis and univariate outliers. Three outliers were removed from the analysis.
3. All variables were tested univariately and then entered together into a single model. A process of backward elimination was employed to remove the least significant variable at each step until all variables left in the final model were statistically significant at the 0.05 level. As shown in Table 23.1 two of the dependent variables entered into the model were "subregion" and "CSA definition". Variability in prevalence across different World Health Organization (WHO) regions and across the broad and narrow definitions of CSA was considered to be important. Thus, these variables were left in the multivariate model even if they were non-significant.

## Results

The final model examining the multivariate effects of methodological factors accounted for $27 \%$ of the variance in prevalence of CSA in males and $22 \%$ of the variance in prevalence of CSA in females (beta weights and significance levels for the statistically significant variables in the multivariate model are shown in Table 23.1). The variables remaining in the "males" model were "subregion" $(F=0.68, d f=3, P=0.5688)$, "CSA definition" $(F=1.98, d f=1, P=0.1637)$ and "sample type" $(F=6.13$, $d f=4, P=0.0002$ ). These results indicate that studies of male college samples and general practitioner (GP) attendees report a significantly higher overall prevalence of CSA than studies of community samples ( $12 \%$ and $12 \%$ vs $6 \%$ ). The variables remaining in the "females" model

Table 23.1 Characteristics of the dependent variables used in the regression analysis

| Dependent variable | Categories (for categorical variables) | Variables left in final model for males |  | Variables left in final model for females |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Beta weights | P -value | Beta weights | P -value |
| Subregion | $\mathrm{I}=$ AMR-A | 0.00 | - | 0.00 | - |
|  | $2=$ EUR-A | -1.16 | 0.4616 | -8.18 | 0.0002* |
|  | 3 = WPR-A | 0.95 | 0.6499 | 1.57 | 0.5897 |
|  | $4=$ All other subregions | 3.63 | 0.2988 | 5.75 | 0.2553 |
| CSA definition | I = Broad (non-contact, contact or intercourse) | 0.0 | - | 0.00 | - |
|  | 2 = Narrow (contact or intercourse) | -1.70 | 0.1637 | -4.60 | 0.0247* |
| Sample type | I = Community volunteers | 0.00 | - | Null | Null |
|  | 2 = College students | 4.52 | 0.0126* |  |  |
|  | 3 = School students | -1.54 | 0.1982 |  |  |
|  | $4=$ GP attendees | 5.77 | 0.0145* |  |  |
|  | $5=$ All other samples | 0.69 | 0.7509 |  |  |
| Number of questions | $\mathrm{I}=$ One | Null | Null | -7.72 | 0.0003* |
|  | $2=$ More than one |  |  | 0.00 | - |
| Definition of childhood | $1=<18$ years old | Null | Null | Null | Null |
|  | $2=<16$ years old |  |  |  |  |
|  | $3=<14$ years old |  |  |  |  |
|  | $4=$ Not reported |  |  |  |  |
| Restrictions on CSA definition | I = None | Null | Null | Null | Null |
|  | 2 = Coercion only |  |  |  |  |
|  | 3 = Age only |  |  |  |  |
|  | 4 = Age or coercion |  |  |  |  |
| Sample size | Continuous variable | Null | Null | Null | Null |
| Response rate | Continuous variable | Null | Null | Null | Null |
| Current age | Continuous variable | Null | Null | Null | Null |

Null Non-significant.

- Reference category.
* Significant at the $P<0.05$ level.
were "subregion" $(F=7.18, d f=3, P=0.0002)$, "CSA definition" $(F=$ 5.17, $d f=1, P=0.0247$ ) and "Number of questions" $(F=13.65, d f=$ $1, P=0.0003$ ). These results indicate reported prevalence is lower among women in some European countries compared to North America (15\% vs $22 \%$ ), among studies that employed a narrow definition of CSA
(19\%) than those that employed a broad CSA definition (23\%), and those which used only one question to assess the presence of CSA ( $14 \%$ ) than those with more than one question ( $23 \%$ ).

The finding that CSA definition independently contributed to variance in prevalence for females but not males is unusual. The difference between the broad and narrow prevalence estimates is wholly explained by whether or not non-contact forms of abuse are included in the definition. Therefore, if non-contact abuse is rare as the only form of abuse experienced in males then it is likely that including non-contact abuse in a definition of CSA in males would not substantially alter the overall prevalence. However, the pattern across types of abuse for males does not indicate this to be the case ( $40 \%$ of all CSA in males is non-contact CSA). A more plausible explanation is that the low overall prevalence in males coupled with the smaller number of estimates and their substantial variance contributed to the lack of effect observed for different CSA definitions in males. One might expect that had more estimates been available in males a significant difference may have emerged.

The finding that studies that included more than one question about abuse yielded higher prevalence rates is widely supported in the CSA literature (Fergusson and Mullen 1999; Mullen et al. 2000; Peters et al. 1986; Plunkett and Oates 1990); it is also reported that the higher rates observed in such studies are likely to be more accurate (Bolen and Scannapieco 1999). Within this context, adjusting the prevalence estimates from studies that used only one question to more closely reflect those that used multiple questions (rather than the other way around) is likely to be a more accurate reflection of true prevalence in the population. Once again, the lack of effect in male estimates may be a function of the smaller number of estimates, especially given that the number of questions did significantly predict prevalence in males in the univariate analyses.

More intriguing is the finding that the prevalence in females was lower in EUR-A countries than in North America, from where the majority of the world's prevalence estimates come. Finkelhor (1994) reported on the international epidemiology of CSA and concluded that between-country differences were more likely to be due to methodological differences than reflective of true differences in prevalence. However, the fact that the effect in the present analysis remains after controlling for methodological factors may reflect true differences between these cultural groups. Moreover, in light of the larger number of studies available for the present analysis, it does appear to be true that international differences in prevalence exist. Within this context, it is important to note that there were substantial variations in CSA prevalence within subregions even when controlling for methodological variations, and this is reflected in the modest proportion of explained variance for the final model $(22 \%)$. Therefore, the lower observed prevalence may, in fact, be a function of a non-examined explanatory variable. Moreover, a substantial number
of subregions were either underrepresented or not represented at all in this analysis, and it is therefore difficult to interpret these findings in a broader international context.

## Weighting of prevalence estimates

The results of the above regression analyses were used to weight the overall CSA prevalence estimates for both males and females. Specifically, the unstandardized regression coefficients from the two final models were used to adjust the raw prevalence estimates. This was achieved by subtracting the coefficients from the prevalence for all levels of a variable that were statistically significant (e.g. if the prevalence in males was derived from a college sample the adjusted prevalence would be the raw prevalence minus 4.52 percentage points). The variables of "subregion" and "CSA definition" did not contribute to this adjustment because they do not reflect differences in methodological quality. This produced the desired effect of controlling for any influence these factors had on other variables while at the same time keeping any variance in prevalence due to these factors.

It is worth noting that implicit in this process of adjustment was the assumption that estimates from the group used as the reference group more closely reflected the true population prevalence. Thus, estimates for males from college and GP samples were adjusted to more closely reflect the prevalence observed in community samples. For females, prevalence estimates were adjusted to more closely reflect those observed in studies where more than one question to define CSA was asked.

## EXTrapolation across LEVELS OF EXPOSURE

Once weighted prevalence estimates were obtained for each study they were divided into the mutually exclusive groups of non-contact, contact and intercourse using the following method. The relationship between each level of exposure, expressed as a proportion of the overall prevalence, was calculated for all studies that reported prevalence of CSA by type of exposure. When a study reported prevalence based on a broad definition, this was apportioned into non-contact CSA, contact CSA and intercourse (eight studies for males, 12 studies for females). When a study reported prevalence based on a narrow definition this was apportioned into contact CSA and intercourse ( 12 studies for males, 20 studies for females). These proportions were then applied to those studies only reporting an overall prevalence to obtain an estimate of prevalence for each level of exposure in each study. A small number of studies reported proportions in the three levels of exposure that were opposite in direction to all other studies, and these were excluded from the present calculations. Sensitivity analyses were carried out to assess the impact of these two apportioning fractions (e.g. including all studies, excluding studies where the proportion was opposite to all other studies) on the prevalence in each of the three levels of exposure. The results of these

Table 23.2 The impact of different apportioning fractions on the prevalence of CSA in the three levels of exposure

|  | Method 1 |  |  | Method 2 |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
|  | Apportioning fraction | Prevalence |  | Apportioning fraction | Prevalence |
| Males |  |  |  |  |  |
| Non-contact CSA | 0.297 | 2.6 |  | 0.387 | 3.1 |
| Contact CSA | 0.448 | 4.0 |  | 0.378 | 3.7 |
| Intercourse | 0.255 | 2.0 |  | 0.235 | 1.9 |
| Females |  |  |  |  |  |
| $\quad$ Non-contact CSA | 0.279 | 6.7 | 0.291 | 6.8 |  |
| Contact CSA | 0.500 | 12.7 | 0.512 | 13.2 |  |
| Intercourse | 0.221 | 5.8 | 0.197 | 5.3 |  |

Method I: Proportions derived from all studies ( $n=12$ for males, $n=20$ for females).
Method 2: Proportions derived from all studies ( $n=10$ for males, $n=17$ for females) except those in opposite direction to expected proportions (this method was used in the final calculations).
analyses are shown in Table 23.2. The overall impact of these excluded studies on each level of exposure was minimal. They were therefore not included in the calculation of this apportioning fraction.

The observed relationship between non-contact, contact and intercourse for both males and females indicated that intercourse was the least common form of abuse, a finding that has been widely reported (Fergusson and Mullen 1999; Mullen et al. 2000).

## Extrapolation of combined estimates across age

Exposure to CSA, by definition, occurs in childhood. Thus, during childhood the prevalence of CSA represents current exposure, and reflects cumulative exposure to CSA from birth until current age. The prevalence of CSA in persons aged $\geq 18$ years necessarily represents past exposure and reflects cumulative exposure from birth until the age of 18 years. Therefore prevalence of CSA will vary across age groups from $0-17$ years dependent on both the age at which exposure to CSA usually begins (age at onset) and the duration of CSA throughout childhood. Because exposure is fixed at the age of 18 years, the prevalence of CSA across age groups after 18 years will vary only if the prevalence of CSA is changing over time. For example, if birth cohorts between 1956 and 1970 (current age 30-44 years) were more likely to be exposed to CSA in childhood than birth cohorts before 1956 (current age $\geq 45$ years) then the prevalence of CSA for these age groups would differ.

The age groups required for reporting by WHO therefore necessitated three separate steps for calculation of exposure across age groups. The first step estimated prevalence in the $0-4-$, $5-14-$, and $15-17$-year age groups. The second step estimated prevalence in the 18-29-, 30-44-, $45-59-, 60-69-, 70-79$ - and $\geq 80$-year age groups. The third step esti-
mated a combined prevalence in the 15-17- and 18-29-year age groups to obtain prevalence for the age group 15-29 years.

## Step 1: Estimating prevalence in persons aged 0-17 years

No data exist on the prevalence of CSA in different childhood age categories. Age is usually examined in terms of the age at first onset of the abuse or the duration of abuse. In order to estimate prevalence of CSA across different childhood age categories, all studies that contained information about onset or duration were examined. The main difficulty arose from the use of disparate age categories that did not always conform to those required for the WHO estimates. The following steps were carried out to examine this issue.

Age at onset of abuse. A total of 22 studies presented data on age at onset of abuse. Although categorization of age varied considerably across studies, onset was consistently more prevalent in the $5-14$-year age group. This pattern was the same for both males and females. It is noted that the validity of self-reported abuse with an onset before the age of 5 years should be considered, at best, speculative. Age at onset for different levels of exposure and between different subregions could not be examined with the available data. Therefore, the following procedures were applied equally to males and females, across all levels of exposure and across all subregions.

- Two studies presented data in the $0-4$-year age group. These studies indicate that approximately $6.5 \%$ of abuse begins in this age group. If the age band is extended to include $0-5$ year olds (six studies in total) then the equivalent estimate is $9.7 \%$. The midpoint of these two values indicates that approximately $8.1 \%$ of all cases of CSA have their onset before the age of 5 years.
- Three studies presented data for those aged $\geq 15$ years. These studies show that approximately $19.7 \%$ of abuse begins in this age group. If the age band is restricted to those aged $\geq 16$ years (six studies in total) then the equivalent estimate is $18.8 \%$. The midpoint of these two values indicates that approximately $19.3 \%$ of all cases of CSA have their onset after the age of 14 years.
- The middle age group was derived from the above calculations to give a value of $72.6 \%(100 \%-8.1 \%-19.3 \%)$, which is consistent with the overall pattern observed.

Duration of abuse. If prevalence of CSA in different childhood age groups is based only on the age at onset, then abuse that begins in one age group and continues into the next will not be counted. One way to account for this is to adjust for duration, that is, to include a proportion of cases with long duration in more than one age category.

Finkelhor (1979) estimated that $16 \%$ of those who experience CSA experience it on more than one occasion and for a duration of more than one week. Bagley and Mallick (2000) estimated this figure at approximately $20 \%$. However, of most interest to the current analysis were the studies that reported the number of CSA cases where duration was at least one year. Two studies with such estimates were found (Collings 1997; Risin and Koss 1987), and they placed the prevalence of CSA cases with a duration of more than one year at $6.3 \%$ and $12 \%$ of all CSA cases. The midpoint of these values indicated that approximately $9.2 \%$ or one in ten cases of CSA would continue for more than one year (Risin and Koss 1987). Finkelhor (1979) further examined duration by level of exposure and indicated that non-contact abuse was the least likely to continue for more than one year while intercourse was the most likely ( $4.3 \%$ for noncontact, $5.8 \%$ for contact and $8.8 \%$ for intercourse, representing respective proportions of $0.68,0.92$ and 1.40 , compared to the overall figure of $6.3 \%$ ).

Combining onset and duration. Weights derived from age at abuse ( $0.081,0.726$ and 0.193 ) were combined with information about duration in the following way: $9.2 \%$ of cases in the $0-4$-year age group were carried over to the $5-14$-year age group using the above proportions for each level of abuse $(0.68,0.92$ and 1.40). The same calculations were made in carrying over cases from the 5-14-year age group to the 15-18year age group.

Step 2: Estimating prevalence in persons aged $>18$ years
For the reasons outlined earlier, prevalence of CSA in adults will only vary if the prevalence of CSA is changing over time. Several reviews have attempted to examine cohort effects in order to address this issue (Bagley 1990, 1995; Bagley and Ramsay 1985; Bickerton et al. 1991; Feldman et al. 1991; Fergusson et al. 2000). Three of these reviews concluded that the prevalence of CSA could be increasing over time while three reviews also concluded that there is no evidence to support a change in prevalence over time. This pattern is not explained by the publication dates of these reviews and many of the authors also pointed out that it is difficult to interpret these results without reference to a potential reporting phenomenon. That is, women in older age groups may be less willing or less able to report experiences of sexual abuse in childhood.

This issue was also examined empirically in the current data set. Each estimate for males and females was assigned a year of birth, calculated by subtracting the mean age of the sample from the year of publication (or where available the year the survey was conducted). This variable was then examined in a linear regression to determine whether year of birth explained any of the variance in prevalence. The continuous year of birth variable was also converted to a categorical variable with birth cohort defined according to the current age categories provided by

WHO. Birth cohort did not explain any variance in prevalence estimates for either males or females.

Given both these findings indicated no clear trend in prevalence of CSA over time, it was decided that estimates of prevalence would be combined across adult age groups. This had the effect of providing more stable estimates, particularly from underrepresented subregions.

Step 3: Combining estimates for 15-17 and 18-29 year olds
In order to combine estimates of prevalence in these two groups, it was necessary to determine the proportion of the population that fell into these two age groups for each subregion. This was calculated for each subregion represented in the data and was done separately for males and females. The data were obtained from the estimates provided by WHO and were based on population figures for the year 2000. As these data were only available for 15-19 year olds, as opposed to 15-17 year olds, population proportions were calculated using this age group.

## Combining estimates within countries

Once estimates of prevalence were apportioned into the three levels of exposure and across the eight separate age groups (yielding 24 separate prevalence estimates for each study), prevalences were combined within countries. The prevalence estimates were combined using meta-analysis with STATA Intercooled 7. For ease of calculation the "meta" macro was used (Sharp and Sterne 1997). Heterogeneity between studies within each country was tested using the chi-squared statistic. When only two or three studies were available for combination, the between-study variance was estimated with poor precision (Cooper and Hedges 1994). Countries with less than five estimates were combined using a fixed-effects model, and countries with five or more estimates, and statistical heterogeneity, were combined using a random-effects model.

## Combining estimates within subregions

In order to combine prevalence estimates between countries within each subregion, each country estimate was weighted for the population of that country. This meant that prevalence estimates from countries with large populations were given more weight in the final estimates. A combined estimate was obtained for each subregion by calculating a mean for each level of exposure in each age group for males and females.

### 2.5 Methods for obtaining estimates where no data existed

Extrapolation across subregions was one of the most difficult issues encountered in the construction of the prevalence estimates and arguably represents one of the greatest threats to the validity of the estimates in subregions where no data were available. For two out of the 14 subregions no prevalence data were found (EMR-B and EMR-D). These sub-
regions represent vastly different cultural, socioeconomic and geographic populations. In the absence of data it is impossible to speculate on how these differences might have impacted on the prevalence of CSA. No estimates were obtained for any countries in the Middle East. Therefore the estimate that was used comes from Turkey (EUR-B), which was considered the most appropriate in the absence of any other data. This extrapolation should be considered conjectural and the resultant estimates of prevalence in these subregions should be quoted with caution. It should also be noted that there was an uneven distribution of estimates in the remaining 12 subregions with a small number of countries making up a large proportion of the total number of estimates.

### 2.6 DESCRIPTION OF STUDIES, INCLUDING METHODOLOGICAL QUALITIES

Prevalence studies included in the analysis are presented in Table 23.3. They are presented in three levels according to the type of sample used and the representativeness of this sample. The levels are defined as follows.

Level A: Representative community samples-samples of adolescents or adults where the article explicitly stated that the sample was representative of the population from which it was drawn. In general this was achieved through the use of complex sampling procedures or weighting.

Level B: Other community samples (representativeness not known)samples of adolescents or adults where it was not known whether the samples were truly representative of the population from which they were drawn.

Level C: Community subgroups and convenience samples-samples of adolescents or adults drawn from a subgroup within the community based on factors such as ethnicity, educational or socioeconomic status.

### 2.7 Characteristics of excluded studies

There were 48 studies that were not coded due to their outcome measurement methods. Where the outcome was measured using a continuous measure for which there were established diagnostic cut-off points, authors were contacted and $2 \times 2$ tables were requested. The majority of authors responded and supplied data and these studies were included in the 179 coded studies. The 48 remaining were either not coded because authors could not supply data or authors were not contacted because the measure used could not be mapped to diagnostic criteria.

The characteristics of the 15 studies that could not be obtained are presented in Table 23.4. Inter-library loans were requested where items were only available interstate or overseas but several articles had not arrived at the time this chapter was being compiled.

Table 23.3 Characteristics of studies included in prevalence analysis

| Study authors | Sample |  |  |  | CSA data collection |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Type | Number | $\begin{gathered} \% \\ \text { female } \\ \hline \end{gathered}$ | Mean age (years) | Method | No. of questions | Childhood definition |
| AFR-D (Cameroon $n=1$ ) |  |  |  |  |  |  |  |
| Level C: Community subgr Menick and Ngoh (1998) | oups and conveni <br> School students | ce sample 1688 | 54 | 12 | Self-report | >1 | <15 |
| AFR-E (Ethiopia $n=1$, South Africa $n=3$ ) |  |  |  |  |  |  |  |
| Level B: Other community Mulugeta et al. (1998) | samples (represe School students | tiveness 719 | ot know 100 | n) 16 | Self-report | >1 | <18 |
| Level C: Community subgroups and convenience samples |  |  |  |  |  |  |  |
| Collings (1991) | College students | 284 | 0 | 20 | Postal SR | $>1$ | $<18$ |
| Madu and Peltzer (2000) | School students | 414 | 52 | 19 | Self-report | >1 | $<18$ |
| AMR-A (Canada $n=7$, USA $n=72$ ) |  |  |  |  |  |  |  |
| Level A: Representative community samples |  |  |  |  |  |  |  |
| Davidson et al. (1991) | Adults | 2985 | 54 | 42 | Face-to-face | >1 | $<16$ |
| Vogeltanz et al. (1999) | Adults | 733 | 100 | - | Face-to-face | >1 | $<18$ |
| Molnar et al. (2001) | Adults | 5877 | 50 | 35 | Self-report | >1 | $<18$ |
| Siegel et al. (1987) | Adults | 3132 | 53 | 42 | Face-to-face | >1 | $<16$ |
| Finkelhor et al. (1990) | Adults | 2626 | 56 | - | Telephone | >1 | $<18$ |
| MacMillan et al. (1997) | Adults | 9953 | 51 | 40 | Face-to-face | >1 | $<18$ |
| Wyatt (1985) | Adults | 248 | 100 | 27 | Face-to-face | $>1$ | $<18$ |
| Wyatt et al. (1999) | Adults | 338 | 100 | 30 | Face-to-face | >1 | $<18$ |
| Saunders et al. (1999) | Adults | 4008 | 100 | 45 | Telephone | >1 | $<18$ |
| Brown et al. (1999) | Adults | 639 | 48 | 18 | Face-to-face | 1 | $<18$ |
| Finkelhor and Dziuba-Leatherman (1994) | Adolescents | 2072 | 48 | 13 | Telephone | >1 | $<17$ |
| Boney-McCoy and Finkelhor (1996) | Adolescents | 1457 | 47 | 13 | Telephone | >1 | $<16$ |
| Kilpatrick et al. (2000) | Adolescents | 4023 | 49 | 15 | Telephone | >1 | $<17$ |
| Risin and Koss (1987) | College students | 2972 | 0 | 21 | Self-report | >1 | <14 |
| Fromuth and Burkhart (1989) | College students | 253 | 0 | 20 | Self-report | >1 | $<17$ |
| Harrison et al. (1997) | School students | 122824 | 51 | 15 | Self-report | >1 | $<18$ |
| Bagley et al. (1995) | School students | 2112 | 49 | 15 | Self-report | >1 | <18 |
| Blum et al. (1988) | School students | 36283 | 49 | 15 | Self-report | 1 | $<18$ |
| American School Health Association (1989) | School students | 3490 | 50 | 13 | Self-report | 1 | $<18$ |
| Nelson et al. (1994) | School students | 2332 | 51 | 16 | Self-report | >1 | <18 |
| Hernandez (1992) | School students | 3179 | 48.3 | 14 | Self-report | >1 | <15 |
| Bensley et al. (1999) | School students | 4790 | 48 | 16 | Self-report | 1 | $<18$ |


|  | Prevalence in males (\%) |  |  |  |  | Prevalence in females (\%) |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Restriction on CSA definition | Any broad | Any narrow | Noncontact only | Contact only | Intercourse | Any broad | Any narrow | Noncontact only | Contact only | Intercourse |
| No | 9.6 | - | - | - | - | 21.3 | - | - | - | - |
| Yes | - | - | - | - | - | - | - | - | - | 5.2 |
| Yes | - | - | - | - | - | - | 34.8 | - | 29.0 | 5.8 |
| No | 28.9 | 9.2 | 19.7 | 4.3 | 4.9 | - | - | - | - | - |
| Yes | - | $56.0{ }^{\text {a }}$ | - | 42.5 | 13.5 | - | 53.2 | - | 35.6 | 17.6 |
| No | - | - | - | - | $1.1{ }^{\text {b }}$ | - | - | - | - | $1.1{ }^{\text {b }}$ |
| Yes | - | - | - | - | - | 24.0 | 17.7 | 6.3 | - | - |
| No | - | 2.5 | - | 1.9 | 0.9 | - | 13.5 | - | 8.5 | 5.0 |
| No | - | 3.8 | - | - | - | - | 6.8 | - | - | - |
| No | 16.0 | - | - | - | - | 27.0 | - | - | - | - |
| Yes | 6.7 | - | - | - | - | 19.5 | - | - | - | - |
| Yes | - | - | - | - | - | $62.0{ }^{\text {a }}$ | 46.0 | 16.0 | - | - |
| Yes | - | - | - | - | - | - | 34.0 | - | - | - |
| No | - | - | - | - | - | - | - | - | - | 8.5 |
| Yes | - | $3.4{ }^{\text {b }}$ | - | - | - | - | $3.4{ }^{\text {b }}$ | - | - | - |
| No | 5.9 | - | - | - | 0.0 | 15.3 | - | - | - | 1.3 |
| No | 3.1 | - | - | - | - | 9.7 | - | - | - | - |
| Yes | - | $8.0^{\text {b }}$ | - | - | - | - | $8.0^{\text {b }}$ | - | - | - |
| No | 7.3 | 4.7 | 2.5 | 2.5 | 2.2 | - | - | - | - | - |
| Yes | 15.0 | - | - | - | - | - | - | - | - | - |
| Yes | - | 4.3 | - | - | - | - | 11.5 | - | - | - |
| No | - | 9.8 | - | - | - | - | 23.6 | - | - | - |
| Yes | - | 2.0 | - | - | - | - | 14.0 | - | - | - |
| Yes | - | - | - | - | 6.2 | - | - | - | - | 18.5 |
| No | 8.1 | - | - | - | - | 33.1 | - | - | - | - |
| Yes | - | $10.0{ }^{\text {b }}$ | - | - | - | - | $10.0{ }^{\text {b }}$ | - | - | - |
| Yes | - | 5.0 | - | - | - | - | 23.8 | - | - | - |

Table 23.3 Characteristics of studies included in prevalence analysis (continued)

| Study authors | Sample |  |  |  | CSA data collection |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Type | Number | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | Mean age (years) | Method | No. of questions | Childhood definition |
| Level B: Other community samples (representativeness not known) |  |  |  |  |  |  |  |
| L. George and I. Winfield-Laird, unpublished document, 1986 | Adults | 1157 | 100 | 41 | Face-to-face | 1 | $<16$ |
| Keckley Market <br> Research, unpublished document, 1983 | Adults | 603 | - | - | Telephone | 1 | $<18$ |
| Murphy (1997) | Adults | 818 | 51 | - | Telephone | 1 | $<18$ |
| Wolf (1992) | Adults | 637 | 56 | - | Telephone | 1 | $<16$ |
| Essock-Vitale and McGuire (1985) | Adults | 300 | 100 | 40 | Face-to-face | 1 | $<18$ |
| Saunders et al. (1992) | Adults | 391 | 100 | 42 | Face-to-face | 1 | $<18$ |
| Bagley and Ramsay (1985) | Adults | 377 | 100 | 40 | Face-to-face | >1 | $<17$ |
| Bagley (1991) | Adults | 750 | 100 | 23 | Face-to-face | >1 | $<17$ |
| Kercher and McShane (I984) | Adults | 1054 | 56 | - | Postal SR | I | $<18$ |
| Russell (1983) | Adults | 930 | 100 | - | Face-to-face | >1 | $<18$ |
| Bagley et al. (1994) | Adults | 750 | 0 | 23 | Self-report | >1 | $<17$ |
| Roosa et al. (1998) | Adults | 2003 | 100 | 20 | Self-report | >1 | $<18$ |
| Bagley (1995) | Adults | 1833 | 56 | - | Self-report | >1 | $<17$ |
| Watts and Ellis (1993) | School students | 670 | 100 | 15 | Self-report | >1 | $<18$ |
| Erickson and Rapkin (1991) | School students | 1197 | 50 | 15 | Self-report | >1 | $<18$ |
| Lodico et al. (1996) | School students | 6224 | 48 | 16 | Self-report | >1 | $<18$ |
| Level C: Community subgroups and convenience samples |  |  |  |  |  |  |  |
| Kendler et al. (2000) | Adults twins | 1411 | 100 | 40 | Postal SR | >1 | $<16$ |
| Silverman et al. (1996) | Adults (long study) | 375 | 50 | 21 | Face-to-face | >1 | $<18$ |
| White and Strange (1993) | College students | 131 | 100 | 20 | Postal SR | >1 | $<17$ |
| Peters and Range (1995) | College students | 266 | 51 | - | Self-report | >1 | $<12$ |
| Thakkar et al. (2000) | College students | 707 | 100 | 19 | Self-report | >1 | $<15$ |
| Schaaf and McCanne (1998) | College students | 238 | 100 | 19 | Self-report | >1 | $<15$ |
| Finkelhor (1979) | College students | 796 | 67 | 21 | Self-report | >1 | $<17$ |
| deLahunta (1996) | College students | 787 | 38 | - | Postal SR | - | - |
| Arroyo (1997) | College students | 221 | 100 | 25 | Face-to-face | >1 | $<18$ |
| Briere and Runtz (1988) | College students | 278 | 100 | 20 | Self-report | >1 | $<15$ |
| Wellman (1993) | College students | 824 | 80 | 20 | Self-report | $>1$ | $<18$ |
| Edwards and Alexander (1992) | College students | 103 | 100 | 23 | Self-report | >1 | $<18$ |
| Fritz et al. (1981) | College students | 952 | 57 | - | Self-report | >1 | $<14$ |


|  | Prevalence in males (\%) |  |  |  |  | Prevalence in females (\%) |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Restriction <br> on CSA <br> definition | $\begin{gathered} \text { Any } \\ \text { broad } \end{gathered}$ | $\begin{gathered} \text { Any } \\ \text { narrow } \end{gathered}$ | Noncontact only | Contact only | Intercourse | $\begin{aligned} & \text { Any } \\ & \text { broad } \end{aligned}$ | $\begin{gathered} \text { Any } \\ \text { narrow } \end{gathered}$ | Noncontact only | Contact only | Intercourse |
| - | - | - | - | - | - | - | 2.0 | - | - | - |
| Yes | 7.0 | - | - | - | - | 11.0 | - | - | - | - |
| Yes | - | 3.0 | - | - | - | - | 13.0 | - | - | - |
| No | 9.0 | - | - | - | - | 27.0 | - | - | - | - |
| No | - | - | - | - | - | - | 17.3 | - | - | - |
| No | - | - | - | - | - | 33.5 | 24.6 | 9.0 | 14.6 | 10.0 |
| Yes | - | - | - | - | - | - | 21.7 | - | - | - |
| No | - | - | - | - | - | - | 32.0 | - | - | - |
| Yes | 3.0 | - | - | - | - | 11.6 | - | - | - | - |
| Yes | - | - | - | - | - | 54.0 | 38.0 | 16.0 | - | - |
| Yes | - | 15.5 | - | - | - | - | - | - | - | - |
| No | - | - | - | - | - | - | 39.0 | - | 16.0 | 23.0 |
| Yes | - | 8.2 | - | - | - | - | 17.6 | - | - | - |
| No | - | - | - | - | - | - | 14.5 | - | - | - |
| No | - | $15.0{ }^{\text {b }}$ | - | - | - | - | $15.0{ }^{\text {b }}$ | - | - | - |
| Yes | 4.2 | - | - | - | - | 16.5 | - | - | - | - |
| Yes | - | - | - | - | - | 30.3 | 22.5 | 7.8 | 14.1 | 8.4 |
| Yes | - | $1.1{ }^{\text {c }}$ | - | - | - | - | $12.3{ }^{\text {c }}$ | - | - | - |
| Yes | - | - | - | - | - | 33.5 | - | - | - | - |
| Yes | 19.1 | 12.2 | 6.9 | - | - | 31.9 | 19.3 | 12.6 | - | - |
| Yes | - | - | - | - | - | - | 13.5 | - | - | - |
| No | - | - | - | - | - | 12.2 | - | - | - | - |
| Yes | 8.6 | - | - | - | - | 19.2 | - | - | - | - |
| No | - | $3.5{ }^{\text {c }}$ | - | - | - | - | $10.8{ }^{\text {c }}$ | - | - | - |
| No | - | - | - | - | - | 31.2 | - | - | - | - |
| Yes | - | - | - | - | - | - | 14.8 | - | - | - |
| No | 23.0 | 13.4 | 9.6 | - | - | 15.0 | 5.6 | 9.4 | - | - |
| Yes | - | - | - | - | - | - | 43.6 | - | - | - |
| Yes | - | 4.8 | - | - | - | - | 7.7 | - | - | - |

Table 23.3 Characteristics of studies included in prevalence analysis (continued)

| Study authors | Sample |  |  |  | CSA data collection |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Type | Number | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | Mean age (years) | Method | No. of questions | Childhood definition |
| Haugaard and Emery (1989) | College students | 1089 | 61 | 19 | Self-report | >1 | $<17$ |
| Duane et al. (1997) | College students | 958 | 61 | - | Self-report | >1 | $<13$ |
| Stepakoff (1998) | College students | 393 | 100 | 20 | Postal SR | >1 | $<17$ |
| Boudewyn and Liem (1995) | College students | 438 | 61 | 25 | Self-report | >1 | $<14$ |
| Bryant and Range (1997) | College students | 486 | 74 | 24 | Self-report | >1 | $<18$ |
| Bolstad and Zinbarg (1997) | College students | 117 | 100 | 26 | Self-report | >1 | $<15$ |
| Fromuth (1986) | College students | 482 | 100 | 20 | Self-report | >1 | $<16$ |
| Sedney and Brooks (1984) | College students | 301 | 100 | 19 | Self-report | - | - |
| Hibbard et al. (1988) | School students | 712 | 50 | 14 | Self-report | 1 | $<18$ |
| Riggs et al. (1990) | School students | 600 | 52 | 16 | Postal SR | 1 | $<18$ |
| Greenwood et al. (1990) | GP attendees | 100 | 59 | 42 | Face-to-face | >1 | - |
| Walch and Broadhead (1992) | GP attendees | 405 | 100 | 29 | Self-report | - | $<18$ |
| Kellogg and Hoffman (1995) | GP attendees | 142 | 60 | 20 | Self-report | 1 | $<18$ |
| Gould et al. (1994) | GP attendees | 292 | 71 | 48 | Self-report | $>1$ | $<17$ |
| Felitti et al. (1998) | GP attendees | 9508 | 54 | 56 | Postal SR | >1 | - |
| Kilpatrick (1986) | Misc. comm. groups | 501 | 100 | 28 | Self-report | >1 | $<15$ |
| DiVasto et al. (1984) | Misc. comm. groups | 500 | 100 | 27 | Self-report | 1 | $<13$ |
| Moeller et al. (1993) | Clinic sample | 668 | 100 | 34 | Postal SR | - | $<18$ |
| Bayatpour et al. (1992) | Clinic sample | 352 | 100 | 15 | Face-to-face | 1 | $<18$ |
| Descamps et al. (2000) | Lesbian women | 1925 | 100 | 35 | Postal SR | - | - |
| Blum et al. (1992) | American Indian and Alaska Native youth | 13454 | 51 | 15 | Self-report | - | $<18$ |
| Robin et al. (1997) | American Indians | 375 | 58 | 37 | Face-to-face | >1 | $<16$ |
| Greenwald and Leitenberg (1990) | Nurses | 1500 | 100 | - | Self-report | - | $<16$ |
| Hall et al. (1993) | Low income women | 203 | 100 | 27 | Face-to-face | >1 | $<18$ |
| Zuravin and Fontanella (1999) | Low income women | 513 | 100 | 30 | Face-to-face | >1 | $<14$ |
| Wingood and DiClemente (1997) | African-American women | 165 | 100 | 24 | Face-to-face | 1 | <16 |
| Romero et al. (1999) | Latina women | 300 | 100 | 32 | Face-to-face | >1 | $<18$ |

AMR-B (Brazil $n=2$, Costa Rica $n=1$, Dominican Republic $n=1$, El Salvador $n=1$, Mexico $n=1$ )
Level A: Representative community samples

| Ramos-Lira et al. <br> (1998) |
| :--- |


|  | Prevalence in males (\%) |  |  |  |  | Prevalence in females (\%) |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Restriction <br> on CSA <br> definition | Any broad | Any narrow | Noncontact only | Contact only | Intercourse | Any broad | $\begin{gathered} \text { Any } \\ \text { narrow } \end{gathered}$ | Noncontact only | Contact only | Intercourse |
| Yes | 5.0 | - | - | - | - | 11.9 | - | - | - | - |
| No | - | 4.0 | - | - | - | - | 4.3 | - | - | - |
| Yes | - | - | - | - | - | - | 14.9 | - | 10.1 | 4.8 |
| Yes | - | 16.2 | - | - | - | - | 23.8 | - | - | - |
| No | $28.2{ }^{\text {b }}$ | - | - | - | - | $28.2^{\text {b }}$ | - | - | - | - |
| Yes | - | - | - | - | - | 31.6 | - | - | - | - |
| Yes | - | - | - | - | - | 22.0 | - | - | - | - |
| - | - | - | - | - | - | - | $16.9{ }^{\text {c }}$ | - | - | - |
| No | - | $8.0{ }^{\text {b,c }}$ | - | - | - | - | $8.0{ }^{\text {b,c }}$ | - | - | - |
| No | $8.1{ }^{\text {b }}$ | - | - | - | - | $8.1{ }^{\text {b }}$ | - | - | - | - |
| No | 0.0 | - | - | - | - | 16.9 | - | - | - | - |
| Yes | - | - | - | - | - | 35.6 | - | - | - | 6.1 |
| No | - | 16.0 | - | - | - | - | 39.0 | - | - | - |
| No | - | 12.0 | - | - | - | - | 30.0 | - | - | - |
| Yes | - | $22.0{ }^{\text {b }}$ | - | $15.1{ }^{\text {b }}$ | $6.9{ }^{\text {b }}$ | - | $22.0{ }^{\text {b }}$ | - | $15.1{ }^{\text {b }}$ | $6.9{ }^{\text {b }}$ |
| No | - | - | - | - | - | 55.0 | - | - | - | 1.8 |
| No | - | - | - | - | - | - | - | - | - | 1.8 |
| Yes | - | - | - | - | - | 19.8 | - | - | - | - |
| No | - | - | - | - | - | 14.8 | - | - | - | - |
| - | - | - | - | - | - | - | 28.7 | - | - | - |
| Yes | - | 10.0 | - | - | - | - | 21.6 | - | - | - |
| Yes | 14.0 | - | - | - | - | 49.0 | - | - | - | - |
| Yes | - | - | - | - | - | 3.6 | - | - | - | 0.7 |
| No | - | - | - | - | - | - | 22.0 | - | - | - |
| No | - | - | - | - | - | - | 20.7 | - | 8.0 | 12.7 |
| No | - | - | - | - | - | - | - | - | - | 12.7 |
| No | - | - | - | - | - | 33.0 | - | - | - | 8.6 |



Table 23.3 Characteristics of studies included in prevalence analysis (continued)

| Study authors | Sample |  |  |  | CSA data collection |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Type | Number | $\begin{gathered} \% \\ \text { female } \\ \hline \end{gathered}$ | Mean age (years) | Method | No. of questions | Childhood definition |
| Level B: Other community samples (representativeness not known) |  |  |  |  |  |  |  |
| WHO (2001) | Adults | 1172 | 100 | 32 | Face-to-face | >1 | $<15$ |
| WHO (200I) | Adults | 1473 | 100 | 32 | Face-to-face | >1 | $<15$ |
| Z.A. Ruiz et al., unpublished document, 1986 | College students | 893 | 54 | 23 | Self-report | - | - |
| Level C: Community subgroups and convenience samples |  |  |  |  |  |  |  |
| Barthauer and Leventhal (1999) | Adults | 83 | 100 | 35 | Face-to-face | >1 | $<18$ |
| Krugman et al. (1992) | College students | 497 | 45 | 20 | Self-report | >1 | $<19$ |
| AMR-D (Nicaragua $n=1$, Peru $n=2$ ) |  |  |  |  |  |  |  |
| Level A: Representative community samples |  |  |  |  |  |  |  |
| Level B: Other community samples (representativeness not known) |  |  |  |  |  |  |  |
| WHO (2001) | Adults | 1415 | 100 | 32 | Face-to-face | >1 | $<15$ |
| WHO (2001) | Adults | 1847 | 100 | 32 | Face-to-face | >1 | $<15$ |
| EMR-B (no estimates available) |  |  |  |  |  |  |  |
| EMR-D (no estimates available) |  |  |  |  |  |  |  |
| EUR-A (Austria $n=I$, Belgium $n=I$, Czech Republic $n=I$, Denmark $n=I$, Finland $n=I$, France $n=3$, Germany $n=3$, Greece $n=I$, Ireland $n=I$, Israel $n=I$, Netherlands $n=I$, Norway $n=4$, Spain $n=2$, Sweden $n=4$, Switzerland $n=2$, United Kingdom $n=9$ ) |  |  |  |  |  |  |  |
| Level A: Representative community samples |  |  |  |  |  |  |  |
| Vandewege et al. (1988) | Adults | 956 | 100 | 35 | Face-to-face | - | - |
| Cawson et al. (2000) | Adults | 2869 | 57 | 21 | Computer | - | $<13$ |
| Baker and Duncan (1985) | Adults | 2019 | 52 | 40 | Face-to-face | >1 | $<16$ |
| Edgardh and Ormstad (2000) | Adults | 1943 | 58 | 17 | Self-report | >1 | $<17$ |
| Ernst et al. (1993) | Adults | 421 | 47 | 28 | Face-to-face | 1 | $<16$ |
| Halperin et al. (1996) | Adults | 1116 | 51 | 15 | Self-report | >1 | $<17$ |
| Spak et al. (1998) | Adults | 316 | 100 | - | Face-to-face | 1 | $<18$ |
| Lopez et al. (1995) | Adults | 1821 | 47 | 39 | Face-to-face | >1 | $<17$ |
| Garnefski and Arends (1998) | Adults | 13894 | 50 | 15 | Self-report | >1 | $<19$ |
| Weiss and Zverina (1997) | Adults | 1719 | 50 | 38 | Postal SR | >1 | $<15$ |
| Bouhet et al. (1992) | Adults | 1511 | 51 | 39 | Postal SR | >1 | $<18$ |
| Hill et al. (2000) | Adults | 862 | 100 | 31 | Postal SR | >1 | $<16$ |
| Bendixen et al. (1994) | College students | 996 | 51 | 23 | Self-report | >1 | - |
| Choquet et al. (1997) | School students | 8140 | 51 | 16 | Self-report | >1 | $<16$ |
| Sariola and Uutela (1994) | School students | 6913 | 52 | 16 | Self-report | >1 | $<15$ |


|  | Prevalence in males (\%) |  |  |  |  | Prevalence in females (\%) |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Restriction <br> on CSA <br> definition | Any broad | $\begin{gathered} \text { Any } \\ \text { narrow } \end{gathered}$ | Noncontact only | Contact only | Intercourse | Any broad | $\begin{gathered} \text { Any } \\ \text { narrow } \end{gathered}$ | Noncontact only | Contact only | Intercourse |
| - | - | - | - | - | - | 7.8 | - | - | - | - |
| - | - | - | - | - | - | 5.8 | - | - | - | - |
| - | $33.0{ }^{\text {b }}$ | - | - | - | - | $33.0{ }^{\text {b }}$ | - | - | - | - |
| No | - | - | - | - | - | 17.0 | - | - | - | 9.6 |
| No | - | 12.8 | - | - | - | - | 32.2 | - | - | - |
| Yes | 20.0 | - | - | - | - | 26.0 | - | - | - | - |
| - | - | - | - | - | - | 19.5 | - | - | - | - |
| - | - | - | - | - | - | 7.9 | - | - | - | - |

$\qquad$

| - | - | - | - | - | - | 19.0 | - | - | - | - |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Yes | - | 11.0 | - | - | - | - | 21.0 | - | - | - |
| No | 8.0 | - | - | - | 0.7 | 12.0 | - | - | - | 0.8 |
| Yes | 3.1 | - | - | - | - | 11.2 | - | - | - | - |
| Yes | 1.8 | - | - | - | - | 4.9 | - | - | - | - |
| No | 10.9 | 3.3 | 7.7 | 2.2 | 1.1 | 33.8 | 20.4 | 13.4 | 14.8 | 5.6 |
| No | - | - | - | - | - | 9.8 | - | - | - | - |
| - | 14.0 | 12.0 | 2.0 | 8.2 | 3.8 | 27.0 | 19.7 | 7.3 | 12.7 | 7.0 |
| Yes | - | 2.2 | - | - | - | - | 8.2 | - | - | - |
| No | - | 4.6 | - | 4.3 | 0.3 | - | 8.4 | - | 7.5 | 0.9 |
| No | 4.6 | 3.1 | 1.4 | - | - | 7.8 | 5.2 | 2.6 | - | - |
| No | - | - | - | - | - | - | 17.5 | - | 11.9 | 5.6 |
| Yes | 3.5 | - | - | - | - | 19.4 | - | - | - | - |
| Yes | - | - | - | - | 0.9 | - | - | - | - | 0.7 |
| Yes | 3.3 | - | - | - | - | 7.6 | - | - | - | - |

Table 23.3 Characteristics of studies included in prevalence analysis (continued)

| Study authors | Sample |  |  |  | CSA data collection |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Type | Number | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | Mean age (years) | Method | No. of questions | Childhood definition |
| Level B: Other community samples (representativeness not known) |  |  |  |  |  |  |  |
| Leth (2001) | Adults | 1235 | 54 | - | Postal SR | - | $<18$ |
| H. Holter, unpublished document, 1990 | Adults | 1017 | - | - | Postal SR | 1 | - |
| Rönström (1985) | Adults | 938 | - | - | Self-report | - | - |
| Schei (1990) | Adults | 118 | 100 | 33 | Face-to-face | 1 | - |
| J. Kinzl and W. Biebl, unpublished data | College students | 1125 | - | - | Self-report | - | - |
| Agathonos et al. (1992) | College students | 746 | - | - | Self-report | >1 | - |
| de Paul et al. (1995) | College students | 403 | 74 | 21 | Postal SR | >1 | $<13$ |
| Lazartigues et al. (1989) | College students | 963 | 58 | 20 | Postal SR | >1 | $<16$ |
| Schoetensack et al. (1992) | College students | 1841 | 48 | 21 | Self-report | >1 | - |
| Kelly et al. (1991) | College students | 1244 | 62 | 19 | Self-report | >1 | $<18$ |
| Pederson and Aas (1995) | School students | 465 | 54 | 19 | Postal SR | >1 | $<13$ |
| Schein et al. (2000) | GP attendees | 1005 | 65 | 36 | Self-report | >1 | $<18$ |
| Level C: Community subgroups and convenience samples |  |  |  |  |  |  |  |
| Market Research Bureau of Ireland (1987) | Adults | 500 | - | - | Self-report | >1 | <18 |
| Richter-Appelt and Tiefensee (1996) | College students | 1068 | 42 | 24 | Self-report | - | $<12$ |
| Raupp and Eggers (1993) | College students | 1009 | 50 | - | Self-report | - | $<18$ |
| Bickerton et al. (1991) | GP attendees | 1232 | 100 | - | Face-to-face | >1 | - |
| Coxell et al. (1999) | GP attendees | 2474 | 0 | 46 | Computer | $>1$ | $<16$ |
| Palmer et al. (1994) | GP attendees | 115 | 0 | 32 | Self-report | >1 | $<16$ |
| Risberg et al. (1999) | GP attendees | 175 | 100 | 41 | Postal SR | 1 | $<18$ |
| Palmer et al. (1993) | GP attendees | 120 | 100 | 30 | Face-to-face | >1 | $<16$ |
| Brown and Harris (1993) | Low income women | 404 | 100 | - | Face-to-face | >1 | $<17$ |
| EUR-B (Turkey $n=1$ ) |  |  |  |  |  |  |  |
| Level C: Community subgroups and convenience samples |  |  |  |  |  |  |  |
| Elal et al. (2000) | College students | 1597 | 62 | - | Self-report | >1 | $<18$ |
| EUR-C (Russia $n=1$ ) |  |  |  |  |  |  |  |
| Level C: Community subgroups and convenience samples |  |  |  |  |  |  |  |
| N. Lvoff and V. Lvoff, unpublished document, 1998 | College students | 723 | 50 | - | Self-report | - | <18 |


| Restriction <br> on CSA <br> definition | Prevalence in males (\%) |  |  |  |  | Prevalence in females (\%) |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Any broad | Any narrow | Noncontact only | Contact only | Intercourse | Any broad | $\begin{gathered} \text { Any } \\ \text { narrow } \end{gathered}$ | Noncontact only | Contact only | Intercourse |
| Yes | 7.0 | 6.5 | 0.4 | 2.2 | 4.3 | 14.0 | 12.1 | 2.0 | 5.2 | 6.9 |
| No | - | 9.0 | - | - | - | - | 19.0 | - | - | - |
| - | 3.0 | - | - | - | - | 11.0 | - | - | - | - |
| No | - | - | - | - | - | 14.0 | - | - | - | - |
| - | 19.0 | - | - | - | - | 36.0 | - | - | - | - |
| No | 6.0 | - | - | - | - | 16.0 | - | - | - | - |
| No | 3.9 | - | - | - | - | 6.4 | - | - | - | - |
| Yes | 3.4 | - | - | - | - | 10.2 | - | - | - | - |
| Yes | 5.8 | 3.9 | 1.9 | 3.0 | 0.9 | 16.1 | 10.8 | 5.3 | 8.9 | 1.9 |
| No | 27.0 | 11.0 | 16.0 | - | - | $59.0{ }^{\text {a }}$ | 27.0 | 32.0 | - | - |
| Yes | 0.5 | 0.0 | 0.5 | - | - | 6.8 | 6.0 | 0.8 | - | - |
| Yes | 15.7 | - | - | - | - | 30.7 | - | - | - | - |
| - | 5.0 | - | - | - | - | 7.0 | - | - | - | - |
| No | 4.0 | - | - | - | - | 23.0 | - | - | - | - |
| No | 6.2 | - | 2.4 | 2.3 | 1.5 | 25.2 | - | 11.2 | 11.7 | 2.3 |
| Yes | - | - | - | - | - | - | 3.5 | - | - | - |
| Yes | - | 13.0 | - | - | - | - | - | - | - | - |
| Yes | 13.9 | - | - | - | 4.3 | - | - | - | - | - |
| No | - | - | - | - | - | 6.8 | 5.7 | 1.1 | - | - |
| Yes | - | - | - | - | - | 33.3 | 20.8 | 12.5 | 15.8 | 5.0 |
| Yes | - | - | - | - | - | - | 6.9 | - | - | - |

$\qquad$

| Yes | 16.0 | - | - | - | - | 28.0 | - | - | - | - |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |


| - | 9.0 | - | - | - | - | 27.0 | - | - | - | - |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |

Table 23.3 Characteristics of studies included in prevalence analysis (continued)

| Study authors | Sample |  |  |  | CSA data collection |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Type | Number | $\begin{gathered} \% \\ \text { female } \\ \hline \end{gathered}$ | Mean age (years) | Method | No. of questions | Childhood definition |
| SEAR-B (Indonesia $n=1$, Sri Lanka $n=1$, Thailand $n=2$ ) |  |  |  |  |  |  |  |
| Level B: Other community samples (representativeness not known) |  |  |  |  |  |  |  |
| WHO (2001) | Adults | 1536 | 100 | 32 | Face-to-face | >1 | $<15$ |
| WHO (2001) | Adults | 1282 | 100 | 32 | Face-to-face | >1 | $<15$ |
| WHO (2001) | Adults | 765 | 100 | 32 | Face-to-face | >1 | $<15$ |
| Level C: Community subgroups and convenience samples |  |  |  |  |  |  |  |
| Miles (2000) | School students | 145 | 43 | 15 | - | 1 | $<18$ |
| SEAR-D (India $n=3$ ) |  |  |  |  |  |  |  |
| Level C: Community subgroups and convenience samples |  |  |  |  |  |  |  |
| Interventions for support (2001) | School students | 198 | 75 | 15 | Self-report | - | $<18$ |
| Interventions for support (2001) | School students | 426 | 100 | 15 | Self-report | - | $<18$ |
| Castelino (1985) | College students | 133 | 50 | 25 | Self-report | - | $<13$ |
| WPR-A (Australia $n=8$, New Zealand $n=4$ ) |  |  |  |  |  |  |  |
| Level A: Representative community samples |  |  |  |  |  |  |  |
| Anderson et al. (1993) | Adults | 497 | 100 | 42 | Face-to-face | >1 | $<16$ |
| Mullen et al. (1988) | Adults | 314 | 100 | - | Face-to-face | $>1$ | $<13$ |
| Goldman and Goldman (1988) | College students | 991 | 61 | 22 | Self-report | >1 | $<16$ |
| Mazza et al. (1996) | GP attendees | 2181 | 100 | - | Self-report | >1 | $<16$ |
| Level B: Other community samples (representativeness not known) |  |  |  |  |  |  |  |
| Fleming (1997) | Adults | 710 | 100 | 40 | Postal SR | $>1$ | $<16$ |
| Martin et al. (1993) | Adults | 1376 | 100 | - | Postal SR | >1 | $<16$ |
| Level C: Community subgroups and convenience samples |  |  |  |  |  |  |  |
| Dinwiddie et al. (2000) | Aduls | 5946 | 65 | 43 | Telephone | 1 | $<18$ |
| Nelson et al. (2002) | Adult twins | 3982 | 58 | 30 | Telephone | $>1$ | $<16$ |
| Fergusson et al. (1996a) | Adolescents | 1019 | 51 | 18 | Face-to-face | >1 | $<16$ |
| Higgins and McCabe (1994) | College students | 253 | 79 | 21 | Self-report | >1 | $<18$ |
| Goldman and Padayachi (1997) | College students | 427 | 67 | 21 | Self-report | >1 | $<17$ |
| Martin (1996) | School students | 352 | 43 | 15 | Self-report | 1 | $<18$ |
| WPR-B (China $n=1$, Malaysia $n=1$ ) |  |  |  |  |  |  |  |
| Level C: Community subgroups and convenience samples |  |  |  |  |  |  |  |
| So-kum Tang (2000) | College students | 2038 | 57 | 21 | Self-report | $>1$ | $<17$ |
| Singh et al. (1996) | College students | 616 | 77 | 22 | Self-report | $>1$ | $<18$ |

Postal SR Postal self-report.

- No data.
a Prevalence estimate not included in regression analysis based on outlier analysis ( $>3$ standard deviations away from the mean).

|  | Prevalence in males (\%) |  |  |  |  | Prevalence in females (\%) |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Restriction on CSA definition | Any broad | $\begin{gathered} \text { Any } \\ \text { narrow } \end{gathered}$ | Noncontact only | Contact only | Intercourse | Any broad | $\begin{gathered} \text { Any } \\ \text { narrow } \end{gathered}$ | Noncontact only | Contact only | Intercourse |
| - | - | - | - | - | - | 7.6 | - | - | - | - |
| - | - | - | - | - | - | 4.7 | - | - | - | - |
| - | - | - | - | - | - | 6.1 | - | - | - | - |
| No | $10.0{ }^{\text {b }}$ | - | - | - | - | $10.0^{\text {b }}$ | - | - | - | - |
| - | 59.0 | - | - | - | - | 78.0 | - | - | - | - |
| - | - | - | - | - | - | - | 65.0 | - | - | - |
| - | 26.0 | - | - | - | - | 26.0 | - | - | - | - |
| Yes | - | - | - | - | - | 31.9 | 24.3 | 6.8 | 17.0 | 7.3 |
| No | - | - | - | - | - | - | 9.9 | - | - | - |
| Yes | 9.0 | - | - | - | - | 27.6 | - | - | - | - |
| No | - | - | - | - | - | 39.0 | - | - | - | 6.0 |
| Yes | - | - | - | - | - | 32.3 | 20.3 | 12.0 | 18.3 | 2.0 |
| Yes | - | - | - | - | - | 31.9 | 25.1 | 6.8 | 22.0 | 3.1 |
| No | - | 2.5 | - | - | - | - | 5.9 | - | - | - |
| Yes | - | 5.4 | - | - | - | - | 16.7 | - | - | - |
| Yes | 3.4 | 3.0 | 0.4 | 1.6 | 1.4 | 17.3 | 13.0 | 4.3 | 7.4 | 5.6 |
| Yes | 22.0 | - | - | - | - | 24.0 | - | - | - | - |
| Yes | 18.6 | 13.0 | 5.6 | - | - | 45.0 | 39.0 | 6.0 | - | - |
| No | - | 4.5 | - | - | - | - | 13.2 | - | - | - |
| No | $33.3{ }^{\text {a }}$ | - | - | - | 3.0 | 28.2 | - | - | - | 5.8 |
| No | 2.1 | 2.1 | 0.0 | 2.1 | 0.0 | 8.8 | 5.5 | 2.3 | 5.1 | 0.4 |

[^90]Table 23.4 Characteristics of 15 excluded studies

| Subregion | Country | Author(s) | Type of study | Sample size | Sample type |
| :---: | :---: | :---: | :---: | :---: | :---: |
| AMR-A | Canada | Berry (1997) | Prevalence | 327 | College students |
| AMR-A | Canada | Committee on Sexual Offences Against Children and Youths (1984) | Prevalence | 2008 | Community adults |
| AMR-A | USA | Hernandez et al. (1993) | Prevalence and risk factor | 2973 | School students |
| AMR-A | USA | Hibbard et al. (1990) | Prevalence and risk factor | 3998 | School students |
| AMR-A | USA | Lenihan (1996) | Prevalence | 1687 | College students |
| AMR-A | USA | Locke (1996) | Risk factor | - | College students |
| AMR-A | USA | Priest (1991) | Prevalence | - | College students |
| EUR-A | Austria | Friedrich et al. (1997) | Prevalence | - | Adult women |
| EUR-A | Italy | Meledandri et al. (1996) | Prevalence | - | - |
| EUR-B | Czech <br> Republic | Pothe et al. (2000) | Prevalence | 1112 | Community adults |
| WPR-A | Australia | Baldini (1996) | - | - | Aboriginal communities |
| WPR-A | Australia | Barton (1987) | Prevalence | $>1000$ | College students |
| WPR-B | China | Wang et al. (1994) | Prevalence | - | - |
| NA | NA | Garabedian (1994) | Risk factor | - | - |
| NA | NA | Garabedian (1994) | Risk factor | - | Adult women |

NA Not applicable.

- No data.


### 2.8 Estimates by age, sex and subregion

Table 23.5 presents the final estimates of CSA prevalence by age, sex, level of exposure and subregion. It is presented in the format required for the WHO estimates. Table 23.6 presents the same data with level of abuse and age categories collapsed. This allows an easier comparison across subregions for both males and females. From Table 23.6 a number of interesting findings emerge. First, it can be seen that, on average, the prevalence of CSA is higher in females than in males. This is a commonly reported phenomenon. As demonstrated in the regression analysis, differences in prevalence also exist between subregions. Due to the paucity of prevalence estimates in subregions other than AMR-A, EUR-A and WPR-A it is not possible to look at differences between all subregions. However, the pattern of results does suggest that a high prevalence of CSA is found in AFR-E and SEAR-D. It should be noted that the

Table 23.5 CSA prevalence estimates (\%) by subregion, sex, level of exposure and age group

| Subregion | Sex | Level | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| AFR-D | Male | I | 0.3 | 2.7 | 2.5 | 3.7 | 3.7 | 3.7 | 3.7 | 3.7 |
|  |  | 2 | 0.3 | 2.6 | 2.5 | 3.6 | 3.6 | 3.6 | 3.6 | 3.6 |
|  |  | 3 | 0.2 | 1.6 | 1.6 | 2.3 | 2.3 | 2.3 | 2.3 | 2.3 |
|  | Female | 1 | 0.5 | 4.5 | 4.3 | 6.2 | 6.2 | 6.2 | 6.2 | 6.2 |
|  |  | 2 | 0.9 | 7.9 | 7.5 | 10.9 | 10.9 | 10.9 | 10.9 | 10.9 |
|  |  | 3 | 0.3 | 3.0 | 2.9 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 |
| AFR-E | Male | 1 | 1.3 | 12.1 | 11.8 | 16.6 | 16.6 | 16.6 | 16.6 | 16.6 |
|  |  | 2 | 0.5 | 4.8 | 4.6 | 7.1 | 7.1 | 7.1 | 7.1 | 7.1 |
|  |  | 3 | 0.5 | 4.2 | 4.2 | 5.9 | 5.9 | 5.9 | 5.9 | 5.9 |
|  | Female | 1 | 0.5 | 4.5 | 4.3 | 6.2 | 6.2 | 6.2 | 6.2 | 6.2 |
|  |  | 2 | 2.5 | 22.2 | 21.6 | 30.5 | 30.5 | 30.5 | 30.5 | 30.5 |
|  |  | 3 | 0.5 | 4.3 | 4.1 | 6.0 | 6.0 | 6.0 | 6.0 | 6.0 |
| AMR-A | Male | I | 0.1 | 1.9 | 1.9 | 2.7 | 2.7 | 2.7 | 2.7 | 2.7 |
|  |  | 2 | 0.2 | 1.8 | 1.8 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |
|  |  | 3 | 0.1 | 1.0 | 1.0 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 |
|  | Female | 1 | 0.4 | 5.4 | 5.3 | 7.5 | 7.5 | 7.5 | 7.5 | 7.5 |
|  |  | 2 | 0.9 | 9.7 | 9.7 | 13.6 | 13.6 | 13.6 | 13.6 | 13.6 |
|  |  | 3 | 0.4 | 3.9 | 4.0 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 |
| AMR-B | Male | I | 0.5 | 4.3 | 4.2 | 5.9 | 5.9 | 5.9 | 5.9 | 5.9 |
|  |  | 2 | 0.3 | 2.3 | 2.3 | 3.2 | 3.2 | 3.2 | 3.2 | 3.2 |
|  |  | 3 | 0.1 | 1.2 | 1.2 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 |
|  | Female | 1 | 0.2 | 1.8 | 1.8 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 |
|  |  | 2 | 0.3 | 3.0 | 3.0 | 4.2 | 4.2 | 4.2 | 4.2 | 4.2 |
|  |  | 3 | 0.1 | 1.3 | 1.3 | 1.7 | 1.7 | 1.7 | 1.7 | 1.7 |
| AMR-D | Male | I | 0.6 | 5.6 | 5.2 | 7.7 | 7.7 | 7.7 | 7.7 | 7.7 |
|  |  | 2 | 0.6 | 5.5 | 5.1 | 7.6 | 7.6 | 7.6 | 7.6 | 7.6 |
|  |  | 3 | 0.4 | 3.4 | 3.2 | 4.7 | 4.7 | 4.7 | 4.7 | 4.7 |
|  | Female | I | 0.3 | 2.8 | 2.7 | 3.9 | 3.9 | 3.9 | 3.9 | 3.9 |
|  |  | 2 | 0.5 | 4.9 | 4.8 | 6.8 | 6.8 | 6.8 | 6.8 | 6.8 |
|  |  | 3 | 0.2 | 1.9 | 1.8 | 2.6 | 2.6 | 2.6 | 2.6 | 2.6 |
| EMR-B | Male | I | 0.4 | 3.2 | 3.2 | 4.4 | 4.4 | 4.4 | 4.4 | 4.4 |
|  |  | 2 | 0.4 | 3.2 | 3.2 | 4.3 | 4.3 | 4.3 | 4.3 | 4.3 |
|  |  | 3 | 0.2 | 2.0 | 2.0 | 2.7 | 2.7 | 2.7 | 2.7 | 2.7 |
|  | Female | I | 0.7 | 5.9 | 5.9 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 |
|  |  | 2 | 1.2 | 10.4 | 10.4 | 14.3 | 14.3 | 14.3 | 14.3 | 14.3 |
|  |  | 3 | 0.4 | 4.0 | 4.0 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 |
| EMR-D | Male | 1 | 0.4 | 3.2 | 3.2 | 4.4 | 4.4 | 4.4 | 4.4 | 4.4 |
|  |  | 2 | 0.4 | 3.2 | 3.2 | 4.3 | 4.3 | 4.3 | 4.3 | 4.3 |
|  |  | 3 | 0.2 | 2.0 | 2.0 | 2.7 | 2.7 | 2.7 | 2.7 | 2.7 |
|  | Female | 1 | 0.7 | 5.9 | 5.9 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 |
|  |  | 2 | 1.2 | 10.4 | 10.4 | 14.3 | 14.3 | 14.3 | 14.3 | 14.3 |
|  |  | 3 | 0.4 | 4.0 | 4.0 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 |

Table 23.5 CSA prevalence estimates (\%) by subregion, sex, level of exposure and age group (continued)

| Subregion | Sex | Level | Age group (years) |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 0-4 | 5-14 | 15-29 | 30-44 | 45-59 | 60-69 | 70-79 | $\geq 80$ |
| EUR-A | Male | I | 0.1 | 0.9 | 0.9 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 |
|  |  | 2 | 0.1 | 1.2 | 1.2 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 |
|  |  | 3 | 0.1 | 0.7 | 0.7 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 |
|  | Female | 1 | 0.4 | 3.5 | 3.6 | 4.9 | 4.9 | 4.9 | 4.9 | 4.9 |
|  |  | 2 | 0.7 | 6.0 | 6.2 | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 |
|  |  | 3 | 0.2 | 1.9 | 2.0 | 2.7 | 2.7 | 2.7 | 2.7 | 2.7 |
| EUR-B | Male | I | 0.4 | 3.2 | 3.2 | 4.4 | 4.4 | 4.4 | 4.4 | 4.4 |
|  |  | 2 | 0.4 | 3.2 | 3.2 | 4.3 | 4.3 | 4.3 | 4.3 | 4.3 |
|  |  | 3 | 0.2 | 2.0 | 2.0 | 2.7 | 2.7 | 2.7 | 2.7 | 2.7 |
|  | Female | 1 | 0.7 | 5.9 | 5.9 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 |
|  |  | 2 | 1.2 | 10.4 | 10.4 | 14.3 | 14.3 | 14.3 | 14.3 | 14.3 |
|  |  | 3 | 0.4 | 4.0 | 4.0 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 |
| EUR-C | Male | I | 0.3 | 2.5 | 2.5 | 3.5 | 3.5 | 3.5 | 3.5 | 3.5 |
|  |  | 2 | 0.3 | 2.5 | 2.4 | 3.4 | 3.4 | 3.4 | 3.4 | 3.4 |
|  |  | 3 | 0.2 | 1.5 | 1.5 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 |
|  | Female | 1 | 0.6 | 5.7 | 5.6 | 7.9 | 7.9 | 7.9 | 7.9 | 7.9 |
|  |  | 2 | 1.1 | 10.0 | 9.9 | 13.8 | 13.8 | 13.8 | 13.8 | 13.8 |
|  |  | 3 | 0.4 | 3.9 | 3.8 | 5.3 | 5.3 | 5.3 | 5.3 | 5.3 |
| SEAR-B | Male | 1 | 0.2 | 1.7 | 1.6 | 2.3 | 2.3 | 2.3 | 2.3 | 2.3 |
|  |  | 2 | 0.2 | 1.6 | 1.6 | 2.3 | 2.3 | 2.3 | 2.3 | 2.3 |
|  |  | 3 | 0.1 | 1.0 | 1.0 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 |
|  | Female | 1 | 0.2 | 1.5 | 1.5 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 |
|  |  | 2 | 0.3 | 2.6 | 2.6 | 3.6 | 3.6 | 3.6 | 3.6 | 3.6 |
|  |  | 3 | 0.1 | 1.0 | 1.0 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 |
| SEAR-D | Male | I | 1.1 | 9.8 | 9.5 | 13.6 | 13.6 | 13.6 | 13.6 | 13.6 |
|  |  | 2 | 1.0 | 9.5 | 9.3 | 13.3 | 13.3 | 13.3 | 13.3 | 13.3 |
|  |  | 3 | 0.7 | 5.9 | 5.7 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 |
|  | Female | 1 | 1.1 | 10.5 | 10.2 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 |
|  |  | 2 | 3.0 | 27.7 | 26.8 | 39.1 | 39.1 | 39.1 | 39.1 | 39.1 |
|  |  | 3 | 1.1 | 10.0 | 9.7 | 13.9 | 13.9 | 13.9 | 13.9 | 13.9 |
| WPR-A | Male | 1 | 0.2 | 1.4 | 1.4 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 |
|  |  | 2 | 0.2 | 1.8 | 1.8 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |
|  |  | 3 | 0.1 | 0.9 | 0.9 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 |
|  | Female | 1 | 0.8 | 6.3 | 6.4 | 8.6 | 8.6 | 8.6 | 8.6 | 8.6 |
|  |  | 2 | 1.0 | 11.4 | 11.6 | 15.8 | 15.8 | 15.8 | 15.8 | 15.8 |
|  |  | 3 | 0.3 | 3.3 | 3.4 | 4.6 | 4.6 | 4.6 | 4.6 | 4.6 |
| WPR-B | Male | I | 1.1 | 9.6 | 9.9 | 13.2 | 13.2 | 13.2 | 13.2 | 13.2 |
|  |  | 2 | 1.0 | 9.2 | 9.5 | 12.7 | 12.7 | 12.7 | 12.7 | 12.7 |
|  |  | 3 | 0.2 | 1.9 | 1.9 | 2.6 | 2.6 | 2.6 | 2.6 | 2.6 |
|  | Female | 1 | 0.6 | 5.8 | 6.0 | 8.0 | 8.0 | 8.0 | 8.0 | 8.0 |
|  |  | 2 | 1.1 | 10.3 | 10.6 | 14.1 | 14.1 | 14.1 | 14.1 | 14.1 |
|  |  | 3 | 0.5 | 4.1 | 4.3 | 5.7 | 5.7 | 5.7 | 5.7 | 5.7 |


| Table 23.6 | CSA prevalence estimates (\%) by subregion and sex |  |
| :--- | :---: | :---: |
|  | Broad estimate |  |
| Subregion | Males | Females |
| AFR-D | 9.6 | 21.3 |
| AFR-E | 29.8 | 42.7 |
| AMR-A | 6.7 | 26.5 |
| AMR-B | 10.7 | 8.4 |
| AMR-D | 20.0 | 13.3 |
| EMR-B | 11.5 | 28.0 |
| EMR-D | 11.5 | 28.0 |
| EUR-A | 3.8 | 15.8 |
| EUR-B | 11.5 | 28.0 |
| EUR-C | 9.0 | 27.0 |
| SEAR-B | 6.0 | 7.1 |
| SEAR-D | 35.0 | 67.7 |
| WPR-A | 5.9 | 29.1 |
| WPR-B | 28.6 | 27.8 |

a Subregions for which no data were available. Estimates were used from EUR-C.
estimates from AFR-E and SEAR-D come from very few studies that were relatively poor methodologically. This makes the estimates for these subregions at best highly uncertain. More studies are obviously needed to confirm whether or not the prevalence is higher in these subregions of the world. Some of these subregional differences appear dependent on sex. For example, in AMR-B, AMR-D and WPR-B the studies reported a higher prevalence in males than females. Only a few studies contributed to the estimates in these subregions and without measures of variability around the estimates it was difficult to draw firm conclusions.

### 2.9 Quantitative and qualitative sources of uncertainty

Uncertainty in the current analysis came from several sources. Studies of the prevalence of CSA varied in terms of methodological characteristics. Regression analyses demonstrated that several methodological factors contributed to the variability in prevalence estimates. This was the major quantitative source of uncertainty and was reduced by adjusting the prevalence estimates to more closely reflect ideal methodology. Metaanalysis was used as a method of quantifying the uncertainty around the final prevalence estimates, taking into account sample size or variability between studies, whichever is appropriate given the homogeneity of the estimates being combined. Other sources of uncertainty arose from
decisions made regarding inclusion and exclusion criteria, and methods of extrapolation across age, sex and subregion. These decisions and their rationale have been documented in the relevant sections. It should also be noted that the estimates of uncertainty will necessarily be underestimated, as data do not exist for every country in all subregions.

## 3. Estimating risk factor-Disease RELATIONSHIPS

### 3.1 Outcomes to be assessed: including evidence FOR CAUSALITY AND REASONS FOR EXCLUSION OF RELATED OUTCOMES

## OUTCOMES INCLUDED AND EXCLUDED

The choice of outcomes to be assessed was guided by two principles. First, choice of outcomes was limited to those diseases or outcomes that were included in the Global Burden of Disease (GBD) study. Second, outcomes were limited to those for which there was sufficient evidence of a causal relationship with CSA, the exposure variable. Section 3.4 presents a detailed assessment of the evidence for causality between CSA and six mental disorders (depression, panic disorder, agoraphobia, PTSD, and alcohol and drug abuse or dependence) and suicide attempts.

Child sexual abuse has also been linked to other mental disorders, including OCD, eating disorders and personality disorders. Three studies examined the relationship between CSA and OCD (Arata 1999; Saunders et al. 1992; Stein et al. 1988), however the evidence was equivocal and further research is required to confirm any association or lack thereof. Eating disorders have long been conceptualized as a response to a dysfunctional family environment. While it was originally thought that CSA played a pivotal role in the development of eating disorders, recent community studies have reported only a modest association between CSA and subsequent development of eating disorders after controlling for confounding influences of the family and social environment (Mullen et al. 1996; Wonderlich et al. 1997). CSA has also been linked to personality disorders, with a couple of studies finding positive associations between CSA and antisocial personality disorder (Scott 1992) and borderline personality disorder (Johnson et al. 1999). These disorders are currently not included in the GBD project and therefore were not examined here.

CSA does not only produce an increased risk of mental disorder. There is anecdotal and epidemiological evidence that CSA increases the probability of negative psychological outcomes such as poor self-esteem (Romans et al. 1997), lack of a sense of control or agency (Mullen and Fleming 1998), difficulties with intimacy and continuing sexual difficul-
ties that demonstrate the far reaching damage that occurs in some individuals (Mullen et al. 1996). These outcomes were outside the scope of this analysis and several meta-analyses examining the relationship between CSA and aspects of psychological adjustment offer a more comprehensive review of these psychological correlates (Jumper 1995; Paolucci et al. 2001; Rind et al. 1998). But this does not infer that they are in any way less harmful or less important to the person than the mental disorders identified in this report. The impact of CSA on adult life has also been studied in terms of non-mental health consequences. These include the increased risk of developing sexually transmitted diseases, teenage pregnancies, multiple sexual partnerships and sexual revictimization (Gorcey et al. 1986; Nagy et al. 1995). It has been suggested that a history of CSA, particularly of the more intrusive types, interrupts the child's development of sexuality and normal sexual relationships (Fleming et al. 1999). However, these were also outside the scope of this analysis.

## GENERAL ASSESSMENT OF CAUSALITY

Below, Hill's (1965) criteria for causality were applied to the case of CSA and adult mental disorders. The remainder of this section outlines each of these criteria as they relate to research in the area of child sexual abuse. A more detailed assessment of causality for each outcome is included in section 3.4.

## Temporality

To satisfy the requirement of temporality, exposure to child sexual abuse must occur prior to the onset of adult mental disorder. By definition, some mental disorders are neither present nor easy to diagnose in children; therefore the time of onset of these disorders relative to child sexual abuse is a moot point. However, it is generally acknowledged that preexisting vulnerabilities or predisposition to adult mental disorder exist even in childhood (Caspi et al. 1996). Long-term prospective studies that follow children from a young age and control for these vulnerabilities therefore provide the best evidence regarding temporality. While a number of research projects have been prospective in design, the majority of studies of child sexual abuse have been cross-sectional and retrospective. Traditionally, cross-sectional studies are not considered sources of evidence for temporality. However, by definition, exposure to child sexual abuse occurs during childhood and therefore prior to onset of an adult psychiatric disorder. It is for this reason that cross-sectional studies that indicate a relationship between child sexual abuse and adult mental disorder can still be indicators of temporality.

It is important to point out that even in prospective studies, data on CSA are gathered retrospectively. It is unethical, and in many countries illegal, to prospectively identify CSA and not intervene. However, retrospective studies rely on recall of memories and can therefore suffer from
unreliable data. Moreover, reports of CSA are often ascertained contemporaneously with assessment of disorder, which leaves recall open to bias in which those with disorder are more prone to recall CSA (Mullen et al. 2000). The issue of potentially unreliable recall represents one of the central threats to the validity of the published literature on CSA. Unfortunately, given that very few cases of CSA are reported to other adults, and even fewer to authorities, the validity of retrospective of CSA is very difficult to establish. Fergusson and Mullen (1999) recommended that one way of approaching this issue was to question the same individuals on multiple occasions to examine the consistency of their reports. Although this issue has rarely been examined, evidence suggests moder-ate-to-good consistency of CSA reports over time (Fergusson and Mullen 1999). Moreover, evidence also indicates that unreliability commonly arises from false negative reports (Fergusson et al. 2000) rather than false positive reports. Presumably greater validity would also be achieved if the presence of disorder were not determined at the same time as reports of CSA were obtained. As almost none of the prospective studies of CSA separated the ascertainment of disorder and reports of CSA, it is difficult to comment on what effect, if any, this may have had on results published to date. In summary, given that retrospective reports of CSA are virtually the only measure of CSA available in the literature, they must be accepted within the context of the caveats stated.

## Strength

Several prospective studies (Fergusson et al. 1996b; Silverman et al. 1996) and several large studies with representative samples (Molnar et al. 2001; Saunders et al. 1999; Stein et al. 1988; Wilsnack et al. 1997) have found an association between CSA and mental disorders. These studies have reported odds ratios (ORs) of between 1.1 for depression (Kendler et al. 2000) and 10.2 for PTSD (Molnar et al. 2001). Despite this variability, CSA has been found to have at least a moderate effect on the outcomes studied (see Table 23.7 for a summary of the evidence). The strength of the relationship between CSA and mental disorders or suicide attempts is generally reduced when the effects of mediating variables are taken into account. This is particularly evident for non-contact forms of abuse.

## Elimination of other possible causes

Child sexual abuse often co-occurs within the context of other family dysfunction, social deprivation, emotional and physical abuse and other environmental stressors that are also associated with mental disorders (Fergusson and Mullen 1999). The interaction between these additional stressors, CSA and adult mental disorders is not likely to be simple or linear (Mullen et al. 1996, 2000; Rutter 1999). Furthermore, it has been argued that the apparent association between CSA and mental disorders can in fact be attributed to family dysfunction rather than to CSA (Rind
et al. 1998). It is therefore important to establish that the effect of CSA on adult functioning remains after controlling for some of these cooccurring factors. The following section contains a discussion of the mediating factors that are commonly reported in studies of the effects of CSA on adult functioning.

Sociodemographic characteristics. CSA is not evenly distributed across sex and socioeconomic groups (Mullen et al. 2000), factors that have been found to be independently associated with mental disorders in adulthood (Andrews et al. 2001; Kessler et al. 1994). However, after controlling for these factors, several studies have demonstrated an independent association between CSA and mental disorders (see Table 23.7). Additionally, a recent meta-analysis found that sex or socioeconomic status did not mediate the relationship between CSA and depressive symptoms, PTSD or suicide attempts (Paolucci et al. 2001).

Family environment. Family environment is one of the most commonly reported mediating factors in the CSA literature. Aspects of family environment studied are myriad but include parental functioning and relationships, domestic violence, parental separation during childhood, growing up away from parents, poor parental health—both physical and emotional-and parental drug and alcohol use (Conte and Schuerman 1987; Fergusson and Mullen 1999; Fromuth 1986; Jumper 1995; Kendall-Tackett et al. 1993; Martin 1996; Mullen et al. 2000; Neumann et al. 1996; Rind et al. 1998; Wyatt and Newcomb 1990). Although measures of adverse family environment vary substantially, it is generally considered to be one of the most important mediators of the effect of CSA on adult functioning (Chandler and Jackson 1997; Fergusson and Mullen 1999; Rind et al. 1998). Moreover, most studies have found that although the independent effect of CSA on adult functioning is substantially reduced once family environment is controlled for CSA, particularly abuse involving penetration is significantly and independently associated with negative outcome (Mullen et al. 2000).

Other abuse. Children who have experienced CSA are at considerably greater risk of experiencing other types of abuse such as physical abuse and neglect (Bifulco et al. 1991; Briere and Runtz 1990; Fergusson and Mullen 1999; Fergusson et al. 1996a; Hibbard et al. 1990; Mullen et al. 1996, 2000; Paradise et al. 1994). There is also some evidence that psychopathology increases with the number of abuse types experienced. In a retrospective study of adult women in New Zealand, the chances of being assigned a clinical diagnosis increased to $24 \%$ for a single type of abuse (sexual, physical or emotional), $41 \%$ for two types of abuse and to $60 \%$ for three types of abuse (Mullen et al. 1996). As such it is difficult to isolate the independent contribution of each of these types of abuse to adult psychopathology. Nonetheless, of the studies included in
the present report that controlled for other types of abuse the majority supported an independent effect of CSA on outcome (Molnar et al. 2001; Mullen et al. 1993, 1996; Yama et al. 1995; Zuravin and Fontanella 1999).

Temperament. The area of CSA deals specifically with human behaviour and therefore with substantial gene environment interaction (Kendler et al. 2000; Rutter 1999). Genetic factors act to enhance vulnerability to mental disorders in general, and may also act to enhance or reduce the risk of developing mental disorders following CSA. Twin studies provide one of the best ways to examine the interplay between genetic and environmental influences. There have been three studies examining the effects of CSA on mental disorders in twins (Dinwiddie et al. 2000; Kendler et al. 2000; Nelson et al. 2002). All three studies concluded that CSA is independently associated with most mental disorders and suicide attempts; two reported that the effect was found, even in twins discordant for CSA, when genetic vulnerabilities and many family factors were controlled for. Unfortunately small numbers of CSA-discordant twins in these analyses meant that, although these studies were included in the final estimates for this report, we could not include a separate genetic adjustment factor.

Protective factors. Much of the discussion around variables that mediate the effects of CSA on adult mental disorder focuses on negative or destructive environmental influences rather than protective ones. However, a recent prospective study examined external protective factors that can modify a child's psychiatric trajectory. Lynskey and Fergusson (1997) developed a regression model that demonstrated factors that protected against the development of psychiatric disorders. These were higher levels of paternal care during childhood and having fewer affiliations with delinquent or substance abusing peers. Once the model adjusted for both of these factors, severity of the sexual abuse (ranging from none, non-contact, contact through to intercourse) was not a significant predictor of outcome. Similarly, having a warm and supportive relationship with the non-offending parent and lower levels of abuserelated stress have been shown to predict resilience in sexually abused girls (Spaccarelli and Kim 1995). In essence, while the majority of research has shown that other negative factors contribute to the risk of developing adult mental disorder having experienced sexual abuse as a child, the converse is also likely to be true. That is, certain positive mediating variables are likely to reduce the risk of negative outcomes following CSA. This concept is often referred to as resilience (Fergusson and Mullen 1999). Unfortunately very few studies have measured protective factors in a systematic way and as such they could not be quantified for the present report.

## Covariation or biologic gradient

For biological risk factors such as blood pressure levels this criterion is typically established by the presence of a dose-response relationship between risk factor and outcome. In the area of child sexual abuse, it is difficult to determine the presence of a dose-response relationship because it is difficult to define a "dose". A dose most closely relates to the severity of abuse to which an individual is subjected. The literature generally defines severity of abuse in five ways: type, frequency, duration, age of onset of abuse, and relationship of victim to offender. Regardless of how it is defined, there is broadly supportive evidence relating the severity of the abuse to the degree of psychiatric or psychological disturbance.

Type of abuse. Using this definition, severity of abuse is generally taken to express the spectrum that ranges from non-contact forms of sexual abuse (e.g. verbal sexual invitations, showing pornography), to contact forms of abuse (touching), through to intercourse. In those studies that have presented risk for disorder according to exposure to different types of abuse, risk for disorder increases as exposure to more severe types of abuse occurs (Baynard 1999; Fergusson et al. 1996b; Kendler et al. 2000; Saunders et al. 1992). In general the literature supports the notion that CSA involving contact or intercourse is associated with a more negative outcome in adulthood than non-contact CSA. It is also the most widely reported definition of the severity of exposure to CSA and is therefore used to obtain estimates for the current report.

Frequency of abuse. It has been demonstrated that experiencing one episode of child sexual abuse is often associated with further sexual victimization. In one study, for example, $16.8 \%$ of children were reabused in the 61-72 months prior to follow-up, with the greatest risk period occurring in the two years immediately after the initial abuse (Levy et al. 1995). In another study of 24507 children with substantiated abuse/ neglect who were monitored up to four years after the initial maltreatment incident, $9.3 \%$ of the children experienced abuse or neglect in the follow-up period (Fryer and Miyoshi 1994). For these children, the risk of reabuse continued to be greater than the risk of abuse in the general population and was greatest immediately following the first notified abuse/neglect incident. For example, 24\% of abused/neglected children were revictimized in the first month following the index event. Bentovim et al. (1987) followed up families who were referred to a treatment programme for sexual abuse and found that $16 \%$ of children had experienced revictimization; in $15 \%$ of children it was unclear whether children had been reabused or not. In a sample of children and adolescents, Boney-McCoy and Finkelhor (1995) found that $39 \%$ of children who had histories of prior sexual victimization had been sexually abused in
the last year. Similar outcomes can occur in adults. Women with a history of CSA were significantly more likely to experience rape as an adult and to be victims of domestic violence (Fleming et al. 1999), which raises the question as to whether child sexual abuse is a vulnerability factor to further sexual abuse in and of itself.

Several studies have reported that not only is reabuse common, it is also associated with poorer outcome. In their meta-analysis, KendallTackett et al. (1993) observed that poor psychological and behavioural outcomes in children were related to a variety of abuse-related variables, including greater abuse frequency. Increases in frequency of abuse have also been shown to be significantly associated with greater severity of psychological disorder, such as making more numerous suicide attempts (Bagley et al. 1995). While it is generally acknowledged that frequency of abuse is associated with more negative outcome, very few studies report outcome for varying abuse frequencies.

Duration of abuse. Duration of CSA has been shown to significantly affect psychological outcome, both in meta-analyses (Kendall-Tackett et al. 1993) and other research. Peters (1988) found that the greater the duration of the abuse, the more mental disorders and suicide attempts in adulthood. This is consistent with the notion that cumulative trauma has a more substantial effect than a single or less frequent abusive event. Once again, outcome of CSA is rarely reported for varying duration of abuse.

Age at onset of the abuse. At first glance, the evidence of the effect of age at abuse onset on mental disorder appears to be conflicting in terms of its direction. Peters (1988) found that women aged 18-36 years who had been older at the time they were sexually abused were diagnosed with more mental disorders or had made more suicide attempts. In contrast, Lynskey and Fergusson (1997) found a significant relationship between being younger at the time of the sexual abuse and increasing rates of mental disorders in a sample of 18 year olds. In both of these analyses, once adjusted for confounders, the relationship between age at abuse onset and mental disorders was no longer significant. Consistent with the conclusion of Browne and Finkelhor (1986) there appears to be no solid evidence for a relationship between age at onset and mental disorders after controlling for other aspects of the abuse and relationship variables.

Relationship of the offender to the child. A meta-analysis of the child sexual abuse research found no significant association between the relationship of the offender to the child, and mental disorders (Paolucci et al. 2001). This relationship is surprising given that intrafamilial abuse may occur over a longer period of time and with greater frequency than extrafamilial abuse (Fergusson and Mullen 1999). Browne and

Finkelhor (1986) suggest two reasons why the relationship of the child to the offender may not be a consistent predictor of negative outcomes. First, that lack of a consistent association may reflect variations in the degree of betrayal, rather than whether the victim and perpetrator are related. Second, while abuse by someone who is trusted may involve betrayal, abuse by a stranger may involve more fear and therefore be more aversive to the victim.

## Consistency

The literature has consistently reported that psychiatric disorders are frequently found to be more common among those subjected to CSA compared to their non-abused peers. This finding persists across a range of populations that have included college students, community samples, school students, children, adolescents and adults, and cohorts in a number of different developed countries. Few studies, however, have been conducted in developing countries. Further research is required to confirm that the deleterious consequences associated with CSA reported in the literature so far also applies to the rest of the world and is not mediated by cultural and social factors.

## Plausibility

It is acknowledged that psychiatric disorder arises from an interaction between adverse environmental influences and an individual's genetic make-up (Rutter 1999). Genetic influences aside, it is accepted that childhood adversity is a potent influence on psychiatric outcome (Brown and Moran 1994; Kessler and Magee 1993). This milieu of adversity has been described as a matrix of disadvantage (Mullen et al. 2000) and includes a variety of socioeconomic, familial and other environmental factors, as outlined earlier. Child sexual abuse falls at the more severe end of the spectrum of this adversity.

In terms of the effect that child sexual abuse has on the individual, it is logical that a child exposed to a traumatic event such as sexual assault may function less well psychologically and may develop phobic responses and anxiety-related symptoms, including PTSD (Green 1988). It has been proposed that the sexual abuse, regardless of type, involves four traumagenic dynamics (Finkelhor and Browne 1988). These are betrayal, powerlessness, traumatic sexualization and stigmatization. Synthesizing the child sexual abuse literature, Polusny and Follette (1995) placed the various outcomes associated with child sexual abuse in the context of emotional avoidance, suggesting that these outcomes are the result of maladaptive coping behaviour. Within this framework, a spectrum of avoidance, anxiety, despair and attempts to control becomes evident. When that fails, it produces anxiety disorders, alcohol and substance abuse, depression and other psychopathology, and suicide at the extreme. Within this context, despite the lack of a biological link between CSA and mental disorders, a causal relationship would certainly be plausible.

### 3.2 DESCRIPTION OF STUDIES INCLUDING METHODOLOGICAL QUALITIES

Table 23.7 presents the characteristics of the studies that contributed to the risk factor-disease relationship grouped by psychiatric outcome. Also reported in Table 23.7 are the range of odds ratios and significance for each study.

### 3.3 Overview of methods

The criteria for identifying relevant studies and the characteristics of excluded studies have been reported on previously. All articles that met the inclusion criteria were coded against each outcome measured. The majority of studies measured more than one of our chosen outcomes. Where possible both unadjusted and adjusted measures of risk (RRs and ORs) were coded from articles. The majority of studies did not present estimates of risk adjusted for relevant confounders. Moreover, many presented proportions only. Regardless of which estimates of risk or association were quoted in articles, $2 \times 2$ tables were coded from every article included in the analysis. Where possible $2 \times 2$ tables were coded for each exposure level and where these data were not available 2 x 2 tables were coded for overall exposure. Where studies presented risks or proportions for both lifetime and current levels of outcome, both were coded. Many articles presented outcome as a continuous rather than a categorical variable. In these cases and where studies had used a measure of outcome that could be mapped to diagnostic criteria via a validated cut-off point, authors were contacted and asked to supply 2 x 2 tables. Unadjusted RRs and ORs and corresponding $95 \%$ confidence intervals were calculated from each $2 \times 2$ table using conventional formulae (Gardner and Altman 1989; Streiner 1998).

EXtrapolations Across SEX, DIAGNOSTIC TIME FRAME AND LEVELS of EXPOSURE

Given the requirements for data presentation (relative risk by age and sex for each level of exposure for each subregion) the biggest source of error in the data arose from having a small number of studies from which to derive estimates. The following sections detail the decisions that were made for each of these extrapolations. However, several general rules applied.

- Consider any theoretical implications of the extrapolation (e.g. is there any reason to expect relative risk will vary with age or sex or that confounding factors will differ for different mental disorder outcomes? Is there a plausible hypothesis or explanation?).
- Assume that no difference exists between the groups of interest (e.g. across sex, age or diagnostic time frame) unless there is clear evidence of a consistent pattern.
Table 23.7 Characteristics of studies included in the risk factor-disease relationship analysis

| Level of evidence | Sample |  |  |  | Adjusted for confounders |  |  |  | Outcome measure ${ }^{a}$ | Diagnosis time frame | Childhood definition | CSA definition ${ }^{\text {b }}$ | OR | $\mathrm{P}<0.05$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Type | $N$ | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | $\begin{gathered} \text { Age } \\ \text { (years) } \end{gathered}$ | Demographics | Family function | Other abuse | Other disorders |  |  |  |  |  |  |
| Depression |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Level I: Twin studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Dinwiddie et al. (2000) | Adult twins | 5946 | 65 | 43 | Yes | No | No | No | SSAGA | Lifetime | $<18$ | Narrow | 2.2-3.9 | Yes |
| Kendler et al. (2000) <br> (10-year follow-up) | Adult twins | 1411 | 100 | 40 | No | Yes | No | No | SCID | Lifetime | $<16$ | Broad | I.1-2.8 | Yes ${ }^{\text {c }}$ |
| Nelson et al. (2002) | Adult twins | 33892 | 58 | 30 | Yes | Yes | No | Yes | SSAGA | Lifetime | $<16$ | Narrow | 1.3-1.7 | Yes ${ }^{\text {c }}$ |
| Level 2: Prospective studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Brown and Harris (1993) (8-year follow-up) | Low income women | 404 | 100 | - | No | Yes | Yes | No | PSE | 12 months | $<17$ | Broad | - | Yes |
| Brown et al. (1999) (13-year follow-up) | Community | 639 | 48 | - | Yes | Yes | No | Yes | DISC | - | $<18$ | Narrow | 3.2 | Yes |
| Ernst et al. (1993) (10-year follow-up) | Community | 421 | 47 | 28 | No | No | No | No | SPIKE <br> interview | - | $<16$ | Broad | - | No |
| Fergusson et al. (I996b) (Follow-up U/K) | Community | 1019 | 51 | 18 | Yes | Yes | No | Yes | CIDI | 2 years | $<16$ | Broad | 3.0-5.4 | Yes |
| Silverman et al. (1996) (17-year follow-up) | Community | 375 | 50 | - | No | No | No | No |  <br> DIS | 2 weeks <br> \& lifetime | $<18$ | - | 2.0 | Yes |

Table 23.7 Characteristics of studies included in the risk factor-disease relationship analysis (continued)

| Level of evidence | Sample |  |  |  | Adjusted for confounders |  |  |  | Outcome measure ${ }^{\text {a }}$ | Diagnosis time frame | Childhood definition | CSA <br> definition ${ }^{\text {b }}$ | OR | $\mathrm{P}<0.05$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Type | $N$ | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | $\begin{gathered} \text { Age } \\ \text { (years) } \end{gathered}$ | Demographics | Family function | Other abuse | Other disorders |  |  |  |  |  |  |
| Level 3: Cross-sectional studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| A. Representative community samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Molnar et al. (2001) | Adults | 5877 | 50 | - | Yes | Yes | Yes | No | CIDI | Lifetime | $<18$ | Broad | 1.8 | Yes ${ }^{\text {c }}$ |
| Saunders et al. (1999) | Adults | 4008 | 100 | 45 | Yes | No | No | No | Clinical interview | Lifetime \& 12 months | $<18$ | Intercourse | 2.5-2.6 | Yes |
| Stein et al. (1988) | Adults | 2683 | 51 | - | Yes | No | No | No | DIS | Lifetime \& 6 months | $<16$ | Narrow | 2.0-2.64 | Yes |
| Wilsnack et al. (1997) | Community | 641 | 100 | - | Yes | No | No | No | DIS | Lifetime | $<18$ | Broad | 2.51 | Yes |
| B. Other community samples (representativeness not known) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Bagley and Ramsay (1985) | Adults | 377 | 100 | 40 | No | No | No | No | CESD | 1 month | $<17$ | Narrow | eta $=0.25$ | Yes |
| Bagley et al. (1994) | Adults | 750 | 0 | 23 | No | No | No | No | CESD $\geq 28$ | 1 month | $<17$ | Narrow | - | Yes |
| Lopez et al. (1995) | Adults | 1821 | 47 | - | - | - | - | - | - | - | $<17$ | Broad | - | Yes |
| Mullen et al. (1996) | Adults | 497 | 100 | - | No | Yes | Yes | No | PSE | Lifetime | $<16$ | Broad | 1.8 | Yes |
| Peters (1988) | Adults | 119 | 100 | - | Yes | No | No | No | SADS | Lifetime | $<18$ | Narrow | - | Yes |
| Saunders et al. (1992) | Adults | 391 | 100 | 42 | No | No | No | No | DIS | Lifetime \& I month | $<18$ | Broad | 1.65-1.75 | Yes ${ }^{\text {c }}$ |
| C. Community subgroups or convenience samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Arata (1999) | College | 92 | 100 | 24 | No | No | No | No | SCID | Lifetime \& I month | <14 | Broad | - | No |


| Chandler and Jackson (1997) | College | 266 | 100 | 19 | No | No | No | No | BDI | 2 weeks | $<18$ | Broad | - | Yes |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Jackson et al. (1990) | College | 40 | 100 | 23 | No | No | No | No | BDI | 2 weeks | $<18$ | Narrow | - | No |
| Zuravin and Fontanella (1999) | Low income women | 513 | 100 | 30 | Yes | Yes | Yes | No | DIS | 1 month | $<14$ | Narrow | 3.36 | Yes |
| Level 4 studies: Prospective case-control studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Swanston et al. (1997) (5-year follow-up) | CPU \& controls from community | 143 | 100 | 15 | Yes | Yes | No | No | $\begin{aligned} & \text { BDI \& } \\ & \text { CDI } \end{aligned}$ | 2 weeks | $<16$ | Narrow | 48.1 | Yes |
| Level 5: Single wave case-control studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Mullen et al. (1993) | Community | 492 | 100 | - | No | No | No | No | PSE | Lifetime | $<16$ | Broad | 2.6-5.2 | Yes |
| Wise et al. (2001) | Community | 732 | 100 | - | Yes | Yes | No | No | SCID | Lifetime | $<18$ | Narrow | 2.2 | Yes |
| Panic disorder |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Level I: Twin studies Dinwiddie et al. (2000) | Adult twins | 5946 | 65 | 43 | Yes | No | No | No | SSAGA | Lifetime | $<18$ | Narrow | 3.5-5.0 | Yes |
| Kendler et al. (2000) (10-year follow-up) | Adult twins | 1411 | 100 | 40 | No | Yes | No | No | SCID | Lifetime | $<16$ | Broad | 1.3-2.6 | Yes ${ }^{\text {c }}$ |
| Level 2: Prospective studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Brown and Harris (1993) (8-year follow-up) | Low income women | 404 | 100 | - | No | Yes | Yes | No | PSE | 12 months | $<17$ | Broad | - | Yes |
| Ernst et al. (1993) (10-year follow-up) | Community | 421 | 47 | 28 | No | No | No | No | SPIKE <br> interview | - | $<16$ | Broad | - | No |

Table 23.7 Characteristics of studies included in the risk factor-disease relationship analysis (continued)

|  | Sample |  |  |  | Adjusted for confounders |  |  |  | Outcome measure ${ }^{\text {a }}$ | Diagnosis time frame | Childhood definition | CSA definition ${ }^{\text {b }}$ | OR | $\mathrm{P}<0.05$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Level of evidence | Type | $N$ | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | $\begin{gathered} \text { Age } \\ \text { (years) } \end{gathered}$ | Demographics | Family function | Other abuse | Other disorders |  |  |  |  |  |  |
| Level 3: Cross-sectional studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| A. Representative community samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Molnar et al. (2001) | Adults | 5877 | 50 | - | Yes | Yes | Yes | No | CIDI | Lifetime | $<18$ | Broad | 0.8-1.4 | Yes ${ }^{\text {c }}$ |
| Stein et al. (1988) | Adults | 2683 | 51 | - | Yes | No | No | No | DIS | Lifetime \& 6 months | $<16$ | Narrow | 3.4-3.9 | Yes ${ }^{\text {c }}$ |
| B. Other community samples (representativeness not known) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Saunders et al. (1992) | Adults | 391 | 100 | 42 | No | No | No | No | DIS | Lifetime \& I month | $<18$ | Broad | 5.0 | Yes ${ }^{\text {c }}$ |
| C. Community subgroups or convenience samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Arata (1999) | College | 92 | 100 | 24 | No | No | No | No | SCID | Lifetime \& I month | $<14$ | Broad | - | No |
| Obsessive-compulsive disorder |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Level 3: Cross-sectional studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| A. Representative community samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Stein et al. (1988) | Adults | 2683 | 51 | - | Yes | No | No | No | DIS | Lifetime \& 6 months | $<16$ | Narrow | - | No |
| B. Other community samples (representativeness not known) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Saunders et al. (1992) | Adults | 391 | 100 | 42 | No | No | No | No | DIS | Lifetime \& I month | $<18$ | Broad | $4.5-\geq 6$ | Yes ${ }^{\text {c }}$ |
| C. Community subgroups or convenience samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Arata (1999) | College | 92 | 100 | 24 | No | No | No | No | SCID | Lifetime \& I month | $<14$ | Broad | - | No |

Post-traumatic stress disorder

| Level 2: Prospective studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Silverman et al. (1996) (17-year follow-up) | Community | 375 | 50 | - | No | No | No | No | DIS | Lifetime | <18 | - | - | Yes |
| Level 3: Cross-sectional studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| A. Representative community samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Davidson et al. (1991) | Adults | 2985 | 54 | - | No | No | No | No | DIS | Lifetime | <16 | Intercourse | 9.5 | Yes |
| Molnar et al. (2001) | Adults | 5877 | 50 | - | Yes | Yes | Yes | No | CIDI | Lifetime | $<18$ | Broad | 5.3-10.2 | Yes |
| Saunders et al. (1999) | Adults | 4008 | 100 | 45 | Yes | No | No | No | Clinical interview | Lifetime \& 12 months | <18 | Intercourse | 3.2-2.4 | Yes |
| C. Community subgroups or convenience samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Arata (1999) | College | 92 | 100 | 24 | No | No | No | No | SCID | Lifetime \& I month | <14 | Broad | - | Yes |
| Hien and Bukszpan (1999) | Obs/gyn. clinic | 98 | 100 | 33 | No | No | No | No | SCID | Lifetime | <18 | Broad | - | Yes |
| Robin et al. (1997) | American Indians | 375 | 58 | 37 | No | No | No | No | SADS-I | Lifetime \& I month | <16 | Broad | 1.6-8.7 | Yes ${ }^{\text {c }}$ |
| Schaaf and McCanne (1998) | College | 269 | 100 | 18 | No | No | No | No | Clinical | Current | <15 | Broad Interview | - | Yes |
| Alcohol abuse or dependence |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Level I:Twin studies Dinwiddie et al. (2000) | Adult twins | 5946 | 65 | 43 | Yes | No | No | No | SSAGA | Lifetime | $<18$ | Narrow | 1.9-2.8 | Yes |
| Kendler et al. (2000) (10-year follow-up) | Adult twins | 1411 | 100 | 40 | No | Yes | No | No | SCID | Lifetime | <16 | Broad | 1.9-6.5 | Yes ${ }^{\text {c }}$ |
| Nelson et al. (2002) | Adult twins | 3892 | 58 | 30 | Yes | Yes | No | Yes | SSAGA | Lifetime | <16 | Narrow | 1.3-1.7 | Yes ${ }^{\text {c }}$ |

Table 23.7 Characteristics of studies included in the risk factor-disease relationship analysis (continued)

| Level of evidence | Sample |  |  |  | Adjusted for confounders |  |  |  | Outcome measure ${ }^{\text {a }}$ | Diagnosis time frame | Childhood definition | CSA definition ${ }^{\text {b }}$ | OR | $\mathrm{P}<0.05$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Type | $N$ | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | $\begin{gathered} \text { Age } \\ \text { (years) } \end{gathered}$ | Demographics | Family function | Other abuse | Other disorders |  |  |  |  |  |  |
| Level 2: Prospective studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Fergusson et al. (I996b) (Follow-up U/K) | Community | 1019 | 51 | 18 | Yes | Yes | No | Yes | CIDI | 2 years | $<16$ | Broad | 1.9-2.7 | Yes ${ }^{\text {c }}$ |
| Silverman et al. (1996) (17-year follow-up) | Community | 375 | 50 | - | No | No | No | No | Various | Lifetime | $<18$ | - | - | Yes |
| Widom and White (1997) (20-year follow-up) | Community | 1190 | 49 | 29 | No | No | No | No | DIS | Lifetime | <11 | Narrow | - | No |
| Level 3: Cross-sectional studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| A: Representative community samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Fleming et al. (1998) | Adults | 710 | 100 | 40 | Yes | Yes | Yes | No | AUDIT | - | $<16$ | Narrow | 0.61 | No |
| Kilpatrick et al. (2000) | Adults | 4023 | 49 | - | Yes | Yes | Yes | No | Clinical interview | 12 month | <17 | Narrow | 2.4 | Yes |
| Molnar et al. (2001) | Adults | 5877 | 50 | - | Yes | Yes | Yes | No | CIDI | Lifetime | $<18$ | Broad | 1.5-1.7 | Yes |
| Saunders et al. (1999) | Adults | 4008 | 100 | 45 | Yes | No | No | No | Clinical interview | Lifetime \& 12 months | $<18$ | Intercourse | 2.0-2.4 | Yes |
| Spak et al. (1998) | Adults | 316 | 100 | - | Yes | No | No | Yes | CIDI-SAM | Lifetime | $<18$ | Broad | 3.5 | Yes ${ }^{\text {c }}$ |
| Stein et al. (1988) | Adults | 2683 | 51 | - | Yes | No | No | No | DIS | Lifetime \& 6 months | $<16$ | Narrow | 1.8 | Yes ${ }^{\text {c }}$ |
| B. Other community samples (representativeness not known) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Peters (1988) | Adults | 119 | 100 | - | Yes | No | No | No | DIS | Lifetime | $<18$ | Narrow | - | Yes |


| . Community subgrou | ni | sample |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Arata (1999) | College | 92 | 100 | 24 | No | No | No | No | SCID | Lifetime \& I month | $<14$ | Broad | - | No |
| Robin et al. (1997) | American Indians | 375 | 58 | 37 | No | No | No | No | SADS-I | Lifetime \& I month | <16 | Broad | 1.8-2.8 | Yes ${ }^{\text {c }}$ |
| Drug abuse or dependence |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Level I: Twin studies Kendler et al. (2000) (10-year follow-up) | Adult twins | 1411 | 100 | 40 | No | Yes | No | No | SCID | Lifetime | <16 | Broad | 1.2-6.6 | Yes ${ }^{\text {c }}$ |
| Level 2: Prospective stud Fergusson et al. (1996b) (Follow-up U/K) |  | 1019 | 51 | 18 | Yes | Yes | No | Yes | CIDI | 2 years | <16 | Broad | 0.1-2.9 | Yes ${ }^{\text {c }}$ |
| Silverman et al. (1996) (17-year follow-up) |  | 375 | 50 | - | No | No | No | No | DIS | Lifetime | $<18$ | - | - | No |
| Widom and White (1997) (20-year follow-up) | Community | 1190 | 49 | 29 | No | No | No | No | DIS | Lifetime | $<11$ | Narrow | - | No |
| Level 3: Cross-sectional studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| A. Representative community samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Molnar et al. (2001) | Adults | 5877 | 50 | - | Yes | Yes | Yes | No | CIDI | Lifetime | $<18$ | Broad | 2.0-2.0 | Yes |
| Stein et al. (1988) | Adults | 2683 | 51 | - | Yes | No | No | No | DIS | Lifetime \& 6 months | <16 | Narrow | 1.8-2.1 | Yes ${ }^{\text {c }}$ |
| C. Community subgroups or convenience samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Arata (1999) | College | 92 | 100 | 24 | No | No | No | No | SCID | Lifetime \& I month | <14 | Broad | - | No |
| Robin et al. (1997) | American Indians | 375 | 58 | 37 | No | No | No | No | SADS-I | Lifetime \& I month | <16 | Broad | 1.6-4.8 | Yes ${ }^{\text {c }}$ |

Table 23.7 Characteristics of studies included in the risk factor-disease relationship analysis (continued)

| Level of evidence | Sample |  |  |  | Adjusted for confounders |  |  |  | Outcome measure ${ }^{2}$ | Diagnosis time frame | Childhood definition | CSA definition ${ }^{\text {b }}$ | OR | $\mathrm{P}<0.05$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Type | N | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | $\begin{gathered} \text { Age } \\ \text { (years) } \end{gathered}$ | Demographics | $\begin{aligned} & \text { Family } \\ & \text { function } \\ & \hline \end{aligned}$ | Other abuse | Other disorders |  |  |  |  |  |  |
| Suicide attempts |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Level I: Twin studies Dinwiddie et al. (2000) | Adult twins | 5946 | 65 | 43 | Yes | No | No | No | SSAGA | Lifetime | <18 | Narrow | 7.1-7.7 | Yes |
| Nelson et al. (2002) | Adult twins | 3892 | 58 | 30 | Yes | Yes | No | Yes | - | Lifetime | $<16$ | Narrow | 0.97-1.1 | Yes ${ }^{\text {c }}$ |
| Level 2: Prospective studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Brown et al. (1999) (I3-year follow-up) | Community | 639 | 48 | - | Yes | Yes | No | Yes | DISC | - | <18 | Narrow | 5.7 | Yes |
| Ernst et al. (1993) (I0-year follow-up) | Community | 421 | 47 | 28 | No | No | No | No | SPIKE interview | - | $<16$ | Broad | - | No |
| Fergusson et al. (1996b) (Follow-up U/K) | Community | 1019 | 51 | 18 | Yes | Yes | No | Yes | CIDI | 2 years | $<16$ | Broad | 0.8-5.0 | Yes ${ }^{\text {c }}$ |
| Silverman et al. (1996) (17-year follow-up) | Community | 375 | 50 | - | No | No | No | No | Various | Lifetime | <18 | - | 10.7 | Yes |
| Level 3: Cross-sectional studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| A. Representative community samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Bagley et al. (1995) | Adolescents | 2112 | 49 | - | No | No | No | No | Studyspecific questions | 6 month | <18 | Broad | 10-37 | Yes |
| Bensley et al. (1999) | Adults | 4790 | 48 | 16 | Yes | No | Yes | No | YRBS | 12 month | $<18$ | Broad | 2.7-47.1 | Yes |
| Garnefski and Arends (1998) | Adults | 13894 | 50 | 15 | Yes | No | No | No | Studyspecific questions | Lifetime | $<19$ | Narrow | - | Yes ${ }^{\text {c }}$ |


| Bagley and Ramsay (1985) | Community | 377 | 100 | 40 | No | No | No | No | Paykel <br> (I972) | - | $<17$ | Narrow | eta $=0.16$ | Yes |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Bagley et al. (1994) | Adults | 750 | 0 | 23 | No | No | No | No | Studyspecific questions | Lifetime | $<17$ | Narrow | - | Yes |
| Leth (2001) | Adults | 1235 | 54 | - | No | No | No | No | - | Lifetime | $<18$ | Broad | - | Yes |
| Martin (1996) | Adolescents | 352 | 43 | 15 | No | No | No | No | Smith and Crawford (1986) | 6 month | $<18$ | Narrow | - | Yes |
| Mullen et al. (1996) | Community | 497 | 100 | - | No | Yes | Yes | No | Studyspecific questions | Lifetime | $<16$ | Broad | 3.6 | Yes |
| Saunders et al. (1992) | Adults | 391 | 100 | 42 | No | No | No | No | DIS | Lifetime \& I month | $<18$ | Broad | 3.0 | Yes ${ }^{\text {c }}$ |
| C. Community subgroups or convenience samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Bendixen et al. (1994) | College | 996 | 51 | 23 | No | No | No | No | Studyspecific questions | Lifetime | - | Broad | - | No |
| Boudewyn and Liem (1995) | College | 438 | 61 | 25 | No | No | No | No | Studyspecific questions | Lifetime | $<14$ | Narrow | - | Yes |
| Chandy et al. (1996) | School students | 2022 | 100 | 15 | No | No | No | No | Studyspecific question | Lifetime | $<18$ | Narrow | - | Yes |
| Chandy et al. (1997) | School students | 740 | 0 | 15 | No | No | No | No | Studyspecific question | Lifetime | $<18$ | Narrow | - | Yes |

Table 23.7 Characteristics of studies included in the risk factor-disease relationship analysis (continued)

| Level of evidence | Sample |  |  |  | Adjusted for confounders |  |  |  | Outcome measure ${ }^{a}$ | Diagnosis time frame | Childhood definition | CSA <br> definition ${ }^{\text {b }}$ | OR | $\mathrm{P}<0.05$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Type | N | $\begin{gathered} \% \\ \text { female } \end{gathered}$ | $\begin{gathered} \text { Age } \\ \text { (years) } \end{gathered}$ | Demographics | Family function | Other <br> abuse | Other disorders |  |  |  |  |  |  |
| Borowsky et al. (1999) | American Indian and Alaska Native youth | 11666 | 52 | 15 | Yes | Yes | Yes | Yes | Studyspecific question | Lifetime | $<18$ | Narrow | - | Yes |
| Lazartigues et al. (1989) | College | 963 | 57 | - | No | No | No | No | - | Lifetime | <16 | Broad | - | Yes |
| Robin et al. (1997) | American Indians | 375 | 58 | 37 | No | No | No | No | Clinical | Lifetime interview | <16 | Broad | 3.1-6.9 | Yes |
| Sedney and Brooks (1984) | College | 102 | 100 | 19 | No | No | No | No | Studyspecific questions | Lifetime | - | Broad | - | No |
| Hibbard et al. (1988) | School students | 712 | 50 | 15 | No | No | No | No | Studyspecific questions | Lifetime | <15 | Broad | 3.1 | Yes |
| Hibbard et al. (1990) | School students | 3998 | 51 | 15 | No | No | No | No | Studyspecific questions | Lifetime | $<15$ | Broad | 9.2 | Yes |
| Yama et al. (1995) | College | 379 | 100 | 20 | No | Yes | No | No | Studyspecific questions | Lifetime | <16 | Narrow | - | Yes |


| Plunkett et al. (2001) (9-year follow-up) | CPU \& controls from community | 259 | 75 | 19 | No | No | No | No | Studyspecific questions | Lifetime | <15 | Narrow | 1.9 | Yes |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Level 5: Single wave case-control studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Briere and Runtz (1986) | Crisis centre | 195 | 100 | 27 | No | No | No | No | Studyspecific questions | Lifetime | <17 | Narrow | - | Yes |
| Molnar et al. (I998) | Street youth | 775 | 35 | 18 | Yes | No | No | No | Studyspecific questions | Lifetime | <18 | Broad | 3.2-4.3 | Yes |
| Mullen et al. (1993) | Community | 492 | 100 | - | No | Yes | No | No | Studyspecific questions | Lifetime | <16 | Intercourse | 8.6-25.6 | Yes |
| Completed suicide |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Level 4 studies: Prospective case-control studies |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Plunkett et al. (2001) | CPU \& controls from community | 259 | 75 | 19 | No | No | No | No | Death certificate | NA | <15 | Narrow | $\begin{aligned} & \text { Rate = } \\ & 0.18 \% \end{aligned}$ | No |
| Key: CPU, Hospital Child Protection Units; NA, not applicable; U/K, unknown. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| No data. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Diagnostic instrument definitions: AUDIT = Alcohol Use Disorders Identification Test; BDI = Beck Depression Inventory (Score $\geq 16$ unless otherwis Depression Inventory (Score $\geq 20$ unless otherwise specified); CESD = Center for Epidemiologic Studies Scale (Depression); CIDI = Composite In (C) = Diagnostic Interview Schedule (for children) ; PSE = Present State Examination; SADS = Schedule for the Affective Disorders and Schizophr Interview for DSM-III-R; SPIKE = Structured Interview to assess psychiatric \& psychosomatic symptoms \& syndromes, social relationships, coping structured assessment for the genetics of alcohol; YRBS $=$ Youth Risk Behaviour Schedule. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Broad: non-contact, contact or intercourse; Narrow: contact or intercourse only. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Significant for some relationships but not others. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

- In the absence of a large number of estimates assume that reporting data from fewer sources is likely to be a greater source of error than extrapolation of data to fit the categories required for reporting (e.g. risk by type of exposure or by sex).


## Sex

Within the literature, a large proportion of the research has examined the sequelae of CSA in females, leaving males underrepresented. This trend was reflected in the data set with many estimates for females and relatively few for males. Additionally, for males, there were no estimates in the severity categories of non-contact, contact and intercourse. Those studies that did examine sequelae in males only gave estimates for the categories of narrow or broad CSA. Therefore, no data were available in the categories required for analysis for males. Given the assumption that more error would be introduced if no estimates were available for males, a decision was made to extrapolate from data for females. To examine the validity of this decision a comparison between males and females was made between those studies that measured outcome in both (restricted to the narrow and broad categories since males only had data for these).

Comparisons were available across all disorder categories and $95 \%$ CIs were compared between male and female estimates within studies to determine if male and female estimates differed significantly. Confidence intervals for the RRs overlapped for depression, agoraphobia, panic, drug and alcohol abuse/dependence for all the studies available for comparison. For PTSD there were significant differences for both of the studies, but as the relative risk was higher for males in one study, and this effect was reversed in the other, no difference between males and females was assumed. For suicide attempts two studies found differences between males and females with the relative risk for males being higher in both. However, in the other six studies available for comparison for suicide attempts no significant differences were observed. Additionally, reviewers have implied that no difference exists between males and females in terms of consequences of CSA (Urquiza and Capra 1990; Watkins and Bentovim 1992). Overall, given that no significant differences were found between the sexes and no theoretical reason presents, no difference between the sexes for the relationship between CSA and mental disorders was assumed.

In light of this decision and in order to maximize the number of estimates available for analysis a hierarchy for selection of studies based on sex was constructed. Where estimates for males and females combined were available from a study these were selected first. Female estimates were then chosen followed by male estimates. This ensured that one estimate was available for each outcome that each study reported.

## Diagnostic time frame

There was considerable variation in the time frame used to measure outcome across studies. While some studies determined the presence of mental disorders over an individual's lifetime, others determined the presence of disorder over the past 12 months, six months or one month. While current or one month estimates might be considered the most accurate measure of current risk, these were only presented for a small number of studies (see Table 23.7). In order to examine the differences between lifetime and current estimates the following analysis was undertaken.

Lifetime and current estimates of risk were compared in the six studies that presented both. All six samples were in the $30-44$-year age group. There were five sets of estimates for depression, two for panic disorder, two for drug dependence, three for alcohol dependence and two for PTSD. Only one study presented estimates for males. Ratios of current to lifetime relative risk ranged from 0.57 to 3.21 with a trend for current risks to be greater than lifetime (ratios of $>1$ ). However, when confidence intervals around RRs were examined very few comparisons were significant.

A second set of comparisons was also undertaken. Relative risks across all studies were grouped according to diagnostic time frame and averaged within age groups and outcomes. Again no clear pattern emerged. On the basis of these analyses and in order to maximize the number of estimates it was decided to include all estimates of relative risk regardless of whether the diagnostic time frame used was lifetime, current or 12 months. Where studies presented both, current risk was used.

## Levels of exposure

Many studies presented relative risk for exposure vs non-exposure only, rather than by levels of exposure. Moreover these studies varied in terms of whether they presented relative risk for broad CSA (non-contact, contact or intercourse) or for narrow CSA (contact or intercourse only). There is strong evidence in the literature to suggest that outcome varies with level of exposure, risk being the highest for those who have experienced abuse involving intercourse (Fergusson and Mullen 1999). In order to examine the relationship between relative risks for each level of exposure the following analysis was undertaken.

Relative risks for overall exposure to CSA were calculated for all studies $(N=5)$ that presented risk by level of exposure. Relative risks for each level of exposure were then expressed as a ratio of the overall risk. Estimates and ratios are presented in Table 23.8.

The ratios in Table 23.8 were then applied to those studies that only reported risk for exposed vs non-exposed. More specifically, the ratios derived from those studies reporting risk estimates for the contact and intercourse categories of abuse were applied to those studies reporting

Table 23.8 Relative risks of each disorder for each level of exposure, as a ratio of overall relative risk ${ }^{\text {a }}$

| Outcome | Study characteristics |  | Exposure level |  |  | Any CSA |  | Ratios |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | RR | $R R_{\text {any }}$ | $R R_{1} /$ | RR2/ | $R R_{3} /$ |
|  | Sample | $N$ |  |  |  | $R R_{1}$ | $R R_{2}$ | $R R_{3}$ | broad | narrow | $R R_{\text {any }}$ | $R R_{\text {any }}$ | $R R_{\text {any }}$ |
| Depression |  |  |  |  |  |  |  |  |  |  |
| Saunders et al. (1992) | Community | 391 | 1.24 | 1.65 | 1.76 | 1.57 | - | 0.79 | 1.05 | 1.12 |
| Fergusson et al. (I996b) | Community | 1019 | 2.19 | 2.04 | 3.83 | 2.56 | - | 0.86 | 0.80 | 1.50 |
| Kendler et al. (2000) | Community | 1411 | 1.25 | 1.39 | 1.83 | 1.48 | - | 0.84 | 0.94 | 1.24 |
| Mullen et al. (1993) | Community | 492 | - | - | 4.38 | 2.62 | - | - | - | 1.67 |
| Panic disorder Saunders et al. (1992) | Community | 391 | 1.49 | 0.91 | 2.67 | 1.59 | - | 0.94 | 0.57 | 1.68 |
| Kendler et al. (2000) | Community | 1411 | 1.41 | 1.76 | 2.42 | 1.86 | - | 0.76 | 0.95 | 1.30 |
| Alcohol dependence |  |  |  |  |  |  |  |  |  |  |
| Fergusson et al. (1996b) | Community | 1019 | 1.63 | 2.19 | 2.02 | 2.01 | - | 0.81 | 1.09 | 1.00 |
| Kendler et al. (1996b) | Community | 1411 | 2.32 | 2.29 | 3.43 | 2.61 | - | 0.89 | 0.88 | 1.31 |
| Drug dependence <br> Fergusson et al. <br> (1996b) | Community | 1019 | 0.78 | 1.64 | 3.66 | 2.13 | - | 0.37 | 0.77 | 1.72 |
| Kendler et al. (2000) | Community | 1411 | 2.63 | 2.03 | 5.19 | 3.05 | - | 0.86 | 0.67 | 1.70 |
| Suicide |  |  |  |  |  |  |  |  |  |  |
| Saunders et al. (1992) | Community | 391 | 0.50 | 2.74 | 3.11 | 2.25 | - | 0.22 | 1.22 | 1.38 |
| Fergusson et al. (1996b) | Community | 1019 | 1.03 | 2.15 | 3.53 | 2.33 | - | 0.44 | 0.92 | 1.47 |
| Mean |  |  |  |  |  |  |  | 0.71 | 0.90 | 1.42 |
| Depression Saunders et al. (1992) | Community | 391 | - | 1.65 | 1.76 | - | 1.69 | - | 0.98 | 1.04 |
| Fergusson et al. (I996b) | Community | 1019 | - | 2.04 | 3.83 | - | 2.68 | - | 0.76 | 1.43 |
| Kendler et al. (2000) | Community | 1411 | - | 1.39 | 1.83 | - | 1.56 | - | 0.89 | 1.17 |
| Banyard (1999) | Low income women | 518 | - | 1.17 | 2.38 | - | 2.04 | - | 0.57 | 1.17 |
| Panic disorder Saunders et al. (1992) | Community | 391 | - | 0.91 | 2.67 | - | 1.63 | - | 0.56 | 1.57 |

Table 23.8 Relative risks of each disorder for each level of exposure, as a ratio of overall relative risk ${ }^{\text {a }}$ (continued)

| Outcome | Study characteristics |  | Exposure level |  |  | Any CSA |  | Ratios |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | $R R_{\text {any }}$ | $R R_{\text {ony }}$ | $R R_{1} /$ | RR ${ }_{2} /$ | $R R_{3} /$ |
|  | Sample | $N$ |  |  |  | $R R_{1}$ | $R R_{2}$ | $R R_{3}$ | broad | narrow | $R R_{\text {any }}$ | $R R_{\text {any }}$ | $R R_{\text {any }}$ |
| Kendler et al. (2000) | Community | 1411 | - | 1.76 | 2.42 | - | 2.01 | - | 0.88 | 1.20 |
| Agoraphobia Saunders et al. (1992) | Community | 391 | - | 2.03 | 5.19 | - | 3.31 | - | 0.61 | 1.57 |
| Alcohol dependenc Fergusson et al. (1996b) | Community | 1019 | - | 2.19 | 2.02 | - | 2.12 | - | 1.03 | 0.95 |
| Kendler et al. (2000) | Community | 1411 | - | 2.29 | 3.43 | - | 2.71 | - | 0.85 | 1.27 |
| Drug dependence Fergusson et al. (1996b) | Community | 1019 | - | 1.64 | 3.66 | - | 2.53 | - | 0.65 | 1.45 |
| Kendler et al. (2000) | Community | 1411 | - | 2.03 | 5.19 | - | 3.20 | - | 0.63 | 1.62 |
| Suicide <br> Saunders et al. (1992) | Community | 391 | - | 2.74 | 3.11 | - | 2.89 | - | 0.95 | 1.08 |
| Fergusson et al. (I996b) | Community | 1019 | - | 2.15 | 3.43 | - | 2.71 | - | 0.79 | 1.27 |
| Mean |  |  |  |  |  |  |  |  | 0.78 | 1.29 |
| $R R_{1}, R R_{2}, R R_{3}$ refers to RRs for level I exposure (non-contact), level 2 (contact) and level 3 (intercourse). |  |  |  |  |  |  |  |  |  |  |
| $95 \% \mathrm{Cls}$ were calculated but are not presented here. All estimates are for females or all persons. None are for males. |  |  |  |  |  |  |  |  |  |  |

risk estimates for the narrow category of abuse. Ratios from studies reporting risk estimates in the non-contact, contact and intercourse categories were applied to the risk estimates for the broad category of abuse. This process generated relative risks for the three categories of abuse required for analysis in relation to the single estimate reported by each study. In this way, the extrapolated relative risks reflect an approximation of what the risk may have been if each study had reported risks for levels of exposure, rather than only risks for exposure vs non-exposure. This ensured that data from studies reporting risk in terms of exposed vs non-exposed, rather than risk by levels of exposure, could still be included in the analysis.

The accuracy of the extrapolated relative risks relies on the premise that risk increases with level of exposure to the same degree across
studies. It was decided, however, that the error introduced by the extrapolation process was less than the error introduced by pooling relative risk estimates from a small number of studies. After the extrapolation process each study that reported an estimate only for the broad category of abuse now had relative risks for the non-contact, contact and intercourse categories of abuse. Accordingly, each study that reported an estimate for the narrow category of abuse now had a relative risk for the contact and intercourse categories of abuse.

In order to calculate confidence intervals for those relative risks that had been extrapolated, standard errors had to be estimated. In this instance, the standard errors derived for the relative risks for the broad and narrow categories of abuse were used. This may not accurately reflect-indeed may underestimate-the true variance around the extrapolated estimates, but for the purposes of this analysis it was assumed to be a reasonable approximation.

## ADJUSTMENT FOR CONFOUNDERS

There is strong evidence within the literature that child sexual abuse is often comorbid with other forms of child abuse and also that child abuse occurs within the context of other family dysfunction (Fergusson and Mullen 1999). Given this, any studies that do not control for these other childhood adversities may inflate the contribution child sexual abuse makes to the onset of our chosen outcomes. Only 13 studies in the data set controlled for confounders. In order to adjust uncontrolled estimates for the potential contribution of confounders a method proposed by Rothman and Greenland (1998) was utilized.

Each of the 13 studies varied in terms of outcomes measured and confounders controlled for. For each study the confounders were recorded and grouped into four categories: sociodemographic, other psychopathology, other abuse and family dysfunction. The studies varied according to the types of confounders controlled for and the measures used. In particular, the category of family dysfunction represented a wide variety of factors. The measures used varied from questionnaires concerning parental attachment and parent/child bonding through to those measuring markers of dysfunction, such as whether the subject grew up in a nuclear family or whether there was parental psychopathology. While the measures varied greatly it was assumed that a common underlying dimension was being measured, that is, the degree to which the family environment was impoverished or dysfunctional and so they were grouped together.

At this stage five studies were excluded from the analysis, as they were not consistent with the pattern of confounders measured by the other seven studies. Saunders et al. (1999), Molnar et al. (1998) and Wilsnack et al. (1997) were excluded since they only controlled for sociodemographic variables and might dilute the adjustment factor if they contributed to the average estimate. Stein et al. (1988) was excluded since
it did not control for any childhood adversity, instead only controlling for subsequent adult sexual abuse. Spak et al. (1998) controlled for childhood behavioural difficulties and childhood psychopathology. These may be important factors confounding the relationship between CSA and alcohol dependence, but, it is unclear to what extent childhood behavioural difficulties and childhood psychopathology are additive to the effects of other confounders. It was therefore excluded.

The eight studies left in the analysis controlled for family dysfunction or other types of abuse, including physical and emotional abuse. Therefore the adjustment factor derived from these studies reflects an adjustment for childhood adversity stemming from dysfunctional home and family environments and is the confound identified as being most important in the literature (Chandler and Jackson 1997; Fergusson and Mullen 1999).

For each of the eight studies adjusted odds ratios and unadjusted odds ratios could be derived. Using the odds ratios an adjustment factor was calculated using the following formula:

$$
U=O R_{u} / O R_{a}
$$

$O R_{u}$ represents the unadjusted odds ratio, $O R_{a}$ represents the odds ratio adjusted for confounders and U is the bias produced from failure to control for the confounders. Since there were only eight studies some assumptions about the commonality of effect across sex, disorder categories and abuse categories had to be made. Table 23.9 presents the final adjustment factors by study and disorder averaged across abuse category and sex.

## Extrapolation across sex

Of the eight studies, seven provided estimates for females but only two provided estimates for males. One study provided estimates for males and females combined. Of the two studies that presented data for both males and females there was no clear pattern of differences between the sexes. Additionally, there is no theoretical reason to expect that the confounders in question would differentially affect the relationship between exposure to CSA and disorder; and due to the paucity of data for males we assumed no difference.

## Extrapolation across levels of exposure

With the exception of three studies, all of the studies provided estimates for only the narrow or broad categories of abuse. In the absence of any data it was assumed that the effects of confounders would be the same across abuse categories.

Table 23.9 Adjustment factors for family dysfunction according to disorder type and study

| Adjustment factor (U) |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Study | Sample type | Depression | Panic | Drug | Alcohol | PTSD | Disorder mean | Suicide |
| Molnar et al. $(2001)^{a}$ | Community | 1.68 | 1.69 | 1.84 | 1.55 | 1.53 | 1.62 | - |
| Kendler et al. $(2000)^{\mathrm{b}}$ | Community | 1.21 | 1.02 | 1.19 | 0.92 | - | 1.08 | - |
| Fergusson et al. $(1996 b)^{\text {b }}$ | Community | 1.43 | - | 0.86 | 1.06 | - | 1.12 | 1.64 |
| Mullen et al. $(1996)^{a}$ | Community | 2.11 | - | - | - | - | 2.11 | 5.48 |
| Zuravin and Fontanella (1999) ${ }^{2}$ | Low income mothers | 1.28 | - | - | - | - | 1.28 | - |
| Mullen et al. $(1993)^{\mathrm{a}}$ | Community | - | - | - | - | - | - | 2.39 |
| Yama et al. $(1995)^{\mathrm{a}}$ | College | - | - | - | - | - | - | I.71 |
| Borowsky et al. $(1999)^{\mathrm{b}}$ | American <br> Indian <br> and Alaska <br> Native youth | - | - | - | - | - | - | 2.08 |
| Mean |  | 1.54 | 1.35 | 1.29 | 1.18 | 1.53 | $1.39^{\text {c }}$ | 2.66 |
| - No data. |  |  |  |  |  |  |  |  |
| Controlled fo <br> Controlled on <br> Excludes suici | both other abus for family facto e. | and family fa |  |  |  |  |  |  |

## Extrapolation across outcomes

Most of the eight studies in the analysis only provided estimates for one or two disorders. There were three studies that gave estimates for most of the disorders, excluding suicide attempts; and within these studies there was no definitive variation across disorder. This conclusion is again limited by paucity of data, but there is no theoretical reason for the confounders to act differentially according to disorder. Suicide attempts appear to be an exception since when the means across disorder and study are compared the adjustment factor for suicide is higher. This may indicate that the confounding variables are more predictive of suicide attempts than the other disorders considered and hence, to be conservative, a different adjustment factor is applied to the suicide estimates.

## Generalizability of the adjustment factor

The validity of adjusting estimates from uncontrolled studies using an adjustment factor derived from studies that do control for confounders will be accurate only to the extent that the confounding effects of the covariates are similar across both the controlled and uncontrolled studies (Greenland 1987). While there is no way to assess this issue quantitatively, we can consider the samples from which our adjustment factors were derived. If the samples from controlled studies are drawn from significantly different groups within the community then generalizability to the uncontrolled studies may be limited. In the current analysis the controlled samples included four community and one college sample that were representative of the samples from uncontrolled studies. Moreover, the two samples from community subgroups, low-income mothers, and American Indian and Alaska Native youth, provided estimates that are comparable to the community estimates.

## Application of adjustment factors

Adjustment factors were applied differentially across the risk estimates according to several criteria. For those eight studies reporting both adjusted and unadjusted risk estimates from which an adjustment factor could be calculated, its own individual adjustment factor was applied to each. For those studies that reported no adjusted estimates and from which an adjustment factor could not be derived, the average of the adjustment factors was applied. The exception was where the unadjusted relative risks were non-significant, and in this instance the relative risks were not adjusted for the presence of confounders. This ensured that a significant protective relationship between CSA and psychiatric outcome was not created artificially.

## META-ANALYSIS

The relative risks were combined using meta-analysis with STATA Intercooled 7. For ease of calculation the macro "meta" was utilized (Sharp and Sterne 1997). Estimates were grouped according to psychiatric outcome and then combined. In most cases the studies combined within each group were significantly heterogeneous according to Cochrane's Q statistic, indicating that moderator variables other than psychiatric outcome were still accounting for significant variation. However the small number of studies prevented further partitioning according to other hypothesized moderator variables.

The presence of significant, unexplained heterogeneity generally indicates preference for a random-effects model to take into account the between-study heterogeneity (Cooper and Hedges 1994; Rothman and Greenland 1998). However, when only two or three studies are available for combination, the between-study variance is estimated with poor precision (Cooper and Hedges 1994). In this instance it was decided that

Table 23.10 Number of studies contributing to each estimate within disorder category

|  | No. of studies |  |  |
| :--- | :---: | :---: | :---: |
| Outcome | Non-contact | Contact | Intercourse |
| Depression | 14 | 23 | 25 |
| Panic disorder | 5 | 8 | 8 |
| Alcohol abuse/dependence | 7 | 13 | 15 |
| Drug abuse/dependence | 4 | 7 | 7 |
| PTSD | 5 | 6 | 8 |
| Suicide attempts | 13 | 29 | 29 |

groups with five or more studies would be combined using a randomeffects model and those with less than five would use a fixed-effects model.

### 3.4 Assessment of causality for each outcome

The following summarizes the evidence of a causal relationship between CSA and each of the seven outcomes examined. A schema or hierarchy was developed to organize this evidence and is outlined below. All of the studies that reported relevant data on proportions of persons in the clinical ranges for each outcome have been tabulated by level of evidence. For some, the $\chi^{2}$ statistic was calculated to test for significance of the odds ratios where these were not available from the studies themselves. Where research on a given psychiatric outcome focused on a subsample of a study group that has been described previously, the findings of the main sample are reported.

Level 1: Studies controlling for both genetic background and family environment.

Level 2: Prospective studies where family environment measured prospectively was used to control confounding of deprivation and CSA.

Level 3: Retrospective cross-sectional studies in which the occurrence of CSA was determined at the time illness was ascertained. Family environment measured reliably was used to control confounding of deprivation and CSA.

Level 3a: Representative community samples: samples of adolescents or adults where either sampling strategy or weighting procedures ensured representativeness of sample.

Level 3b: Non-representative community samples: samples of adolescents or adults where methodology did not necessarily ensure representativeness of sample.

Level 3c: Community subgroup samples: samples of college students, general practice attendees or other community subgroups.

Level 4: Prospective case-control studies in which a CSA group was compared with controls matched for family environment (actually or statistically) and followed over time to measure the onset of mental disorders.

Level 5: Single wave case-control studies in which a CSA group was compared with controls matched for family environment (actually or statistically).

Level 6: Studies of special groups such as foster care children, street youth and juvenile detainees.

## DEPRESSION

The most convincing evidence for a relationship between child sexual abuse and adult depression is from three level 1 studies of adult twins. Dinwiddie et al. (2000) adjusted for demographic factors and found significant relationships between contact or intercourse CSA and depression for female ( $O R=2.20$ ) and male twins $(O R=3.93)$. Restricting the sample to twin pairs discordant for CSA (and thereby substantially reducing the power of the analysis), the relationship between CSA and depression was no longer statistically significant for either males or females even though both were at increased risk for depression.

Kendler et al. (2000) adjusted for family functioning and parental psychopathology in a sample of female twins. They found that odds ratios for the risk of depression in abused twins compared to non-abused twins were modest but not significant for non-genital abuse ( $\mathrm{OR}=1.08$ ), but increased with the severity of the abuse. For abuse involving genital contact $(\mathrm{OR}=1.58)$ and intercourse $(\mathrm{OR}=2.79)$, odds ratios were significant. When analyses were restricted to twins discordant for CSA, odds ratios were of similar magnitude to the previous analysis but only risk in the intercourse category of abuse remained statistically significant (Kendler et al. 2000). It was noted by the author that the discordant twin analyses were limited by small sample size.

In a sample consisting of monozygotic and dizygotic twins who were discordant for CSA, the risk of developing major depression was 1.68 for women and 1.25 in men; however, this relationship was significant only for women (Nelson et al. 2002). Of note was the non-significant trend for non-abused co-twins to also be at greater risk of having a history of depression in comparison to twin pairs with no history of abuse, providing evidence for the contribution of familial factors to the onset of disorder. Therefore, the finding that abused co-twins have a higher risk of depression than their non-abused co-twins demonstrates the increased risk that CSA contributes over and above family background.

Across the three twin studies a significant relationship between CSA and depression has been found. Particularly, abuse involving contact and penetration has been found to significantly increase the risk of a depressive disorder. Discordant twin analyses were conducted for three studies and have the advantage of controlling more tightly for familial and genetic factors. These analyses also found CSA to contribute significantly to onset of depression, but in the case of Dinwiddie et al. (2000) and Kendler et al. (2000) only when the abuse involved intercourse. While the small sample sizes of the discordant twins restricted the power of the analyses they do provide strong support for a causal relationship between depression and CSA involving intercourse.

To conclude, there have been three level 1, and 19 level 2 and 3 studies examining the relationship between CSA and depression; 19 of them supported a significant relationship. There is strong, consistent evidence that, after adjustment for confounders, there is a significant relationship between CSA and depression in adults, particularly for women who experienced more severe forms of abuse.

## PANIC DISORDER

There is evidence from level 1 twin studies for a relationship between CSA and panic disorder. Defining sexual abuse as contact or intercourse, Dinwiddie et al. (2000) adjusted for demographic factors and found significant relationships between CSA and panic disorder for female (OR = 3.54 ) and male twins ( $\mathrm{OR}=5.02$ ). In the twin pairs discordant for CSA, the relationship between CSA and panic disorder was no longer significant for women even though they were at increased risk for the disorder ( $\mathrm{OR}=2.00$ ). Similar analyses were unable to be conducted for men because of sample size.

In their analyses of panic disorder, Kendler et al. (2000) adjusted for family functioning (but not parental psychopathology) and found that odds ratios in abused twins compared to non-abused twins were modest but not significant for either non-genital abuse ( $O R=1.25$ ) or for abuse involving genital contact ( $\mathrm{OR}=1.92$ ). Odds ratios for intercourse (OR $=2.62$ ) were significant. Discordant co-twin analyses were not performed due to small sample sizes. Prospective studies have found equivocal evidence of a relationship (Brown and Harris 1993; Ernst et al. 1993) and studies using community samples have found significant associations between some forms of CSA but not others (Molnar et al. 2001; Saunders et al. 1992; Stein et al. 1988). Although few studies have examined panic disorder as an outcome of CSA, there is evidence from two level 1 studies, one level 2 study and three level 3 studies that the rates of panic disorder are increased in sexually abused adults and are more strongly predicted by abuse involving penetration.

## ObSESSIVE-COMPULSIVE DISORDER (OCD)

Only three studies explored the relationship between CSA and OCD and only one of these controlled for confounders. This representative community sample (level 3a) adjusted for demographics and other abuse but found no significant relationships between CSA and lifetime or 6-month history of OCD (Stein et al. 1988). Results from the two samples that did not adjust for confounders were mixed. In a sample of female college students (level 3c) (Arata 1999), no significant relationships were reported between CSA and lifetime or 1-month history of OCD. However, childhood was defined as aged $<14$ years and was therefore quite restrictive. Saunders et al. (1992) (level 3b) defined childhood more broadly, as aged $<18$ years. While results for non-contact abuse were not significant, women were 4.5 times more likely to have a lifetime history of OCD if they had experienced CSA in the form of contact abuse and over six times more likely if they had experienced intercourse; and these findings were significant. This pattern of significant results was also the case for the relationship between CSA and women who met the diagnostic criteria for OCD at interview. In summary, only one study supported a relationship between CSA and OCD. Further research is required in order to confirm this association. For this reason the risk for OCD will not be calculated in this report.

## Post-traumatic stress disorder (PTSD)

There was one study of PTSD that provided level 2 evidence (Silverman et al. 1996). This was a 17-year follow-up of a community sample of children in a working class area. There was no adjustment for confounders; however, a significant relationship was found between CSA and lifetime history of PTSD for females at age 21 years.

Of the remaining studies of PTSD, all were from level 3 evidence and only two of these adjusted for confounders. These studies were representative community samples (level 3a). One study restricted CSA to intercourse only, and only adjusted for age, but found that CSA significantly increased the risk of lifetime $(\mathrm{OR}=3.42)$ and recent $(\mathrm{OR}=3.17)$ PTSD (Saunders et al. 1999). The other study adjusted for demographic and family variables such as parental substance abuse and psychopathology, and the presence of physical abuse (Molnar et al. 2001). Child sexual abuse was defined as contact sexual abuse or intercourse and was found to be significantly related to lifetime history of PTSD in women ( $\mathrm{OR}=10.2$ ) and men ( $\mathrm{OR}=5.3$ ). A notable finding was that the risk of PTSD was significantly higher for penetrative abuse than contact abuse.

In summary, there were eight level 2 and 3 studies; all showed a significant relationship between child sexual abuse and adult PTSD. There is strong, consistent evidence that, after adjustment for confounders, there is a significant relationship between CSA and PTSD in adults, particularly for those who have experienced more severe forms of abuse.

## AlCOHOL ABUSE OR DEPENDENCE

The evidence for a relationship between childhood sexual abuse and adult alcohol abuse/dependence is from the three level 1 studies of adult twins. Dinwiddie et al. (2000) adjusted for demographic factors and found significant relationships between contact or intercourse CSA and alcohol abuse/dependence for female ( $\mathrm{OR}=2.81$ ) and male twins (OR $=1.91)$. Restricting the sample to twin pairs discordant for CSA, the relationship between CSA and alcohol abuse/dependence was no longer significant for either males or females; however females remained at increased risk ( $\mathrm{OR}=2.50$ ).

Kendler et al. (2000) adjusted for family functioning and parental psychopathology in a sample of female twins. The odds ratios for the risk of alcohol abuse/dependence in abused twins compared to non-abused twins were modest but not significant for genital contact abuse ( $\mathrm{OR}=$ 1.91). For non-genital abuse ( $\mathrm{OR}=3.20$ ) and intercourse ( $\mathrm{OR}=6.48$ ), odds ratios were significant. Analyses for twins discordant for CSA, or where the co-twin had experienced a less severe form of abuse, showed that intercourse significantly increased the risk of alcohol abuse/dependence. In another sample that also presented results for CSA among discordant twins the risk of developing alcohol abuse/dependence was 1.73 for women and 1.25 in men; however this relationship was significant only for women (Nelson et al. 2002). In particular, the findings from the discordant twin analyses suggest a causal relationship between CSA and alcohol abuse/dependence, and indicate that CSA increases the risk over and above that arising from family background. Low sample size of the discordant twin analyses may have produced the non-significant results for contact forms of abuse but the significant finding for intercourse in spite of low power makes this finding notable. Other prospective and community studies also provide evidence for a relationship between CSA and alcohol abuse/dependence with two out of three level 2 studies and seven out of nine level 3 studies presenting significant odds ratios.

## DRUG ABUSE OR DEPENDENCE

There was only one level 1 study that explored drug abuse or dependence. In their study of female twins, Kendler et al. (2000) adjusted for family functioning and parental psychopathology. They found that odds ratios for the risk of drug abuse/dependence in abused twins compared to non-abused twins were modest but not significant for genital contact abuse ( $\mathrm{OR}=1.21$ ). For non-genital abuse ( $\mathrm{OR}=3.57$ ) and intercourse ( $\mathrm{OR}=6.55$ ), odds ratios were significant. Analyses of twins discordant for CSA, or where the co-twin had experienced a less severe form of abuse, showed a similar pattern of results with non-genital CSA and CSA involving intercourse placing subjects at increased risk (ORs 4.29 and 2.85 , respectively). However, these results were not significant, and this is likely to be a function of small sample sizes.

Of the three level 2 studies, only one showed a significant association and was significant only for the intercourse category of abuse ( $\mathrm{OR}=5.1$ ) (Fergusson et al. 1996b). The two other prospective studies failed to show a relationship (Silverman et al. 1996; Widom and White 1997). Of the four level 3 studies using community samples, three demonstrated a significant association but not across all three levels of abuse. In summary, there is evidence to suggest a relationship between CSA and drug dependence but only for the more severe forms of abuse. However, the evidence is more equivocal compared to other outcomes and more research is required to confirm this relationship.

## Suicide attempts

There were two level 1 studies of adult twins that addressed suicide attempts. In the first, Dinwiddie et al. (2000) adjusted for demographic factors and found significant relationships between contact or intercourse CSA and serious suicide attempts for female ( $\mathrm{OR}=7.74$ ) and male twins $(\mathrm{OR}=7.07)$. Restricting the sample to twin pairs discordant for CSA, the relationship between CSA and suicide attempts was no longer significant for females; however they were still at increased risk ( $\mathrm{OR}=2.33$ ). Risk estimates could not be computed for males in the discordant twin analysis due to small numbers. In the second study of twins who were discordant for CSA (Nelson et al. 2002) the risk of suicide attempts was 2.33 for women and 4.50 in men. This relationship was significant for women, and almost reached significance in men (95\% CI 0.97-20.83), even after adjustment for demographic and family factors. Furthermore, three of four prospective studies (level 2) and 16 of 18 level 3 studies also found significant associations between CSA and suicide attempts. There is strong evidence for a relationship between CSA and suicide attempts.

## Completed suicide

Completed suicide, by definition, can only be studied prospectively. Only one study to date has examined the relationship between completed suicide and child sexual abuse. This Australian study was based on a relatively small sample of sexually abused young people, the majority of them female, who presented to hospitals for the abuse, and a control group of non-abused young people from the community (Plunkett et al. 2001). Without controlling for confounders, CSA was found to increase the risk of completed suicide; and the rate was found to be very high, 179.5 per 100000 person-years, or $1.8 \%$, but this relationship was not significant. It should be noted that the number of completed suicides was 3 out of a sample of 259 , and none of the controls had committed suicide, so analyses were somewhat limited. The national suicide death rates for 15-24 year olds during the same time period ranged from 13.8 to 16.7 , so the observed rate in the study was 10.7 to 13.0 times that of the Australian national suicide death rate (Dudley et al. 1998).

## RELATIONSHIP BETWEEN ATTEMPTED AND COMPLETED SUICIDE

Attempted suicide has consistently been shown to be a strong predictor of completed suicide (see Graham et al. 2000 for review). Estimates of the magnitude of this risk however, vary considerably. The prevalence of completed suicides among those who have attempted has been estimated at $1 \%$ to $19 \%$ in the 12 months after the attempt (Diekstra 1992; Graham et al. 2000), $2.8 \%$ after 8 years (Hawton and Fagg 1988) and $10 \%$ after 10 years (Tejedor et al. 1999). Lifetime prevalence of completed suicide in those who have ever been hospitalized for suicidality has been estimated at $8.6 \%$ (Bostwick and Pankratz 2000). Although difficult to quantify on the basis of available data, the evidence suggests that approximately 1 in 10 individuals who attempt suicide will die by suicide at some point following the attempt. These reviews also pointed out, however, that the majority of completed suicides are not preceded by a suicide attempt (Graham et al. 2000). Estimates of risk for completed suicide that are derived only from estimates of previous attempts are therefore likely to underestimate the risk of completed suicide.

The issue is further complicated in that mental disorders have also been shown to be strong and consistent predictors of both suicide attempts and completed suicide (Brent et al. 1999; Graham et al. 2000; Harris and Barraclough 1997; Hawton and Fagg 1988; Kessler et al. 1999). Within this context, the exact nature of the relationship between CSA, mental disorder, suicide attempts and completed suicide is likely to be complex (Beautrais 2000). Unfortunately, none of the studies in the present report that examined the relationship between CSA and suicide attempts controlled for concurrent psychopathology.

From the above it is concluded that there is evidence of a relationship between attempted suicide and completed suicide. However, for the purposes of the current analysis it is necessary to determine whether a person who has been subject to CSA has any higher or lower chance of completing suicide conditional on having attempted it. To our knowledge these data do not exist. In the absence of these data the relative risk for suicide attempts will be used as proxies for the relative risk for completed suicide. This assumes a constant relationship between attempted and completed suicide given exposure to CSA. The extent to which this underestimates or overestimates the relative risk is unknown.

### 3.5 Estimates of risk factor-disease relationships by age, SEX AND SUbREGION

## Estimates of RISK

Table 23.11 summarizes the results of the meta-analysis and reports the relative risks for the psychiatric outcomes examined. No estimates are available for different age and sex groups. All studies were from AMR-A, WPR-A or EUR-A. No studies were available from other subregions that may have different cultural norms and socioeconomic
Table 23.1I Unadjusted and adjusted relative risks across disorder and abuse categories

|  | Non-contact abuse |  |  |  | Contact abuse |  |  |  | Intercourse |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Adjusted ${ }^{\text {a }}$ |  | Unadjusted |  | Adjusted ${ }^{\text {a }}$ |  | Unadjusted |  | Adjusted ${ }^{\text {a }}$ |  | Unadjusted |  |
|  | Pooled RR | 95\% Cl | Pooled RR | 95\% Cl | Pooled RR | 95\% Cl | Pooled RR | 95\% Cl | Pooled RR | 95\% Cl | Pooled RR | 95\% Cl |
| Depression | 1.06 | 0.91-1.24 | 1.37 | 1.16-1.61 | 1.32 | 1.16-1.51 | 1.80 | 1.61-2.02 | 2.04 | 1.78-2.35 | 2.80 | 2.44-3.22 |
| Panic disorder | 1.01 | 0.76-1.35 | 1.52 | 1.13-2.04 | 1.64 | 1.12-2.42 | 2.27 | 1.56-3.32 | 2.60 | 1.70-3.97 | 3.58 | 2.34-5.48 |
| Alcohol abuse/dependence | 1.19 | 0.85-1.67 | 1.35 | 0.91-2.02 | 1.32 | 1.07-1.63 | 1.61 | 1.24-2.09 | 1.87 | 1.47-2.39 | 2.58 | 2.04-3.24 |
| Drug abuse/dependence | $1.03^{\text {b }}$ | 0.84-1.26 | $1.57{ }^{\text {b }}$ | 1.28-1.93 | 1.31 | 0.90-1.91 | 1.68 | 1.03-2.72 | 2.40 | 1.46-3.96 | 3.04 | 1.83-5.04 |
| PTSD | 1.95 | 0.95-3.98 | 2.70 | 1.32-5.54 | 2.95 | 1.53-5.68 | 4.10 | 2.12-7.90 | 4.48 | 2.33-8.65 | 6.23 | 3.23-12.02 |
| Suicide attempts | 1.02 | 0.71-1.45 | 2.80 | 1.89-4.15 | 1.32 | $1.08-1.60$ | 3.25 | 2.53-4.18 | 2.21 | 1.77-2.76 | 5.56 | 4.32-7.16 |
| These relative risks have been adjusted for family dysfunction and other types of abuse. The RRs for suicide attempts have been divided by 2.66 and have been adjusted by I.39. |  |  |  |  |  |  |  |  |  |  |  |  |
| b Fixed-effects model used to combine estimates due to small number of studies available. |  |  |  |  |  |  |  |  |  |  |  |  |

circumstances. No differences in risk were assumed across sex and age breakdowns since numbers were few and many studies used lifetime diagnoses.

On the whole several conclusions can be drawn from these results. First, the relative risks were not significantly different across types of mental disorder, suggesting that CSA is not particularly associated with any one disorder. Rather the risk appears to be pervasive across the whole spectrum of mental disorders examined. This lack of specificity makes CSA particularly damaging, putting individuals at risk for a wide range of mental disorders.

Second, with the exception of suicide attempts, risks did not vary significantly across categories of abuse. This may reflect the small number of studies available for analysis, as there is a general trend for increased risk to be associated with "increased" exposure to CSA. That is, as more severe forms of CSA are experienced the risks for developing a mental disorder increase. This may indicate that those exposed to CSA do not represent a homogeneous group but instead reflect a group that varies in terms of exposure and subsequent risk for psychiatric disorder (Fergusson and Mullen 1999). However, this finding is largely an artefact of the extrapolation process. Further research will enable this "dose-response relationship" to be confirmed.

Non-contact abuse was not a significant predictor of risk after external adjustment for confounders. Non-contact abuse may constitute a more heterogeneous category of abuse compared to the contact and intercourse categories. Certainly, in the studies that contributed to this analysis the category of non-contact abuse encompassed a variety of acts. Such heterogeneity may make the results of the analysis hard to interpret. The question is whether non-contact does not place individuals at increased risk for disorders or whether some non-contact forms of abuse are more or less harmful than others. Cultural factors are likely to play a role in this and further investigation is required. Across the majority of disorders (excluding drug abuse/dependence) the relative risks for contact and intercourse forms of abuse remained significant after external adjustment. These results are more easily interpreted as the two categories are a more homogeneous group of acts and can be more tightly defined.

Results of the current meta-analysis were consistent with the only other review to look systematically at psychiatric diagnosis as an outcome of CSA. Fergusson and Mullen (1999) collated and reanalysed data from 12 studies reporting on the relationship between CSA and psychiatric dysfunction. While Fergusson and Mullen (1999) did not statistically combine the estimates, consistent and pervasive relationships between CSA and adult psychopathology were apparent. Other metaanalyses have been conducted in this area (Jumper 1995; Paolucci et al. 2001; Rind 1997; Rind et al. 1998) but have focused on continuous measures of adjustment. Jumper (1995) and Paolucci et al. (2001) have com-
mented on the difficulty in quantifying the impact that confounders have since many large-scale community studies do not measure them. Rind et al. (1998) was the only meta-analysis to control for the confounding effect of family environment using college samples since confounds are more likely to be measured in studies using these samples. Results of their analysis showed that the relationship between CSA and adjustment in adulthood disappeared after controlling for family environment. Adjustment in their analysis refers to the various psychological correlates of CSA measured in the studies that contributed to their analysis. Eighteen categories of psychological correlates were coded from the studies and included anxiety and depressive symptoms, plus broader areas of adjustment such as sexual adjustment and social adjustment. When commenting on Rind et al. (1998) results, Kendler et al. (2000) stated that CSA appears to be more related to lifetime psychiatric disorder rather than cross-sectional measures of adjustment which focus solely on current well-being. Certainly, the results of the current metaanalysis do suggest a significant relationship between CSA and psychiatric disorder after adjustment for important confounders.

Several caveats should be mentioned in regard to the current analysis. The external method of adjustment used was limited by the methodology of the studies from which the adjustment factors were derived. Across the studies family dysfunction was not defined or measured consistently and the measures that were used may not be reliable. It should also be noted that current research methodologies make it impossible to determine whether any antecedent signs or symptoms of psychiatric disorders mean that the child is more vulnerable to abuse, thus inflating the risk factor-disease relationship. Additionally, the majority of studies contributing to this analysis were conducted during a time when CSA was ignored and the victims were stigmatized. While no cohort effects were found for reporting episodes of CSA the stigmatization could possibly have had an impact on the degree to which people were affected.

To conclude, it appears that CSA is particularly damaging with effects evident over and above other forms of childhood adversity. Contact and intercourse categories of abuse remained significant after controlling for confounders but the category of non-contact was non-significant. Further research is required to understand the effects of various forms of noncontact abuse.

## RISK REVERSIBILITY

By definition, exposure to CSA stops at the age of 17 years. The reversibility of risk therefore refers simply to decline in the risk of developing mental disorder given exposure to CSA over the lifespan. Problems in quantifying the decline arise from both paucity of data and competing theoretical arguments. There are two arguments. Either assume that risk remains constant throughout the lifespan or assume that
risk subsides over time and eventually becomes that of people who were never abused. One can argue that risk remains constant over time as there is evidence to suggest that CSA alters your life trajectory such that you are more likely to experience problems with relationships, selfesteem and sexual adjustment (Mullen et al. 2000). These problems in themselves are associated with mental disorder and so potentially mediate the relationship between CSA and disorder in later life. Therefore, those exposed to CSA remain at increased risk for mental disorder compared to those who have not been exposed. Alternatively one can argue that risk decreases over the lifespan. As a person moves away from a traumatic life event, its power to inflict psychological harm is lessened and so those exposed to CSA eventually have the same risk for mental disorders as those not exposed.

The data collected for this report do not inform either argument further. Preliminary analyses divided the relative risks into two age groups, 15-29 and 30-44 years. Confidence intervals overlapped suggesting no significant difference between the two age groups. Two explanations can be given. First, the numbers in the groups were small limiting the robustness of the estimates. Second, most studies reported estimates of risk for lifetime mental disorders thereby preventing any clear distinction in onset of mental disorders between the two age groups from being made. In the absence of data we have assumed that risk remains constant over the lifespan, as it becomes mere speculation if one tries to estimate the amount risk reduces over time. In order to better inform this decision, data on risk for current disorder are required, broken down into appropriate age groups. Further analyses of the large community and prospective samples are required to address this.

### 3.6 Reasons and implications for extrapolation of risk FACTOR-DISEASE RELATIONSHIPS FROM ONE SUBREGION TO ANOTHER

This represents perhaps the biggest threat to validity to the present study. The studies in the risk analysis were overrepresented by samples from Australia, Canada, New Zealand, some European countries and the United States of America. The prevalence of psychiatric disorders varies from country to country and this variability represents the myriad social, economic and cultural factors that interplay with the development of disorder. The degree to which these same factors mediate the relationship between CSA and mental disorder cannot be quantified and is difficult to even speculate on. The answer to this is not likely to be reached in the near future as extensive research is required before sound conclusions can be drawn. The paucity of data and theoretical complexity necessitated the assumption that the relative risks remain constant across subregions.

### 3.7 Quantitative and Qualitative sources of uncertainty

Uncertainty in the current analysis came from several sources. Metaanalysis was used as a method of quantifying the quantitative uncertainty around the final prevalence estimates taking into account sample size or variability between studies, whichever is appropriate given the homogeneity of the estimates being combined.

The methods of extrapolation also introduce uncertainty into the analysis. In particular, the extrapolation conducted to get relative risks for the three levels of exposure will have introduced error. The ratios used for extrapolation came from only five studies and the relationship between the relative risks for each level of exposure may vary according to the population studied.

The external method employed to adjust for confounders introduces three sources of uncertainty. First, the instruments used to measure the confounders may have variable reliability. Second, the 12 studies that contributed to the adjustment factor all measured family dysfunction differently thereby creating another source of uncertainty. Third, the ability of the adjustment factor to be generalized to other studies conducted on different populations can be questioned. It is possible that confounders may mediate the relationship between CSA and outcomes differentially depending upon the population studied. Not enough studies that controlled for confounds were available to see if the adjustment factor varied in any systematic way.

Additionally, while all estimates came from AMR-A, WPR-A and EUR-A not all countries from those subregions were represented. Inter-country variations within subregions are possible and may affect the generalizability of our relative risk estimates. Even greater uncertainty is introduced when estimates from one subregion are extrapolated to another, a decision necessitated by the paucity of data from other subregions.

## 4. <br> Results and discussion

The discussion of the estimates of the burden in mental disorder outcome that is attributable to CSA will focus on differences between mental disorders, followed by an examination of sex, age and subregional differences. Across the world, CSA contributed to between $4 \%$ and $5 \%$ of the disability-adjusted life years (DALYs) in males and between $7 \%$ and $8 \%$ of the DALYs in females for each of depression, alcohol abuse/dependence and drug abuse/dependence. The attributable fractions were higher for panic disorder ( $7 \%$ for males and $13 \%$ for females) and higher still for PTSD ( $21 \%$ for males and $33 \%$ for females). For suicide attempts the attributable fractions were $6 \%$ for males and $11 \%$ for females. As discussed above, the confidence intervals around the relative risks of each of the mental disorders overlap and thus the apparent differences may not be real. The same applies to the attributable fractions.

On the whole CSA contributes to a higher percentage of DALYs for females than for males. This difference is driven by the difference in the prevalence of CSA for females and males. There are slight subregional variations in the attributable fractions. In particular, AFR-E and SEARD have higher attributable fractions than the other subregions. This is a function of the higher prevalence rates for these subregions. However, data for these subregions came from a few studies that were poor methodologically.

The number of DALYs attributable to CSA varies as this is a function of both the attributable fractions and the amount of burden of disease accounted for by psychiatric disorders in the various subregional groups. One pattern is evident, however: the number of DALYs are greater in the younger age groups and decline in the older age groups. Since risk was assumed to be constant across age, this merely reflected the distribution of DALYs for mental disorders, which impact largely in the younger age groups due to their early onset and chronic nature.

## 5. METHODS FOR PROJECTION OF EXPOSURE FORWARD

Estimates of avoidable burden, the amount of burden that could be prevented if exposure to CSA was curtailed, are dependent upon the expected prevalence of CSA in the future. As outlined above, we have assumed that CSA will not vary over time. The reviews conducted on cohort effects have provided equivocal evidence (Bagley 1990, 1995; Bagley and Ramsay 1985; Bickerton et al. 1991; Feldman et al. 1991; Fergusson et al. 2000). Three reviews concluded that the prevalence of CSA could increase over time while three reviews also conclude that there is no evidence to support a change in prevalence over time. Analysis of this phenomenon in our own data set also provided no evidence of a change in prevalence over time. Additionally, it is difficult to speculate on what future factors may arise to influence prevalence. Therefore, the estimates of avoidable burden should be the same as those for attributable burden.

## Acknowledgements

Steering Committee: Majid Ezzati, David Fergusson, Bronwyn Gould, Beth Kotze, Ron Kessler, Alan Lopez, Michael Lynskey, Judy Martin, Paul Mullen, Anthony Rodgers, Bedirhan Üstün.

## Note

1 See preface for an explanation of this term.

## References

Agathonos H, Alexandridis A, Fereti I (1992) A retrospective study of child sexual abuse among Greek university students: implications for prevention and future research. (Proceedings of the Ninth International Congress on Child Abuse and Neglect.) Chicago, IL.
American School Health Association (1989) The national student health survey: a report on the health of America's youth. Third Party Publishing Co., Oakland, CA.

Anderson J, Martin J, Mullen P, Romans S, Herbison P (1993) Prevalence of childhood sexual abuse experiences in a community sample of women. Journal of the American Academy of Child and Adolescent Psychiatry, 32:911-919.

Andrews G, Henderson S, Hall W (2001) Australia's mental health: an overview of the general population survey. British Journal of Psychiatry, 178:145153.

Arata CM (1999) Repeated sexual victimization and mental disorders in women. Journal of Child Sexual Abuse, 7(3):1-17.

Arroyo JA (1997) Childhood sexual abuse among Hispanic and nonHispanic white college women. Hispanic Journal of Behavioral Sciences, 19:57-68.

Bagley C (1990) Is the prevalence of child sexual abuse decreasing? Evidence from a random sample of 750 young adult women. Psychological Reports, 66:1037-1038.

Bagley C (1991) The prevalence and mental health sequels of child sexual abuse in a community sample of women aged 18 to 27. Canadian Journal of Community Mental Health, 10:103-116.
Bagley C (1995) Child sexual abuse in Canada: further analysis of the 1983 national survey. In: Child sexual abuse and mental health in adolescents and adults: British and Canadian perspectives. Bagley C, ed. Ashgate Publishing, Brookfield, VT.

Bagley C, Bolitho F, Bertrand L (1995) Mental health profiles, suicidal behavior, and community sexual assault in 2112 Canadian adolescents. Crisis, 16: 126-131.

Bagley C, Mallick K (2000) Prediction of sexual, emotional and physical maltreatment and mental health outcomes in a longitudinal cohort of 290 adolescent women. Child Maltreatment, 5:218-226.
Bagley C, Ramsay R (1985) Sexual abuse in childhood: psychosocial outcomes and implications for social work practice. Journal of Social Work and Human Sexuality, 4:33-47.

Bagley C, Wood M, Young L (1994) Victim to abuser: mental health and behavioral sequels of child sexual abuse in a community survey of young adult males. Child Abuse and Neglect, 18:683-697.

Baker AW, Duncan SP (1985) Child sexual abuse: a study of prevalence in Great Britain. Child Abuse and Neglect, 9:457-467.

Baldini G (1996) Rape and sexual abuse within the Aboriginal communities. Sexual Assault Referral Centre, Perth, WA.

Banyard VL (1999) Childhood maltreatment and the mental health of lowincome women. American Journal of Orthopsychiatry, 69:161-171.

Barthauer LM, Leventhal JM (1999) Prevalence and effects of child sexual abuse in a poor, rural community in El Salvador: a retrospective study of women after 12 years of civil war. Child Abuse and Neglect, 23:1117-1126.
Barton G (1987) Child sexual abuse: an Australian view; the statistics. IANUA, The Official Journal of the Kindergarten Teachers' Association of Victoria, 1:15.
Bayatpour M, Wells RD, Holford S (1992) Physical and sexual abuse as predictors of substance use and suicide among pregnant teenagers. Journal of Adolescent Health, 13:128-132.

Beautrais AL (2000) Risk factors for suicide and attempted suicide among young people. Australian and New Zealand Journal of Psychiatry, 34:420-436.

Bendixen M, Muus KM, Schei B (1994) The impact of child sexual abuse-a study of a random sample of Norwegian students. Child Abuse and Neglect, 18:837-847.

Bensley LS, Van Eenwyk J, Spieker SJ, Schoder J (1999) Self-reported abuse history and adolescent problem behaviors. I. Antisocial and suicidal behaviors. Journal of Adolescent Health, 24:163-172.
Bentovim A, Boston P, van Elburg A (1987) Child sexual abuse-children and families referred to a treatment project and the effects of intervention. British Medical Journal, 295:1453-1457.
Berry C (1997) Attitudes toward child sexual abuse, gender differences and prevalence: a comparative study. MAI, 35:1905.

Bickerton D, Hall R, Williams ALJ (1991) Women's experiences of sexual abuse in childhood. Public Health, 105:447-453.

Bifulco A, Brown GW, Adler Z (1991) Early sexual abuse and clinical depression in adult life. British Journal of Psychiatry, 159:115-122.
Blum RW, Geer L, Hutton L et al. (1988) The Minnesota adolescent health survey: implications for physicians. Minnesota Medicine, 71:143-149.
Blum RW, Harmon B, Harris L, Bergeisen L, Resnick MD (1992) American Indian-Alaska Native youth health. Journal of the American Medical Association, 267:1637-1644.
Bolen RM, Scannapieco M (1999) Prevalence of child sexual abuse: a corrective meta-analysis. Social Service Review, 73:281-313.
Bolstad BR, Zinbarg RE (1997) Sexual victimization, generalized perception of control, and posttraumatic stress disorder symptom severity. Journal of Anxiety Disorders, 11:523-540.

Boney-McCoy S, Finkelhor D (1995) Psychosocial sequelae of violent victimization in a national youth sample. Journal of Consulting and Clinical Psychology, 63:726-736.

Boney-McCoy S, Finkelhor D (1996) Is youth victimization related to trauma symptoms and depression after controlling for prior symptoms and family relationships? A longitudinal, prospective study. Journal of Consulting and Clinical Psychology, 64:1406-1416.
Borowsky IW, Resnick MD, Ireland M, Blum RW (1999) Suicide attempts among American Indian and Alaska Native youth: risk and protective factors. Archives of Pediatrics and Adolescent Medicine, 153:573-580.
Bostwick JM, Pankratz VS (2000) Affective disorders and suicide risk: a reexamination. American Journal of Psychiatry, 157:1925-1932.
Boudewyn AC, Liem JH (1995) Childhood sexual abuse as a precursor to depression and self-destructive behavior in adulthood. Journal of Traumatic Stress, 8:445-459.

Bouhet B, Perard D, Zorman M (1992) De l'importance des abus sexuels en France. In: Les enfants victimes d'abus sexuels. Gabel M, ed. Puf, Paris.

Brent DA, Baugher M, Bridge J, Chen T, Chiappetta L (1999) Age- and sexrelated risk factors for adolescent suicide. Journal of the American Academy of Child and Adolescent Psychiatry, 38:1497-1505.
Briere J, Runtz M (1986) Suicidal thoughts and behaviours in former sexual abuse victims. Canadian Journal of Behavioural Science, 18:413-423.
Briere J, Runtz M (1988) Symptomatology associated with childhood sexual victimization in a nonclinical adult sample. Child Abuse and Neglect, 12:51-59.

Briere J, Runtz M (1990) Differential adult symptomatology associated with three types of child abuse histories. Child Abuse and Neglect, 14:357-364.

Brown GW, Harris TO (1993) Aetiology of anxiety and depressive disorders in an inner-city population: 1—Early adversity. Psychological Medicine, 23: 143-154.

Brown GW, Moran P (1994) Clinical and psychosocial origins of chronic depressive episodes. I: a community survey. British Journal of Psychiatry, 165:447-456.

Brown J, Cohen P, Johnson JG, Smailes EM (1999) Childhood abuse and neglect: specificity of effects on adolescent and young adult depression and suicidality. Journal of the American Academy of Child and Adolescent Psychiatry, 38:1490-1496.

Browne A, Finkelhor DA (1986) Impact of child sexual abuse: a review of the research. Psychological Bulletin, 99:66-77.

Bryant SL, Range LM (1995) Suicidality in college women who were sexually and physically abused and physically punished by parents. Violence and Victims, 10:195-201.

Bryant SL, Range LM (1997) Type and severity of child abuse and college students' lifetime suicidality. Child Abuse and Neglect, 21:1169-1176.

Caspi A, Moffitt TE, Newman DL, Silva PA (1996) Behavioral observations at age 3 years predict adult psychiatric disorders. Longitudinal evidence from a birth cohort. Archives of General Psychiatry, 53:1033-1039.

Castelino CT (1985) Child sexual abuse: a retrospective study. Tate Institute of Social Sciences, Bombay.

Cawson P, Wattam C, Brooker S, Kelly G (2000) Child maltreatment in the United Kingdom: a study of the prevalence of child abuse and neglect. National Society for the Prevention of Cruelty to Children (NSPCC), London.

Chandler RK, Jackson JL (1997) Family environment and childhood sexual victimization: a test of the buffering hypothesis. Journal of Interpersonal Violence, 12:3-17.
Chandy JM, Blum RW, Resnick MD (1996) Female adolescents with a history of sexual abuse: risk outcome and protective factors. Journal of Interpersonal Violence, 11:503-518.

Chandy JM, Blum RW, Resnick MD (1997) Sexually abused male adolescents: how vulnerable are they? Journal of Child Sexual Abuse, 6:1-16.

Choquet M, Darves-Bornoz JM, Ledoux S, Manfredi R, Hassler C (1997) Selfreported health and behavioral problems among adolescent victims of rape in France: results of a cross-sectional survey. Child Abuse and Neglect, 21:823-832.
Collings SJ (1991) Childhood sexual abuse in a sample of South African university males: prevalence and risk factors. South African Journal of Psychology, 21:153-158.
Collings SJ (1997) Child sexual abuse in a sample of South African women students: prevalence, characteristics, and long-term effects. South African Journal of Psychology, 27:37-42.

Committee on Sexual Offences Against Children and Youths (1984) Sexual offences against children. Volume 1. Minister of Supply and Services, Ottawa.

Conte JR, Schuerman JR (1987) Factors associated with an increased impact of child sexual abuse. Child Abuse and Neglect, 11:201-211.

Cooper H, Hedges LV (1994) The handbook of research synthesis. Cooper H, Hedges LV, eds. Russell Sage Foundation, New York.

Coxell A, King M, Mezey G, Gordon D (1999) Lifetime prevalence, characteristics, and associated problems of non-consensual sex in men: cross sectional survey. British Medical Journal, 318:846-850.
Davidson JRT, Hughes D, Blazer DG, George LK (1991) Post-traumatic stress disorder in the community: an epidemiological study. Psychological Medicine, 21:713-721.
deLahunta EA (1996) Personal exposure of faculty and medical students to family violence. Journal of the American Medical Association, 275: 1903-1906.
de Paul J, Milner JS, Mugica P (1995) Childhood maltreatment, childhood social support, and child abuse potential in a Basque sample. Child Abuse and Neglect, 19:907-920.
Descamps MJ, Rothblum ED, Bradford J, Ryan C (2000) Mental health impact of child sexual abuse, rape, intimate partner violence, and hate crimes in the

National Lesbian Health Care Survey. Journal of Gay and Lesbian Social Services, 11:27-55.

Diekstra RF (1992) The prevention of suicidal behaviour: evidence for the efficacy of clinical community-based programs. International Journal of Mental Health, 21:69-87.
Dinwiddie S, Heath AC, Dunne MP et al. (2000) Early sexual abuse and lifetime psychopathology: a co-twin-control study. Psychological Medicine, 30:41-52.

DiVasto PV, Kaufman A, Rosner L et al. (1984) The prevalence of sexually stressful events among females in the general population. Archives of Sexual Behavior, 13:59-67.

Duane EA, Stewart CS, Bridgeland WM (1997) Consequences of childhood sexual abuse for college students. Journal of College Student Development, 38:13-23.

Dudley MJ, Kelk NJ, Florio TM, Howard JP, Waters BG (1998) Suicide among young Australians, 1964-1993: an interstate comparison of metropolitan and rural trends. Medical Journal of Australia, 169:77-80.
Edgardh K, Ormstad K (2000) Prevalence and characteristics of sexual abuse in a national sample of Swedish seventeen-year-old boys and girls. Acta Paediatrica, 89:310-319.
Edwards JJ, Alexander PC (1992) The contribution of family background to the long-term adjustment of women sexually abused as children. Journal of Interpersonal Violence, 7:306-320.
Elal G, Oral G, Atamer A, Sabol E (2000) A comparative study of child sexual abuse among university students. (Paper presented at the 16th Annual Meeting of Traumatic Stress Studies). San Antonio, TX.
Erickson PI, Rapkin AJ (1991) Unwanted sexual experiences among middle and high school youth. Journal of Adolescent Health, 12:319-325.

Ernst C, Angst J, Foldenyi M (1993) The Zurich Study. XVII. Sexual abuse in childhood. Frequency and relevance for adult morbidity data of a longitudinal epidemiological study. European Archives of Psychiatry and Clinical Neuroscience, 242:293-300.
Essock-Vitale SM, McGuire MT (1985) Women’s lives viewed from an evolutionary perspective. I. Sexual histories, reproductive success, and demographic characteristics of a random sample of American women. Ethology and Sociobiology, 6:137-154.

Feldman W, Feldman E, Goodman JT et al. (1991) Is childhood sexual abuse really increasing in prevalence? An analysis of the evidence. Pediatrics, 88:29-33.

Felitti VJ, Anda RF, Nordenberg D et al. (1998) Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: the adverse childhood experiences (ACE) study. American Journal of Preventive Medicine, 14:245-258.

Fergusson DM, Horwood LJ, Lynskey MT (1996b) Childhood sexual abuse and psychiatric disorder in young adulthood: II. Psychiatric outcomes of childhood sexual abuse. Journal of the American Academy of Child and Adolescent Psychiatry, 35:1365-1374.
Fergusson DM, Horwood LJ, Woodward LJ (2000) The stability of child abuse reports: a longitudinal study of the reporting behaviour of young adults. Psychological Medicine, 30:529-544.
Fergusson DM, Lynskey MT, Horwood LJ (1996a) Childhood sexual abuse and psychiatric disorder in young adulthood: I. Prevalence of sexual abuse and factors associated with sexual abuse. Journal of the American Academy of Child and Adolescent Psychiatry, 35:1355-1364.
Fergusson DM, Mullen PE (1999) Childhood sexual abuse: an evidence based perspective. Sage Publications, Thousand Oaks, CA.

Finkelhor D (1979) Sexually victimized children. The Free Press, New York.
Finkelhor D (1994) The international epidemiology of child sexual abuse. Child Abuse and Neglect, 18:409-417.

Finkelhor D, Browne A (1988) Assessing the long-term impact of child sexual abuse: a review and conceptualization. In: Handbook on sexual abuse of children. Assessment and treatment issues. Walker LEA, ed. Springer Publishing Company, New York.
Finkelhor D, Dziuba-Leatherman J (1994) Children as victims of violence: a national survey. Pediatrics, 94:413-420.
Finkelhor D, Hotaling G, Lewis IA, Smith C (1990) Sexual abuse in a national survey of adult men and women: prevalence, characteristics, and risk factors. Child Abuse and Neglect, 14:19-28.

Fleming J, Mullen PE, Sibthorpe B, Attewell R, Bammer G (1998) The relationship between childhood sexual abuse and alcohol abuse in women-a case-control study. Addiction, 93:1787-1798.
Fleming J, Mullen PE, Sibthorpe B, Bammer G (1999) The long-term impact of childhood sexual abuse in Australian women. Child Abuse and Neglect, 23:145-159.
Fleming JM (1997) Prevalence of childhood sexual abuse in a community sample of Australian women. Medical Journal of Australia, 166:65-68.
Friedrich E, Klupp N, Akkaya T (1997) Child abuse in Vienna over the last 10 years. Journal de Médecine Legale Droit Medical, 40:337-338.
Fritz GS, Stoll K, Wagner NN (1981) A comparison of males and females who were sexually molested as children. Journal of Sex and Marital Therapy, 7:54-59.

Fromuth ME (1986) The relationship of childhood sexual abuse with later psychological and sexual adjustment in a sample of college women. Child Abuse and Neglect, 10:5-15.

Fromuth ME, Burkhart BR (1989) Long-term psychological correlates of childhood sexual abuse in two samples of college men. Child Abuse and Neglect, 13:533-542.

Fryer GE, Miyoshi TJ (1994) A survival analysis of the revictimization of children: the case of Colorado. Child Abuse and Neglect, 18:1063-1071.

Garabedian MJ (1994) Relationship of child sexual, physical, and psychological abuse to eating disorders and post-traumatic stress disorder in adult women. Dissertation Abstracts International Section A: Humanities and Social Sciences, 54:2465-2465.
Gardner MJ, Altman DG, eds. (1989) Statistics with confidence: confidence intervals and statistical guidelines. British Medical Journal Books, London.
Garnefski N, Arends E (1998) Sexual abuse and adolescent maladjustment: differences between male and female victims. Journal of Adolescence, 21:99-107.

Goldman JDG, Padayachi UK (1997) The prevalence and nature of child sexual abuse in Queensland, Australia. Child Abuse and Neglect, 21:489-498.
Goldman RJ, Goldman JDG (1988) The prevalence and nature of child sexual abuse in Australia. Australian Journal of Sex, Marriage and Family, 9:94-106.

Gorcey M, Santiago JM, McCall-Perez F (1986) Psychological consequences for women sexually abused in childhood. Social Psychiatry, 21:129-133.

Gorey KM, Leslie DR (1997) The prevalence of child sexual abuse: integrative review adjustment for potential response and measurement biases. Child Abuse and Neglect, 21:391-398.
Gould DA, Stevens NG, Ward NG, Carlin AS, Sowell HE, Gustafson B (1994) Self-reported childhood abuse in an adult population in a primary care setting. Prevalence, correlates, and associated suicide attempts. Archives of Family Medicine, 3:252-256.
Graham A, Reser J, Scuderi C, Zubrick S, Smith M, Turley B (2000) Suicide: an Australian Psychological Society discussion paper. Australian Psychologist, 35:1-28.

Green AH (1988) Child maltreatment and its victims. A comparison of physical and sexual abuse. Psychiatric Clinics of North America, 11:591-610.
Greenland S (1987) Quantitative methods in the review of epidemiologic literature. Epidemiologic Reviews, 9:1-30.
Greenwald E, Leitenberg H (1990) Posttraumatic stress disorder in a nonclinical and nonstudent sample of adult women sexually abused as children. Journal of Interpersonal Violence, 5:217-228.
Greenwood CL, Tangalos EG, Maruta T (1990) Prevalence of sexual abuse, physical abuse, and concurrent traumatic life events in a general medical population. Mayo Clinic Proceedings, 65:1067-1071.
Hall LA, Sachs B, Rayens MK, Lutenbacher M (1993) Childhood physical and sexual abuse: their relationship with depressive symptoms in adulthood. Image—The Journal of Nursing Scholarship, 25:317-323.

Halperin DS, Bouvier P, Jaffe PD et al. (1996) Prevalence of child sexual abuse among adolescents in Geneva: results of a cross sectional survey. British Medical Journal, 312:1326-1329.

Harris EC, Barraclough B (1997) Suicide as an outcome for mental disorders. A meta-analysis. British Journal of Psychiatry, 170:205-228.

Harrison PA, Fulkerson JA, Beebe TJ (1997) Multiple substance use among adolescent physical and sexual abuse victims. Child Abuse and Neglect, 21:529-539.

Haugaard JJ, Emery RE (1989) Methodological issues in child sexual abuse research. Child Abuse and Neglect, 13:89-100.
Hawton K, Fagg J (1988) Suicide, and other causes of death, following attempted suicide. British Journal of Psychiatry, 152:359-366.
Hernandez JT (1992) Substance abuse among sexually abused adolescents and their families. Journal of Adolescent Health, 13:658-662.
Hernandez JT, Lodico M, DiClemente RJ (1993) The effects of child abuse and race on risk-taking in male adolescents. Journal of the National Medical Association, 85:593-597.
Hibbard RA, Brack CJ, Rauch S, Orr DP (1988) Abuse, feelings, and health behaviors in a student population. American Journal of Diseases of Children, 142:326-330.

Hibbard RA, Ingersall GM, Orr DP (1990) Behavioral risk, emotional risk, and child abuse among adolescents in a nonclinical setting. Pediatrics, 86:896-901.

Hien D, Bukszpan C (1999) Interpersonal violence in a "normal" low-income control group. Women and Health, 29:1-16.
Higgins DJ, McCabe MP (1994) The relationship of child sexual abuse and family violence to adult adjustment: toward an integrated risk-sequelae model. Journal of Sex Research, 31:255-266.
Hill AB (1965) The environment and disease: association or causation? Proceedings of the Royal Society of Medicine, 58:295-300.
Hill J, Davis R, Byatt M, Burnside E, Rollinson L, Fear S (2000) Childhood sexual abuse and affective symptoms in women: a general population study. Psychological Medicine, 30:1283-1291.

Interventions for support (2001) Child sexual abuse: statistics from IFSHA, New Delhi, 1996 to 2001. Available at http://www.ifsha.org.
Jackson JL, Calhoun KS, Amick AE, Maddever HM, Habif VL (1990) Young adult women who report childhood intrafamilial sexual abuse: subsequent adjustment. Archives of Sexual Behavior, 19:211-221.
Johnson JG, Cohen P, Brown J, Smailes EM, Bernstein DP (1999) Childhood maltreatment increases risk for personality disorders during early adulthood. Archives of General Psychiatry, 56:600-606.
Jumper SA (1995) A meta-analysis of the relationship of child sexual abuse to adult psychological adjustment. Child Abuse and Neglect, 19:715-728.

Kellogg ND, Hoffman TJ (1995) Unwanted and illegal sexual experiences in childhood and adolescence. Child Abuse and Neglect, 19:1457-1468.

Kelly L, Regan L, Burton S (1991) An exploratory study of the prevalence of sexual abuse in a sample of 16-21 year olds. Child Abuse Studies Unit, University of North London, London.

Kendall-Tackett KA, Williams LM, Finkelhor D (1993) Impact of sexual abuse on children: a review and synthesis of recent empirical studies. Psychological Bulletin, 113:164-180.
Kendler KS, Bulik CM, Silberg J, Hettema JM, Myers J, Prescott CA (2000) Childhood sexual abuse and adult psychiatric and substance use disorders in women: an epidemiological and cotwin control analysis. Archives of General Psychiatry, 57:953-959.

Kercher GA, McShane M (1984) The prevalence of child sexual abuse victimization in an adult sample of Texas residents. Child Abuse and Neglect, 8:495-501.

Kessler RC, Borges G, Walters EE (1999) Prevalence of and risk factors for lifetime suicide attempts in the National Comorbidity Survey. Archives of General Psychiatry, 56:617-626.
Kessler RC, Davis CG, Kendler KS (1997) Childhood adversity and adult psychiatric disorder in the US National Comorbidity Survey. Psychological Medicine, 27:1101-1119.
Kessler RC, Magee WJ (1993) Childhood adversities and adult depression: basic patterns of association in a US national survey. Psychological Medicine, 23:679-690.

Kessler RC, McGonagle KA, Zhao S et al. (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Archives of General Psychiatry, 51:8-19.

Kilpatrick AC (1986) Some correlates of women's childhood sexual experiences: a retrospective study. Journal of Sex Research, 22:221-242.
Kilpatrick DG, Schnurr PP, Acierno R, Resnick HS, Best CL (2000) Risk factors for adolescent substance abuse and dependence: data from a national sample. Journal of Consulting and Clinical Psychology, 68:19-30.
Krugman S, Mata L, Krugman R (1992) Sexual abuse and corporal punishment during childhood: a pilot retrospective survey of university students in Costa Rica. Pediatrics, 90:157-161.
Lazartigues A, Perard D, Lisandre H, Pailleux T (1989) Sexual abuse. Study of a population of 1000 students. Neuropsychiatrie de l'enfance et de l'adolescence, 37:223-229.

Lenihan GO (1996) Childhood sexual abuse: gender differences in prevalence, experience, and outcomes. (Presented at the 104th Annual Convention of the American Psychological Association.) Toronto.
Leth I (2001) Sexual abuse of children and adolescents: results from a Scandinavian research concerning the extent and character of sexual abuse. International Planned Parenthood Federation, London.
Levy HB, Markovic J, Chaudhry U, Ahart S, Torres H (1995) Reabuse rates in a sample of children followed for 5 years after discharge from a child abuse inpatient assessment program. Child Abuse and Neglect, 19:1363-1377.

Locke SD (1996) The long-term psychological effects of childhood sexual abuse among black and white female college students (A retrospective analysis of socio-demographic factors). Dissertation Abstracts International, A: The Humanities and Social Sciences, 57:875.

Lodico MA, Gruber E, DiClemente RJ (1996) Childhood sexual abuse and coercive sex among school-based adolescents in a Midwestern State. Journal of Adolescent Health, 18:211-217.
Lopez F, Carpintero E, Hernandez A, Martin MJ (1995) The prevalence and consequences of child sexual abuse in Spain. Child Abuse and Neglect, 19:1039-1050.

Lynskey MT, Fergusson DM (1997) Factors protecting against the development of adjustment difficulties in young adults exposed to childhood sexual abuse. Child Abuse and Neglect, 21:1177-1190.
MacMillan HL, Fleming JE, Trocme N et al. (1997) Prevalence of child physical and sexual abuse in the community. Results from the Ontario Health Supplement. Journal of the American Medical Association, 278:131-135.
Madu SN, Peltzer K (2000) Risk factors and child sexual abuse among secondary school students in the Northern Province (South Africa). Child Abuse and Neglect, 24:259-268.
Market Research Bureau of Ireland (1987) Child sexual abuse in Dublin (Pilot survey report). Market Research Bureau of Ireland, Ltd., Dublin.
Martin $G$ (1996) Reported family dynamics, sexual abuse, and suicidal behaviours in community adolescents. Archives of Suicide Research, 2:183-195.

Martin J, Anderson J, Romans S, Mullen P, O’Shea M (1993) Asking about child sexual abuse: methodological implications of a two stage survey. Child Abuse and Neglect, 17:383-392.

Mazza D, Dennerstein L, Ryan V (1996) Physical, sexual and emotional violence against women: a general practice-based prevalence study. Medical Journal of Australia, 164:14-17.
Meledandri G, Zantedeschi E, Cattaruzza MS, Signorelli C, Fara GM (1996) Child sexual abuse: Italian aspects of the phenomenon. Annali d'Igiene, 8:239-244.
Menick DM, Ngoh F (1998) Child sexual abuse at school: results from an action research study in Cameroon [in French]. Médecine Tropicale, 58:249-252.

Miles GM (2000) Children don't do sex with adults for pleasure: Sri Lankan children's views on sex and sexual exploitation. Child Abuse and Neglect, 24:995-1003.

Moeller TP, Bachmann GA, Moeller JR (1993) The combined effects of physical, sexual, and emotional abuse during childhood: long-term health consequences for women. Child Abuse and Neglect, 17:623-640.
Molnar BE, Buka SL, Kessler RC (2001) Child sexual abuse and subsequent psychopathology: results from the National Comorbidity survey. American Journal of Public Health, 91:753-760.

Molnar BE, Shade SB, Kral AH, Booth RE, Watters JK (1998) Suicidal behavior and sexual-physical abuse among street youth. Child Abuse and Neglect, 22:213-222.

Mullen PE, Fleming J (1998) Long-term effects of child sexual abuse. Issues in Child Abuse Prevention, 9. National Child Protection Clearinghouse, Australia.

Mullen PE, King NJ, Tonge BJ (2000) Child sexual abuse: an overview. Behaviour Change, 17:2-14.
Mullen PE, Martin JL, Anderson JC, Romans SE, Herbison GP (1996) The long-term impact of the physical, emotional, and sexual abuse of children: a community study. Child Abuse and Neglect, 20:7-21.

Mullen PE, Martin JL, Anderson JC, Romans SE, Herbison GP (1993) Childhood sexual abuse and mental health in adult life. British Journal of Psychiatry, 163:721-732.

Mullen PE, Romans-Clarkson SE, Walton VA, Herbison GP (1988) Impact of sexual and physical abuse on women's mental health. The Lancet, 1:841-845.

Mulugeta E, Kassaye M, Berhane Y (1998) Prevalence and outcomes of sexual violence among high school students. Ethiopian Medical Journal, 36:167-174.

Murphy J (1997) Prevalence of child sexual abuse and consequent victimization in the general population. (Paper presented at the Third National Family Violence Research Conference, Durham.) University of New Hampshire, Durham, NH.

Nagy S, DiClemente R, Adcock AG (1995) Adverse factors associated with forced sex among southern adolescent girls. Pediatrics, 96:944-946.

Nelson DE, Higginson GK, Grant-Worley JA (1994) Using the youth risk behavior survey to estimate prevalence of sexual abuse among Oregon high school students. Journal of School Health, 64:413-416.

Nelson EC, Heath AC, Madden PA et al. (2002) Association between selfreported childhood sexual abuse and adverse psychosocial outcomes: results from a twin study. Archive of General Psychiatry, 59:139-145.
Neumann DA, Houskamp BM, Pollock VE, Briere J (1996) The long-term sequelae of childhood sexual abuse in women: a meta-analytic review. Child Maltreatment, 1:6-16.
Olsson A, Ellsberg M, Berglund S et al. (2000) Sexual abuse during childhood and adolescence among Nicaraguan men and women: a population-based anonymous survey. Child Abuse and Neglect, 24:1579-1589.

Palmer RL, Bramble D, Metcalfe M, Oppenheimer R, Smith J (1994) Childhood sexual experiences with adults: adult male psychiatric patients and general practice attenders. British Journal of Psychiatry, 165:675-679.

Palmer RL, Coleman L, Chaloner D, Oppenheimer R, Smith J (1993) Childhood sexual experiences with adults. A comparison of reports by women psychiatric patients and general-practice attenders. British Journal of Psychiatry, 163:499-504.

Paolucci EO, Genuis ML, Violato C (2001) A meta-analysis of the published research on the effects of child sexual abuse. Journal of Psychology, 135:17-36.

Paradise JE, Rose L, Sleeper LA, Nathanson M (1994) Behavior, family function, school performance, and predictors of persistent disturbance in sexually abused children. Pediatrics, 93:452-459.
Pederson W, Aas H (1995) Sexual victimization in Norwegian children and adolescents: victims, offenders, assaults. Scandinavian Journal of Social Medicine, 23:173-178.

Peters DK, Range LM (1995) Childhood sexual abuse and current suicidality in college women and men. Child Abuse and Neglect, 19:335-341.
Peters SD (1988) Child sexual abuse and later psychological problems. In: Lasting effects of child sexual abuse. Wyatt GE, Powell GL, eds. Sage Publications, Thousand Oaks, CA.

Peters SD, Wyatt GE, Finkelhor D (1986) Prevalence. In: Sourcebook on child sexual abuse. Finkelhor DA, ed. Sage Publications, Thousand Oaks, CA.

Plunkett A, Oates RK (1990) Methodological considerations in research on child sexual abuse. Paediatric and Perinatal Epidemiology, 4:351-360.
Plunkett A, O’Toole BI, Swanston H, Oates RK, Shrimpton S, Parkinson P (2001) Suicide risk following child sexual abuse. Ambulatory Pediatrics, 1:262-266.
Polusny MA, Follette VM (1995) Long-term correlates of child sexual abuse: theory and review of the empirical literature. Applied and Preventive Psychology, 4:143-166.
Pothe P, Csemy L, Halfarova H, Bosak V (2000) Sexual abuse of children in the Czech Republic—retrospective study. Ceska a Slovenska Psychiatrie, 96:131-135.

Priest R (1991) An assessment of the prevalence of child sexual victimization and the utilization of mental health services in selected African-American college samples. Dissertation Abstracts International, 51:2641.

Ramos-Lira L, Saldivar-Hernandez G, Medina-Mora ME, Rojas-Guiot E, Villatoro-Velazquez J (1998) Prevalence of sexual abuse in students and its relationship with drug consumption. Salud Publica de Mexico, 40:221-233.
Raupp U, Eggers C (1993) Sexual abuse of children. A regional study of the prevalence and characteristics. Monatsschrift Kinderbeilkunde, 141:316-322.
Richter-Appelt H, Tiefensee J (1996) Soziale und familiäre Gegebenheiten bei körperlichen Misshandlungen und sexuellen Missbrauchserfahrungen in der Kindheit aus der Sicht junger Erwachsener. Psychotherapie, Psychosomatik und Medizinische Psychologie, 46:367-378.

Riggs S, Alario AJ, McHorney C (1990) Health risk behaviors and attempted suicide in adolescents who report prior maltreatment. Journal of Pediatrics, 116:815-821.

Rind B (1997) A meta-analytic review of findings from national samples on psychological correlates of child sexual abuse. The Journal of Sex Research, 34:237-255.

Rind B, Tromovitch P, Bauserman R (1998) A meta-analytic examination of assumed properties of child sexual abuse using college samples. Psychological Bulletin, 124:22-53.

Risberg G, Lundgren E, Westman G (1999) Prevalence of sexualized violence among women. A population-based study in a primary healthcare district. Scandinavian Journal of Public Health, 27:247-253.
Risin LI, Koss MP (1987) The sexual abuse of boys: prevalence and descriptive characteristics of childhood victimizations. Journal of Interpersonal Violence, 2:309-323.

Robin RW, Chester B, Rasmussen JK, Jaranson JM, Goldman D (1997) Prevalence, characteristics, and impact of childhood sexual abuse in a Southwestern American Indian tribe. Child Abuse and Neglect, 21:769-787.

Romans S, Martin J, Mullen P (1997) Childhood sexual abuse and later psychological problems: neither necessary, sufficient nor acting alone. Criminal Behaviour and Mental Health, 7:327-338.

Romero GJ, Wyatt GE, Loeb TB, Carmona JV, Solis BM (1999) The prevalence and circumstances of child sexual abuse among Latina women. Hispanic Journal of Behavioral Sciences, 21:351-365.
Roosa MW, Reyes L, Reinholtz C, Angelini PJ (1998) Measurement of women's child sexual abuse experiences: an empirical demonstration of the impact of choice of measure on estimates of incidence rates and of relationships with pathology. Journal of Sex Research, 35:225-233.

Rothman KJ, Greenland S (1998) Modern epidemiology. 2nd edn. LippincottRaven Publishers, Philadelphia, PA.

Rönström A (1985) Sexual abuse of children in Sweden: perspectives on research, interventions and consequences, Radda Barnen, Stockholm.

Russell DE (1983) The incidence and prevalence of intrafamilial and extrafamilial sexual abuse of female children. Child Abuse and Neglect, 7: 133-146.

Rutter ML (1999) Psychosocial adversity and child psychopathology. British Journal of Psychiatry, 174:480-493.
Sariola H, Uutela A (1994) The prevalence of child sexual abuse in Finland. Child Abuse and Neglect, 18:827-835.
Saunders BE, Kilpatrick DG, Hanson RF, Resnick HS, Walker ME (1999) Prevalence, case characteristics, and long-term psychological correlates of child rape among women: a national survey. Child Maltreatment, 4:187200.

Saunders BE, Villeponteaux LA, Lipovsky JA, Kilpatrick DG (1992) Child sexual assault as a risk factor for mental disorders among women: a community survey. Journal of Interpersonal Violence, 7:189-204.

Schaaf KK, McCanne TR (1998) Relationship of childhood sexual, physical, and combined sexual and physical abuse to adult victimization and posttraumatic stress disorder. Child Abuse and Neglect, 22:1119-1133.
Schei B (1990) Prevalence of sexual abuse history in a random sample of Norwegian women. Scandinavian Journal of Social Medicine, 18:63-68.

Schein M, Biderman A, Baras M et al. (2000) The prevalence of a history of child sexual abuse among adults visiting family practitioners in Israel. Child Abuse and Neglect, 24:667-675.
Schoetensack K, Elliger T, Gross A, Nissen G (1992) Prevalence of sexual abuse of children in Germany. Acta Paedopsychiatrica, 55:211-216.
Scott KD (1992) Childhood sexual abuse: impact on a community's mental health status. Child Abuse and Neglect, 16:285-295.
Sedney MA, Brooks B (1984) Factors associated with a history of childhood sexual experience in a nonclinical female population. Journal of the American Academy of Child and Adolescent Psychiatry, 23:215-218.
Sharp SJ, Sterne JAC (1997) Meta-analysis. Stata Technical Bulletin, 38:9-14.
Siegel JM, Sorenson SB, Golding JM, Burnam MA, Stein JA (1987) The prevalence of childhood sexual assault. The Los Angeles Epidemiologic Catchment Area Project. American Journal of Epidemiology, 126:1141-1153.

Silverman AB, Reinherz HZ, Giaconia RM (1996) The long-term sequelae of child and adolescent abuse: a longitudinal community study. Child Abuse and Neglect, 20:709-723.

Singh HS, Yiing WW, Nurani HN (1996) Prevalence of childhood sexual abuse among Malaysian paramedical students. Child Abuse and Neglect, 20:487-492.
So-kum Tang C (2000) Survey study on child sexual abuse among Chinese college students. End Child Sexual Abuse Foundation, Hong Kong.
Spaccarelli S, Kim S (1995) Resilience criteria and factors associated with resilience in sexually abused girls. Child Abuse and Neglect, 19:1171-1182.
Spak L, Spak F, Allebeck P (1998) Sexual abuse and alcoholism in a female population. Addiction, 93:1365-1373.
Stein JA, Golding JM, Siegel JM, Burnam MA, Sorenson SB (1988) Long-term psychological sequelae of child sexual abuse: the Los Angeles Epidemiologic Catchment Area study. In: Lasting effects of child sexual abuse. Wyatt GE, Powell GL, eds. Sage Focus Editions, Vol. 100. Sage Publications, Newbury Park, CA.
Stepakoff S (1998) Effects of sexual victimization on suicidal ideation and behavior in US college women. Suicide and Life-Threatening Behavior, 28:107-126.

Streiner DL (1998) Risky business: making sense of estimates of risk. Canadian Journal of Psychiatry, 43:411-415.
Swanston HY, Tebbutt JS, O’Toole BI, Oates RK (1997) Sexually abused children 5 years after presentation: a case-control study. Pediatrics, 100:600-608.
Tejedor MC, Diaz A, Castillon JJ, Pericay JM (1999) Attempted suicide: repetition and survival-findings of a follow-up study. Acta Psychiatrica Scandanavica, 100:205-211.

Thakkar RR, Gutierrez PM, Kuczen CL, McCanne TR (2000) History of physical and/or sexual abuse and current suicidality in college women. Child Abuse and Neglect, 24:1345-1354.

Urquiza AJ, Capra M (1990) The impact of sexual abuse: initial and long-term effects. In: The sexually abused male. Vol. I: Prevalence, impact and treatment. Hunter M, ed. Lexington Books, New York.

Vandewege R, Bruynooghe R, Opdebeeck S (1988) Ervaringen van vrouwen met fysiek en sexsueel geweld: prevalentie en gevolgen. Limburgs Universitair Centrum, Diepenbeek.
Vogeltanz ND, Wilsnack SC, Harris TR, Wilsnack RW, Wonderlich SA, Kristjanson AF (1999) Prevalence and risk factors for childhood sexual abuse in women: national survey findings. Child Abuse and Neglect, 23:579-592.
Walch AG, Broadhead WE (1992) Prevalence of lifetime sexual victimization among female patients. Journal of Family Practice, 35:511-516.

Wang J, Liu XZ, Mai QG, Xu JL, Chen LB (1994) The analysis of the frequency of child abuse. Chinese Social Medicine, 50:23-27.

Watkins B, Bentovim A (1992) The sexual abuse of male children and adolescents: a review of current research. Journal of Child Psychology and Psychiatry and Allied Disciplines, 33:197-248.

Watts WD, Ellis AM (1993) Sexual abuse and drinking and drug use: implications for prevention. Journal of Drug Education, 23:183-200.
Weiss P, Zverina J (1997) Prevalence of sexual abuse in childhood in the general population: results of a national survey. Cesko-Slovenska Psychiatrie, 93:66-74.

Wellman MM (1993) Child sexual abuse and gender differences: attitudes and prevalence. Child Abuse and Neglect, 17:539-547.
White K, Strange C (1993) Effects of unwanted childhood sexual experiences on psychosocial development of college women. Journal of College Student Development, 34:289-294.

WHO (2001) WHO multi-country study on women's health and domestic violence. (Preliminary results.) World Health Organization, Geneva.

Widom CS, White HR (1997) Problem behaviours in abused and neglected children grown up: prevalence and co-occurrence of substance abuse, crime and violence. Criminal Behaviour and Mental Health, 7:287-310.
Wilsnack SC, Vogeltanz ND, Klassen AD, Harris TR (1997) Childhood sexual abuse and women's substance abuse: national survey findings. Journal of Studies on Alcohol, 58:264-271.
Wingood GM, DiClemente RJ (1997) Child sexual abuse, HIV sexual risk, and gender relations of African-American women. American Journal of Preventive Medicine, 13:380-384.

Wise LA, Zierler S, Krieger N, Harlow BL (2001) Adult onset of major depressive disorder in relation to early life violent victimisation: a case-control study. The Lancet, 358:881-887.

Wolf J (1992) Adult reports of sexual abuse during childhood: results of a statewide telephone survey in Kentucky. (Paper presented at the 47th Annual Conference of the American Association for Public Opinion Research.) St Petersburg Beach, FL.

Wonderlich SA, Brewerton TD, Jocic Z, Dansky BS, Abbott DW (1997) Relationship of childhood sexual abuse and eating disorders. Journal of the American Academy of Child and Adolescent Psychiatry, 36:1107-1115.
Wyatt GE (1985) The sexual abuse of Afro-American and white-American women in childhood. Child Abuse and Neglect, 9:507-519.
Wyatt GE, Loeb TB, Solis B, Carmona JV (1999) The prevalence and circumstances of child sexual abuse: changes across a decade. Child Abuse and Neglect, 23:45-60.
Wyatt GE, Newcomb M (1990) Internal and external mediators of women's sexual abuse in childhood. Journal of Consulting and Clinical Psychology, 58:758-767.

Wyatt GE, Powell GJ, eds. (1988) Lasting effects of child sexual abuse. Sage Focus Editions, Vol. 100. Sage Publications, Newbury Park, CA.

Wynkoop TF, Capps SC, Priest BJ (1995) Incidence and prevalence of child sexual abuse: a critical review of data collection procedures. Journal of Child Sexual Abuse, 4:49-66.

Yama MF, Tovey SL, Fogas BS, Morris J (1995) The relationship among childhood sexual abuse, parental alcoholism, family environment and suicidal behavior in female college students. Journal of Child Sexual Abuse, 4:79-93.
Zuravin SJ, Fontanella C (1999) The relationship between child sexual abuse and major depression among low-income women: a function of growing up experiences? Child Maltreatment: Journal of the American Professional Society on the Abuse of Children, 4:3-12.

## Chapter 24

# Distribution of risk factors BY POVERTY 

Tony Blakely, Simon Hales, Charlotte Kieft, Nick Wilson and Alistair Woodward

## Summary

Socioeconomic position is an important distal risk determinant for many health outcomes. While it was not possible (owing to limitations of data and other factors) to directly map socioeconomic position to the burden of disease, it was considered possible to map some risk factors by absolute poverty, which is one measure of socioeconomic position.

The proportions of the population living on $<$ US $\$ 1$, on US\$ 1-2 and on $>$ US $\$ 2$ per day were estimated for each of the 14 subregions ${ }^{1}$ using World Bank estimates of poverty by country. The counterfactual scenario was no absolute poverty in the world (no one living on $<$ US $\$ 2$ per day). The prevalences of risk factors for each subregion were obtained from the relevant risk factor chapters in this publication.

The associations of absolute poverty with eight risk factors (child and maternal underweight, unsafe water and sanitation, unsafe sex, indoor air pollution, outdoor air pollution, tobacco use, alcohol use and overweight [women only]) were determined by an indirect method, using asset scores calculated from demographic and health survey (DHS) data and income from living standards measurement surveys (LSMS) data. First, the joint association of the asset score or income variable with the risk factor was determined for each subregion. Second, the percentage estimates of poverty by subregion were overlaid on the ranked asset scores and income variables (e.g. if $20 \%$ of people in a subregion were estimated to be living on $<$ US $\$ 1$ per day, then the prevalence of each factor among these poor people was assumed to be that observed for the $20 \%$ of people with lowest asset scores). Third, the crude relative risks of each risk factor by level of poverty were estimated based on this overlay. We also undertook selective literature reviews for some risk factors.

Approximately one fifth of the world's population live on $<$ US\$ 1 per day and almost half on $<\mathrm{US} \$ 2$ per day. Of the 14 subregions, three
(EUR-A, AMR-A and WPR-A) had negligible levels of absolute poverty and were excluded from all subsequent analyses. We estimate that $9 \%$ of people in EMR-B were living on $<$ US $\$ 2$ per day ( $2 \%$ on $<$ US $\$ 1$ per day), but the estimates for this subregion were based on sparse data. The estimates for the remaining 10 subregions ranged from $18 \%$ ( $3 \%$ ) for EUR-B to $85 \%(42 \%)$ for SEAR-D and $78 \%$ ( $56 \%$ ) for AFR-D.

Childhood malnutrition, unimproved water and sanitation and indoor air pollution were strongly associated with absolute poverty. The associations of poverty with one indicator of unsafe sex (unprotected sex with a non-marital partner) and tobacco and alcohol consumption were weaker and variable across subregions. Our analyses and literature reviews were consistent with the proposition that tobacco and alcohol consumption, adverse lipid profiles, hypertension and overweight initially affect the non-poor in developing countries.

If the worldwide prevalence of childhood malnutrition among all children living on $<$ US $\$ 2$ per day were changed to that of children living on $>$ US $\$ 2$ per day (i.e. counterfactual scenario 1 ), $37 \%$ of the cases worldwide of underweight would be prevented (assuming a causal relationship). The equivalent percentage reduction from shifting all poor children to exactly US\$ 2 per day (counterfactual scenario 2 ) was $23 \%$. For inadequate water and sanitation these percentage reductions were $51 \%$ and $36 \%$, respectively.

Both poverty and risk factor data were available or were used for only some countries within each subregion, and thus extrapolations had to be made from those countries with data. Second, some subregions had no data at all for some risk factors (e.g. unsafe sex), thus limiting the number of subregions for which we could conduct analyses. Third, all results were based on survey data with their associated errors. Another notable limitation with our analyses was the assumption that the ranking by asset score was comparable to the unobserved ranking by income poverty. Nevertheless, our findings confirm that there are currently severe inequalities in the distribution of childhood malnutrition, inadequate water and sanitation and indoor air pollution by income poverty worldwide, with these risks concentrated among the poorest sectors of society.

## 1. Introduction

Income may affect health outcomes directly or through exposure to other risk factors (Figure 24.1). In this work, we mapped the crude associations of poverty with some of the health risk factors detailed elsewhere in this book. The risk factors included were childhood underweight, unsafe water and sanitation, unsafe sex, alcohol use, tobacco use, overweight (women only), indoor air pollution and outdoor air pollution.

There were insufficient data at the time to present analyses for the other risk factors. Also, some of the analyses for the risk factors men-

Figure 24.I Diagrammatic representation of poverty in relation to other risk factors and the global burden of disease

tioned above were more prone to bias because of unavailability of data, data being available or used for only a few countries within a given subregion, or proxy data being used (e.g. expenditure on tobacco and alcohol used as a proxy for actual consumption). Literature reviews were therefore conducted for both (i) the risk factors for which we conducted quantitative analyses but had concerns about the accuracy of our results; and (ii) those remaining risk factors for which there was published literature on the association with poverty or socioeconomic position. These reviews are presented for tobacco use, alcohol use, hypertension/blood pressure, cholesterol level, physical inactivity, overweight, lead and illicit drugs.

This chapter includes neither quantitative nor literature review findings for the following risk factors: child sexual abuse, contaminated injections in health care settings, fruit and vegetable consumption, injuries, micronutrient (vitamin A, iron and zinc) deficiencies, climate change, mental health risk factors and occupational risks.

### 1.1 SOCIOECONOMIC POSITION, RISK FACTORS AND HEALTH STATUS

Socioeconomic position is a multifaceted construct that includes both resource-based and prestige-based measures:

Resource-based measures refer to material and social resources and assets, including income, wealth and educational credentials; terms used to describe inadequate resources include "poverty" and "deprivation". Prestige-based measures refer to individuals' rank or status in
a social hierarchy, typically evaluated with reference to people's access to and consumption of goods, services and knowledge, as linked to their occupational prestige, income and educational level (Krieger 2001).

Health status is strongly determined by socioeconomic position. A burgeoning research literature from developed countries demonstrates that all-cause mortality and most causes of death occur at greater rates among groups with lower socioeconomic position (Blakely 2002; Drever and Whitehead 1997; Howden-Chapman and Tobias 2000; Mackenbach et al. 1997; Pamuk et al. 1998; Sorlie et al. 1995). Emerging research in the developing world points to similar associations within countries (Evans et al. 2001; Leon and Walt 2001; Wagstaff 2000; Wagstaff and Watanabe 2001), although data are still limited for individual-level (as opposed to ecological-level) analyses. However, it is unlikely that the relationship of personal socioeconomic position with health is of similar magnitude in all countries. For example, the association of lung cancer with lower socioeconomic position in developed countries is (mainly) due to the greater prevalence of tobacco smoking among groups with a lower socioeconomic position. However, this picture has arisen over time as tobacco consumption has moved from a pattern of behaviour associated with the more advantaged groups to one associated with those of lower socioeconomic positions in the developed world. As the tobacco epidemic moves through the developing world, we would first expect to see higher tobacco consumption among groups with higher socioeconomic position, and correspondingly higher lung cancer rates among groups with higher socioeconomic position. Thus we would expect to see varying associations of socioeconomic position with health status by subregion, and variation over time. This means that caution should be applied when generalizing the association of socioeconomic position with health status from one country to another, and from one time period to another.

### 1.2 Poverty as a measure of socioeconomic position

The definition of socioeconomic position given above makes it clear that poverty is only one component of the broader construct of socioeconomic position. As with socioeconomic position, poverty too is a multifaceted construct.

> A complex construct, poverty is inherently a normative concept that can be defined-in both absolute and relative terms-in relation to: "need", "standard of living","limited resources", "lack of basic security", "lack of entitlement", "multiple deprivation", "exclusion", "inequality", "class", "dependency" and "unacceptable hardship" (Krieger 2001).

The United Nations has identified two forms of poverty: "human poverty" and "income poverty" (UNDP 2000). Income poverty refers to
deprivation in a single construct-income. While income poverty is only one dimension of poverty, it is undoubtedly a core element. Human poverty, on the other hand, is characterized by impoverishment in multiple dimensions such as health, knowledge, standard of living and participation in society. The World Bank also accepts this view of poverty, which covers not only material deprivation but also low achievement in health and education (World Bank 2001a).

Sen (1999) views income poverty as one component of "capability deprivation", whereby the latter includes health and educational status that may offset income poverty, lead to income poverty, or be affected by income poverty. Also, the instrumental relationship between income and capability deprivation varies by circumstance. For example, someone in poor health may lose his or her source of income and incur extra expenditure.

A World Health Organization (WHO)-sponsored commission has recently underlined the importance at a macroeconomic level of investing in health care and public health services to enhance economic development (Commission on Macroeconomics and Health 2001). Such an approach is in contrast to the assumption that improved health status follows economic development. Thus, at both the individual and the country level, we are forced to think of a bi-directional association between poverty and health.

The advantages of income poverty as the indicator of socioeconomic status (SES) include:

- income poverty is more readily measured in a standardized way across countries and subregions than a more general and multifaceted measure of poverty;
- there has recently been a concerted attempt to measure the prevalence of absolute poverty by country (sponsored by the World Bank) in which the percentages of people in each country living on $<$ US $\$ 1$ and <US\$2 per day have been estimated (Chen and Ravallion 2000; World Bank 2001a);
- while we fully recognize that a broader concept of poverty encompasses health and other capabilities, in mapping health risk factors by poverty we wanted to avoid the endogeneity that would be incurred by directly incorporating health status in our measure of poverty (put another way, we wanted to avoid "self-correlation" whereby we map health risk behaviour by a poverty variable that incorporates health status as one component);
- income poverty might be considered more directly amenable to policy intervention than measures of socioeconomic position incorporating status and prestige; and
- it is not implausible to imagine a world in which people do not live on $<\mathrm{US} \$ 1$ or $<\mathrm{US} \$ 2$ per day-it would take only a relatively small amount of the income of populations in developed countries to be redistributed to poor people to lift everyone out of absolute poverty.
Nevertheless, there are disadvantages in using absolute poverty as the measure of socioeconomic position. First, the use of the absolute limits of US\$ 1 and US\$ 2 per day means that the affluent regions of the world (Western Europe, North America and the Western Pacific) are not included in analyses, since only a negligible proportion of the population in these regions live in such depths of absolute income poverty. Poverty still exists in these regions, but given the higher standards of living and average income it is of a more "relative" than "absolute" nature. One solution to this problem is to use relative poverty cut-offs, e.g. $60 \%$ of the median income in each country. We did not pursue this alternative, although our methods could be adapted to do so and there are World Bank estimates of relative poverty that are consistent with the absolute poverty estimates used in this chapter (Chen and Ravallion 2000). Second, poverty is only one aspect of socioeconomic position; we could, for example, have mapped health risk factors by educational status. Difficulties with such an approach include data limitations and a lack of standardized measures of education between countries and subregions.

Studies in developing countries were examined in order to give an approximation of the degree of concordance between the different socioeconomic measures, e.g. education, income and occupation. Of the 79 papers retrieved, 29 reported on the prevalence of a risk factor by more than one socioeconomic measure. Table 24.1 shows the papers that included more than one socioeconomic measure and the number that then revealed similar trends between the measures used. Some risk factors and socioeconomic measures have distinctive gender or cultural patterns that may affect these results.

Based on these papers, the distribution of risk factors showed similar trends by education and by income. Indeed, 10 of the 13 papers that used both income and education as independent measures of socioeconomic position showed similar relationships with the risk factor in question during analysis. By contrast, four out of the six papers that used both occupation and income independently as measures of socioeconomic position and only five out of 10 papers with occupation and education showed similar trends.

### 1.3 Poverty-Risk association

In addition to the advantages and disadvantages of our choice of absolute poverty as the measure of socioeconomic position, there are other limitations that must be borne in mind when interpreting the findings of this analysis.

Table 24.I Overview of papers from the literature review that used more than one measure of socioeconomic position

| Socioeconomic variable | Reference | Proportion showing similar association |
| :---: | :---: | :---: |
| Occupation, socioeconomic status | Agarwal et al. (1994) | 1/I |
| Education, occupation, income | al-Mannai et al. (1996); Ekpo et al. (I992); Grol et al. (1997a) | 2/3 |
| Education, income | al-Nuaim et al. (1996, 1997); Delpeuch et al. (2000); Kikafunda et al. (1998); Monteiro et al. (1995); Obot (1990); Rossouw et al. (1990); Sakamoto et al. (200I) | 6/8 ${ }^{\text {a }}$ |
| Education, occupation | Chaturvedi et al. (1998); Hodge et al. (1994); Hu and Tsai (2000); Jarallah et al. (1999); Narayan et al. (1996) | 2/5 |
| Water supply type, income, education | Chen (1996) | 0/1 |
| Education, insurance premium level | Chung et al. (1992) | 0/1 |
| Occupation, income | Dhurandhar and Kulkarni (1992); Ge et al. (1994); Naidu and Rao (1994) | $2 / 3^{\text {b }}$ |
| Education, socioeconomic measure | Delpeuch et al. (1994); Martorell et al. (1998) | $1 / 2^{\text {c }}$ |
| Nongovernmental assistance, education, land ownership, occupation, housing, media exposure | Hadi (2000) |  |
| Education, occupation, housing type | Hoa et al. (1995) | 0/1 |
| Education, water, toilet type | Li et al. (1999) | I/I |
| Education, income, land ownership | Radebe et al. (1996) | I/I |
| Television/radio ownership, water source, toilet type, floor type, fuel type, education | Ricci and Becker (1996) | I/I |
| a In women, not men. |  |  |
| In men, not women. |  |  |
| ${ }^{\text {c }}$ Not rural. |  |  |

## Causal inference

There are two major problems with determining causality for poverty-risk factor relationships: endogeneity (i.e. the inseparability of poverty and health owing to dynamic, synergistic and bi-directional causal associations, as described above) and confounding (since poverty may be correlated with other determinants of health such as education) (Kaufman and Cooper 1999). We have taken the approach of simply
reporting ("mapping") crude associations of poverty and health risk factors. While this approach is limited by problems of endogeneity and confounding, it also avoids tenuous assumptions about controlling for confounders that are part of the constellation of factors that accompany poverty.

## Time-Lags

Any changes in health status or health risk factors caused by poverty will take time to occur. Some effects may be relatively rapid, such as an improvement in income that may immediately lead to the ability to buy food and avoid malnutrition. Other effects may take a long time. For example, it may take many years before political decisions utilize the collective wealth of the community to invest in a supply of clean water and adequate sewage disposal.

We do not account for time-lags, so that current distributions of risk factors may be a consequence of poverty levels in the past. This is consistent with the "mapping" nature of our project that determines the crude associations of poverty and risk factors.

## Heterogeneity of income within poverty levels

People living on $<\mathrm{US} \$ 1$ per day in one part of the world will differ in many ways from those living on $<$ US $\$ 1$ per day in other parts of the world. Some of this variation will be due to the accompanying (and possibly confounding) factors. For example, two countries may differ greatly in average educational level, despite having similar levels of income poverty.

However, there is likely to be variation in the depth of income poverty between countries and subregions. Most of the people living on $<$ US $\$ 1$ per day in one country may be living on an amount just less than US\$ 1 per day, whereas in another country the depth of poverty may be much greater. Having determined the prevalence of poverty by cut-offs such as US\$ 1 and US\$ 2 per day, economists then go on to characterize variation in depth of poverty by measures such as the poverty gap (and the squared poverty gap), which take account of the absolute shortfall (squared) in income beneath the poverty cut-off.

We have not accounted directly for varying levels of depth of poverty by subregion. However, by conducting analysis of people ranked by value of socioeconomic position (e.g. asset scores calculated at a subregional level), and using both a <US\$ 1 and <US\$ 2 per day cut-off, we have captured some of this variation.

Heterogeneity of the association of poverty with risk factors
In addition to the varying depths of poverty, there will be a range of reasons for heterogeneity in the association between poverty and health or health risk factors across the world, including:

- macroeconomic policies may mitigate against the expected health consequences of income poverty (for example, welfare system development and infrastructure may reduce the impact on risk factors of being poor);
- health care and public health services, if well organized and effective, should alter socioeconomic gradients in risk factors; and
- the effect of one socioeconomic factor (e.g. education and cultural factors) may interact with another (e.g. poverty), which if their joint distribution between countries/regions varies (as would the joint distribution of education and poverty), would manifest as heterogeneity of effect for either education or poverty considered in isolation.

For example, the cross-national comparison studies by Kunst and Mackenbach and colleagues demonstrate that the all-cause mortality differentials are surprisingly similar across European countries. However, the cause-specific gradients vary notably (Kunst et al. 1998a, 1998b; Mackenbach et al. 1997). For example, there are stronger cancer gradients in the south of Europe, and stronger cardiovascular disease gradients in the north. This variation by cause of death points to underlying variations between the north and south in the distribution of traditional lifestyle risk factors by socioeconomic position.

## 2. Methods

### 2.1 ESTIMATING ABSOLUTE POVERTY AND RISK FACTOR LEVELS

The proportion of total population in each country in absolute poverty was taken from the World Development Indicators 2001 CD-ROM (World Bank 2001b). These estimates were made by Chen and Ravallion (2000) using 265 national sample surveys for 83 countries, allowing for purchasing power parity (1993 base). In brief, these estimates of absolute poverty:

- extended previous work by Ravillion et al. (1991) for 1985 that found US\$1 a day to be representative of poverty lines found among lowincome countries;
- were based on consumption data where possible, although income data (with adjustment) were used when consumption data were missing;
- considered local as well as global estimates of poverty lines in order to arrive at one poverty level applicable to all countries;
- used data from different time periods, but adjusted all dollar estimates to one point in time (1993) and adjusted for purchasing power parity; and
- applied to consumption per person per day (i.e. not household consumption).

Chen and Ravallion set the main poverty line at US\$ 1.08 per person per day, using 1993 as the base year. These authors and the World Bank simply referred to this cut-off as US\$ 1 per day, and we followed this method. We also included an additional income range of US\$ 1-2 per day.

To allow comprehensiveness, we used 1993 estimates of poverty as provided on the World Development Indicators 2001 CD-ROM (World Bank 2001b) and extracted data on the percentage of the population with a daily income of $<\mathrm{US} \$ 1$ and $<\mathrm{US} \$ 2$ per day for all available countries. We used the estimates of country populations provided on the CDROM to calculate the distribution of the population by subregion within the trichotomous absolute poverty variable.

To estimate the percentage of people living in absolute poverty by subregion, we simply extrapolated the estimated percentages of people living in absolute poverty based on countries with data to the countries without data. Figure 24.2 shows the variation in the percentages of the total population of each subregion living in countries with World Bank estimates

Figure 24.2 Percentages of the population in each subregion living in countries with available World Bank income poverty estimates


Subregion

[^91]of absolute poverty. For example, in AFR-D the countries for which there were poverty data comprised $79.7 \%$ of the total population of that subregion.

The three subregions with $<5 \%$ of their populations represented by countries with absolute poverty estimates were the three most affluent subregions: AMR-A, EUR-A and WPR-A. As these three subregions contain no (or very few) people living in absolute poverty, they were not included in the analyses.

Ten of the remaining 11 subregions had country-level poverty data covering $>70 \%$ of the total population in the subregion. The subregion with the least data was EMR-B. For this subregion only $11 \%$ of the total population lived in countries for which there were absolute poverty estimates from the World Bank. Accordingly, all estimates presented in this chapter of an association between absolute poverty and risk factors in EMR-B, should be treated with considerable caution.

The country-level (World Bank) and subregional-level estimates (our extrapolations) of the percentages of people in absolute poverty are shown in Table 24.2. We estimated that $9.1 \%$ of people in EMR-B were living on $<$ US $\$ 2$ per day ( $2.1 \%$ on $<$ US\$ 1 per day), although the estimates for this subregion were based on sparse data, as described above. The poverty estimates for the remaining 10 subregions ranged from $17.6 \%$ of people living on $<$ US\$ 2 per day ( $3.0 \%$ on $<$ US $\$ 1$ per day) for EUR-B to $85.2 \%$ ( $42.4 \%$ ) for SEAR-D and $77.9 \%$ ( $55.5 \%$ ) for AFR-D.

Table 24.3 presents estimates of absolute poverty for the whole world and for developing countries. The first panel of estimates is that provided by Chen and Ravallion (2000) for the member 2 (i.e. developing, nondonor) countries of the World Bank. These estimates, covering the period 1987-1998, generally indicate that a quarter of the population of the developing world live on $<$ US $\$ 1$ per day, that over half live on $<$ US $\$ 2$ per day, and that there was some reduction in the prevalence of absolute poverty between 1993 and 1996. However, the estimates for 1998 were provisional.

The second panel of results in Table 24.3 shows our estimates of absolute poverty for the 11 "non-A" subregions, i.e. the pooled estimates based on the data in Table 24.2. The World Bank data used in Table 24.2 were mostly for the late 1990 s. We have therefore placed our estimates under 1996-1998 and they can be seen to be very close to those for a similar time period for the non-donor countries of the World Bank.

The third panel of results in Table 24.3 shows our estimates of absolute poverty for the whole world, assuming nil absolute poverty in EUR-A, AMR-A and WPR-A. We estimated that $20.2 \%$ of the world's population live on $<\mathrm{US} \$ 1$ per day and $48.6 \%$ live on $<\mathrm{US} \$ 2$ per day, representing 1.2 billion and 2.9 billion people, respectively. These estimates agree well with those of 1.2 and 2.8 billion, respectively, reported in the World development report 2000/2001 (World Bank 2001a).
Table 24.2 Populations, poverty estimates and availability of survey data by country and subregion ${ }^{2}$

| Subregion | Country | Population in 1999 (000s) ${ }^{\text {b }}$ | Percentage living on <US\$ I per day ${ }^{\text {c }}$ | Percentage living on US\$ I-2 per day ${ }^{\text {c }}$ | Percentage living on >US\$ 2 per day ${ }^{\text {c }}$ | Demographic and health survey (DHS) data | Living standards measurement survey (LSMS) data |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | Algeria | 29950.0 | 2.0 | 13.1 | 84.9 |  |  |
|  | Benin | 6114.0 |  |  |  | Yes |  |
|  | Burkina Faso | 10995.7 | 61.2 | 24.6 | 14.2 | Yes |  |
|  | Cameroon | 14691.0 |  |  |  | Yes |  |
|  | Chad | 7486.0 |  |  |  | Yes |  |
|  | Comoros | 544.0 |  |  |  | Yes |  |
|  | Gambia | 1251.0 | 53.7 | 30.4 | 16.0 |  |  |
|  | Ghana | 18784.5 | 38.8 | 35.8 | 25.4 | Yes | Yes |
|  | Guinea | 7251.0 |  |  |  | Yes |  |
|  | Liberia | 3044.0 |  |  |  | Yes |  |
|  | Madagascar | 15050.5 | 63.4 | 25.7 | 11.0 | Yes |  |
|  | Mali | 10583.7 | 72.8 | 17.8 | 9.5 | Yes |  |
|  | Mauritania | 2598.3 | 28.6 | 40.0 | 31.3 |  |  |
|  | Niger | 10495.6 | 61.4 | 23.9 | 14.7 | Yes |  |
|  | Nigeria | 123896.5 | 70.2 | 20.6 | 9.2 | Yes |  |
|  | Senegal | 9285.3 | 26.3 | 41.5 | 32.2 | Yes |  |
|  | Sierra Leone | 4949.3 | 57.0 | 17.4 | 25.5 |  |  |
|  | Togo | 4567.0 |  |  |  | Yes |  |
|  | Total for subregion ${ }^{\text {d }}$ | 286129.7 | 55.5 | 22.4 | 22.1 | NA | NA |
| AFR-E | Botswana | 1588.1 | 33.3 | 28.1 | 38.7 |  |  |
|  | Burundi | 6678.0 |  |  |  | Yes |  |
|  | Central African Republic | 3539.8 | 66.6 | 17.4 | 16.0 | Yes |  |
|  | Côte d'lvoire | 15545.5 | 12.3 | 37.1 | 50.6 | Yes | Yes |
|  | Ethiopia | 62782.0 | 31.3 | 45.2 | 23.6 | Yes |  |
|  | Kenya | 29410.0 | 26.5 | 35.8 | 37.7 | Yes |  |
|  | Lesotho | 2105.0 | 43.1 | 22.6 | 34.3 |  |  |
|  | Malawi | 10788.0 |  |  |  | Yes |  |


|  | Mozambique | 17299.0 | 37.9 | 40.5 | 21.6 | Yes |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Namibia | 1701.3 | 34.9 | 20.9 | 44.2 | Yes |  |
|  | Rwanda | 8310.0 | 35.7 | 48.8 | 15.5 | Yes |  |
|  | South Africa | 42106.2 | 11.5 | 24.3 | 64.2 |  | Yes |
|  | United Republic of Tanzania | 32922.6 | 19.9 | 39.8 | 40.4 | Yes |  |
|  | Zambia | 9881.2 | 63.7 | 23.8 | 12.6 | Yes |  |
|  | Zimbabwe | 11903.7 | 36.0 | 28.3 | 35.8 | Yes |  |
|  | Total for subregion ${ }^{\text {d }}$ | 330084.7 | 27.3 | 36.2 | 36.5 | NA | NA |
| AMR-B | Brazil | 167966.7 | 9.0 | 16.4 | 74.7 | Yes |  |
|  | Chile | 15017.8 | 2.0 | 16.4 | 81.6 |  |  |
|  | Colombia | 41539.0 | 11.0 | 17.7 | 71.3 | Yes |  |
|  | Costa Rica | 3589.0 | 6.9 | 16.4 | 76.7 |  |  |
|  | Dominican Republic | 8404.4 | 3.2 | 12.8 | 84.0 | Yes |  |
|  | El Salvador | 6153.9 | 26.0 | 28.0 | 46.0 |  |  |
|  | Honduras | 6317.7 | 40.5 | 28.3 | 31.2 |  |  |
|  | Jamaica | 2598.0 | 3.2 | 22.1 | 74.8 |  |  |
|  | Mexico | 96585.7 | 12.2 | 22.6 | 65.2 | Yes |  |
|  | Panama | 2811.0 | 10.3 | 14.8 | 74.9 |  | Yes |
|  | Paraguay | 5358.8 | 19.5 | 29.8 | 50.7 | Yes |  |
|  | Trinidad and Tobago | 1292.8 | 12.4 | 26.6 | 61.0 | Yes |  |
|  | Uruguay | 3313.0 | 2.0 | 4.6 | 93.4 |  |  |
|  | Venezuela | 23707.0 | 18.7 | 25.9 | 55.4 |  |  |
|  | Total for subregion ${ }^{\text {d }}$ | 424396.0 | 11.0 | 19.1 | 69.8 | NA | NA |
| AMR-D | Bolivia | 8138.0 | 29.4 | 22.0 | 48.6 | Yes |  |
|  | Ecuador | 12412.0 | 20.2 | 32.1 | 47.7 | Yes | Yes |
|  | Guatemala | 11088.4 | 10.0 | 23.8 | 66.2 |  |  |
|  | Nicaragua | 4919.0 |  |  |  | Yes |  |
|  | Peru | 25230.0 | 15.5 | 25.9 | 58.6 | Yes |  |
|  | Total for subregion ${ }^{\text {d }}$ | 69897.5 | 17.4 | 26.3 | 56.3 | NA | NA |

Table 24.2 Populations, poverty estimates and availability of survey data by country and subregion ${ }^{\text {a }}$ (continued)

| Subregion | Country | Population in 1999 (000s) ${ }^{\text {b }}$ | Percentage living on <US\$ I per day ${ }^{\text {c }}$ | Percentage living on US\$ I-2 per day ${ }^{\text {c }}$ | Percentage living on >US\$ 2 per day ${ }^{\text {c }}$ | Demographic and health survey (DHS) data | Living standards measurement survey (LSMS) data |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| EMR-B | Jordan | 4739.9 | 2.0 | 5.4 | 92.6 |  |  |
|  | Tunisia | 9456.7 | 2.0 | 8.0 | 90.0 | Yes |  |
|  | Total for subregion ${ }^{\text {d }}$ | 136797.5 | 2.0 | 7.1 | 90.9 | NA | NA |
| EMR-D | Egypt | 62654.9 | 3.1 | 49.6 | 47.3 | Yes |  |
|  | Morocco | 28238.0 | 2.0 | 5.5 | 92.5 | Yes |  |
|  | Pakistan | 134790.0 | 31.0 | 53.7 | 15.4 | Yes | Yes |
|  | Sudan | 28993.0 |  |  |  | Yes |  |
|  | Yemen | 17047.6 | 15.7 | 29.5 | 54.8 | Yes |  |
|  | Total for subregion ${ }^{\text {d }}$ | 348468.4 | 19.3 | 45.3 | 35.3 | NA | NA |
| EUR-B | Armenia | 3808.9 | 7.8 | 26.2 | 66.0 |  |  |
|  | Azerbaijan | 7983.0 | 2.0 | 7.6 | 90.4 |  | Yes |
|  | Bulgaria | 8208.0 | 2.0 | 19.9 | 78.1 |  | Yes |
|  | Georgia | 5452.0 | 2.0 | 0.0 | 98.0 |  |  |
|  | Krygyzstan | 4865.0 |  |  |  | Yes |  |
|  | Poland | 38654.0 | 2.0 | 0.0 | 98.0 |  |  |
|  | Romania | 22458.0 | 2.8 | 24.7 | 72.5 |  |  |
|  | Tajikistan | 6237.0 |  |  |  |  | Yes |
|  | Turkey | 64385.0 | 2.4 | 15.7 | 82.0 |  |  |
|  | Turkmenistan | 4779.3 | 20.9 | 38.1 | 41.0 |  |  |
|  | Uzbekistan | 24406.3 | 3.3 | 23.2 | 73.5 | Yes |  |
|  | Total for subregion ${ }^{\text {d }}$ | 215275.9 | 3.0 | 14.6 | 82.3 | NA | NA |
| EUR-C | Belarus | 10032.0 | 2.0 | 0.0 | 98.0 |  |  |
|  | Estonia | 1442.4 | 2.0 | 3.2 | 94.8 |  |  |
|  | Hungary | 10068.0 | 2.0 | 5.3 | 92.7 |  |  |
|  | Kazakhstan | 14927.0 | 1.5 | 13.8 | 84.7 | Yes | Yes |


|  | Latria | 2431.1 | 2.0 | 6.3 | 91.7 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Lithuania | 3699.0 | 2.0 | 5.8 | 92.2 |  |  |
|  | Republic of Moldova | 4281.0 | 11.3 | 27.2 | 61.6 |  |  |
|  | Russian Federation | 146200.0 | 7.1 | 18.0 | 74.9 |  | Yes |
|  | Ukraine | 49550.0 | 2.9 | 42.7 | 54.4 |  |  |
|  | Total for subregion ${ }^{\text {d }}$ | 246335.9 | 5.4 | 21.3 | 73.3 | NA | NA |
| SEAR-B | Indonesia | 207021.6 | 7.7 | 47.7 | 44.7 | Yes |  |
|  | Sri Lanka | 18985.0 | 6.6 | 38.8 | 54.7 |  |  |
|  | Thiland | 60245.8 | 2.0 | 26.2 | 71.9 | Yes |  |
|  | Total for subregion ${ }^{\text {d }}$ | 288750.3 | 6.4 | 42.5 | 51.0 | NA | NA |
| SEAR-D | Bangladesh | 127668.8 | 29.1 | 48.8 | 22.2 | Yes |  |
|  |  | 997515.2 | 44.2 | 42.0 | ${ }_{13.8}$ | Yes |  |
|  | Nepal | 23384.2 | 37.7 | 44.8 | 17.5 | Yes |  |
|  | Total for subregion ${ }^{\text {d }}$ | 1219491.8 | 42.4 | 42.8 | 14.8 | NA | NA |
| WPR-B | China | 1253595.0 | 18.5 | 35.2 | 46.3 |  | Yes |
|  | Lao People's Democratic Repulic | 5096.7 | ${ }^{26.3}$ | 46.8 | 26.9 |  |  |
|  | Mongolia | 2378.3 | 13.9 | 36.0 | 50.0 |  |  |
|  | Philippines | 74259.0 |  |  |  | Yes |  |
|  | Repubic of Korea | 46858.0 | 2.0 | 0.0 | 98.0 |  |  |
|  | Total for subregion ${ }^{\text {d }}$ | 1520272.9 | 17.9 | 34.0 | 48.1 | NA | NA |
|  | Total for all subregions ${ }^{\text {c }}$ | 5085900.6 | 23.7 | 33.4 | 42.9 | NA | NA |
| NA Not | plicale. |  |  |  |  |  |  |
|  | ies are included in this able if they have a | one of the follow | verty es | Jan, LSM |  |  |  |
|  | ion couns for country-1vel fliom World d | Sopment indicaors | Word | do for sur | from Un |  |  |
|  | estinates (World Bank 20006). |  |  |  |  |  |  |
|  | onal toals include counries not lised in | table. Povery | are base | conrries | in the | text fo |  |
|  | onal toals salaulated sising country-kevel | erry and popula | from W | (10) and | vel popu | Un |  |

Table 24.3 Estimates of absolute poverty for the whole world and for developing countries

|  | Year |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- |
| Poverty category | 1987 | 1990 | 1993 | 1996 | 1998 |
| Member 2 countries of the World Bank <br> <US\$ 1.08 per day | 28.3 | 29.0 | 28.2 | 24.5 | 24.0 |
| <US\$ 2.15 per day | 61.0 | 61.7 | 60.1 | 56.1 | 56.0 |
| Our estimates—II poorest subregions only <br> <US\$ I per day <br> <US\$ 2 per day | - | - | - | 23.7 |  |
| Our estimates—whole world | - | - | - | 57.1 |  |
| <US\$ I per day |  |  |  |  |  |
| <US\$ 2 per day | - | - | - | 20.2 |  |
| - No data. | - | - | - | 48.6 |  |

The country-level poverty estimates from the World Bank were not disaggregated by sex and age. We assumed a uniform distribution of absolute poverty by sex and age. (We also estimated the crude joint distribution of poverty and risk factors without considering variation by sex and age.) A future improvement to the analyses presented here, therefore, would be to conduct sensitivity analyses on the variation in the poverty distribution by sex and age. This improvement, however, would also require estimating sex- and age-specific joint associations of poverty and each risk factor, for which there are currently very few data.

We obtained the prevalence (or distribution) of risk factors by subregion from the respective chapters in this publication. Wherever possible we used a dichotomous specification of the risk factor for the age and sex strata combined (i.e. crude prevalence estimates by subregion for each risk factor) except for unsafe sex, for which the different sexes were analysed separately. Using a dichotomous variable resulted in the loss of information included in distributions or multiple exposure categories.

### 2.2 Estimating Joint distribution of absolute poverty and RISK FACTOR RELATIONSHIP

Table 24.4 represents a stylized version of the data collection sheets required for each subregion. The marginal totals (in bold) show absolute poverty and risk factor prevalence. Our task was to estimate the cell values, $a, b, c$, $d$, e and $f$. If each of the values " $a$ " to " $f$ " represents the absolute percentage of the population of the total subregion in each cell, then:

$$
a+b+c+d+e+f=100 \% .
$$

Table 24.4 Data collection sheets by subregion: 2 by 3 tables of percentage risk factor by poverty

|  | Exposed to risk factor | Not exposed to risk factor |  |
| :--- | :---: | :---: | :---: |
| <US\$ I per day | a | b | $\mathbf{M}_{\mathbf{1}}$ |
| US\$ I-2 per day | c | d | $\mathbf{M}_{\mathbf{2}}$ |
| Non-poverty | e | f | $\mathbf{M}_{\mathbf{3}}$ |
|  | $\mathbf{P}$ | $\mathbf{Q}$ | $\mathbf{1 0 0 \%}$ |

It was possible to approximate the values for " a " to " f " via the results of the regression analyses and indirect method described above. However, these estimates would have been consistent with the prevalence of the risk factor in the survey data sets used (i.e. DHS or LSMS), but not necessarily consistent with the prevalence of the risk factors determined in the relevant chapters in this publication. Likewise, the fixed poverty percentages came from an external source, and thus all the marginal percentages ( $\mathrm{M}_{1}, \mathrm{M}_{2}, \mathrm{M}_{3}, \mathrm{P}$ and Q ) were fixed. To solve the joint distribution within the table, we used prevalence ratios (referred to as "relative risks" hereafter) estimated using the regression and indirect method, such that:

$$
\begin{align*}
\mathrm{RR}_{1} & =\text { relative risk of risk factor for <US\$1 compared to }>\text { US } \$ 2 \\
& =\left(\mathrm{a} / \mathrm{M}_{1}\right) /\left(\mathrm{e} / \mathrm{M}_{3}\right) \tag{1}
\end{align*}
$$

$\mathrm{RR}_{2}=$ relative risk of risk factor for $>$ US \$ 1 but $<$ US $\$ 2$

$$
\begin{equation*}
\text { compared to >US\$ } 2 \tag{2}
\end{equation*}
$$

$$
=\left(\mathrm{c} / \mathrm{M}_{2}\right) /\left(\mathrm{e} / \mathrm{M}_{3}\right)
$$

Given that $\mathrm{c}=(\mathrm{P}-\mathrm{a}-\mathrm{e})$, then Equation 2 can be rearranged to give:

$$
\begin{align*}
\mathrm{RR}_{2} & =\left((\mathrm{P}-\mathrm{a}-\mathrm{e}) / \mathrm{M}_{2}\right) /\left(\mathrm{e} / \mathrm{M}_{3}\right) \\
& =\frac{(\mathrm{P}-\mathrm{a}-\mathrm{e}) \mathrm{M}_{3}}{\mathrm{eM}} \\
\rightarrow \mathrm{eRR}_{2} \mathrm{eM}_{2} & =\mathrm{PM}_{3}-\mathrm{aM}_{3}-\mathrm{eM}_{3}  \tag{3}\\
\rightarrow \mathrm{e}\left(\mathrm{RR}_{2} \mathrm{M}_{2}+\mathrm{M}_{3}\right) & =\mathrm{PM}_{3}-\mathrm{aM}_{3} \\
\rightarrow \mathrm{e} & =\frac{\mathrm{M}_{3}(\mathrm{P}-\mathrm{a})}{\mathrm{RR}_{2} \mathrm{M}_{2}+\mathrm{M}_{3}}
\end{align*}
$$

Rearranging Equation 1 to solve for "e", and substituting in Equation 3 allows a solution in terms of " a ":

Figure 24.3 Flow diagram of how the distribution of poverty and risk factors might be estimated


$$
\begin{align*}
a M_{3} / R_{1} M_{1} & =M_{3}(\mathrm{P}-\mathrm{a}) /\left(\mathrm{RR}_{2} \mathrm{M}_{2}+\mathrm{M}_{3}\right) \\
\rightarrow \mathrm{a} & =\left(\mathrm{RR}_{1} \mathrm{M}_{1} \mathrm{P}-\mathrm{RR}_{1} \mathrm{M}_{1} \mathrm{a}\right) / \mathrm{RR}_{2} \mathrm{M}_{2}+\mathrm{M}_{3} \\
\rightarrow \mathrm{a}\left(\mathrm{RR}_{2} \mathrm{M}_{2}+\mathrm{M}_{3}\right) & =\mathrm{RR}_{1} \mathrm{M}_{1} \mathrm{P}-\mathrm{RR}_{1} \mathrm{M}_{1} \mathrm{a}  \tag{1}\\
\rightarrow \mathrm{a}\left(\left(\mathrm{RR}_{2} \mathrm{M}_{2}+\mathrm{M}_{3}\right)+\mathrm{RR}_{1} \mathrm{M}_{1}\right) & =\mathrm{RR}_{1} \mathrm{M}_{1} \mathrm{P} \\
\rightarrow \mathrm{a} & =\mathrm{RR}_{1} \mathrm{M}_{1} \mathrm{P} /\left(\left(\mathrm{RR}_{2} \mathrm{M}_{2}+\mathrm{M}_{3}\right)+\mathrm{RR}_{1} \mathrm{M}_{1}\right)
\end{align*}
$$

Having calculated "a" the remaining within-cell percentages in Table 24.4 are easily solved.

Figure 24.3 shows a flow diagram of three methods that might be used to estimate this joint distribution.

Ideally the direct method would have been used, but there were no appropriate survey data using a poverty definition equivalent to that of Chen and Ravallion (2000). We therefore had to employ an indirect estimation procedure, using survey data on the joint distribution of other socioeconomic and risk factors. A necessary and key assumption of this indirect method was that the association of risk factors with various
ranked socioeconomic factors was similar. Put another way, we had to assume that one socioeconomic factor could be substituted for another when measuring socioeconomic gradients in health-related behaviour. This was a strong assumption, but one with some support. We did not use the last option of extrapolation from other subregions or expert opinion.

### 2.3 Data sources and analysis

Table 24.5 provides a summary of the risk factors reported elsewhere in this book, and the extent of inclusion of risk factors in this work.

Indirect estimation of the joint distribution of poverty by risk factor required individual-level data from countries within all subregions. Selected DHS were the main source of these data. We also used LSMS data and survey data specific to China.

## DEmographic and health survey (DHS) data

DHS are nationally representative household surveys with large sample sizes of about 5000 households (see http://www.measuredhs.com). The DHS programme is conducted by Macro International, with support from the US Agency for International Development. A standard set of questionnaires, similar in all countries, is used to collect data at individual, household and community levels. Four rounds of surveys have been conducted in the past two decades, in which several thousand households were sampled at intervals in some 50 countries across Asia, Africa, the Middle East, Latin America and the area covered by the former Soviet Union.

The core DHS consists of questionnaires for the household and for women specifically. A sample of women aged 15-49 years are interviewed on the following topics: household characteristics, lifetime reproduction, contraceptive knowledge and use, maternity and breastfeeding, immunization, children's health, marriage and fertility preferences, husband's background and the woman's work status. In addition, other country-specific questions may be asked. The DHS contains data for five of the comparative risk assessment (CRA) risk factors:

- underweight in children (weight-for-age $z$-score, DHS variable code hw8);
- improved water and sanitation (DHS variable codes v113 and v116);
- unsafe sex (DHS variable codes v502, 525, 531, 761, 851, 852, 853, 872);
- mother's body mass index (DHS variable code v445); and
- cooking fuel (country-specific variables).

We constructed these variables, in some cases by using several of the original DHS variables. We obtained unit-level data for 53 countries,
Table 24.5 Summary of the CRA risk factors

| Risk factor | Continuous/categorical | Survey data available for socioeconomic position by risk factor | Inclusion in data analysis | Inclusion in systematic literature review |
| :---: | :---: | :---: | :---: | :---: |
| Air pollution, indoor | Prevalence of solid fuel exposure | Yes-LSMS | Yes | No |
| Air pollution, urban | PM $10\left(\mu \mathrm{~g} / \mathrm{m}^{3} ; 1-10,11-50,5 \mathrm{l}-100, \geq 100\right)$ | Partial | Yes | No |
| Alcohol use | Amount weighted by pattern of drinking | Yes ${ }^{\text {a }}$ | Yes | Yes |
| Child sexual abuse |  | No | No | No |
| Contaminated injections in health care settings | Proportion receiving injection contaminated with HIV, HCV, HBV | Insufficient | No | No |
| Global climate change |  | Insufficient | No | No |
| High blood pressure |  | Insufficient | No | Yes |
| High cholesterol |  | Insufficient | No | Yes |
| Illicit drug use |  | Insufficient | No | Yes |
| Lead exposure | Mean (SD), prevalence $>0.016 \mu \mathrm{~g} / \mathrm{dl}$ | Insufficient | No | Yes |
| Low fruit and vegetable consumption | Mean (g/day), prevalence could be derived | Insufficient | No | No |
| Malnutrition: protein-energy (0-4 years) | Prevalence underweight (weight-for-age z-score <-2) | Yes-DHS | Yes | No |
| Micronutrient deficiencies | (Vitamin A, haemoglobin, iodine) | Insufficient-UNICEF MICS2 not yet available | No | No |
| Occupational risks (selected) | Prevalence chemical, physical, biological, human factors, safety risk | Insufficient | No | No |
| Overweight and obesity | Prevalence of overweight and obesity | Yes (women only)—DHS | Yes (women only) | Yes |
| Physical inactivity |  | No | No | Yes |
| Tobacco use | Used indirect method | Yes ${ }^{\text {a }}$ | Yes | Yes |
| Unsafe sex | United Nations variables | Yes-DHS | Yes | No |
| Unsafe water, sanitation and hygiene | Prevalence of five exposure scenarios | Yes-DHS | Yes | No |

a Alcohol and tobacco data on LSMS were household expenditure data only. Further, it was not possible to "weight" alcohol data by pattern of drinking as in chapter 12.
taking the most recent survey if the country had been surveyed more than once in the period 1986-2000. Of these, an asset score could be calculated for 51 countries, as listed in Table 24.2.

## Calculation of asset score

Income was not directly elicited by the standard DHS. However, the World Bank has undertaken extensive work to create asset scores using DHS data (Filmer and Pritchett 1988). Data from the first round of surveys were re-coded to ensure comparability with the subsequent three rounds of surveys. Core questions from all surveys were combined into a single data set. Four categorical variables were constructed as component variables for a DHS asset score: urban-rural status, housing construction material (usually floor material), educational status and availability of electricity. If these variables were missing for a particular country a substitute variable was used where possible, as follows:

- wall material was substituted for floor material in Pakistan;
- number of rooms was substituted for floor material in India; and
- possession of a radio was substituted for electricity supply in Burundi, the Dominican Republic, Liberia and Tunisia.

For two countries (Sri Lanka and Turkey) it was not possible to find suitable substitute variables and asset scores could not be calculated. We considered including more variables in the factor analysis in order to increase the resolution of the asset scores. However, no further variables were available for all countries; thus the addition of further variables would have meant that fewer countries could be included in the subsequent analyses. We considered it important that the asset score be applicable to as many countries as possible. Also, unlike some other asset scores, ours did not include access to safe water to avoid self-correlation in the analyses of water and sanitation by asset score.

Figures 24.4-24.7 show the distribution of DHS observations by the four variables used to calculate the asset scores. (The actual numbers of observations are shown in Appendix A, Table A.1.) There was a predominance of rural people in the DHS data sets (Figure 24.4). There was notable variation between subregions in floor type (Figure 24.5), but there was also some possible variation in coding patterns between countries with regard to the use of the intermediate "rudimentary" floor type. The education variable had four levels, with reasonable spread across respondents and variation in patterns between subregions (Figure 24.6). There was marked variation between subregions in the dichotomous variable for access to electricity.

The distributions of these variables varied markedly between different countries, even within subregions (Appendix A, Table A.1).

A global ${ }^{2}$ asset score was calculated using factor analysis. The four variables used to make up the asset scores were coded within the DHS

Figure 24.4 Distribution of DHS asset score variables by country: urban-rural index


Figure 24.5 Distribution of DHS asset score variables by country: floor type


Figure 24.6 Distribution of DHS asset score variables by country: highest education level achieved


AFR-D


AFR-E

AMR-B





 (g-ydM) souidd!!ud $\mathbb{B} \quad$ Almmenm


AMR-D EMR-B EMR-D
EUR-B EUR-C SEAR-B SEAR-D

Figure 24.7 Distribution of DHS asset score variables by country: access to electricity


Table 24.6 Principal factors, one factor retained

| Factor | Eigenvalue | Difference | Proportion | Cumulative |
| :--- | ---: | :---: | :---: | :---: |
| 1 | 1.53227 | 1.61037 | 1.3339 | 1.3339 |
| 2 | -0.07810 | 0.05096 | -0.0680 | 1.2659 |
| 3 | -0.12906 | 0.04731 | -0.1124 | 1.1535 |
| 4 | -0.17637 |  | -0.1535 | 1.0000 |

Table 24.7 Loadings for factor I (from Table 24.6)

| Variable | Factor loading | Uniqueness |
| :--- | :---: | :---: |
| Education | 0.60216 | 0.63740 |
| Urban-rural | -0.58536 | 0.65735 |
| Floor type | 0.59736 | 0.64316 |
| Electricity | 0.68570 | 0.52982 |

Table 24.8 Scoring coefficients for factor I (from Table 24.6)

| Variable | Scoring coefficient |
| :--- | :---: |
| Education | 0.25759 |
| Urban-rural | -0.24446 |
| Floor type | 0.25310 |
| Electricity | 0.34373 |

data sets with numeric values corresponding to different categories or variable values:

- urban-rural status: $0=$ large city, $1=$ small city, $2=$ town, $3=$ rural area
- floor type: 1 = natural, 2 = rudimentary, 3 = finished
- highest education level achieved: $0=$ none, $1=$ primary, $2=$ secondary
- access to electricity: $0=$ no, $1=$ yes.

Principal factor analysis was performed in Stata (an interactive data analysis program) for these four variables. The results are shown in Tables 24.6 and 24.7.

Given that only four variables (each with relatively few values) were available for the factor analysis, only 96 discrete asset score values were generated, ranging from -1.12 (more poor) to 1.60 (least poor). The scoring coefficients are shown in Table 24.8.

A full account of the resulting asset score values, and the number of DHS observations with each score by subregion, is given in Appendix A, Table A.2.

Figure 24.8 shows the number of DHS observations by asset score by subregion. The size of the circles represents the number of observations at that asset score value. The number of DHS observations with asset scores is fewer for EMR-B $(n=4184)$ and EUR-C $(n=4704)$ than for subregions such as AFR-D $(n=94811)$ and SEAR-D $(n=107044)$. For AFR-D and AFR-E the asset scores are skewed towards the lower end (as expected) and for AMR-B and AMR-D are skewed to the higher end.

We did not calculate subregional-level asset scores. As the eventual use of asset scores was to rank individuals by socioeconomic position, based on the assumption that this ranking was comparable to that by absolute poverty (also a global construct), we did not consider the construction of subregional-level asset scores justified. Moreover, the range of discrete values of the asset score and sample sizes within some subregions would have introduced instability in asset score indices between subregions that may not have been a function of genuinely varying "asset score constructs" but rather introduced greater random error.

## Weighting of the DHS data sets

The objective of the weighting was to ensure that the analyses represented as closely as possible the distribution of the whole population in each subregion. A two-step weighting procedure was thus required.

Figure 24.8 Distribution of asset scores by subregion


Table 24.9 Examples of weights for DHS analyses

| Country in given subregion | Number of people <br> in DHS sample ( $n$ ) | Number of people <br> in country (N) | Weight (N/n) |
| :--- | :---: | :---: | :---: |
| Country A | 1000 | 500000 | 500 |
| Country B | 5000 | 5000000 | 1000 |
| Country C | 2000 | 100000 | 50 |
| Total of countries sampled | 8000 | 5600000 |  |

1. Owing to the sampling characteristics of the DHS, each observation in the DHS is assigned a country-level weight. The sum of these weights is the same as the total sample size for that particular country, multiplied by 1 million. We simply converted that weight back to a distribution with a mean of 1.0 by dividing by 1 million. We called this component weight $\mathrm{w}_{\mathrm{i}}$, being the individual-level DHS weight for the $\mathrm{i}^{\text {th }}$ observation in each country.
2. We then weighted the individual DHS observations to be representative of country population by dividing the total population of the $j^{\text {th }}$ country $\left(N_{j}\right)$ by $\sum^{i} w_{1}$-the country-level weight. (Note that $\sum^{i} w_{i}$ is simply the DHS sample size for that particular country.) The product of the individual-level and country-level weights, $\mathrm{w}_{\mathrm{ij}}=\left(\mathrm{N}_{\mathrm{i}} / \sum^{\mathrm{i}} \mathrm{w}_{\mathrm{i}}\right) \times \mathrm{w}_{\mathrm{i}}$, was then assigned to the $\mathrm{i}^{\text {th }}$ observation in the $\mathrm{j}^{\text {th }}$ DHS country (within subregions). Note that within each country, the sum of $\mathrm{w}_{\mathrm{ij}}$ is simply the total population of that country.

To illustrate the necessity for the second weighting step, Table 24.9 provides hypothetical data for a given subregion. If each observation were entered unweighted into any analysis, the results would be skewed towards the pattern of risk factor by poverty present for the individuals "most represented". In Table 24.9, that would be the individuals in country C, where each observation represents only 50 people. To rectify this, the weights in step 2 above (i.e. those in the final column of Table 24.9) were combined with the DHS weights.

The application of the weights for the analysis using the indirect method was not simply a matter of weighting each individual observation. Rather, the asset score was modelled as a cumulative proportion of the population, i.e. a ranking. Thus, we used the weights to make the ranking representative of the total subregional population's ranking, not just the ranking in the available data samples, as described below.

Table 24.10 shows a hypothetical data set for 10 DHS observations (record numbers i , for $\mathrm{i}=1-10$ ). Without considering weights, each of these 10 observations represents $10 \%$ (or a proportion of 0.1 ) of all

Table 24.10 Hypothetical example of IO DHS observations and weighted "midpoint" of cumulative proportion

| Record number (i) | Midpoint of cumulative proportion distribution of $i$ | Weight | Sum of weights | Cumulative proportion of weights | Midpoint of cumulative proportion of weights |
| :---: | :---: | :---: | :---: | :---: | :---: |
| I | 0.05 | 200 | 200 | 0.040 | 0.020 |
| 2 | 0.15 | 700 | 900 | 0.180 | 0.110 |
| 3 | 0.25 | 900 | 1800 | 0.360 | 0.270 |
| 4 | 0.35 | 150 | 1950 | 0.390 | 0.375 |
| 5 | 0.45 | 1900 | 3850 | 0.770 | 0.580 |
| 6 | 0.55 | 200 | 4050 | 0.810 | 0.790 |
| 7 | 0.65 | 100 | 4150 | 0.830 | 0.820 |
| 8 | 0.75 | 400 | 4550 | 0.910 | 0.870 |
| 9 | 0.85 | 300 | 4850 | 0.970 | 0.940 |
| 10 | 0.95 | 150 | 5000 | 1.000 | 0.985 |

observations. The midpoint of the cumulative proportion distribution is therefore $0.05,0.15, \ldots, 0.95$, as shown in the second column of Table 24.10. Assume, however, that each of these 10 observations represents varying numbers of the total population, as represented by the weights in third column. The cumulative sum, cumulative proportion and "midpoint" of the cumulative proportion of the weights are shown in the final three columns of Table 24.10. For example, the weight for record 2 was 700 , and the sum of weights for records 1 and 2 was $200+700=900$. The sum of all weights was 5000 . Therefore, the cumulative proportion of weights up to record 2 was $900 / 5000=0.180$. The midpoint of the cumulative proportion distribution was half way between the cumulative proportion of weights for records 1 and 2, i.e. $0.040+((0.180-$ $0.040) / 2$ ) $=0.110$.

Thus, weighting of the midpoints of the cumulative proportion "repositions" the rank value between 0 and 1 . This repositioning aims to place each observation at about the position it would have been if the entire population of the subregion had been sampled.

## Comparison of asset scores and income in Pakistan

The Pakistan Integrated Household Survey 1991, one of a series of LSMS conducted by the World Bank, included direct estimates of household income as well as variables required to generate an asset score. Thus, we were able to examine the distribution of asset scores by income at the individual level within one country. Note that the available income data
were not consumption data. Also, the asset score variables were not identical to those on the DHS data sets.

We calculated an asset score similar to the method described above. It was not possible, however, to include a variable for urban-rural status, as there were no comparable data readily available from the survey.

Figure 24.9 is a whisker plot of asset scores by rank of household income. Each box-whisker plot is for a given asset score. The boxes show the 25 th and 75 th percentiles of rank income for a given asset score, and the whiskers show the 5th and 95th percentile ranks of income for that asset score. It is evident that, while there is an association in the expected direction of rank asset score and rank household income, there is also considerable variation of household income ranks within a given asset score.

It is possible that the wide distribution of income data in Figure 24.9 is due to measurement error. Also, incomes tend to be volatile whereas assets are a more stable indicator of long-term income. Therefore, while income and asset scores are not as closely associated as one might wish, this does not render asset scores an unsuitable proxy for income poverty. First, in developed countries we find similar associations between a range of socioeconomic factors and health despite considerable imperfections

Figure 24.9 Box-whisker plot of the normalized rank of estimated household income (NRinc) by normalized rank of household asset score for Pakistan, using LSMS data


Table 24.1 I Comparison of assigning income poverty level in Pakistan using asset score and household income score according to LSMS data

|  |  | Allocation to level of poverty by <br> household asset score |  |  |
| :--- | :---: | :---: | :---: | :---: |
|  |  | <US\$ 1 | US\$ I-2 | $>$ US\$2 |
| Allocation to level of poverty | <US\$ I | $17 \%$ | $14 \%$ | $1 \%$ |
| by household income score | US\$ I-2 | $16 \%$ | $35 \%$ | $3 \%$ |
|  | $>$ US\$2 | $1 \%$ | $10 \%$ | $3 \%$ |

in the correlation of these socioeconomic factors at individual level (Blakely and Pearce 2002; Blakely et al. 2002). Second, we cannot tell, from Figure 24.9, which is the better proxy for income poverty. Third, we treat income poverty here as a categorical variable. It may thus be more reasonable to compare household income score and asset score rank in assigning people to $<$ US $\$ 1$, US $\$ 1-2$ and $>$ US $\$ 2$ per day using the World Bank estimates of $31 \%$, $54 \%$ and $15 \%$, respectively (Table 24.11). Fifty-five per cent of households are correctly assigned to an income category using asset score rankings.

## Asset scores by gross national product (GNP)

There was a moderately strong correlation between the rank of weighted average asset score for each country and the rank of GNP per capita by country (Spearman rank correlation coefficient 0.68), providing further (albeit highly aggregated) evidence that the asset score is a proxy measure of income (Figure 24.10).

## Indirect method using DHS asset scores

We identified three possible pathways for estimating the joint distribution of poverty and risk factors at the subregional level.

1. Pool DHS data by subregion, directly estimate association of asset scores with risk factors and then derive relative risks of each risk factor by poverty categories.
2. Determine the association of asset scores and risk factors for each country, derive relative risks of each risk factor by poverty categories at the country level, then either: (a) pool (e.g. Maentel-Haenzel) the country-level relative risk estimates; or (b) estimate the number of people in each cell of the two by three table (risk factor by three-level poverty variable) for each country using population data, sum the numbers of people in each cell for all countries in the subregion with data, and then calculate subregional-level relative risks. (Note: 2 b estimates the "crude" relative risks by poverty at the subregional

Figure 24.10 Scatter plot of rank of a country's per capita GNP against the rank of a country's average asset score

level, whereas 2 a estimates the relative risks unconfounded by country.)

A priori we chose method 1 as our task was to estimate subregionallevel associations. We were not confident that the data would support country-level calculations for each risk factor, and we were concerned about instability arising from country-level working. However, a strength of method 2 was that it did not assume that a given asset score equated to the same income poverty across countries within subregions.

The principle of the indirect method applied directly to subregionallevel data is first illustrated using a simplified example of aggregate (quintile) data before the more detailed non-parametric regression techniques are considered.

Table 24.12 shows some hypothetical data for the distribution of childhood malnutrition by quintile of asset score at the subregional level. The midpoints of each quintile on a cumulative proportional distribution are shown in the last column.

Figure 24.11 shows a scatter plot of the prevalence of childhood malnutrition by these midpoints. If the plot is nearly linear (as in this case), a regression line is fitted. This method is similar to that used to calculate the relative and slope indices of inequalities (Mackenbach and Kunst 1997).

Table 24.12 Hypothetical distribution of childhood malnutrition by asset score

| Quintile of asset score | Prevalence of childhood <br> malnutrition (\%) | Midpoint on cumulative <br> proportion distribution |
| :--- | :---: | :---: |
| (poorest) | 22 | 0.1 |
| 2 | 17 | 0.3 |
| 3 | 12 | 0.5 |
| 4 | 9 | 0.7 |
| 5 | 6 | 0.9 |

Figure 24.II Percentage childhood malnutrition plotted against cumulative proportion of population ranked by asset score


Assuming that this regression line is similar to the unobserved line for poverty by underweight, it can be used to estimate the prevalence of malnutrition by level of poverty. For example, if in a particular subregion the percentage of children living on $<$ US $\$ 1$ per day is $20 \%$, the percentage living on US\$1-2 per day is $30 \%$, leaving $50 \%$ non-poor. The midpoints on the cumulative proportional distribution for these poverty groups would accordingly be $0.10,0.35$ and 0.75 . Solving the regression equation (shown in Figure 24.11) for each of these midpoints, we estimate the prevalence of underweight to be $21.2 \%, 16.2 \%$ and $8.2 \%$, respectively. Treating the non-poor as the reference group, these per-
centages equate to relative risks of 2.59 for children living on $<$ US $\$ 1$ per day (i.e. 21.2/8.2) and 1.98 for children living on US\$ 1-2 per day (i.e. 16.2/8.2). Also, the relative risk of underweight for children living on US $\$ 2$ per day compared to all those living on $>$ US $\$ 2$ can be estimated as 1.61 .

Locally linear kernel regression. While illustrative, this example is an oversimplification for two reasons. First, we needed to allow for nonlinear associations. Second, we wanted to use the fact that more than five values of the asset score were captured in the unit-level DHS data. These objectives can be met using a non-parametric regression technique: locally linear kernel regression. This method essentially fits separate linear regression curves for each unit increase in the independent variable(s). In so doing, observations closest to the midpoint are given more weight, and observations further away are given less weight. The outcome is a fitted curve comprising many linear segments that, together, give a smoothed regression line.

Given the large size of the data set, we were able to use the unit-level DHS data for this regression method only for the unsafe sex risk factor. For these unit-level analyses we modelled the logit in Stata. However, for underweight, and for water and sanitation, the observations were too many for efficient computation. Therefore, we calculated the average prevalence of each risk factor within categories of asset score. These aggregate data were then analysed with Proc Loess in SAS software. The procedure for a single risk factor within one subregion was as follows.

1. The normalized rank of each asset score value was calculated based on the DHS and country-level weights as before (section 2.3).
2. The average prevalence of the risk factor was estimated for each asset score value.
3. The prevalence of the risk factor was regressed on the asset score using locally linear kernel regression.
4. This analysis was weighted according to the number of DHS observations represented by each data point ("weight" option of Proc LOESS in SAS—otherwise the default options were used).

This method was repeated for each subregion, and the whole process repeated for each of the risk factors.

Estimating the joint association of poverty and risk factors. To estimate the joint association of absolute poverty and each risk factor, we determined the average prevalence of the risk factor over the range of asset score rankings equivalent to the poverty estimate.

1. Dummy observations were appended to the data set, representing $0.1 \%$ increments of asset score rank between the poorest (rank 0.00) and the least poor (rank 1.00).
2. The results of the locally linear kernel regression were linearly interpolated to provide estimated risk factor prevalences for each of the $0.1 \%$ increments of asset score ranking.
3. Average risk factor prevalences were estimated for the appropriate range of the dummy observations, to the nearest $0.1 \%$.

For example, if $20.5 \%$ of people in a given subregion lived on $<$ US $\$ 1$ per day, we estimated the percentage underweight among those living on $<$ US\$1 as the average prevalence under the interpolated curve representing asset score ranks of $0.000-0.205$. Thus, for each subregion, we estimated the prevalence of the risk factor for:

- people living on $<\mathrm{US} \$ 1$ per day;
- people living on (equal to) US\$ 1 per day;
- people living on US\$ 1-2 per day;
- people living on (equal to) US\$2 per day; and
- people living on $>$ US $\$ 2$ per day.

Using these prevalence estimates, we then estimated the following relative risks of each risk factor for each subregion:

- <US\$ 1 per day compared to $>$ US $\$ 2$ per day;
- US\$ 1-2 per day compared to >US\$2 per day;
- (equal to) US\$2 per day compared to $>$ US\$ 2 per day;
- <US\$ 1 per day compared to $>$ US\$ 1 per day; and
- (equal to) US\$ 1 per day compared to $>$ US\$ 1 per day.


## Living standards measurement survey (LSMS) Data

The LSMS protocol and questionnaires were developed by the World Bank in 1980 and have been carried out in a number of countries. The main purpose of the LSMS is to provide a tool for monitoring living standards and household behaviour in developing countries, with a focus on measuring and understanding poverty. There are three levels of questionnaires within the LSMS: household (household-level income, expenditure and possibly behavioural questions), community (information common to all members of the local community, such as availability of schools, water and electricity) and price (local market prices for basic commodities).

Table 24.13 LSMS countries available for analysis: total population of country and year surveyed

| Country | Subregion | Total population (000s) | Year surveyed $^{\text {a }}$ |
| :--- | :--- | :---: | :---: |
| Ghana | AFR-D | 18785 | $1998 / 1999$ |
| Côte d'lvoire | AFR-E | 15545 | 1988 |
| South Africa | AFR-E | 42106 | 1994 |
| Panama | AMR-B | 2811 | 1997 |
| Ecuador | AMR-D | 12412 | 1995 |
| Pakistan | EMR-D | 134790 | 1991 |
| Azerbaijan | EUR-B | 7983 | 1995 |
| Bulgaria | EUR-B | 8208 | 1995 |
| Tajikistan | EUR-B | 6237 | 1999 |
| Kazakhstan | EUR-C | 14927 | 1996 |
| Russian Federation | EUR-C | 146200 | 1992 |

a Where the country was surveyed in more than one year only the latest year's data were taken.

The questionnaires are made up of modules (not all of which are used by every country) and are often modified to suit the needs and the situation of the country in question. However, all LSMS surveys are run according to similar protocols, with rigorous quality control measures. They are designed to be quickly and easily administered in the field and rapidly entered into data entry programs.

There are currently about 25 countries listed on the World Bank web site as having LSMS data, some of which have been surveyed more than once. There are different access policies for data from the different countries: some have open data policies, while others allow only restricted access to data. The data set collected varies from country to country, but may include information on income, expenditure, cooking fuel, tobacco use, alcohol consumption, height and weight and many other variables. We were restricted in our analyses to countries for which we could access LSMS data within the given time frame, and to those for which income or expenditure information as well as appropriate risk factor information was available (Table 24.13). Unless otherwise stated, it was assumed that every LSMS survey was representative of the country; hence no survey weights were used in the analyses.

Table 24.14 summarizes the risk factor information available by country. As information on hypertension and physical inactivity were available for only one country each (Bulgaria and the Russian Federation, respectively), these risk factors were not analysed.

For Bulgaria, Ghana, South Africa and Tajikistan there was no question in the surveys asking about individual tobacco or alcohol consumption. The prevalence of this risk factor thus had to be calculated

Table 24.14 LSMS countries and risk factor information available from their data sets

| Country | Subregion | Indoor air <br> pollution | Tobacco <br> use | Alcohol <br> use | Hypertension | Physical <br> inactivity |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Ghana | AFR-D | Yes | Yes* | Yes* | No | No |
| Côte d'lvoire | AFR-E | Yes | No | No | No | No |
| South Africa | AFR-E | Yes | Yes* | Yes* | No | No |
| Panama | AMR-B | Yes | Yes | Yes | No | No |
| Ecuador | AMR-D | Yes | Yes | Yes | No | No |
| Pakistan | EMR-D | Yes | Yes | No | No | No |
| Azerbaijan | EUR-B | Yes | Yes** | Yes** | No | No |
| Bulgaria | EUR-B | Yes | Yes* | Yes* | Yes | No |
| Tajikistan | EUR-B | Yes | Yes* | Yes* | No | No |
| Kazakhstan | EUR-C | No | Yes | Yes | No | No |
| Russian Federation | EUR-C | Yes | Yes | Yes | No | Yes |

Yes Variable is available.
No Variable is not available.

* Variable is available only in form of household spending on item.
** Alcohol and tobacco information available only in combined variable.
using household expenditure information for alcohol and for tobacco. A household was classified as using tobacco if any money was recorded as having been spent on cigarettes or tobacco within the expenditure section of the questionnaire. Likewise, a household was classified as using alcohol if any money had been spent on alcohol. While this is a rather crude measure, it should at least give us an approximate idea of the proportion of households with smokers and/or alcohol consumers.

For Azerbaijan the measure was even cruder, as there was only one variable to cover both alcohol and tobacco expenditure. Hence it could not be determined whether any money spent was on tobacco or alcohol or both.

Finally, it must be emphasized that the results based on LSMS data must be treated somewhat cautiously owing to the limited number of countries (and subregions) represented and the proxy nature of some of the risk factor variables.

## Calculation of income

LSMS data sets included information on income. A pre-generated aggregate per capita income variable was used if one was provided with the data set; otherwise an income variable of this type was generated using the available data. All income variables were equivalized by dividing by the square root of the number of members of the household, if this had
not already been performed within the generated variable. For Ecuador, only aggregated per capita expenditure was used.

As these income variables were generated in different ways, and to allow for purchasing power parities between countries, each country's income data were treated separately and ranked on a continuous scale from 0 to 1 for further analysis. For example, the household with the median LSMS income in a given country was assigned a ranked income value of 0.5 .

## Indirect method using LSMS income data

Since we were concerned about purchasing power parities between subregions, we did not pool the data by subregion to estimate the association of poverty with risk factors. Rather, we estimated associations at country level and aggregated these to subregions. Unlike the DHS data, the LSMS analyses were unweighted. Also, the smaller data sets allowed regression analyses on unit-level data (rather than using aggregated data). The associations between income and the various risk factors were determined by categorizing each risk factor into a dichotomous variable and then using the indirect method at the country level to estimate relative risks by poverty (see section 2.3).

Having obtained country-level joint distributions of the total population for poverty by the risk factor, we then aggregated these to estimate the association at the subregional level. Table 24.15 shows, as an example, the distribution of indoor air pollution by countries within subregion EUR-B.

Thus, the estimated subregional prevalence of indoor air pollution in EUR-B for those living on $<$ US $\$ 1$ per day is $0.36 \times 64 \%+0.37 \times 20 \%$ $+0.28 \times 74 \%=0.51$. The final subregional estimates are based only on those countries for which we had data (as for the DHS analyses).

Table 24.15 Distribution of indoor air pollution by countries in EUR-B for which data were available

| Country | Population (000s) | Proportion of population | Prevalence of indoor air pollution by poverty category (\%) |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | <US\$ I/day | US\$ I-2/day | >US\$ 2/day |
| Azerbaijan | 7983 | 0.36 | 64 | 62 | 43 |
| Bulgaria | 8208 | 0.37 | 20 | 22 | 17 |
| Tajikistan | 6237 | 0.28 | 74 | 76 | 74 |
| Total | 22428 | 1.00 |  |  |  |

## Data for China

We used data from the 1993 China Health and Nutrition Survey (CHNS). The information on this survey that follows is adapted from the web site of the Carolina Population Center at the University of North Carolina at Chapel Hill (http://www.cpc.unc.edu/china, accessed 23 June 2003).

The study population is drawn from the provinces of Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning and Shandong. This sample is diverse, with variation in a wide range of socioeconomic factors (income, employment, education, modernization) and other related health, nutritional and demographic measures.

A multistage, random cluster process was used to draw the sample surveyed in each of the provinces. Counties in the nine provinces were stratified by income (low, middle, high) and a weighted sampling scheme was used to randomly select four counties in each province. In addition, the provincial capital and a lower-income city were selected. Villages and townships within the counties, and urban and suburban neighbourhoods within the cities, were selected randomly. The 190 primary sampling units consisted of 32 urban neighbourhoods, 30 suburban neighbourhoods, 32 towns and 96 villages. In 1989-1993 there were 190 primary sampling units and about 3800 households in the overall survey, covering some 16000 individuals.

All household members in 1993 provided individual data on dietary intake, body fat distribution, blood pressure, medical history and healthrelated behaviour (e.g. smoking, beverage consumption, medication, key chronic diseases). The following risk factor variables were available.

- Underweight ( $n=702$ children)

Weight-for-age $z$-scores, calculated using the EpiNut program, ranged from -5.08 to 9.98 (mean -0.313 ). Using a cut-off $z$-score of $\leq-2.0$, $12.4 \%$ of children were malnourished.

- Water and sanitation $(n=3422)$

Data on how households obtained water and type of toilet facilities were used to assign $63.9 \%$ of households as having "improved water and sanitation".

- Indoor air pollution $(n=3412)$

Data on fuel used for cooking was used to assign $81.4 \%$ of households as using "smoky" fuels (coal, wood, sticks, straw and charcoal).

- Tobacco ( $n=8617$ )

A total of $31.9 \%$ of subjects were assigned as smokers.

- Alcohol ( $n=8659$ )

A total of $34.5 \%$ of subjects were assigned as alcohol drinkers.

- Body weight ( $n=4465$ )


## Calculation of income

Questions on income and time allocation look for any possible activity that each person might have engaged in during the previous year, both inside and outside the formal market. Full income from market and nonmarket activities can be imputed. The variable we used for income was the aggregated deflated total per capita income variable. We considered this a better option than creating a new equivalized household income variable. Also, it was not exactly clear how many people there were in each household to allow for equivalization. Some exploratory analyses (not reported) suggested that the results for an equivalized household income variable were very similar to the deflated per capita income variable. Estimated deflated per capita household income was used to generate income rankings.

The association between income and the various risk factors was determined by categorizing each risk factor into a dichotomous variable. The indirect method, using locally weighted sum of squares regression in the same manner as described below for DHS asset scores and LSMS income data, was then used to estimate the relationship between the normalized rank of income and the dichotomous risk factor variable. The income-poverty cut-offs were then superimposed to estimate the prevalence of each risk factor within each income-poverty category. A total of $18.5 \%$ of the Chinese population were assumed to be living on <US\$ 1 per day, $35.2 \%$ on US $\$ 1-2$ per day and $46.3 \%$ on $>$ US $\$ 2$ per day.

Other methods adopted for specific risk factors

## Ambient air pollution

Asset scores were used to estimate income distribution in urban and rural populations. For this purpose, the urban-rural variable in the DHS data was reclassified as follows: 0 or $1=$ urban; 2 or $3=$ rural.

We assumed that the distribution of asset scores in the DHS data was representative of the distribution in the relevant subregion. The approximate income distribution of urban and rural populations was estimated based on subregional asset-score cut-offs as before. Because the estimates of urban and rural income were approximate, we used only two income categories: <US $\$ 1$ per day and $>$ US $\$ 1$ per day.

We assumed that rural populations (proportions c and d in Table 24.16) were uniformly exposed to a particulate level of $5 \mu \mathrm{~g} / \mathrm{m}^{3}$ in all subregions. We further assumed that, within subregions, the exposure of urban populations (proportions a and b in Table 24.16) was independent of income. Average urban air pollution exposures by subregion were obtained from chapter 17. The exposure estimates were weighted accord-

Table 24.16 Proportion of population within urban/rural category

|  | Income |  |
| :--- | :---: | :---: |
|  | <US\$I per day | $>$ US\$ I per day |
| Urban | a | b |
| Rural | c | d |

ing to the proportions in Table 24.16. For example, the average exposure in populations living on $<\mathrm{US} \$ 1$ per day $=$ ( a xexposure in urban areas) + (cxexposure in rural areas).

## Body weight

The only body weight (body mass index) data we had available were for women in the DHS data set with children aged $<5$ years. Thus our analyses were restricted to this group.

## Systematic literature reviews

We conducted eight literature reviews focusing on the developing world for tobacco, alcohol, illicit drugs, hypertension, cholesterol, body weight, physical inactivity and lead (see section 4). We also conducted some less rigorous literature reviews for malnutrition, water and sanitation and unsafe sex (see Appendix C).

An initial literature search was conducted using Medline. Papers were limited to those published in English between 1990 and 2001. Four search domains were specified, as shown in Table 24.17. Search domains 1-3 were common to all, while domain 4 was risk-factor-specific.

There was substantial heterogeneity among the retrieved papers with regard to:

- subregions included;
- socioeconomic factor(s) used;
- measurement of the risk factor (e.g. not many of the protein malnutrition papers used weight-for-age [underweight]; they used rather height-for-age [stunting] or weight-for-height [wasting]);
- time period (a particular problem if the prevalence of poverty or the patterning of risk factor [e.g. tobacco] by socioeconomic position was rapidly changing during the 1980s and 1990s);
- study population sampling (e.g. many studies were conducted among poor communities, not as random population samples);
- study population demographics (e.g. limited to certain age groups); and

Table 24.17 Search domains and Medline search strategy

| Search domain | Medline search strategy |
| :--- | :--- |
| I. Socioeconomic | Poverty.af or socioeconomic:.af or income:.af <br> position or poverty <br> MeSH headings <br> Socioeconomic factors (exploded to include subheadings: <br> career mobility, poverty, poverty areas, social class, social mobility) |
| 2. Survey or |  |
| quantitative data | Prevalence.af or survey:.af or review.af <br> MeSH headings |
| Data collection, health surveys, health status indicators, nutrition <br> surveys, diet surveys, population surveillance, sentinel surveillance, <br> nutrition assessment, questionnaires, records (exploded), <br> registries, vital statistics, interviews and meta-analysis |  |
| 3. Developing countries | Developing countr:.af or third-world.af <br> MeSH headings |
| Africa (exploded), Caribbean region (exploded), Central America |  |
| (exploded), Latin America (exploded), South America (exploded), |  |
| Asia (exploded), Pacific Islands, Melanesia (exploded), Micronesia |  |
| (exploded) or Polynesia (exploded) |  |
| Where possible, we used the same Medline search strategy as |  |
| that used for other chapters of this publication |  |

a The suffix ".af" in the Medline search strategy refers to "all fields".

- reporting of results (e.g. only odds ratios reported with no prevalence data, only multivariate results reported, or only $P$ values reported with no effect estimate).

For the more rigorous literature reviews, additional searches were performed as follows.

The Cochrane Library was searched for relevant systematic reviews. While Cochrane reviews have been shown to be of higher quality and to be less biased on average than other systematic reviews, they nevertheless have limitations. Other systematic reviews were searched for in the Database of Abstracts of Reviews of Effectiveness and by the recommended methods for review identification (Glanville and Lefebvre 2000).

Medline searches were conducted for the period 1980 to February 2002. The key search terms for the topics included: smoking, tobacco, alcohol, illicit, cannabis, marijuana, opiate, amphetamine, cocaine, inhalant, hypertension, blood pressure, cholesterol, obesity, body weight, inactivity, exercise, lead and poisoning, and blood lead. The key search terms for socioeconomic status included: poverty, inequality and socioeconomic factors. Where more detailed searches were undertaken, additional search terms included: social class, education, occupation and income.

Other references were identified from the reference lists of key review articles and from a WHO key informant.

## 3. Quantitative results

### 3.1 Underweight

Underweight in children was defined as a weight-for-age $z$-score $<-2$, using the NCHS or Harvard reference populations (see chapter 2). For the developing world, excluding WPR-B, underweight information was ascertained using DHS data sets. For WPR-B, for which we had no applicable DHS data set, underweight prevalence information was obtained using the CHNS data set.

As shown in Table 24.18, there were reasonable numbers of children available for analysis in all subregions except EUR-C and WPR-B. Thus, results for those two subregions should be treated with caution.

Locally linear kernel regression smooth plots ("Loess plots") of underweight according to weighted asset score ranking at the subregional level are shown in Figure 24.12. Each sub-figure plots the proportion of malnourished children (y-axis ranging from 0 to 1 ) by asset score rank for the subregion (x-axis ranging from 0 [poorest] to 1 [least poor in subregion]). The centre point of each plotted circle depicts the prevalence of underweight children at that given asset score rank, and the size of each circle is proportional to the number of observations. The middle line is the fitted curve and the upper and lower lines are the $95 \%$ confidence bands. (More weight is given to those observations with larger numbers as indicated by a larger circle size.)

From Figure 24.12 it is clear that children living in households with low asset scores (and by inference also living in absolute income poverty) were substantially more likely to be underweight in all subregions. This is consistent with expectation and the findings of Wagstaff and

Table 24.18 Sample sizes for child underweight by subregion

| Subregion | Low weight-for-age (z-score <-2) |  | Weight-for-age (z-score >-2) |  |
| :---: | :---: | :---: | :---: | :---: |
|  | $N$ | Percentage | N | Percentage |
| AFR-D | 10142 | 34 | 19723 | 66 |
| AFR-E | 9417 | 29 | 23551 | 71 |
| AMR-B | 683 | 6 | 11333 | 94 |
| AMR-D | 1900 | 10 | 17499 | 90 |
| EMR-B | 175 | 10 | 1498 | 90 |
| EMR-D | 1707 | 24 | 5391 | 76 |
| EUR-B | 238 | 14 | 1467 | 86 |
| EUR-C | 13 | 3 | 427 | 97 |
| SEAR-B | 365 | 21 | 1333 | 79 |
| SEAR-D | 17323 | 49 | 17905 | 51 |
| WPR-B | 321 | 14.2 | 1943 | 85.8 |

Figure 24.12 Loess plots of the prevalence of underweight children ( $y$ axis) by normalized rank of asset score (scale $0-I$, x-axis)


Watanabe (2001). Other than the overall differences in prevalence of underweight by subregion, the only other notable difference between subregions was a prevalence of underweight that did not vary greatly among those with lowest ranked asset scores for EMR-D and SEAR-D, but then dropped off more rapidly among those with higher ranked asset scores (i.e. a "shoulder effect"). The EMR-D analyses were based on only two countries-Morocco and Pakistan. With nearly five times the population of Morocco, Pakistan dominates the empirical estimates for EMR-D. An inspection of the Pakistan-only Loess plot (not shown) demonstrates a shoulder effect similar to that for EMR-D in Figure 24.12, although not as marked. Morocco, with only $2 \%$ of its population living on <US\$ 1 per day, had much lower levels of underweight children than Pakistan, further exaggerating the apparent shoulder effect. Using only the Pakistan data, we would have estimated relative risks of 2.3 and 1.8 for those living on <US $\$ 1$ per day and on US $\$ 1-2$ per day, respectively, compared to >US\$2 per day. The overall relative risk estimates for EMR-D were both 1.7 (Table 24.19). The sensitivity analysis presented below also suggests that the results for EMR-D set out in Table 24.19 should be treated with considerable caution.

The SEAR-D analyses were based on three countries-Bangladesh, India and Nepal. India makes up about $87 \%$ of the population of these three countries combined, and therefore dominates the SEAR-D estimates. The Loess plot for India (not shown) also demonstrates a shoulder effect. Some $86 \%$ of the Indian population lives on $<$ US $\$ 2$ per day ( $44 \%$ on <US $\$ 1$ per day). It appears that the steeply declining levels of underweight at the right-hand end of the SEAR-D plot in Figure 24.12 are due to the non-poor (and often financially quite well-off) Indian population living on $>$ US $\$ 2$ per day.

Inspection of the remaining country-level Loess plots (not shown) did not disclose any other countries with a distinctly different pattern of underweight by asset score rank. Consequently, we conclude that the remaining subregional-level plots in Figure 24.12 provide an approximate summary of the country-level associations.

Based on the smoothed curves in Figure 24.12 and a subregional-level analysis of asset scores by underweight, we estimated the relative risks of underweight by level of absolute poverty, as shown in the final column of Table 24.19. For example, we estimated that children in AFR-D living on $<$ US $\$ 1$ per day were 2.3 times more likely to be underweight than children living on >US $\$ 2$ per day. Likewise, children living on US \$1-2 per day were estimated to be 1.4 times more likely to be underweight, and children living on exactly US $\$ 2$ per day were estimated to be 1.2 times more likely to be underweight. Using these first two relative risk estimates and the fixed marginal prevalences of absolute poverty and underweight (shown in italics to the right of the two by three tables), we calculated the values for each of the cells within the two by three tables. Finally, the middle column of Table 24.19 simply presents the prevalence

Table 24.19 Cell prevalence of weight-for-age by poverty, prevalence of underweight by poverty, and relative risk of underweight by poverty, by subregion, for children aged $0-5$ years for a US\$2 per day cut-off

| Subregion | Weight-for-age |  |  |  | Prevalence of underweight by poverty (\%) |  | Relative risk of underweight |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | <-2 SD | >-2 SD |  |  |  |  |  |
| AFR-D | <US\$ 1/day | 22.5 | 33.0 | 55.5 | <US\$ I/day | 40.5 | <US\$ I/day | 2.3 |
|  | US\$ I-2/day | 5.8 | 16.6 | 22.4 | US\$ I-2/day | 25.7 | US\$ I-2/day | 1.4 |
|  | >US\$ 2/day | 4.0 | 18.1 | 22.1 | >US\$ 2/day | 17.9 | US\$ 2/day | 1.2 |
|  |  | 32.2 | 67.8 | 100.0 |  |  | >US\$ 2/day | 1 |
| AFR-E | <US\$ I/day | 12.6 | 14.7 | 27.3 | <US\$ I/day | 46.1 | <US\$ I/day | 2.6 |
|  | US\$ I-2/day | 11.9 | 24.3 | 36.2 | US\$ I-2/day | 32.8 | US\$ I-2/day | 1.8 |
|  | >US\$ 2/day | 6.5 | 30.0 | 36.5 | >US\$ 2/day | 17.9 | US\$ $2 /$ day | 1.4 |
|  |  | 31.0 | 69.0 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-B | <US\$ I/day | 1.0 | 10.0 | 11.0 | <US\$ I/day | 9.3 | <US\$ I/day | 2.4 |
|  | US\$ I-2/day | 1.3 | 17.8 | 19.1 | US\$ I-2/day | 7.0 | US\$ I-2/day | 1.8 |
|  | >US\$ 2/day | 2.6 | 67.2 | 69.8 | >US\$ 2/day | 3.8 | US\$ 2/day | 1.6 |
|  |  | 5.0 | 95.0 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-D | <US\$ I/day | 4.5 | 12.9 | 17.4 | <US\$ I/day | 25.8 | <US\$ I/day | 3.7 |
|  | US\$ I-2/day | 3.9 | 22.4 | 26.3 | US\$ I-2/day | 15.0 | US\$ I-2/day | 2.1 |
|  | >US\$ 2/day | 4.0 | 52.3 | 56.3 | >US\$ 2/day | 7.0 | US\$ 2/day | 1.6 |
|  |  | 12.4 | 87.6 | 100.0 |  |  | >US\$ 2/day | 1 |
| EMR-B | <US\$ I/day | 0.3 | 1.7 | 2.0 | <US\$ I/day | 15.3 | <US\$ I/day | 2.1 |
|  | US\$ I-2/day | 1.0 | 6.1 | 7.1 | US\$ I-2/day | 14.3 | US\$ I-2/day | 1.9 |
|  | >US\$ 2/day | 6.8 | 84.1 | 90.9 | >US\$ 2/day | 7.5 | US\$ 2/day | 1.8 |
|  |  | 8.1 | 91.9 | 100.0 |  |  | >US\$ 2/day | 1 |
| EMR-D ${ }^{\text {a }}$ | <US\$ I/day | 5.7 | 13.6 | 19.3 | <US\$ I/day | 29.6 | <US\$ I/day | 1.7 |
|  | US\$ I-2/day | 13.4 | 31.9 | 45.3 | US\$ I-2/day | 29.5 | US\$ I-2/day | 1.7 |
|  | >US\$ 2/day | 6.0 | 29.3 | 35.3 | >US\$ 2/day | 17.0 | US\$ 2/day | 1.5 |
|  |  | 25.1 | 74.9 | 100.0 |  |  | >US\$ 2/day | 1 |
| EUR-B | <US\$ I/day | 0.4 | 2.6 | 3.0 | <US\$ I/day | 13.1 | <US\$ I/day | 1.9 |
|  | US\$ I-2/day | 1.6 | 13.0 | 14.6 | US\$ I-2/day | 10.8 | US\$ I-2/day | 1.6 |
|  | >US\$ 2/day | 5.6 | 76.7 | 82.3 | >US\$ 2/day | 6.8 | US\$ $2 /$ day | 1.3 |
|  |  | 7.6 | 92.4 | 100.0 |  |  | >US\$ 2/day | 1 |
| EUR-C | <US\$ I/day | 0.3 | 5.1 | 5.4 | <US\$ I/day | 4.8 | <US\$ I/day | 2.4 |
|  | US\$ I-2/day | 0.9 | 20.4 | 21.3 | US\$ I-2/day | 4.2 | US\$ I-2/day | 2.1 |
|  | >US\$ 2/day | 1.5 | 71.8 | 73.3 | >US\$ $2 /$ day | 2.0 | US\$ 2/day | 1.8 |
|  |  | 2.6 | 97.4 | 100.0 |  |  | >US\$ 2/day | 1 |
| SEAR-B | <US\$ I/day | 3.2 | 3.2 | 6.4 | <US\$ I/day | 50.4 | <US\$ I/day | 3.3 |
|  | US\$ I-2/day | 14.7 | 27.8 | 42.5 | US\$ I-2/day | 34.5 | US\$ I-2/day | 2.2 |
|  | >US\$ 2/day | 7.9 | 43.1 | 51.0 | >US\$ 2/day | 15.5 | US\$ $2 /$ day | 1.7 |
|  |  | 25.8 | 74.2 | 100.0 |  |  | >US\$ 2/day | 1 |
| SEAR-D | <US\$ I/day | 22.9 | 19.5 | 42.4 | <US\$ I/day | 54.1 | <US\$ I/day | 2.1 |
|  | US\$ I-2/day | 19.1 | 23.7 | 42.8 | US\$ I-2/day | 44.7 | US\$ I-2/day | 1.7 |
|  | >US\$ 2/day | 3.9 | 10.9 | 14.8 | >US\$ $2 /$ day | 26.1 | US\$ $2 /$ day | 1.3 |
|  |  | 45.9 | 54.1 | 100.0 |  |  | >US\$ 2/day | 1 |

Table 24.19 Cell prevalence of weight-for-age by poverty, prevalence of underweight by poverty, and relative risk of underweight by poverty, by subregion, for children aged $0-5$ years for a US\$2 per day cut-off (continued)

| Subregion | Weight-for-age |  |  |  | Prevalence of underweight by poverty (\%) |  | Relative risk of underweight |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | $<-2$ SD | >-2 SD |  |  |  |  |  |
| WPR-B | <US\$ I/day | 2.7 | 15.2 | 17.9 | <US\$ I/day | 15.1 | <US\$ I/day | 1.1 |
|  | US\$ I-2/day | 6.6 | 27.4 | 34.0 | US\$ I-2/day | 19.4 | US\$ I-2/day | 1.4 |
|  | >US\$ $2 /$ day | 6.7 | 41.4 | 48.1 | >US\$ $2 /$ day | 13.9 | US\$ $2 /$ day | 1.3 |
|  |  | 16.0 | 84.0 | 100.0 |  |  | >US\$ 2/day | 1 |
| Total ${ }^{\text {b }}$ | <US\$ 1/day | 10.6 | 15.5 | 26.1 | <US\$ 1/day | 40.5 | <US\$ I/day | 3.1 |
|  | US\$ I-2/day | 10.4 | 23.5 | 33.9 | US\$ I-2/day | 30.6 | US\$ I-2/day | 2.4 |
|  | >US\$ $2 /$ day | 5.2 | 34.8 | 40.0 | >US\$ $2 /$ day | 13.0 | US\$ $2 /$ day | - |
|  |  | 26.1 | 73.9 | 100.0 |  |  | >US\$ $2 /$ day | I |

[^92]of underweight children within strata (or "rows") of poverty level. For example, the prevalence of underweight children among those living on $<$ US\$ 1 per day in AFR-D was 22.5/55.5 = 40.5 \% .

The results in Table 24.19 demonstrate a reasonably consistent association of underweight with absolute poverty across subregions. Aggregating the 11 subregions to give a total developing world summary, we estimated that children living on $<$ US $\$ 1$ per day were 3.1 times more likely to be malnourished than children living on $>$ US $\$ 2$ per day. Table 24.20 shows the calculations for a cut-off of US\$ 1 per day.

## SENSITIVITY ANALYSIS

An alternative way of estimating the association of poverty and DHS risk factors for subregions was to conduct analyses for each country (i.e. to apply the indirect method at the country level to estimate the joint distribution of income poverty and risk factors), then aggregate up to subregions. The advantage of this method is that it allows for instances in which a given asset score does not equate to a similar income poverty level in different countries. The main disadvantages are that for many risk factors country-level estimates of the association of asset score and risk factors are unstable, and that countries without a World Bank estimate of income poverty could not be included. In terms of the latter, we had to exclude six countries from AFR-D (Benin, Cameroon, Chad, the Comoros, Guinea and Togo), Malawi from AFR-E, Nicaragua from AMR-D and Kyrgyzstan from EUR-B from country-level analyses. (See section 2 for discussion on the relative advantages of the country-level and subregional-level alternatives.) The country-level estimates, together

Table 24.20 Cell prevalence of weight-for-age by poverty, prevalence of underweight by poverty, and relative risk of underweight by poverty, by subregion, for children aged $0-5$ years for a US\$ I per day cut-off

| Subregion | Weight-for-age |  |  |  | Prevalence of underweight by poverty (\%) |  | Relative risk of underweight |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | $<-2$ SD | $>-2 S D$ |  |  |  |  |  |
| AFR-D | <US\$ I/day | 22.2 | 33.3 | 55.5 | <US\$ I/day | 39.9 | <US\$ I/day | 1.8 |
|  | >US\$ I/day | 10.0 | 34.5 | 44.5 | >US\$ I/day | 22.6 | US\$ I/day | 1.3 |
|  |  | 32.2 | 67.8 | 100.0 |  |  | >US\$ I/day | 1 |
| AFR-E | <US\$ I/day | 12.3 | 15.0 | 27.3 | <US\$ 1/day | 44.9 | <US\$ I/day | 1.7 |
|  | >US\$ I/day | 18.7 | 54.0 | 72.7 | >US\$ 1/day | 25.8 | US\$ I/day | 1.5 |
|  |  | 31.0 | 69.0 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-B | <US\$ I/day | 1.0 | 10.0 | 11.0 | <US\$ 1/day | 9.3 | <US\$ I/day | 2.1 |
|  | >US\$ I/day | 4.0 | 85.0 | 89.0 | >US\$ 1/day | 4.5 | US\$ I/day | 1.7 |
|  |  | 5.0 | 95.0 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-D | <US\$ I/day | 4.5 | 12.9 | 17.4 | <US\$ 1/day | 25.6 | <US\$ I/day | 2.7 |
|  | >US\$ I/day | 7.9 | 74.7 | 82.6 | >US\$ I/day | 9.6 | US\$ I/day | 2.0 |
|  |  | 12.4 | 87.6 | 100.0 |  |  | >US\$ I/day | 1 |
| EMR-B | <US\$ I/day | 0.3 | 1.7 | 2.0 | <US\$ I/day | 15.2 | <US\$ I/day | 1.9 |
|  | >\$1/day | 7.8 | 90.2 | 98.0 | >\$1/day | 8.0 | US\$ I/day | 1.9 |
|  |  | 8.1 | 91.9 | 100.0 |  |  | >US\$ I/day | 1 |
| EMR-D ${ }^{\text {a }}$ | <US\$ I/day | 5.7 | 13.6 | 19.3 | <US\$ 1/day | 29.5 | <US\$ I/day | 1.2 |
|  | >US\$ I/day | 19.4 | 61.3 | 80.7 | >US\$ I/day | 24.0 | US\$ I/day | 1.2 |
|  |  | 25.1 | 74.9 | 100.0 |  |  | >US\$ I/day | 1 |
| EUR-B | <US\$ I/day | 0.4 | 2.6 | 3.0 | <US\$ 1/day | 12.5 | <US\$ I/day | 1.7 |
|  | >US\$ I/day | 7.2 | 89.8 | 97.0 | >US\$ I/day | 7.4 | US\$ I/day | 1.6 |
|  |  | 7.6 | 92.4 | 100.0 |  |  | >US\$ I/day | 1 |
| EUR-C | <US\$ I/day | 0.3 | 5.1 | 5.4 | <US\$ 1/day | 5.1 | <US\$ I/day | 2.1 |
|  | >US\$ I/day | 2.3 | 92.3 | 94.6 | >US\$ I/day | 2.5 | US\$ I/day | 2.1 |
|  |  | 2.6 | 97.4 | 100.0 |  |  | >US\$ I/day | 1 |
| SEAR-B | <US\$ I/day | 3.2 | 3.2 | 6.4 | <US\$ 1/day | 50.2 | <US\$ I/day | 2.1 |
|  | >US\$ I/day | 22.6 | 71.0 | 93.6 | >US\$ 1/day | 24.1 | US\$ I/day | 2.0 |
|  |  | 25.8 | 74.2 | 100.0 |  |  | >US\$ I/day | 1 |
| SEAR-D | <US\$ I/day | 22.9 | 19.5 | 42.4 | <US\$ 1/day | 54.1 | <US\$ I/day | 1.4 |
|  | >US\$ I/day | 23.0 | 34.6 | 57.6 | >US\$ 1/day | 39.9 | US\$ I/day | 1.3 |
|  |  | 45.9 | 54.1 | 100.0 |  |  | >US\$ I/day | 1 |
| WPR-B | <US\$ I/day | 2.7 | 15.2 | 17.9 | <US\$ 1/day | 15.0 | <US\$ I/day | 0.9 |
|  | >US\$ I/day | 13.3 | 68.8 | 82.1 | >US\$ I/day | 16.2 | US\$ I/day | 1.0 |
|  |  | 16.0 | 84.0 | 100.0 |  |  | >US\$ I/day | 1 |
| Total ${ }^{\text {b }}$ | <US\$ I/day | 10.5 | 15.6 | 26.1 | <US\$ I/day | 40.3 | <US\$ I/day | 1.9 |
|  | >US\$ I/day | 15.6 | 58.3 | 73.9 | >US\$ 1/day | 21.1 | US\$ I/day | - |
|  |  | 26.1 | 73.9 | 100.0 |  |  | >US\$ I/day | 1 |

[^93]Table 24.2 I Country- and subregional-level relative risk estimates

| Subregion | $<U S \$ 1 \quad c .>U S \$ 2$ | US $\$ I-2 \quad c f .>U S \$ 2$ |
| :--- | :---: | :---: |
| AFR-D | $1.3(1.4)$ | $1.7(2.3)$ |
| AFR-E | $1.8(1.8)$ | $2.3(2.6)$ |
| AMR-B | $3.0(2.4)$ | $2.1(1.8)$ |
| AMR-D | $4.2(3.7)$ | $2.6(2.1)$ |
| EMR-D | $3.6(1.7)$ | $2.8(1.7)$ |
| EUR-B | $2.0(1.9)$ | $1.7(1.6)$ |
| EUR-C | $2.3(2.4)$ | $2.1(1.9)$ |
| SEAR-D | $2.0(2.1)$ | $1.6(1.7)$ |

with the subregional-level estimates from Table 24.20 (in parentheses), are shown in Table 24.21.

The country and subregional-level estimates vary notably for AFR-D, but any conclusion as to which method is more valid is difficult owing to the absence of six countries from the country-level sensitivity analysis. The other notable disagreement was for EMR-D (Morocco and Pakistan only). Closer scrutiny revealed a low estimate of absolute poverty in Morocco compared to Pakistan ( $7.5 \%$ and $85 \%$ <US $\$ 2$ per day, respectively), yet the asset scores in Morocco and Pakistan largely overlapped, and the prevalence of underweight children was much less in Morocco than in Pakistan. Correspondingly, the subregional-level relative risk estimates of 1.7 and 1.7 should be treated with some caution. More generally, the subregional-level estimates for EMR-D for water and sanitation, and overweight should also be treated with caution. For other subregions there was insufficient reason to prefer the country-level to the subregional-level estimates presented in sections 3.1, 3.2, 3.3 and 3.9 using DHS data. However, the findings for EMR-D illustrate the importance of scale. Relationships between income, asset score and risk factors may vary, not only between countries but also within national populations.

### 3.2 Unsafe water and poor sanitation

Water and sanitation availability was defined as a two-level categorical variable:

- no water supply and/or no sanitation (categories ${ }^{3} \mathrm{Va}, \mathrm{Vb}$ and VI in Prüss et al. 2001); and
- improved water and improved sanitation (category IV in Prüss et al. 2001).

Table 24.22 shows the actual samples sizes for the water and sanitation analyses. CHNS data were used in WPR-B (see Appendix B).

Table 24.22 Sample sizes for water and sanitation by subregion

| Subregion | No water and/or no sanitation |  | Improved water and sanitation |  |
| :---: | :---: | :---: | :---: | :---: |
|  | n | Percentage | n | Percentage |
| AFR-D | 49323 | 55 | 39968 | 45 |
| AFR-E | 43815 | 53 | 38781 | 47 |
| AMR-B | 6554 | 14 | 40314 | 86 |
| AMR-D | 16422 | 32 | 35647 | 68 |
| EMR-B | 1154 | 28 | 3030 | 72 |
| EMR-D | 13217 | 33 | 26650 | 67 |
| EUR-B | 826 | 10 | 7336 | 90 |
| EUR-C | 74 | 2 | 4472 | 98 |
| SEAR-B | 9346 | 74 | 3339 | 26 |
| SEAR-D | 67293 | 64 | 38647 | 36 |
| WPR-B | 2185 | 64 | 1237 | 36 |

Smoothed scatter-plots of water and sanitation according to household asset score rankings showed increasing availability of water and sanitation with increasing asset score, despite considerable variation in the shape of the curves (Figure 24.13). Conversely, households living in absolute poverty were more likely to lack water and/or sanitation.

Table 24.23 presents the relative risks derived from the smoothed plots in Figure 24.13 and the two by three tables. Note that the data in Table 24.23 are expressed for unimproved water and sanitation, i.e. 1 category IV, where category IV is according to Prüss et al. (2001). The strength of the association was even stronger than that for childhood underweight, with relative risks for people living on $<$ US $\$ 1$ per day compared to people living on $>$ US $\$ 2$ per day ranging from 2.0 to 15.1. Given this wide range of relative risks, the aggregated global results should be treated with caution. Full tabular results for a US\$ 1 per day cut-off are presented in Table 24.24.

### 3.3 Unsafe sex

To derive estimates of unsafe sex, we estimated the prevalence of sex with a non-marital partner (variable UN1 on DHS data) and, of those having sex with a non-marital partner, the proportion using a condom (variable UN2 on DHS data).

| UNI Higher-risk sex in <br> the last year | All who had sex in the last year <br> (denominator) | Sex with non-marital partner <br> in the last year (numerator) |
| :--- | :--- | :--- |
| UN2 Condom use at | All who had higher-risk sex in the | People who used condom at <br> latest higher-risk sex <br> last year and were asked about <br> condom use at the latest time <br> (denominator) | | latest higher-risk sex |
| :--- |
| (numerator) |

Figure 24.13 Loess plots of the prevalence of improved water and sanitation ( $y$-axis) by normalized rank of asset score (x-axis)


Table 24.23 Cell prevalence of unimproved water supply by poverty, prevalence of unimproved water supply by poverty, and relative risk of unimproved water supply by poverty, by subregion, for a US\$ 2 per day cut-off

| Subregion | Unimproved water supply |  |  |  | Prevalence of unimproved water supply (\%) |  | Relative risk of unimproved water supply |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ 1/day | 39.3 | 16.2 | 55.5 | <US\$ I/day | 70.8 | <US\$ I/day | 9.4 |
|  | US\$ I-2/day | 4.9 | 17.5 | 22.4 | US\$ I-2/day | 21.8 | US\$ I-2/day | 2.9 |
|  | >US\$ 2/day | 1.7 | 20.4 | 22.1 | >US\$ 2/day | 7.6 | US\$ $2 /$ day | 1.6 |
|  |  | 45.8 | 54.2 | 100.0 |  |  | >US\$ 2/day | 1 |
| AFR-E | <US\$ I/day | 25.1 | 2.2 | 27.3 | <US\$ I/day | 92.1 | <US\$ I/day | 4.6 |
|  | US\$ I-2/day | 25.2 | 11.0 | 36.2 | US\$ I-2/day | 69.6 | US\$ I-2/day | 3.5 |
|  | >US\$ 2/day | 7.3 | 29.2 | 36.5 | >US\$ 2/day | 20.0 | US\$ 2/day | 2.7 |
|  |  | 57.6 | 42.4 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-B | <US\$ I/day | 10.4 | 0.6 | 11.0 | <US\$ I/day | 94.8 | <US\$ I/day | 12.3 |
|  | US\$ I-2/day | 8.4 | 10.7 | 19.1 | US\$ I-2/day | 44.2 | US\$ I-2/day | 5.7 |
|  | >US\$ 2/day | 5.4 | 64.4 | 69.8 | >US\$ 2/day | 7.7 | US\$ 2/day | 3.4 |
|  |  | 24.2 | 75.8 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-D | <US\$ I/day | 15.6 | 1.8 | 17.4 | <US\$ I/day | 89.9 | <US\$ I/day | 8.9 |
|  | US\$ I-2/day | 10.7 | 15.6 | 26.3 | US\$ I-2/day | 40.8 | US\$ I-2/day | 4.0 |
|  | >US\$ $2 /$ day | 5.7 | 50.6 | 56.3 | >US\$ $2 /$ day | 10.1 | US\$ $2 /$ day | 2.3 |
|  |  | 32.1 | 67.9 | 100.0 |  |  | >US\$ 2/day | I |
| EMR-B | <US\$ I/day | 1.0 | 1.0 | 2.0 | <US\$ I/day | 48.8 | <US\$ I/day | 3.6 |
|  | US\$ I-2/day | 3.1 | 4.0 | 7.1 | US\$ I-2/day | 43.6 | US\$ I-2/day | 3.2 |
|  | >US\$ 2/day | 12.5 | 78.4 | 90.9 | >US\$ 2/day | 13.7 | US\$ 2/day | 2.9 |
|  |  | 16.6 | 83.4 | 100.0 |  |  | >US\$ 2/day | 1 |
| EMR-D ${ }^{\text {a }}$ | <US\$ I/day | 15.0 | 4.3 | 19.3 | <US\$ I/day | 77.9 | <US\$ I/day | 15.1 |
|  | US\$ I-2/day | 16.9 | 28.4 | 45.3 | US\$ I-2/day | 37.2 | US\$ I-2/day | 7.2 |
|  | >US\$ 2/day | 1.8 | 33.5 | 35.3 | >US\$ 2/day | 5.2 | US\$ 2/day | 3.2 |
|  |  | 33.7 | 66.3 | 100.0 |  |  | >US\$ 2/day | 1 |
| EUR-B | <US\$ I/day | 1.4 | 1.6 | 3.0 | <US\$ I/day | 47.8 | <US\$ I/day | 3.1 |
|  | US\$ I-2/day | 6.0 | 8.6 | 14.6 | US\$ I-2/day | 41.3 | US\$ I-2/day | 2.7 |
|  | >US\$ 2/day | 12.7 | 69.6 | 82.3 | >US\$ 2/day | 15.4 | US\$ 2/day | 2.3 |
|  |  | 20.2 | 79.8 | 100.0 |  |  | >US\$ 2/day | I |
| EUR-C | <US\$ I/day | 1.2 | 4.2 | 5.4 | <US\$ I/day | 22.0 | <US\$ I/day | 11.8 |
|  | US\$ I-2/day | 3.4 | 17.9 | 21.3 | US\$ I-2/day | 15.8 | US\$ I-2/day | 8.4 |
|  | >US\$ 2/day | 1.4 | 71.9 | 73.3 | >US\$ 2/day | 1.9 | US\$ $2 /$ day | 5.4 |
|  |  | 5.9 | 94.1 | 100.0 |  |  | >US\$ 2/day | 1 |
| SEAR-B | <US\$ I/day | 2.8 | 3.6 | 6.4 | <US\$ I/day | 43.3 | <US\$ I/day | 2.0 |
|  | US\$ I-2/day | 15.9 | 26.6 | 42.5 | US\$ I-2/day | 37.5 | US\$ I-2/day | 1.7 |
|  | >US\$ 2/day | 11.3 | 39.7 | 51.0 | >US\$ 2/day | 22.1 | US\$ $2 /$ day | 1.4 |
|  |  | 30.0 | 70.0 | 100.0 |  |  | >US\$ 2/day | 1 |
| SEAR-D | <US\$ I/day | 34.2 | 8.2 | 42.4 | <US\$ I/day | 80.7 | <US\$ I/day | 5.0 |
|  | US\$ I-2/day | 28.3 | 14.5 | 42.8 | US\$ I-2/day | 66.0 | US\$ I-2/day | 4.1 |
|  | >US\$ 2/day | 2.4 | 12.4 | 14.8 | >US\$ 2/day | 16.0 | US\$ 2/day | 2.1 |
|  |  | 64.9 | 35.2 | 100.0 |  |  | >US\$ 2/day | I |

Table 24.23 Cell prevalence of unimproved water supply by poverty, prevalence of unimproved water supply by poverty, and relative risk of unimproved water supply by poverty, by subregion, for a US\$2 per day cut-off (continued)

| Subregion | Unimproved water supply |  |  |  | Prevalence of unimproved water supply (\%) |  | Relative risk of unimproved water supply |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| WPR-B | <US\$ I/day | 14.0 | 3.9 | 17.9 | <US\$ I/day | 78.1 | <US\$ I/day | 1.7 |
|  | US\$ I-2/day | 21.5 | 12.5 | 34.0 | US\$ I-2/day | 63.2 | US\$ I-2/day | 1.3 |
|  | >US\$ 2/day | 22.6 | 25.5 | 48.1 | >US\$ 2/day | 47.1 | US\$ 2/day | 1.0 |
|  |  | 58.1 | 41.9 | 100.0 |  |  | >US\$ 2/day | 1 |
| Total ${ }^{\text {b }}$ | <US\$ I/day | 18.6 | 5.1 | 23.7 | <US\$ I/day | 78.7 | <US\$ I/day | 3.3 |
|  | US\$ I-2/day | 18.5 | 14.8 | 33.4 | US\$ I-2/day | 55.5 | US\$ I-2/day | 2.4 |
|  | >US\$ 2/day | 10.1 | 32.8 | 42.9 | >US\$ 2/day | 23.6 | US\$ $2 /$ day | - |
|  |  | 47.3 | 52.7 | 100.0 |  |  | >US\$ 2/day | 1 |

a The results for EMR-D should be interpreted with caution.
b Summary of the subregions in the table, including the subregional-level estimates for EMR-D.

For males, the relatively low sample sizes meant that it was possible to perform locally linear kernel regression directly on the unit-level (binary) data using logistic regression (see Figure 24.14). Thus, in Figure 24.14 (unlike the other figures of smooth plots) there are no "bubbles", since there was no need to use aggregate data. For females, the same aggregated procedure was used as for the other risk factors in this chapter, using the prevalence of each risk factor at the discrete values of the asset score. From the smoothed curves, we estimated the distribution of UN1 and UN2 within categories of income and by subregion, as before. Table 24.25 gives sample sizes by subregion.

Sample sizes were inadequate for several subregions. In EUR-B and EUR-C, adequate data were available only for women and detailed results for these subregions are not presented. Sex with a non-marital partner and condom use among those having sex with a non-marital partner were each less common among poor males and females within subregions (see Figures 24.14-24.16).

The prevalence of higher-risk (non-marital) sex by poverty for a US\$2 per day cut-off is shown in Tables 24.26 (males) and 24.27 (females). The prevalence of condom use during higher-risk sex by poverty for a US $\$ 2$ per day cut-off is shown in Tables 24.28 (males) and 24.29 (females). The corresponding values for a US\$1 per day cut-off are given in Tables 24.30-24.33.

Table 24.24 Cell prevalence of unimproved water supply by poverty, prevalence of unimproved water supply by poverty, and relative risk of unimproved water supply by poverty, by subregion, for a US\$ I per day cut-off


[^94]Figure 24.14 Loess plots of the prevalence of higher-risk sex (y-axis) by normalized rank of asset score (x-axis): men
AFR-D

Figure 24.15 Loess plots of the prevalence of higher-risk sex (y-axis) by normalized rank of asset score (x-axis): women


Table 24.25 Sample sizes for unsafe sex analyses by subregion

| Subregion | No non-marital sex |  | Non-marital sex |  | Those having non-marital sex who: |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | did not use a condom | used a condom |  |
|  | $n$ | Percentage |  |  | $n$ | Percentage | n | Percentage | n | Percentage |
| Males |  |  |  |  |  |  |  |  |
| AFR-D | 13036 | 73 | 4820 | 27 | 2995 | 67 | 1487 | 33 |
| AFR-E | 4043 | 51 | 3897 | 49 | 1883 | 64 | 1045 | 36 |
| AMR-B | 220 | 26 | 638 | 74 | 265 | 42 | 373 | 58 |
| AMR-D | 2360 | 54 | 1999 | 46 | 1193 | 60 | 792 | 40 |
| EUR-C | 147 | 48 | 159 | 52 | 60 | 38 | 99 | 62 |
| Females |  |  |  |  |  |  |  |  |
| AFR-D | 48735 | 82 | 11050 | 18 | 7250 | 84 | 1350 | 16 |
| AFR-E | 27056 | 69 | 12131 | 31 | 6873 | 83 | 1401 | 17 |
| AMR-B | 13720 | 68 | 6385 | 32 | 3950 | 77 | 1154 | 23 |
| AMR-D | 33841 | 88 | 4472 | 12 | 3605 | 90 | 382 | 10 |
| EUR-B | 2297 | 92 | 197 | 8 |  |  |  |  |
| EUR-C | 1205 | 67 | 598 | 33 | 408 | 79 | 107 | 21 |

Figure 24.16 Loess plots of the prevalence of higher-risk sex (with condom use) ( $y$-axis) by normalized rank of asset score (x-axis): women


Table 24.26 Cell prevalence of non-marital sex in the last year (UNI) by poverty, prevalence of non-marital sex by level of income poverty and relative risk of non-marital sex among males, by subregion, for a US\$ 2 per day cut-off

| Subregion | Non-marital sex |  |  |  | Prevalence of non-marital sex by poverty (\%) |  | Relative risk of non-marital sex |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 16.6 | 38.9 | 55.5 | <US\$ I/day | 29.9 | <US\$ I/day | 0.6 |
|  | US\$ I-2/day | 10.3 | 12.1 | 22.4 | US\$ I-2/day | 45.8 | US\$ I-2/day | 0.9 |
|  | >US\$ 2/day | 11.8 | 10.3 | 22.1 | >US\$ 2/day | 53.6 | US\$ $2 /$ day | 0.9 |
|  |  | 38.7 | 61.3 | 100.0 |  |  | >US\$ 2/day | I |
| AFR-E | <US\$ I/day | 8.6 | 18.7 | 27.3 | <US\$ I/day | 31.4 | <US\$ I/day | 0.9 |
|  | US\$ I-2/day | 14.1 | 22.1 | 36.2 | US\$ I-2/day | 38.9 | US\$ I-2/day | 1.1 |
|  | >US\$ 2/day | 13.2 | 23.3 | 36.5 | >US\$ 2/day | 36.0 | US\$ 2/day | 1.0 |
|  |  | 35.8 | 64.2 | 100.0 |  |  | >US\$ 2/day | I |
| AMR-B | <US\$ I/day | 3.7 | 7.3 | 11.0 | <US\$ I/day | 33.8 | <US\$ I/day | 0.7 |
|  | US\$ I-2/day | 7.8 | 11.3 | 19.1 | US\$ I-2/day | 41.0 | US\$ I-2/day | 0.9 |
|  | >US\$ 2/day | 32.9 | 36.9 | 69.8 | >US\$ 2/day | 47.1 | US\$ 2/day | 1.0 |
|  |  | 44.4 | 55.6 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-D | <US\$ I/day | 5.5 | 11.9 | 17.4 | <US\$ I/day | 31.6 | <US\$ I/day | 0.7 |
|  | US\$ I-2/day | 11.1 | 15.2 | 26.3 | US\$ I-2/day | 42.3 | US\$ I-2/day | 0.9 |
|  | >US\$ 2/day | 26.6 | 29.7 | 56.3 | >US\$ 2/day | 47.3 | US\$ 2/day | 1.0 |
|  |  | 43.3 | 56.7 | 100.0 |  |  | >US\$ 2/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ I/day | 8.2 | 18.2 | 26.4 | <US\$ I/day | 31.1 | <US\$ I/day | 0.7 |
|  | US\$ I-2/day | 10.3 | 14.7 | 25.0 | US\$ I-2/day | 41.3 | US\$ I-2/day | 0.9 |
|  | >US\$ 2/day | 22.1 | 26.4 | 48.6 | >US\$ 2/day | 45.6 | US\$ $2 /$ day | - |
|  |  | 40.7 | 59.3 | 100.0 |  |  | >US\$ 2/day | 1 |

[^95]Table 24.27 Cell prevalence of non-marital sex in the last year (UNI) by poverty, prevalence of non-marital sex by level of income poverty and relative risk of non-marital sex among females, by subregion, for a US\$ 2 per day cut-off

| Subregion | Non-marital sex |  |  |  | Prevalence of non-marital sex by poverty (\%) |  | Relative risk of non-marital sex |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ 1/day | 5.5 | 50.0 | 55.5 | <US\$ I/day | 9.9 | <US\$ 1/day | 0.4 |
|  | US\$ I-2/day | 4.3 | 18.1 | 22.4 | US\$ I-2/day | 19.0 | US\$ I-2/day | 0.9 |
|  | >US\$ 2/day | 4.9 | 17.2 | 22.1 | >US\$ 2/day | 22.2 | US\$ $2 /$ day | 0.9 |
|  |  | 14.6 | 85.4 | 100.0 |  |  | >US\$ 2/day | 1 |
| AFR-E | <US\$ I/day | 3.2 | 24.1 | 27.3 | <US\$ I/day | 11.6 | <US\$ I/day | 0.6 |
|  | US\$ I-2/day | 5.6 | 30.6 | 36.2 | US\$ I-2/day | 15.3 | US\$ I-2/day | 0.8 |
|  | >US\$ 2/day | 6.9 | 29.6 | 36.5 | >US\$ 2/day | 19.0 | US\$ 2/day | 1.0 |
|  |  | 15.7 | 84.3 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-B | <US\$ I/day | 1.8 | 9.2 | 11.0 | <US\$ I/day | 16.1 | <US\$ I/day | 0.8 |
|  | US\$ I-2/day | 2.4 | 16.7 | 19.1 | US\$ I-2/day | 12.6 | US\$ I-2/day | 0.6 |
|  | >US\$ 2/day | 13.8 | 56.0 | 69.8 | >US\$ 2/day | 19.7 | US\$ 2/day | 0.7 |
|  |  | 18.0 | 82.0 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-D | <US\$ I/day | 0.5 | 16.9 | 17.4 | <US\$ 1/day | 2.9 | <US\$ 1/day | 0.7 |
|  | US\$ I-2/day | 1.1 | 25.2 | 26.3 | US\$ I-2/day | 4.2 | US\$ I-2/day | 1.0 |
|  | >US\$ 2/day | 2.3 | 54.0 | 56.3 | >US\$ 2/day | 4.1 | US\$ $2 /$ day | 1.0 |
|  |  | 3.9 | 96.1 | 100.0 |  |  | >US\$ 2/day | 1 |
| EUR-C | <US\$ I/day | 0.3 | 5.1 | 5.4 | <US\$ I/day | 5.7 | <US\$ 1/day | 0.4 |
|  | US\$ I-2/day | 1.4 | 19.9 | 21.3 | US\$ I-2/day | 6.7 | US\$ I-2/day | 0.4 |
|  | >US\$ 2/day | 11.3 | 62.0 | 73.3 | >US\$ 2/day | 15.4 | US\$ $2 /$ day | 0.5 |
|  |  | 13.0 | 87.0 | 100.0 |  |  | >US\$ 2/day | 11 |
| Total ${ }^{\text {a }}$ | <US\$ I/day | 2.3 | 19.0 | 21.4 | <US\$ 1/day | 10.8 | <US\$ 1/day | 0.6 |
|  | US\$ I-2/day | 3.1 | 21.0 | 24.1 | US\$ I-2/day | 12.8 | US\$ I-2/day | 0.7 |
|  | >US\$ 2/day | 9.6 | 44.9 | 54.6 | >US\$ 2/day | 17.6 | US\$ $2 /$ day | - |
|  |  | 15.0 | 85.0 | 100.0 |  |  | >US\$ 2/day | 1 |

[^96]Table 24.28 Cell prevalence of condom use with higher-risk sex (UN2) by poverty, prevalence of condom use with higher-risk sex by level of income poverty and relative risk of condom use with higher-risk sex among males, by subregion, for a US\$2 per day cut-off

| Subregion | Condom use |  |  |  | Prevalence of condom use by poverty (\%) |  | Relative risk of condom use |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 14.2 | 41.3 | 55.5 | <US\$ I/day | 25.5 | <US\$ I/day | 0.5 |
|  | US\$ I-2/day | 9.6 | 12.8 | 22.4 | US\$ I-2/day | 42.9 | US\$ I-2/day | 0.9 |
|  | >US\$ 2/day | 10.4 | 11.7 | 22.1 | >US\$ 2/day | 46.9 | US\$ 2/day | 1.0 |
|  |  | 34.1 | 65.9 | 100.0 |  |  | >US\$ 2/day | I |
| AFR-E | <US\$ I/day | 4.6 | 22.7 | 27.3 | <US\$ I/day | 17.0 | <US\$ I/day | 0.3 |
|  | US\$ I-2/day | 13.1 | 23.1 | 36.2 | US\$ I-2/day | 36.1 | US\$ I-2/day | 0.7 |
|  | >US\$ 2/day | 19.3 | 17.2 | 36.5 | >US\$ 2/day | 52.9 | US\$ $2 /$ day | 0.8 |
|  |  | 37.0 | 63.0 | 100.0 |  |  | >US\$ 2/day | I |
| AMR-B | <US\$ I/day | 4.3 | 6.7 | 11.0 | <US\$ I/day | 38.9 | <US\$ I/day | 0.6 |
|  | US\$ I-2/day | 9.4 | 9.7 | 19.1 | US\$ I-2/day | 49.5 | US\$ I-2/day | 0.8 |
|  | >US\$ 2/day | 41.8 | 28.0 | 69.8 | >US\$ 2/day | 59.9 | US\$ 2/day | 0.9 |
|  |  | 55.6 | 44.4 | 100.0 |  |  | >US\$ 2/day | I |
| AMR-D | <US\$ I/day | 3.4 | 14.0 | 17.4 | <US\$ I/day | 19.3 | <US\$ I/day | 0.4 |
|  | US\$ I-2/day | 9.5 | 16.8 | 26.3 | US\$ I-2/day | 36.3 | US\$ I-2/day | 0.7 |
|  | >US\$ 2/day | 28.5 | 27.8 | 56.3 | >US\$ $2 /$ day | 50.7 | US\$ $2 /$ day | 0.8 |
|  |  | 41.5 | 58.5 | 100.0 |  |  | >US\$ 2/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ 1/day | 6.7 | 19.8 | 26.4 | <US\$ 1/day | 25.2 | <US\$ I/day | 0.4 |
|  | US\$ I-2/day | 10.5 | 14.5 | 25.0 | US\$ I-2/day | 41.9 | US\$ I-2/day | 0.7 |
|  | >US\$ 2/day | 27.4 | 21.2 | 48.6 | >US\$ 2/day | 56.4 | US\$ 2/day | - |
|  |  | 44.5 | 55.5 | 100.0 |  |  | >US\$ 2/day | I |

[^97]Table 24.29 Cell prevalence of condom use with higher-risk sex (UN2) by poverty, prevalence of condom use with higher-risk sex by level of income poverty and relative risk of condom use with higher-risk sex among females, by subregion, for a US\$2 per day cut-off

| Subregion | Condom use |  |  |  | Prevalence of condom use by poverty (\%) |  | Relative risk of condom use |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 5.1 | 50.4 | 55.5 | <US\$ I/day | 9.3 | <US\$ I/day | 0.3 |
|  | US\$ I-2/day | 5.0 | 17.4 | 22.4 | US\$ I-2/day | 22.4 | US\$ I-2/day | 0.7 |
|  | >US\$ 2/day | 7.0 | 15.1 | 22.1 | >US\$ 2/day | 31.5 | US\$ 2/day | 0.9 |
|  |  | 17.1 | 82.9 | 100.0 |  |  | >US\$ 2/day | 1 |
| AFR-E | <US\$ I/day | 0.8 | 26.5 | 27.3 | <US\$ I/day | 3.1 | <US\$ I/day | 0.1 |
|  | US\$ I-2/day | 5.3 | 30.9 | 36.2 | US\$ I-2/day | 14.7 | US\$ I-2/day | 0.4 |
|  | >US\$ 2/day | 12.7 | 23.8 | 36.5 | >US\$ 2/day | 34.8 | US\$ 2/day | 0.6 |
|  |  | 18.8 | 81.2 | 100.0 |  |  | >US\$ 2/day | I |
| AMR-B | <US\$ I/day | 1.4 | 9.6 | 11.0 | <US\$ I/day | 12.6 | <US\$ I/day | 0.3 |
|  | US\$ I-2/day | 3.6 | 15.5 | 19.1 | US\$ I-2/day | 18.9 | US\$ I-2/day | 0.4 |
|  | >US\$ 2/day | 29.7 | 40.1 | 69.8 | >US\$ 2/day | 42.5 | US\$ 2/day | 0.6 |
|  |  | 34.7 | 65.3 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-D | <US\$ I/day | 2.6 | 14.8 | 17.4 | <US\$ I/day | 15.1 | <US\$ I/day | 0.3 |
|  | US\$ I-2/day | 5.2 | 21.1 | 26.3 | US\$ I-2/day | 19.8 | US\$ I-2/day | 0.3 |
|  | >US\$ 2/day | 33.9 | 22.4 | 56.3 | >US\$ 2/day | 60.1 | US\$ 2/day | 0.5 |
|  |  | 41.7 | 58.3 | 100.0 |  |  | >US\$ 2/day | 1 |
| EUR-C | <US\$ I/day | 1.2 | 4.2 | 5.4 | <US\$ I/day | 21.7 | <US\$ I/day | 1.0 |
|  | US\$ I-2/day | 4.7 | 16.6 | 21.3 | US\$ I-2/day | 22.2 | US\$ I-2/day | 1.0 |
|  | >US\$ 2/day | 16.6 | 56.7 | 73.3 | >US\$ 2/day | 22.6 | US\$ 2/day | 1.0 |
|  |  | 22.5 | 77.5 | 100.0 |  |  | >US\$ 2/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ 1/day | 2.0 | 19.4 | 21.4 | <US\$ I/day | 9.2 | <US\$ I/day | 0.3 |
|  | US\$ I-2/day | 4.6 | 19.5 | 24.1 | US\$ I-2/day | 18.9 | US\$ I-2/day | 0.5 |
|  | >US\$ 2/day | 19.2 | 35.4 | 54.6 | >US\$ 2/day | 35.2 | US\$ 2/day | - |
|  |  | 25.7 | 74.3 | 100.0 |  |  | >US\$ 2/day | 1 |

[^98]Table 24.30 Cell prevalence of non-marital sex in the last year (UNI) by poverty, prevalence of non-marital sex by level of income poverty and relative risk of non-marital sex among males, by subregion, for a US\$ I per day cut-off

| Subregion | Non-marital sex |  |  |  | Prevalence of non-marital sex by poverty (\%) |  | Relative risk of non-marital sex |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 16.8 | 38.7 | 55.5 | <US\$ I/day | 30.3 | <US\$ I/day | 0.6 |
|  | >US\$ I/day | 21.9 | 22.6 | 44.5 | >US\$ I/day | 49.2 | US\$ I/day | 0.9 |
|  |  | 38.7 | 61.3 | 100.0 |  |  | >US\$ I/day | 1 |
| AFR-E | <US\$ 1/day | 8.5 | 18.8 | 27.3 | <US\$ I/day | 31.3 | <US\$ I/day | 0.8 |
|  | >US\$ I/day | 27.3 | 45.4 | 72.7 | >US\$ I/day | 37.5 | US\$ I/day | 1.1 |
|  |  | 35.8 | 64.2 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-B | <US\$ I/day | 3.7 | 7.3 | 11.0 | <US\$ I/day | 33.8 | <US\$ I/day | 0.7 |
|  | >US\$ I/day | 40.7 | 48.3 | 89.0 | >US\$ I/day | 45.7 | US\$ I/day | 0.8 |
|  |  | 44.4 | 55.6 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-D | <US\$ I/day | 5.5 | 11.9 | 17.4 | <US\$ I/day | 31.6 | <US\$ I/day | 0.7 |
|  | >US\$ I/day | 37.8 | 44.8 | 82.6 | >US\$ I/day | 45.7 | US\$ I/day | 0.8 |
|  |  | 43.3 | 56.7 | 100.0 |  |  | >US\$ I/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ I/day | 8.3 | 18.2 | 26.4 | <US\$ I/day | 31.3 | <US\$ I/day | 0.7 |
|  | >US\$ I/day | 32.4 | 41.2 | 73.6 | >US\$ I/day | 44.1 | US\$ I/day | - |
|  |  | 40.7 | 59.3 | 100.0 |  |  | >US\$ I/day | I |

a Summary of the subregions in the table.

Table 24.3I Cell prevalence of non-marital sex in the last year (UNI) by poverty, prevalence of non-marital sex by level of income poverty and relative risk of non-marital sex among females, by subregion, for a US\$ I per day cut-off

| Subregion | Non-marital sex |  |  |  | Prevalence of non-marital sex by poverty (\%) |  | Relative risk of non-marital sex |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 5.5 | 50.0 | 55.5 | <US\$ I/day | 10.0 | <US\$ I/day | 0.5 |
|  | >US\$ I/day | 9.1 | 35.4 | 44.5 | >US\$ I/day | 20.4 | US\$ I/day | 0.9 |
|  |  | 14.6 | 85.4 | 100.0 |  |  | >US\$ I/day | 1 |
| AFR-E | <US\$ I/day | 3.2 | 24.1 | 27.3 | <US\$ I/day | 11.7 | <US\$ I/day | 0.7 |
|  | >US\$ I/day | 12.5 | 60.2 | 72.7 | >US\$ I/day | 17.1 | US\$ I/day | 0.8 |
|  |  | 15.7 | 84.3 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-B | <US\$ I/day | 1.8 | 9.2 | 11.0 | <US\$ I/day | 16.1 | <US\$ I/day | 0.9 |
|  | >US\$ I/day | 16.2 | 72.8 | 89.0 | >US\$ I/day | 18.2 | US\$ I/day | 0.8 |
|  |  | 18.0 | 82.0 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-D | <US\$ I/day | 0.5 | 16.9 | 17.4 | <US\$ I/day | 2.9 | <US\$ I/day | 0.7 |
|  | >US\$ I/day | 3.4 | 79.2 | 82.6 | >US\$ I/day | 4.1 | US\$ I/day | 0.9 |
|  |  | 3.9 | 96.1 | 100.0 |  |  | >US\$ I/day | 1 |
| EUR-C | <US\$ I/day | 0.3 | 5.1 | 5.4 | <US\$ I/day | 5.4 | <US\$ I/day | 0.4 |
|  | >US\$ I/day | 12.7 | 81.9 | 94.6 | >US\$ I/day | 13.4 | US\$ I/day | 0.4 |
|  |  | 13.0 | 87.0 | 100.0 |  |  | >US\$ I/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ I/day | 2.3 | 19.0 | 21.4 | <US\$ I/day | 10.9 | <US\$ I/day | 0.7 |
|  | >US\$ I/day | 12.7 | 65.9 | 78.6 | >US\$ I/day | 16.1 | US\$ I/day | - |
|  |  | 15.0 | 85.0 | 100.0 |  |  | >US\$ I/day | 1 |

[^99]Table 24.32
Cell prevalence of condom use with higher-risk sex (UN2) by poverty, prevalence of condom use with higher-risk sex by level of income poverty and relative risk of condom use with higher-risk sex among males, by subregion, for a US\$ I per day cut-off

| Subregion | Condom use |  |  |  | Prevalence of condom use by poverty (\%) |  | Relative risk of condom use |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ 1/day | 14.3 | 41.2 | 55.5 | <US\$ I/day | 25.7 | <US\$ I/day | 0.6 |
|  | >US\$ I/day | 19.9 | 24.6 | 44.5 | >US\$ I/day | 44.7 | US\$ I/day | 0.9 |
|  |  | 34.1 | 65.9 | 100.0 |  |  | >US\$ I/day | I |
| AFR-E | <US\$ 1/day | 4.8 | 22.5 | 27.3 | <US\$ I/day | 17.4 | <US\$ I/day | 0.4 |
|  | >US\$ I/day | 32.3 | 40.4 | 72.7 | >US\$ I/day | 44.4 | US\$ I/day | 0.7 |
|  |  | 37.0 | 63.0 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-B | <US\$ I/day | 4.3 | 6.7 | 11.0 | <US\$ I/day | 38.9 | <US\$ I/day | 0.7 |
|  | >US\$ I/day | 51.3 | 37.7 | 89.0 | >US\$ I/day | 57.6 | US\$ I/day | 0.7 |
|  |  | 55.6 | 44.4 | 100.0 |  |  | >US\$ I/day | I |
| AMR-D | <US\$ I/day | 3.4 | 14.0 | 17.4 | <US\$ I/day | 19.4 | <US\$ I/day | 0.4 |
|  | >US\$ I/day | 38.1 | 44.5 | 82.6 | >US\$ I/day | 46.1 | US\$ I/day | 0.6 |
|  |  | 41.5 | 58.5 | 100.0 |  |  | >US\$ I/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ 1/day | 6.7 | 19.7 | 26.4 | <US\$ I/day | 25.5 | <US\$ I/day | 0.5 |
|  | >US\$ 1/day | 37.8 | 35.8 | 73.6 | >US\$ I/day | 51.4 | US\$ I/day | - |
|  |  | 44.5 | 55.5 | 100.0 |  |  | >US\$ I/day | 1 |

a Summary of the subregions in the table.

Table 24.33 Cell prevalence of condom use with higher-risk sex (UN2) by poverty, prevalence of condom use with higher-risk sex by level of income poverty and relative risk of condom use with higher-risk sex among females, by subregion, for a US\$ I per day cut-off

| Subregion | Condom use |  |  |  | Prevalence of condom use by poverty (\%) |  | Relative risk of condom use |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 5.3 | 50.2 | 55.5 | <US\$ I/day | 9.6 | <US\$ I/day | 0.4 |
|  | >US\$ I/day | 11.8 | 32.7 | 44.5 | >US\$ I/day | 26.5 | US\$ I/day | 0.7 |
|  |  | 17.1 | 82.9 | 100.0 |  |  | >US\$ I/day | 1 |
| AFR-E | <US\$ I/day | 0.9 | 26.4 | 27.3 | <US\$ I/day | 3.3 | <US\$ I/day | 0.1 |
|  | >US\$ I/day | 17.9 | 54.8 | 72.7 | >US\$ I/day | 24.7 | US\$ I/day | 0.3 |
|  |  | 18.8 | 81.2 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-B | <US\$ I/day | 1.4 | 9.6 | 11.0 | <US\$ I/day | 12.6 | <US\$ I/day | 0.3 |
|  | >US\$ I/day | 33.3 | 55.7 | 89.0 | >US\$ I/day | 37.4 | US\$ I/day | 0.4 |
|  |  | 34.7 | 65.3 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-D | <US\$ I/day | 2.7 | 14.7 | 17.4 | <US\$ I/day | 15.3 | <US\$ I/day | 0.3 |
|  | >US\$ I/day | 39.0 | 43.6 | 82.6 | >US\$ I/day | 47.3 | US\$ I/day | 0.4 |
|  |  | 41.7 | 58.3 | 100.0 |  |  | >US\$ I/day | 1 |
| EUR-C | <US\$ I/day | 1.2 | 4.2 | 5.4 | <US\$ I/day | 21.7 | <US\$ I/day | 1.0 |
|  | >US\$ I/day | 21.3 | 73.3 | 94.6 | >US\$ I/day | 22.5 | US\$ I/day | 1.0 |
|  |  | 22.5 | 77.5 | 100.0 |  |  | >US\$ I/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ I/day | 2.0 | 19.4 | 21.4 | <US\$ I/day | 9.4 | <US\$ I/day | 0.3 |
|  | >US\$ I/day | 23.7 | 54.9 | 78.6 | >US\$ I/day | 30.1 | US\$ I/day | - |
|  |  | 25.7 | 74.3 | 100.0 |  |  | >US\$ I/day | I |

[^100]
### 3.4 Indoor air pollution

LSMS data for 10 countries were used for the indoor air pollution analyses (see Table 24.14). Supplementary data were also used for three DHS countries-Colombia (AMR-B), Indonesia (SEAR-B) and India (SEARD). Two other demographic and health surveys (Ethiopia plus Zimbabwe combined and Panama) had some indoor air pollution data, but could not be used owing to inadequate sample sizes among the non-poor. The combined sample sizes and distribution are shown in Table 24.34.

We used estimates of indoor air pollution unadjusted for ventilation. The survey data questions we used to determine the joint association were compatible with these unadjusted estimates.

Locally linear kernel regression smooth plots of smoke-producing cooking fuel use according to normalized equivalized income ranking are shown in Figure 24.17. Each subfigure plots the proportion of households using smoke-producing cooking fuels ( $y$-axis ranging from 0 to 1 ) by normalized equivalized income rank for the country ( x -axis ranging from 0 [poorest] to 1 [richest in the country]).

Table 24.35 shows the estimated association of absolute poverty and indoor air pollution for a cut-off of US\$ 2 per day, using the area under the curves in Figure 24.17. There is considerable variation between subregions in the average level of exposure to indoor air pollution and in the relative differences within subregions by poverty. In both African subregions, and to a lesser extent in both South-East Asian subregions, there is both an extraordinarily high prevalence of exposure to indoor air

Table 24.34 Sample sizes and distribution of household use of smokeproducing cooking fuels by subregion

| Subregion | Presence of smoke-producing cooking fuels |  | Absence of smoke-producing cooking fuels |  |
| :---: | :---: | :---: | :---: | :---: |
|  | $n$ | Percentage | n | Percentage |
| AFR-D | 5585 | 94 | 332 | 6 |
| AFR-E | 4287 | 42 | 5933 | 58 |
| AMR-B | 1282 | 26 | 3647 | 74 |
| AMR-D | 812 | 14 | 4805 | 86 |
| EMR-B | - | - | - | - |
| EMR-D | 13349 | 71 | 5561 | 29 |
| EUR-B | 2724 | 43 | 3634 | 57 |
| EUR-C | 276 | 8 | 3337 | 92 |
| SEAR-B | 12705 | 44 | 16105 | 56 |
| SEAR-D | 67631 | 77 | 20078 | 23 |
| WPR-B | 2778 | 81 | 634 | 19 |

- No data.

Table 24.35 Cell prevalence of indoor air pollution by poverty, prevalence of indoor air pollution by poverty and relative risk of indoor air pollution by poverty, by subregion, for a US\$2 per day cut-off

| Subregion | Exposure |  |  |  | Prevalence of indoor air pollution by poverty (\%) |  | Relative risk of indoor air pollution |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 42.1 | 13.4 | 55.5 | <US\$ 1/day | 75.9 | <US\$ I/day | 1.1 |
|  | US\$ I-2/day | 16.6 | 5.8 | 22.4 | US\$ I-2/day | 74.0 | US\$ I-2/day | 1.1 |
|  | >US\$ 2/day | 14.7 | 7.4 | 22.1 | >US\$ $2 /$ day | 66.4 | US\$ 2/day | 1.1 |
|  |  | 73.4 | 26.6 | 100.0 |  |  | >US\$ 2/day | 1.0 |
| AFR-E ${ }^{\text {a }}$ | <US\$ I/day | 27.3 | 0.0 | 27.3 | <US\$ 1/day | 100.0 | <US\$ I/day | 2.0 |
|  | US\$ I-2/day | 36.2 | 0.0 | 36.2 | US\$ I-2/day | 100.0 | US\$ I-2/day | 1.9 |
|  | >US\$ 2/day | 19.4 | 17.1 | 36.5 | >US\$ 2/day | 53.3 | US\$ 2/day | 1.7 |
|  |  | 85.8 | 14.2 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-B | <US\$ I/day | 8.2 | 2.8 | 11.0 | <US\$ I/day | 74.6 | <US\$ I/day | 7.2 |
|  | US\$ I-2/day | 9.2 | 9.9 | 19.1 | US\$ I-2/day | 48.1 | US\$ I-2/day | 4.7 |
|  | >US\$ 2/day | 7.2 | 62.6 | 69.8 | >US\$ 2/day | 10.3 | US\$ 2/day | 3.3 |
|  |  | 24.6 | 75.4 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-D ${ }^{\text {a }}$ | <US\$ I/day | 17.4 | 0.0 | 17.4 | <US\$ 1/day | 100.0 | <US\$ I/day | 14.6 |
|  | US\$ I-2/day | 15.0 | 11.3 | 26.3 | US\$ I-2/day | 56.9 | US\$ I-2/day | 4.7 |
|  | >US\$ $2 /$ day | 6.9 | 49.4 | 56.3 | >US\$ 2/day | 12.2 | US\$ $2 /$ day | 2.3 |
|  |  | 52.9 | 47.1 | 100.0 |  |  | >US\$ 2/day | 1 |
| EMR-D | <US\$ I/day | 17.0 | 2.3 | 19.3 | <US\$ 1/day | 88.1 | <US\$ I/day | 4.0 |
|  | US\$ I-2/day | 30.5 | 14.8 | 45.3 | US\$ I-2/day | 67.2 | US\$ I-2/day | 3.1 |
|  | >US\$ 2/day | 7.7 | 27.6 | 35.3 | >US\$ $2 /$ day | 21.9 | US\$ $2 /$ day | 1.7 |
|  |  | 55.2 | 44.8 | 100.0 |  |  | >US\$ 2/day | 1 |
| EUR-B | <US\$ 1/day | 1.5 | 1.5 | 3.0 | <US\$ I/day | 49.1 | <US\$ I/day | 1.2 |
|  | US\$ I-2/day | 7.2 | 7.4 | 14.6 | US\$ I-2/day | 49.5 | US\$ I-2/day | 1.2 |
|  | >US\$ 2/day | 32.8 | 49.5 | 82.3 | >US\$ 2/day | 39.9 | US\$ 2/day | 1.2 |
|  |  | 41.5 | 58.5 | 100.0 |  |  | >US\$ 2/day | 1 |
| EUR-C | <US\$ 1/day | 1.5 | 3.9 | 5.4 | <US\$ 1/day | 26.9 | <US\$ I/day | 1.3 |
|  | US\$ I-2/day | 6.2 | 15.1 | 21.3 | US\$ I-2/day | 29.0 | US\$ I-2/day | 1.4 |
|  | >US\$ 2/day | 15.2 | 58.1 | 73.3 | >US\$ 2/day | 20.7 | US\$ 2/day | 1.3 |
|  |  | 22.8 | 77.2 | 100.0 |  |  | >US\$ 2/day | 1 |
| SEAR-D ${ }^{\text {a }}$ | <US\$ 1/day | 42.4 | 0.0 | 42.4 | <US\$ I/day | 100.0 | <US\$ I/day | 3.5 |
|  | US\$ I-2/day | 35.6 | 7.2 | 42.8 | US\$ I-2/day | 83.3 | US\$ I-2/day | 2.9 |
|  | >US\$ 2/day | 4.3 | 10.5 | 14.8 | >US\$ 2/day | 29.2 | US\$ 2/day | 1.8 |
|  |  | 83.5 | 16.5 | 100.0 |  |  | >US\$ 2/day | 1 |
| WPR-B | <US\$ 1/day | 16.6 | 1.3 | 17.9 | <US\$ 1/day | 92.8 | <US\$ I/day | 1.4 |
|  | US\$ I-2/day | 30.6 | 3.4 | 34.0 | US\$ I-2/day | 89.9 | US\$ I-2/day | 1.4 |
|  | >US\$ 2/day | 30.9 | 17.2 | 48.1 | >US\$ 2/day | 64.3 | US\$ $2 /$ day | 1.3 |
|  |  | 78.1 | 21.9 | 100.0 |  |  | >US\$ 2/day | 1 |
| Total ${ }^{\text {b }}$ | <US\$ 1/day | 23.5 | 1.9 | 25.4 | <US\$ I/day | 92.3 | <US\$ I/day | 2.2 |
|  | US\$ I-2/day | 26.9 | 6.7 | 33.6 | US\$ I-2/day | 80.1 | US\$ I-2/day | 1.9 |
|  | >US\$ 2/day | 17.2 | 23.9 | 41.0 | >US\$ 2/day | 41.8 | US\$ 2/day | - |
|  |  | 68.2 | 31.8 | 100.0 |  |  | >US\$ 2/day | 1 |

[^101]Figure 24.17 Loess plots of the prevalence of use of household smokeproducing cooking fuel (y-axis) by normalized income rank (x-axis)


Figure 24.17 Loess plots of the prevalence of use of household smokeproducing cooking fuel ( $y$-axis) by normalized income rank (x-axis) (continued)

pollution and little relative difference between the poor and non-poor. Analyses for SEAR-B are presented only for a cut-off of US\$ 1 per day (Table 24.36), as the data were insufficient for analyses based on a US\$2 per day cut-off.

For the 11 subregions, it was estimated that indoor air pollution was 2.2 times more likely among those living on <US\$ 1 per day than among those living on $>$ US $\$ 2$ per day. We stress that this estimate obscures considerable heterogeneity between subregions.

### 3.5 OUTDOOR AIR POLLUTION

The estimated exposures to ambient particulate air pollution $\left(\mathrm{PM}_{10}\right)$ by income category and subregion are shown in Table 24.37. The method used assumed that people living in urban areas are all exposed to a constant level of outdoor air pollution, and likewise those living in rural areas. However, the uneven distribution of income poverty between urban and rural areas allowed us to make (crude) estimates of varying

Table 24.36 Cell prevalence of indoor air pollution by poverty, prevalence of indoor air pollution by poverty and relative risk of indoor air pollution by poverty, by subregion, for a US\$ I per day cut-off

| Subregion | Indoor air pollution |  |  |  | Prevalence of indoor air pollution by poverty (\%) |  | Relative risk of indoor air pollution |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 42.0 | 13.5 | 55.5 | <US\$ I/day | 75.7 | <US\$ I/day | 1.1 |
|  | >US\$ I/day | 31.4 | 13.1 | 44.5 | >US\$ I/day | 70.6 | US\$ I/day | 1.1 |
|  |  | 73.4 | 26.6 | 100.0 |  |  | >US\$ I/day | 1 |
| AFR-E ${ }^{\text {a }}$ | <US\$ I/day | 27.3 | 0.0 | 27.3 | <US\$ I/day | 100.0 | <US\$ I/day | 1.6 |
|  | >US\$ I/day | 58.5 | 14.2 | 72.7 | >US\$ I/day | 80.5 | US\$ I/day | 1.6 |
|  |  | 85.8 | 14.2 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-B | <US\$ I/day | 8.8 | 2.2 | 11.0 | <US\$ I/day | 79.8 | <US\$ I/day | 4.5 |
|  | >US\$ I/day | 15.8 | 73.2 | 89.0 | >US\$ I/day | 17.8 | US\$ I/day | 3.8 |
|  |  | 24.6 | 75.4 | 100.0 |  |  | >US\$ I/day | I |
| AMR-D ${ }^{\text {a }}$ | <US\$ I/day | 17.4 | 0.0 | 17.4 | <US\$ I/day | 100.0 | <US\$ I/day | 5.9 |
|  | >US\$ I/day | 35.5 | 47.1 | 82.6 | >US\$ I/day | 43.0 | US\$ I/day | 3.5 |
|  |  | 52.9 | 47.1 | 100.0 |  |  | >US\$ I/day | 1 |
| EMR-D | <US\$ I/day | 14.9 | 4.4 | 19.3 | <US\$ I/day | 77.0 | <US\$ I/day | 1.5 |
|  | >US\$ I/day | 40.3 | 40.4 | 80.7 | >US\$ I/day | 50.0 | US\$ I/day | 1.5 |
|  |  | 55.2 | 44.8 | 100.0 |  |  | >US\$ I/day | I |
| EUR-B | <US\$ I/day | 1.5 | 1.5 | 3.0 | <US\$ I/day | 49.6 | <US\$ I/day | 1.2 |
|  | >US\$ I/day | 40.0 | 57.0 | 97.0 | >US\$ I/day | 41.3 | US\$ I/day | 1.2 |
|  |  | 41.5 | 58.5 | 100.0 |  |  | >US\$ I/day | I |
| EUR-C | <US\$ I/day | 1.5 | 3.9 | 5.4 | <US\$ I/day | 27.2 | <US\$ I/day | 1.2 |
|  | >US\$ I/day | 21.3 | 73.3 | 94.6 | >US\$ I/day | 22.5 | US\$ I/day | 1.3 |
|  |  | 22.8 | 77.2 | 100.0 |  |  | >US\$ I/day | 1 |
| SEAR-B ${ }^{\text {a }}$ | <US\$ I/day | 6.4 | 0.0 | 6.4 | <US\$ I/day | 100.0 | <US\$ I/day | 2.6 |
|  | >US\$ I/day | 60.1 | 33.5 | 93.6 | >US\$ I/day | 64.2 | US\$ I/day | 2.1 |
|  |  | 66.5 | 33.5 | 100.0 |  |  | >US\$ I/day | 1 |
| SEAR-D ${ }^{\text {a }}$ | <US\$ I/day | 42.4 | 0.0 | 42.4 | <US\$ I/day | 100.0 | <US\$ I/day | 1.5 |
|  | >US\$ I/day | 41.1 | 16.5 | 57.6 | >US\$ I/day | 71.4 | US\$ I/day | 1.4 |
|  |  | 83.5 | 16.5 | 100.0 |  |  | >US\$ I/day | 1 |
| WPR-B | <US\$ I/day | 16.5 | 1.4 | 17.9 | <US\$ I/day | 92.4 | <US\$ I/day | 1.2 |
|  | >US\$ I/day | 61.6 | 20.5 | 82.1 | >US\$ I/day | 75.0 | US\$ I/day | 1.2 |
|  |  | 78.1 | 21.9 | 100.0 |  |  | >US\$ I/day | 1 |
| Total ${ }^{\text {b }}$ | <US\$ I/day | 23.3 | 2.1 | 25.4 | <US\$ I/day | 91.8 | <US\$ I/day | 1.5 |
|  | >US\$ I/day | 44.9 | 29.7 | 74.6 | >US\$ I/day | 60.2 | US\$ I/day | - |
|  |  | 68.2 | 31.8 | 100.0 |  |  | >US\$ I/day | I |

[^102]Table 24.37 Exposure to ambient air pollution by poverty, by subregion, for a US\$ I per day cut-off

| Subregion | Average exposure ( $\mu \mathrm{g} / \mathrm{m}^{3}$ ) |  | Average exposure for people living on <US\$ I as a percentage of average exposure for people living on >US\$ I |
| :---: | :---: | :---: | :---: |
| AFR-D | <US\$ I/day | 9 | 28 |
|  | >US\$ I/day | 32 |  |
| AFR-E | <US\$ I/day | 5 | 27 |
|  | >US\$ I/day | 18 |  |
| AMR-B | <US\$ I/day | 6 | 13 |
|  | >US\$ I/day | 47 |  |
| AMR-D | <US\$ I/day | 6 | 10 |
|  | >US\$ I/day | 55 |  |
| EMR-B | <US\$ I/day | 5 | 13 |
|  | >US\$ I/day | 40 |  |
| EMR-D | <US\$ I/day | 5 | 11 |
|  | >US\$ I/day | 40 |  |
| EUR-B | <US\$ I/day | 5 | 27 |
|  | >US\$ I/day | 19 |  |
| EUR-C | <US\$ I/day | 5 | 21 |
|  | >US\$ I/day | 24 |  |
| SEAR-B | <US\$ I/day | 5 | 13 |
|  | >US\$ I/day | 38 |  |
| SEAR-D | <US\$ I/day | 7 | 14 |
|  | >US\$ I/day | 48 |  |
| WPR-B | <US\$ I/day | 6 | 8 |
|  | >US\$ I/day | 78 |  |

exposure to air pollution by income poverty levels. (See section 2 for further details.)

Exposure to outdoor air pollution was estimated to be substantially higher in the higher-income category (those living on >US\$ 1 per day) in all subregions. This is due to a greater proportion of poor people living in rural areas, where there is less outdoor air pollution. These summary estimates inevitably obscure considerable heterogeneity within countries and even within cities, where particular patterns of air pollution emissions and meteorological conditions may combine to produce local variations in air pollution.

### 3.6 Tobacco use

Chapter 11 uses an indirect method based on lung cancer mortality to estimate the contribution of tobacco use to the global burden of disease, and the prevalence of tobacco use by subregion was not estimated. We took estimates of subregional prevalence of tobacco use
from the web site of the WHO Tobacco or Health Programme (http://www.cdc.gov/tobacco/who/whofirst.htm).

Analyses for tobacco were based on LSMS data for 10 countries in seven subregions-AFR-D (Ghana), AFR-E (South Africa), AMR-B (Panama), AMR-D (Ecuador), EMR-D (Pakistan), EUR-B (Azerbaijan, Bulgaria, Tajikistan) and EUR-C (Kazakhstan, the Russian Federation). Data on tobacco for Bulgaria, Ghana, South Africa and Tajikistan were available only in the form of household expenditure data, and for Azerbaijan as a composite variable of combined alcohol and tobacco expenditure. A household was classified as a smoking household if any money was spent on cigarettes or tobacco. While this is a rather crude measure, it should at least give an approximation of the proportion of households containing smokers, assuming that all purchases made by the household were for household use and that these items were not received through other means. For the remaining countries, questions were asked in the surveys on whether the individual was a smoker or not. For these countries, analyses were confined to individuals aged $>15$ years. Because of the incorporation of proxy expenditure data, the quantitative results in this section should be treated cautiously and interpreted in the light of the literature review in section 4.

The sample sizes for tobacco use by subregion are shown in Table 24.38.

Locally linear kernel regression smooth plots of tobacco use according to normalized equivalized income ranking are shown in Figure 24.18. Each subfigure plots the proportion of tobacco use ( y -axis ranging from 0 to 1 ) by normalized equivalized income rank for the country ( x -axis ranging from 0 [poorest] to 1 [richest in the country]).

Table 24.38 Sample sizes for tobacco use by subregion

| Subregion | Persons using tobacco |  | Persons not using tobacco |  |
| :---: | :---: | :---: | :---: | :---: |
|  | $n$ | Percentage | n | Percentage |
| AFR-D | 557 | 9 | 5441 | 91 |
| AFR-E | 3961 | 46 | 4644 | 54 |
| AMR-B | 1310 | 9 | 12669 | 91 |
| AMR-D | 1642 | 10 | 14823 | 90 |
| EMR-B | - | - | - | - |
| EMR-D | 2954 | 16 | 15956 | 84 |
| EUR-B | 2899 | 45 | 3580 | 55 |
| EUR-C | 3679 | 36 | 6653 | 64 |
| SEAR-B | - | - | - | - |
| SEAR-D | - | - | - | - |
| WPR-B | 2747 | 32 | 5870 | 68 |

- No data.

Figure 24.18 Loess plots of the prevalence of tobacco use ( $y$-axis) by normalized income rank (x-axis)


Figure 24.18 Loess plots of the prevalence of tobacco use (y-axis) by normalized income rank (x-axis) (continued)


Full tabular results for US $\$ 2$ and US $\$ 1$ cut-offs are given in Tables 24.39 and 24.40, respectively, using the area under the curves in Figure 24.18.

Perhaps the most important result shown in Table 24.39 is the considerable variation between subregions in overall average tobacco use, but the relatively weak association of tobacco use within subregions by individual-level poverty. Thus the average prevalence of tobacco use in each subregion is a more useful predictor of individual tobacco use than absolute level of poverty.

Variation in the prevalence of tobacco use by poverty level within subregions was of secondary importance, and variable in direction. In AFRE (South Africa data only) and AMR-D (Ecuador data only) there was a suggestion of lower prevalence of tobacco use among the poorest, and in EMR-B (Pakistan only) the converse. No reliable socioeconomic patterning of tobacco use was evident for other subregions. We emphasize that these are crude measures of tobacco use, possibly obscuring

Table 24.39 Cell prevalence of tobacco use by poverty, prevalence of tobacco use by poverty and relative risk of tobacco use by poverty, by subregion, for a US\$2 per day cut-off

| Subregion | Tobacco use |  |  |  | Prevalence of tobacco use by poverty (\%) |  | Relative risk of tobacco use |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 9.2 | 46.3 | 55.5 | <US\$ I/day | 16.5 | <US\$ I/day | . 1 |
|  | US\$ I-2/day | 3.1 | 19.3 | 22.4 | US\$ I-2/day | 13.7 | US\$ I-2/day | 0.9 |
|  | >US\$ 2/day | 3.3 | 18.8 | 22.1 | >US\$ 2/day | 14.7 | US\$ 2/day | 0.9 |
|  |  | 15.5 | 84.5 | 100.0 |  |  | >US\$ 2/day | 1 |
| AFR-E | <US\$ I/day | 5.2 | 22.1 | 27.3 | <US\$ I/day | 19.2 | <US\$ I/day | 0.7 |
|  | US\$ I-2/day | 8.1 | 28.1 | 36.2 | US\$ I-2/day | 22.4 | US\$ I-2/day | 0.8 |
|  | >US\$ $2 /$ day | 9.7 | 26.8 | 36.5 | >US\$ 2/day | 26.5 | US\$ $2 /$ day | 0.9 |
|  |  | 23.0 | 77.0 | 100.0 |  |  | >US\$ 2/day | I |
| AMR-B | <US\$ I/day | 3.4 | 7.6 | 11.0 | <US\$ I/day | 31.2 | <US\$ I/day | 1.1 |
|  | US\$ I-2/day | 6.7 | 12.4 | 19.1 | US\$ I-2/day | 34.8 | US\$ I-2/day | 1.2 |
|  | >US\$ 2/day | 20.6 | 49.2 | 69.8 | >US\$ 2/day | 29.5 | US\$ 2/day | 1.2 |
|  |  | 30.7 | 69.3 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-D | <US\$ I/day | 3.3 | 14.1 | 17.4 | <US\$ I/day | 19.2 | <US\$ 1/day | 0.6 |
|  | US\$ 1-2/day | 6.2 | 20.1 | 26.3 | US\$ I-2/day | 23.6 | US\$ 1-2/day | 0.7 |
|  | >US\$ 2/day | 18.0 | 38.3 | 56.3 | >US\$ 2/day | 31.9 | US\$ $2 /$ day | 0.8 |
|  |  | 27.5 | 72.5 | 100.0 |  |  | >US\$ 2/day | 1 |
| EMR-D | <US\$ I/day | 5.6 | 13.7 | 19.3 | <US\$ I/day | 29.2 | <US\$ I/day | 1.7 |
|  | US\$ I-2/day | 11.2 | 34.1 | 45.3 | US\$ I-2/day | 24.8 | US\$ I-2/day | 1.4 |
|  | >US\$ 2/day | 6.2 | 29.1 | 35.3 | >US\$ 2/day | 17.6 | US\$ 2/day | 1.1 |
|  |  | 23.1 | 76.9 | 100.0 |  |  | >US\$ 2/day | 1 |
| EUR-B | <US\$ I/day | 0.9 | 2.1 | 3.0 | <US\$ I/day | 31.1 | <US\$ I/day | 0.8 |
|  | US\$ I-2/day | 4.2 | 10.4 | 14.6 | US\$ I-2/day | 28.6 | US\$ I-2/day | 0.8 |
|  | >US\$ 2/day | 30.2 | 52.1 | 82.3 | >US\$ 2/day | 36.7 | US\$ 2/day | 0.8 |
|  |  | 35.3 | 64.7 | 100.0 |  |  | >US\$ 2/day | 1 |
| EUR-C | <US\$ I/day | 2.0 | 3.4 | 5.4 | <US\$ I/day | 37.1 | <US\$ I/day | 1.1 |
|  | US\$ I-2/day | 7.5 | 13.8 | 21.3 | US\$ I-2/day | 35.4 | US\$ I-2/day | 1.0 |
|  | >US\$ 2/day | 25.2 | 48.1 | 73.3 | >US\$ 2/day | 34.4 | US\$ 2/day | 1.0 |
|  |  | 34.8 | 65.2 | 100.0 |  |  | >US\$ 2/day | 1 |
| WPR-B | <US\$ I/day | 6.4 | 11.5 | 17.9 | <US\$ I/day | 36.0 | <US\$ I/day | 1.0 |
|  | US\$ I-2/day | 11.7 | 22.3 | 34.0 | US\$ I-2/day | 34.3 | US\$ I-2/day | 1.0 |
|  | >US\$ 2/day | 16.7 | 31.4 | 48.1 | >US\$ 2/day | 34.6 | US\$ 2/day | 1.0 |
|  |  | 34.8 | 65.2 | 100.0 |  |  | >US\$ 2/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ I/day | 5.3 | 13.1 | 18.4 | <US\$ I/day | 28.7 | <US\$ I/day | 0.9 |
|  | US\$ I-2/day | 9.2 | 20.8 | 30.0 | US\$ I-2/day | 30.6 | US\$ I-2/day | 1.0 |
|  | >US\$ 2/day | 16.4 | 35.1 | 51.6 | >US\$ 2/day | 31.9 | US\$ 2/day | - |
|  |  | 30.9 | 69.1 | 100.0 |  |  | >US\$ 2/day | 1 |

[^103]Table 24.40 Cell prevalence of tobacco use by poverty, prevalence of tobacco use by poverty and relative risk of tobacco use by poverty, by subregion, for a US\$ I per day cut-off

| Subregion | Tobacco use |  |  |  | Prevalence of tobacco use <br> by poverty (\%) |  | Relative risk of tobacco use |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 9.2 | 46.3 | 55.5 | <US\$ I/day | 16.5 | <US\$ I/day | 1.2 |
|  | >US\$ I/day | 6.3 | 38.2 | 44.5 | >US\$ I/day | 14.2 | US\$ I/day | 1.1 |
|  |  | 15.5 | 84.5 | 100.0 |  |  | >US\$ I/day | I |
| AFR-E | <US\$ I/day | 5.1 | 22.2 | 27.3 | <US\$ I/day | 18.6 | <US\$ I/day | 0.8 |
|  | >US\$ I/day | 17.9 | 54.8 | 72.7 | >US\$ I/day | 24.6 | US\$ I/day | 0.8 |
|  |  | 23.0 | 77.0 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-B | <US\$ I/day | 3.5 | 7.5 | 11.0 | <US\$ I/day | 31.4 | <US\$ I/day | 1.0 |
|  | >US\$ I/day | 27.2 | 61.8 | 89.0 | >US\$ I/day | 30.6 | US\$ I/day | 1.1 |
|  |  | 30.7 | 69.3 | 100.0 |  |  | >US\$ I/day | I |
| AMR-D | <US\$ I/day | 3.4 | 14.0 | 17.4 | <US\$ I/day | 19.6 | <US\$ I/day | 0.7 |
|  | >US\$ I/day | 24.1 | 58.5 | 82.6 | >US\$ I/day | 29.2 | US\$ I/day | 0.8 |
|  |  | 27.5 | 72.5 | 100.0 |  |  | >US\$ I/day | I |
| EMR-D | <US\$ I/day | 5.3 | 14.0 | 19.3 | <US\$ I/day | 27.6 | <US\$ I/day | 1.3 |
|  | >US\$ I/day | 17.8 | 62.9 | 80.7 | >US\$ I/day | 22.0 | US\$ I/day | 1.2 |
|  |  | 23.1 | 76.9 | 100.0 |  |  | >US\$ I/day | I |
| EUR-B | <US\$ I/day | 0.9 | 2.1 | 3.0 | <US\$ I/day | 31.2 | <US\$ I/day | 0.9 |
|  | >US\$ I/day | 34.4 | 62.6 | 97.0 | >US\$ I/day | 35.5 | US\$ I/day | 0.9 |
|  |  | 35.3 | 64.7 | 100.0 |  |  | >US\$ I/day | I |
| EUR-C | <US\$ I/day | 2.0 | 3.4 | 5.4 | <US\$ I/day | 37.1 | <US\$ I/day | 1.1 |
|  | >US\$ I/day | 32.8 | 61.8 | 94.6 | >US\$ I/day | 34.7 | US\$ I/day | 1.1 |
|  |  | 34.8 | 65.2 | 100.0 |  |  | >US\$ I/day | I |
| WPR-B | <US\$ I/day | 6.4 | 11.5 | 17.9 | <US\$ I/day | 36.0 | <US\$ I/day | 1.0 |
|  | >US\$ I/day | 28.3 | 53.8 | 82.1 | >US\$ I/day | 34.5 | US\$ I/day | 1.0 |
|  |  | 34.8 | 65.2 | 100.0 |  |  | >US\$ I/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ I/day | 5.3 | 13.1 | 18.4 | <US\$ I/day | 28.6 | <US\$ I/day | 0.9 |
|  | >US\$ I/day | 25.7 | 55.9 | 81.6 | >US\$ I/day | 31.4 | US\$ I/day | - |
|  |  | 30.9 | 69.1 | 100.0 |  |  | >US\$ I/day | 1 |

[^104]important differences, for example between men and women or between age groups.

### 3.7 Alcohol use

Analyses for alcohol were based on LSMS data for nine countries in six subregions-AFR-D (Ghana), AFR-E (South Africa), AMR-B (Panama), AMR-D (Ecuador), EUR-B (Azerbaijan, Bulgaria, Tajikistan) and EURC (Kazakhstan, the Russian Federation). As for the data on tobacco (section 3.6), data for Bulgaria, Ghana, South Africa and Tajikistan were available only in the form of household expenditure data, and for Azerbaijan as a composite variable of combined expenditure on alcohol and tobacco. A household was classified as an alcohol-consuming household if any money was spent on alcohol. While this is a rather crude measure, it should at least give an approximation of the proportion of households containing consumers of alcohol, assuming that all purchases made by the household were for household use and that these items were not received through other means. For the remaining countries, questions were asked in the surveys on whether the individual consumed alcohol or not. For these countries, analyses were confined to individuals aged $>15$ years. As with the tobacco analyses, these quantitative results for alcohol should be treated cautiously and interpreted in the light of the literature review presented in section 4.

The sample sizes for alcohol use by subregion are shown in Table 24.41.

Locally linear kernel regression smooth plots of alcohol use according to normalized equivalized income ranking are shown in Figure 24.19.

Table 24.4 I Sample sizes for alcohol use by subregion

|  | Persons using alcohol |  |  | Persons not using alcohol |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Subregion | $n$ | 3583 |  | $n$ |  |
| AFR-D | 3314 | 39 | 2415 | Percentage |  |

- No data.

Figure 24.19 Loess plots of the prevalence of alcohol use ( $y$-axis) by normalized income rank (y-axis)


Figure 24.19 Loess plots of the prevalence of alcohol use ( $y$-axis) by normalized income rank (y-axis) (continued)


Each subfigure plots the proportion of alcohol use (y-axis ranging from 0 to 1) by normalized equivalized income rank for the country ( x -axis ranging from 0 [poorest] to 1 [richest in the country]).

Full tabular results for US $\$ 2$ and US $\$ 1$ cut-offs are given in Tables 24.42 and 24.43, respectively, based on the area under the curves in Figure 24.19. The results presented in Tables 24.42 and 24.43 demonstrate a more marked variation in overall average prevalence of alcohol use between subregions than within subregions by individual-level absolute poverty. In none of the subregions analysed was there a suggestion of increasing prevalence of alcohol use among the poorest. Also, in two subregions-AFR-E (South Africa data only) and AMR-B (Panama only)—poor people had approximately half the prevalence of alcohol use of non-poor people.

### 3.8 OvERWEIGHT (WOMEN ONLY)

For developing countries (excluding those in WPR-B) we used DHS data for maternal weight, and the body weight results reported here are

Table 24.42 Cell prevalence of alcohol use by poverty, prevalence of alcohol use by poverty and relative risk of alcohol use by poverty, by subregion, for a US\$2 per day cut-off

|  | Alcohol use |  |  |  | Prevalence of alcohol | Relative risk of |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Subregion | Yes |  |  | No |  | use by poverty $(\%)$ | alcohol use |

[^105]Table 24.43 Cell prevalence of alcohol use by poverty, prevalence of alcohol use by poverty and relative risk of alcohol use by poverty, by subregion, for a US\$ I per day cut-off

| Subregion | Alcohol use |  |  |  | Prevalence of alcohol use by poverty (\%) |  | Relative risk of alcohol use |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 19.8 | 35.7 | 55.5 | <US\$ I/day | 35.7 | <US\$ I/day | 0.9 |
|  | >US\$ I/day | 18.5 | 26.0 | 44.5 | >US\$ I/day | 41.6 | US\$ I/day | 0.9 |
|  |  | 38.3 | 61.7 | 100.0 |  |  | >US\$ I/day | 1 |
| AFR-E | <US\$ I/day | 7.3 | 20.0 | 27.3 | <US\$ I/day | 26.8 | <US\$ I/day | 0.5 |
|  | >US\$ I/day | 36.7 | 36.0 | 72.7 | >US\$ I/day | 50.5 | US\$ I/day | 0.6 |
|  |  | 44.0 | 56.0 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-B | <US\$ I/day | 3.9 | 7.1 | 11.0 | <US\$ I/day | 35.2 | <US\$ I/day | 0.5 |
|  | >US\$ I/day | 62.3 | 26.7 | 89.0 | >US\$ I/day | 70.0 | US\$ I/day | 0.6 |
|  |  | 66.2 | 33.8 | 100.0 |  |  | >US\$ I/day | 1 |
| AMR-D | <US\$ I/day | 9.6 | 7.8 | 17.4 | <US\$ I/day | 55.1 | <US\$ I/day | 0.9 |
|  | >US\$ I/day | 52.9 | 29.7 | 82.6 | >US\$ I/day | 64.1 | US\$ I/day | 0.9 |
|  |  | 62.5 | 37.5 | 100.0 |  |  | >US\$ I/day | I |
| EUR-B | <US\$ I/day | 1.5 | 1.5 | 3.0 | <US\$ I/day | 51.6 | <US\$ I/day | 0.8 |
|  | >US\$ I/day | 60.3 | 36.7 | 97.0 | >US\$ I/day | 62.2 | US\$ I/day | 0.8 |
|  |  | 61.9 | 38.1 | 100.0 |  |  | >US\$ I/day | 1 |
| EUR-C | <US\$ I/day | 3.6 | 1.8 | 5.4 | <US\$ I/day | 67.5 | <US\$ I/day | 0.8 |
|  | >US\$ I/day | 80.9 | 13.7 | 94.6 | >US\$ I/day | 85.5 | US\$ I/day | 0.8 |
|  |  | 84.5 | 15.5 | 100.0 |  |  | >US\$ I/day | 1 |
| WPR-B | <US\$ I/day | 9.2 | 8.7 | 17.9 | <US\$ I/day | 51.5 | <US\$ 1/day | 0.9 |
|  | >US\$ I/day | 48.3 | 33.8 | 82.1 | >US\$ I/day | 58.9 | US\$ I/day | 0.9 |
|  |  | 57.6 | 42.4 | 100.0 |  |  | >US\$ I/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ I/day | 8.1 | 10.3 | 18.3 | <US\$ I/day | 43.9 | <US\$ 1/day | 0.7 |
|  | >US\$ I/day | 51.0 | 30.7 | 81.7 | >US\$ I/day | 62.4 | US\$ I/day | - |
|  |  | 59.0 | 41.0 | 100.0 |  |  | >US\$ 1/day | I |

[^106]Table 24.44 Sample sizes for overweight mothers by subregion (DHS data)

| Subregion | $n$ |
| :--- | :---: |
| AFR-D | 38005 |
| AFR-E | 43880 |
| AMR-B | 15177 |
| AMR-D | 29398 |
| EMR-B | - |
| EMR-D | 3314 |
| EUR-B | 10722 |
| EUR-C | - |
| SEAR-B | - |
| SEAR-D | 8187 |
| WPR-B | 4469 |
| No data. |  |

therefore for mothers only. Overweight was defined as a body mass index (BMI) of between 25 and $29.9 \mathrm{~kg} / \mathrm{m}^{2}$. The overall proportion of overweight women within each subregion was estimated using data from chapter 8.

For WPR-B, for which we had no applicable DHS data set, prevalence information was obtained using the CHNS data set. To make the WPR-B values comparable with those obtained using the DHS data, analysis of body weight was restricted to adult women (aged $>15$ years) only. As shown in Table 24.44, there were sufficient sample sizes available for analysis in all subregions except EMR-B, EUR-C and SEAR-B. Because all analyses were limited to women with children aged $<5$ years and do not correct for pregnancy, the results must be treated with caution.

Full tabular results for US $\$ 2$ and US\$ 1 cut-offs are given in Tables 24.45 and 24.46, respectively. For both African subregions and SEARD (dominated by India) the prevalence of overweight among poor women was less than half that among non-poor women (Table 24.45). For WPR-B (dominated by China) there was no discernible socioeconomic gradient. For the remaining subregions (Central and South American countries, and the south-eastern European and central Asian countries) there was a tendency for overweight to be about $20 \%$ less common among poor women than among non-poor women.

### 3.9 Results summary - Relative risks and risk differences

## ReLative risks

Table 24.47 shows the relative risks for the prevalence of each risk factor for people living on $<$ US\$ 1, US\$ 1-2 and exactly US\$ 2 per day

Table 24.45 Cell prevalence of overweight by poverty, prevalence of overweight by poverty and relative risk of overweight by poverty, by subregion, for a US\$2 per day cut-off (women only)

| Subregion | Overweight |  |  |  | Prevalence of overweight by poverty (\%) |  | Relative risk of overweight |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 5.4 | 50.1 | 55.5 | <US\$ I/day | 9.8 | <US\$ I/day | 0.4 |
|  | US\$ I-2/day | 4.0 | 18.4 | 22.4 | US\$ I-2/day | 17.9 | US\$ I-2/day | 0.8 |
|  | >US\$ 2/day | 5.1 | 17.0 | 22.1 | >US\$ 2/day | 22.9 | US\$ 2/day | 0.9 |
|  |  | 14.5 | 85.5 | 100.0 |  |  | >US\$ 2/day | 1 |
| AFR-E | <US\$ I/day | 4.0 | 23.3 | 27.3 | <US\$ I/day | 14.8 | <US\$ I/day | 0.4 |
|  | US\$ I-2/day | 9.8 | 26.4 | 36.2 | US\$ I-2/day | 27.2 | US\$ I-2/day | 0.7 |
|  | >US\$ 2/day | 14.2 | 22.3 | 36.5 | >US\$ 2/day | 38.9 | US\$ $2 /$ day | 0.8 |
|  |  | 28.1 | 71.9 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-B | <US\$ I/day | 2.5 | 8.5 | 11.0 | <US\$ I/day | 22.3 | <US\$ I/day | 0.8 |
|  | US\$ I-2/day | 5.1 | 14.0 | 19.1 | US\$ I-2/day | 26.9 | US\$ I-2/day | 1.0 |
|  | >US\$ $2 /$ day | 19.3 | 50.5 | 69.8 | >US\$ $2 /$ day | 27.6 | US\$ 2/day | 1.0 |
|  |  | 26.9 | 73.1 | 100.0 |  |  | >US\$ 2/day | 1 |
| AMR-D | <US\$ I/day | 4.5 | 12.9 | 17.4 | <US\$ I/day | 26.0 | <US\$ I/day | 0.8 |
|  | US\$ I-2/day | 7.9 | 18.4 | 26.3 | US\$ I-2/day | 29.9 | US\$ I-2/day | 0.9 |
|  | >US\$ 2/day | 17.8 | 38.5 | 56.3 | >US\$ 2/day | 31.7 | US\$ $2 /$ day | 1.0 |
|  |  | 30.2 | 69.8 | 100.0 |  |  | >US\$ 2/day | 1 |
| EMR-D | <US\$ I/day | 2.3 | 17.0 | 19.3 | <US\$ I/day | 12.2 | <US\$ I/day | 0.7 |
|  | US\$ I-2/day | 6.4 | 38.9 | 45.3 | US\$ I-2/day | 14.1 | US\$ I-2/day | 0.8 |
|  | >US\$ 2/day | 6.1 | 29.2 | 35.3 | >US\$ 2/day | 17.3 | US\$ $2 /$ day | 0.9 |
|  |  | 14.9 | 85.1 | 100.0 |  |  | >US\$ 2/day | 1 |
| EUR-B | <US\$ I/day | 0.9 | 2.1 | 3.0 | <US\$ I/day | 29.8 | <US\$ I/day | 0.8 |
|  | US\$ I-2/day | 4.7 | 9.9 | 14.6 | US\$ I-2/day | 32.4 | US\$ I-2/day | 0.8 |
|  | >US\$ 2/day | 32.0 | 50.3 | 82.3 | >US\$ 2/day | 38.9 | US\$ 2/day | 1.0 |
|  |  | 37.6 | 62.4 | 100.0 |  |  | >US\$ 2/day | 1 |
| SEAR-D | <US\$ I/day | 0.0 | 42.4 | 42.4 | <US\$ I/day | 0.1 | <US\$ I/day | 0.4 |
|  | US\$ I-2/day | 0.1 | 42.7 | 42.8 | US\$ I-2/day | 0.2 | US\$ I-2/day | 0.7 |
|  | >US\$ 2/day | 0.0 | 14.8 | 14.8 | >US\$ $2 /$ day | 0.2 | US\$ 2/day | 0.9 |
|  |  | 0.1 | 99.9 | 100.0 |  |  | >US\$ 2/day | 1 |
| WPR-B | <US\$ I/day | 3.9 | 14.0 | 17.9 | <US\$ I/day | 21.8 | <US\$ I/day | 1.1 |
|  | US\$ I-2/day | 6.7 | 27.3 | 34.0 | US\$ I-2/day | 19.6 | US\$ I-2/day | 1.0 |
|  | >US\$ 2/day | 9.6 | 38.5 | 48.1 | >US\$ 2/day | 20.1 | US\$ 2/day | I. 1 |
|  |  | 20.2 | 79.8 | 100.0 |  |  | >US\$ 2/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ I/day | 2.6 | 23.1 | 25.7 | <US\$ I/day | 9.9 | <US\$ I/day | 0.4 |
|  | US\$ I-2/day | 4.7 | 29.2 | 33.9 | US\$ I-2/day | 13.8 | US\$ I-2/day | 0.6 |
|  | >US\$ 2/day | 9.2 | 31.2 | 40.4 | >US\$ 2/day | 22.8 | US\$ $2 /$ day | - |
|  |  | 16.0 | 84.0 | 100.0 |  |  | >US\$ 2/day | I |

[^107]Table 24.46 Cell prevalence of overweight by poverty, prevalence of overweight by poverty and relative risk of overweight by poverty, by subregion, for a US\$ I per day cut-off (women only)

| Subregion | Overweight |  |  |  | Prevalence of overweight by poverty (\%) |  | Relative risk of overweight |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Yes | No |  |  |  |  |  |
| AFR-D | <US\$ I/day | 5.5 | 50.0 | 55.5 | <US\$ I/day | 10.0 | <US\$ I/day | 0.5 |
|  | >US\$ I/day | 8.9 | 35.6 | 44.5 | >US\$ I/day | 20.1 | US\$ I/day | 0.8 |
|  |  | 14.5 | 85.5 | 100.0 |  |  | >US\$ I/day | 1 |
| AFR-E | <US\$ I/day | 4.0 | 23.3 | 27.3 | <US\$ I/day | 14.7 | <US\$ I/day | 0.4 |
|  | >US\$ I/day | 24.1 | 48.6 | 72.7 | >US\$ I/day | 33.1 | US\$ I/day | 0.6 |
|  |  | 28.1 | 71.9 | 100.0 |  |  | >US\$ I/day | I |
| AMR-B | <US\$ I/day | 2.5 | 8.5 | 11.0 | <US\$ I/day | 22.3 | <US\$ I/day | 0.8 |
|  | >US\$ I/day | 24.4 | 64.6 | 89.0 | >US\$ I/day | 27.4 | US\$ I/day | 0.9 |
|  |  | 26.9 | 73.1 | 100.0 |  |  | >US\$ I/day | I |
| AMR-D | <US\$ I/day | 4.5 | 12.9 | 17.4 | <US\$ I/day | 26.1 | <US\$ I/day | 0.8 |
|  | >US\$ I/day | 25.7 | 56.9 | 82.6 | >US\$ I/day | 31.1 | US\$ I/day | 0.9 |
|  |  | 30.2 | 69.8 | 100.0 |  |  | >US\$ I/day | I |
| EMR-D | <US\$ I/day | 2.3 | 17.0 | 19.3 | <US\$ I/day | 12.0 | <US\$ I/day | 0.8 |
|  | >US\$ I/day | 12.5 | 68.2 | 80.7 | >US\$ I/day | 15.5 | US\$ I/day | 0.8 |
|  |  | 14.9 | 85.1 | 100.0 |  |  | >US\$ I/day | \| |
| EUR-B | <US\$ I/day | 0.9 | 2.1 | 3.0 | <US\$ I/day | 29.4 | <US\$ I/day | 0.8 |
|  | >US\$ I/day | 36.7 | 60.3 | 97.0 | >US\$ I/day | 37.9 | US\$ I/day | 0.8 |
|  |  | 37.6 | 62.4 | 100.0 |  |  | >US\$ I/day | 1 |
| SEAR-D | <US\$ I/day | 0.0 | 42.4 | 42.4 | <US\$ I/day | 0.1 | <US\$ I/day | 0.5 |
|  | >US\$ I/day | 0.1 | 57.5 | 57.6 | >US\$ I/day | 0.2 | US\$ I/day | 0.6 |
|  |  | 0.1 | 99.9 | 100.0 |  |  | >US\$ I/day | I |
| WPR-B | <US\$ I/day | 3.6 | 14.3 | 17.9 | <US\$ I/day | 20.4 | <US\$ I/day | 1.0 |
|  | >US\$ I/day | 16.6 | 65.5 | 82.1 | >US\$ I/day | 20.2 | US\$ I/day | 1.0 |
|  |  | 20.2 | 79.8 | 100.0 |  |  | >US\$ I/day | 1 |
| Total ${ }^{\text {a }}$ | <US\$ I/day | 2.5 | 23.2 | 25.7 | <US\$ I/day | 9.6 | <US\$ I/day | 0.5 |
|  | >US\$ I/day | 14.0 | 60.3 | 74.3 | >US\$ I/day | 18.8 | US\$ I/day | - |
|  |  | 16.0 | 84.0 | 100.0 |  |  | >US\$ I/day | 1 |

[^108]Table 24.47 Summary of relative risks by poverty: reference category >US \$2 per day

| Subregion |  | Unimproved water and/or sanitation | Underweight (low weight-for-age) | Non-marital sex (men) | Non-marital sex (women) | Condom use (men) | Condom use (women) | Indoor air pollution | Tobacco use | Alcohol use | Body weight (women) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | <US\$ 1/day | 9.4 | 2.3 | 0.6 | 0.4 | 0.5 | 0.3 | 1.1 | 1.1 | 0.8 | 0.4 |
|  | US\$ I-2/day | 2.9 | 1.4 | 0.9 | 0.9 | 0.9 | 0.7 | 1.1 | 0.9 | 0.8 | 0.8 |
|  | US\$ $2 /$ day | 1.6 | 1.2 | 0.9 | 0.9 | 1.0 | 0.9 | 1.1 | 0.9 | 0.9 | 0.9 |
|  | >US\$ 2/day | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| AFR-E | <US\$ 1/day | 4.6 | 2.6 | 0.9 | 0.6 | 0.3 | 0.1 | 2.0 | 0.7 | 0.5 | 0.4 |
|  | US\$ I-2/day | 3.5 | 1.8 | 1.1 | 0.8 | 0.7 | 0.4 | 1.9 | 0.8 | 0.6 | 0.7 |
|  | US\$ $2 /$ day | 2.7 | 1.4 | 1.0 | 1.0 | 0.8 | 0.6 | 1.7 | 0.9 | 0.7 | 0.8 |
|  | >US\$ 2/day | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| AMR-B | <US\$ 1/day | 12.3 | 2.4 | 0.7 | 0.8 | 0.6 | 0.3 | 7.2 | 1.1 | 0.5 | 0.8 |
|  | US\$ I-2/day | 5.7 | 1.8 | 0.9 | 0.6 | 0.8 | 0.4 | 4.7 | 1.2 | 0.7 | 1.0 |
|  | US\$ $2 /$ day | 3.4 | 1.6 | 1.0 | 0.7 | 0.9 | 0.6 | 3.3 | 1.2 | 0.8 | 1.0 |
|  | >US\$ 2/day |  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| AMR-D | <US\$ 1/day | 8.9 | 3.7 | 0.7 | 0.7 | 0.4 | 0.3 | 14.6 | 0.6 | 0.8 | 0.8 |
|  | US\$ I-2/day | 4.0 | 2.1 | 0.9 | 1.0 | 0.7 | 0.3 | 4.7 | 0.7 | 0.9 | 0.9 |
|  | US\$ $2 /$ day | 2.3 | 1.6 | 1.0 | 1.0 | 0.8 | 0.5 | 2.3 | 0.8 | 1.0 | 1.0 |
|  | >US\$ 2/day | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| EMR-B | <US\$ 1/day | 3.6 | 2.1 | - | - | - | - | - | - | - | - |
|  | US\$ 1-2/day | 3.2 | 1.9 | - | - | - | - | - | - | - | - |
|  | US\$2/day | 2.9 | 1.8 | - | - | - | - | - | - | - | - |
|  | >US\$ 2/day | 1 | 1 | - | - | - | - | - | - | - | - |

Table 24.47 Summary of relative risks by poverty: reference category $>$ US $\$ 2$ per day (continued)

| Subregion |  | Unimproved water and/or sanitation | Underweight (low weight-for-age) | Non-marital sex (men) | Non-marital sex (women) | Condom use (men) | Condom use (women) | Indoor air pollution | Tobacco use | Alcohol use | Body weight (women) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| EMR-D | <US\$ 1/day | 15.1 | 1.7 | - | - | - | - | 4.0 | 1.7 | - | 0.7 |
|  | US\$ I-2/day | 7.2 | 1.7 | - | - | - | - | 3.1 | 1.4 | - | 0.8 |
|  | US\$ $2 /$ day | 3.2 | 1.5 | - | - | - | - | 1.7 | 1.1 | - | 0.9 |
|  | >US\$ 2/day | 1 | 1 | - | - | - | - | 1 | I | - | 1 |
| EUR-B | <US\$ 1/day | 3.1 | 1.9 | - | - | - | - | 1.2 | 0.8 | 0.8 | 0.8 |
|  | US\$ I-2/day | 2.7 | 1.6 | - | - | - | - | 1.2 | 0.8 | 0.7 | 0.8 |
|  | US\$ $2 /$ day | 2.3 | 1.3 | - | - | - | - | 1.2 | 0.8 | 0.7 | 1.0 |
|  | >US\$ 2/day | 1 | 1 | - | - | - | - | 1 | 1 | 1 | 1 |
| EUR-C | <US\$ 1/day | 11.8 | 2.4 | - | 0.4 | - | 1.0 | 1.3 | 1.1 | 0.8 | - |
|  | US\$ I-2/day | 8.4 | 2.1 | - | 0.4 | - | 1.0 | 1.4 | 1.0 | 0.8 | - |
|  | US\$2/day | 5.4 | 1.8 | - | 0.5 | - | 1.0 | 1.3 | 1.0 | 0.8 | - |
|  | >US\$2/day | 1 | 1 | - | 1 | - | 1 | 1 | 1 | 1 | - |
| SEAR-B | <US\$ 1/day | 2.0 | 3.3 | - | - | - | - | - | - | - | - |
|  | US\$ I-2/day | 1.7 | 2.2 | - | - | - | - | - | - | - | - |
|  | US\$ $2 /$ day | 1.4 | 1.7 | - | - | - | - | - | - | - | - |
|  | >US\$ 2/day | I | I | - | - | - | - | - | - | - | - |


| SEAR-D | <USS / /ay | 5.0 | 2.1 | - | - | - | - | 3.5 | - |  | 0.4 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Uss 1-2/day | 4.1 | 1.7 |  |  |  |  | 2.9 |  |  | 0.7 |
|  | Us\$2/day | 2.1 | 1.3 |  |  |  | - | 1.8 |  |  | 0.9 |
|  | >US\$2/day | 1 | 1 |  |  |  |  | 1 |  |  |  |
| WPR-B | <USs / /day | 1.7 | 1.1 | - | 2.4 |  | - | 1.4 | 1.0 | 0.8 | 1.1 |
|  | Uss 1-2/day | 1.3 | 1.4 |  | 1.9 |  |  | 1.4 | 1.0 | 0.9 | 1.0 |
|  | US\$2/day | 1.0 | 1.3 | - | 1.5 |  | - | 1.3 | 1.0 | 0.9 | 1.1 |
|  | > 3 S\$2/day | 1 | 1 | - | 1 | - | - | 1 | 1 | 1 | 1 |
| $\overline{\text { Total (crude) }{ }^{\text {a }} \text {, }}$ | <USSI/day | 3.3 | 3.1 | 0.7 | 0.6 | 0.4 | 0.3 | 2.2 | 0.9 | 0.7 |  |
|  | Uss 1-2/day | 2.4 | 2.4 | 0.9 | 0.7 | 0.7 | 0.5 | 1.9 | 1.0 | 0.8 | 0.6 |
|  | US\$2/day |  |  |  |  |  |  |  |  |  |  |
|  | > 3 S 2/day | 1 | 1 | 1 | 1 | 1 | I | । | , | । |  |
| Total (pooled) | <USS / /day | 7.9 | 2.5 | 0.7 | 0.6 | 0.5 |  | ${ }^{3.8}$ |  |  |  |
|  | US\$1-2/day | 4.2 | 1.8 | 0.9 | 0.9 | 0.8 | 0.5 | 2.8 | 1.0 | 0.8 | 0.8 |
|  | US\$2/day |  |  |  |  |  |  |  |  |  |  |
|  | >US\$2/day | 1 | , | 1 | 1 | । | । | । | , | , |  |
| data |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | Othe II Out of |  |  |  | hense |  |  |  |  |  |  |

compared to the baseline of $>$ US $\$ 2$ per day. The equivalent summary relative risks for the dichotomous poverty variable are shown in Table 24.48.

The only difference from the relative risk results presented previously is the inclusion of both a crude and a pooled "total" summary relative risk for the 11 subregions included in this chapter. The crude total relative risk is simply the estimated relative risk comparing people by poverty stratum, regardless of the subregion in which they live. This relative risk estimate is crude in the sense that it does not allow for confounding by subregion; nevertheless, it is easily understood.

The "pooled" relative risk is a Mantel-Haenszel pooled relative risk across subregions. The actual sample sizes for each subregion are used to create the weights. This total relative risk estimate seeks to summarize the within-subregion associations, and will vary from the crude total estimate if there is independent variation across subregions in both poverty levels and risk factor prevalence.

It should be noted that there is marked heterogeneity between subregions in the association of income poverty with most of the risk factors considered here. The association of income poverty with lack of water and sanitation is stronger in AMR-B, AFR-D, EMR-D and EUR-C, and that with indoor air pollution is stronger in AMR-B and AMR-D. Thus, the pooled relative risk estimates for the total developing world should be treated cautiously. Nevertheless, comparing the crude and pooled total relative risks in Table 24.47 reveals substantial differences for unimproved water and sanitation and indoor air pollution. In both these instances, the pooled relative risk is greater than the crude. Thus, there are (on average) stronger associations of poverty with these two risk factors within subregions than those suggested by a crude global analysis. Put another way, subregion confounds the association of individual income poverty with unsafe water and sanitation, since more poor subregions have higher prevalences of these two risk factors independent of personal income poverty.

## RISK DIFFERENCES

All the summary statistics thus far in this chapter are relative risks. However, as the prevalence of risk factors varies by subregion, a constant relative risk across subregions will correspond to differences in the absolute difference in prevalence of risk factor between poverty levels. Tables 24.49 and 24.50 present risk difference estimates by subregion for the three-level and two-level poverty estimates, respectively.

For ease of interpretation, the risk differences are presented as the difference between the poor and the referent group. For example, it was estimated that the prevalence of underweight among 0-4 year olds living on $<$ US $\$ 1$ per day in AFR-D was $22.6 \%$ greater than that among $0-4$ year olds living on $>$ US $\$ 2$ per day (Table 24.49). The actual risk factor prevalence in the referent group is also given. Thus, the prevalence of
Table 24.48 Summary of relative risks by poverty: reference category >US\$ I per day

| Subregion |  | Unimproved water and/or sanitation | Underweight (low weight-for-age) | Non-marital sex (men) | Non-marital sex (women) | Condom use (men) | Condom use (women) | Indoor air pollution | Tobacco use | Alcohol use | Body weight (women) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | <US\$ 1/day | 4.2 | 1.8 | 0.6 | 0.5 | 0.6 | 0.4 | 1.1 | 1.2 | 0.9 | 0.5 |
|  | US\$ I/day | 2.1 | 1.3 | 0.9 | 0.9 | 0.9 | 0.7 | 1.1 | 1.1 | 0.9 | 0.8 |
|  | >US\$ 1/day | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| AFR-E | <US\$ 1/day | 1.9 | 1.7 | 0.8 | 0.7 | 0.4 | 0.1 | 1.6 | 0.8 | 0.5 | 0.4 |
|  | US\$ I/day | 1.7 | 1.5 | 1.1 | 0.8 | 0.7 | 0.3 | 1.6 | 0.8 | 0.6 | 0.6 |
|  | >US\$ 1/day | 1 | 1 | 1 | 1 | 1 | 1 | I | 1 | 1 | 1 |
| AMR-B | <US\$ 1/day | 6.2 | 2.1 | 0.7 | 0.9 | 0.7 | 0.3 | 4.5 | 1.0 | 0.5 | 0.8 |
|  | US\$ I/day | 4.8 | 1.7 | 0.8 | 0.8 | 0.7 | 0.4 | 3.8 | 1.1 | 0.6 | 0.9 |
|  | >US\$ 1/day | 1 | 1 | 1 | 1 | I | I | 1 | 1 | 1 | I |
| AMR-D | <US\$ 1/day | 4.4 | 2.7 | 0.7 | 0.7 | 0.4 | 0.3 | 5.9 | 0.7 | 0.9 | 0.8 |
|  | US\$ I/day | 3.5 | 2.0 | 0.8 | 0.9 | 0.6 | 0.4 | 3.5 | 0.8 | 0.9 | 0.9 |
|  | >US\$ 1/day | 1 | 1 | 1 | I | 1 | I | 1 | 1 | 1 | 1 |
| EMR-B | <US\$ 1/day | 3.0 | 1.9 | - | - | - | - | - | - | - | - |
|  | US\$ I/day | 2.9 | 1.9 | - | - | - | - | - | - | - | - |
|  | >US\$ 1/day | I | 1 | - | - | - | - | - | - | - | - |
| EMR-D | <US\$ 1/day | 3.3 | 1.2 | - | - | - | - | 1.5 | 1.3 | - | - |
|  | US\$ I/day | 2.8 | 1.2 | - | - | - | - | 1.5 | 1.2 | - | - |
|  | >US\$ 1/day | 1 | 1 | - | - | - | - | 1 | 1 | - | - |
| EUR-B | <US\$ 1/day | 2.2 | 1.7 | - | - | - | - | 1.2 | 0.9 | 0.8 | 0.8 |
|  | US\$ I/day | 2.2 | 1.6 | - | - | - | - | 1.2 | 0.9 | 0.8 | 0.8 |
|  | >US\$ 1/day | 1 | 1 | - | - | - | - | 1 | 1 | 1 | 1 |
| EUR-C | <US\$ 1/day | 5.7 | 2.1 | - | 0.4 | - | 1.0 | 1.2 | 1.1 | 0.8 | - |
|  | US\$ I/day | 5.6 | 2.1 | - | 0.4 | - | 1.0 | 1.3 | 1.1 | 0.8 | - |
|  | >US\$ 1/day | 1 | 1 | - | I | - | 1 | I | 1 | 1 | - |

Table 24.48 Summary of relative risks by poverty: reference category >US\$ I per day (continued)

| Subregion |  | Unimproved water and/or sanitation | Underweight (low weight-for-age) | Non-marital sex (men) | Non-marital sex (women) | Condom use (men) | Condom use (women) | Indoor air pollution | Tobacco use | Alcohol use | Body weight (women) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| SEAR-B | <US\$ I/day | 1.5 | 2.1 | - | - | - | - | 2.6 | - | - | - |
|  | US\$ I/day | 1.5 | 2.0 | - | - | - | - | 2.1 | - | - | - |
|  | >US\$ I/day | 1 | 1 | - | - | - | - | 1 | - | - | - |
| SEAR-D | <US\$ I/day | 1.5 | 1.4 | - | - | - | - | 1.5 | - | - | 0.5 |
|  | US\$ I/day | 1.5 | 1.3 | - | - | - | - | 1.4 | - | - | 0.6 |
|  | >US\$ I/day | 1 | 1 | - | - | - | - | 1 | - | - | 1 |
| WPR-B | <US\$ I/day | 2.3 | 0.9 | - | - | - | - | 1.2 | 1.0 | 0.9 | 1.0 |
|  | US\$ I/day | 2.3 | 1.0 | - | - | - | - | 1.2 | 1.0 | 0.9 | 1.0 |
|  | >US\$ I/day | 1 | 1 | - | - | - | - | 1 | 1 | 1 |  |
| Total (crude) ${ }^{\text {a }}$ | <US\$ I/day | 2.4 | 1.9 | 0.7 | 0.7 | 0.5 | 0.3 | 1.5 | 0.9 | 0.7 | 0.5 |
|  | US\$ 1/day | - | - | - | - | - | - | - | - | - | - |
|  | >US\$ I/day | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Total (pooled) ${ }^{\text {b }}$ | <US\$ I/day | 3.0 | 1.7 | 0.7 | 0.6 | 0.5 | 0.3 | 1.6 | 1.0 | 0.8 | 0.6 |
|  | US\$ I/day | - | - | - | - | - | - | - | - | - | - |
|  | >US\$ I/day | 1 | 1 | 1 | I | 1 | 1 | 1 | I | 1 | 1 |
| - No data. |  |  |  |  |  |  |  |  |  |  |  |
| a Total refers to the 11 out of 14 subregions included in this report. The crude "total" estimate is derived by summing the estimated number factor within each poverty stratum, then recalculating the relative risks at this total level. It is crude in so far as it does not allow for confou risk association. |  |  |  |  |  |  |  |  |  |  |  |
| Total refers to the II out of 14 subregions included in this report. Unlike the crude estimate, the pooled estimate uses Mantel-Haenszel wei subregions. The actual DHS or LSMS sample sizes are used to calculate the Mantel-Haenszel weights. |  |  |  |  |  |  |  |  |  |  |  |

Table 24.49 Absolute percentage differences in risk factor prevalence by subregion by three-level poverty variable: reference
Underweight Unimproved
(low weight- water and/or Non-marital Non-marital Condom Condom use Indoor air Body weight
Tobacco use Alcohol use (women)

|  |  | Mồ No 웅 | $\stackrel{\infty}{\sim} \stackrel{\infty}{1} \underset{O}{O}$ |
| :---: | :---: | :---: | :---: |
| ONㅜㄴ운 | $\frac{0}{m} \frac{\sigma}{1} \frac{n}{1} \stackrel{n}{1}_{0}^{0} 0$ | $\underset{\sim}{m} \underset{\sim}{m} \underset{\sim}{\underset{N}{N}} \underset{\sim}{m} 0$ |  |

Table 24.49 Absolute percentage differences in risk factor prevalence by subregion by three-level poverty variable: reference group >US\$2 per day prevalence (continued)

| Subregion |  | Underweight (low weight-for-age) | Unimproved water and/or sanitation | Non-marital sex (men) | Non-marital sex (women) | Condom use (men) | Condom use (women) | Indoor air pollution | Tobacco use | Alcohol use | Body weight (women) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Reference prevalence ${ }^{\text {a }}$ | >US\$ 2/day | 7.0 | 10.1 | 47.3 | 4.1 | 50.7 | 60.1 | 12.2 | 31.9 | 65.9 | 31.7 |
| EMR-B | <US\$ I/day | 7.8 | 35.1 | - | - | - | - | - | - | - |  |
|  | US\$ 1-2/day | 6.8 | 29.8 | - | - | - | - | - | - | - | - |
|  | US\$ 2/day | 6.0 | 26.0 | - | - | - | - | - | - | - | - |
|  | >US\$ 2/day | 0.0 | 0.0 | - | - | - | - | - | - | - | - |
| Reference prevalence ${ }^{\text {a }}$ | >US\$ 2/day | 7.5 | 13.7 | - | - | - | - | - | - | - | - |
| EMR-D | <US\$ I/day | 12.6 | 72.7 |  |  | - | - | 66.2 | 11.5 | - | -5.2 |
|  | US\$ I-2/day | 12.5 | 32.0 | - | - | - | - | 45.3 | 7.2 | - | -3.2 |
|  | US\$ 2/day | 7.7 | 11.2 | - | - | - | - | 15.0 | 2.2 | - | -1.1 |
|  | >US\$ 2/day | 0.0 | 0.0 | - | - | - | - | 0.0 | 0.0 | - | 0.0 |
| Reference prevalence ${ }^{\text {a }}$ | >US\$ 2/day | 17.0 | 5.2 | - | - | - | - | 21.9 | 17.6 | - | 17.3 |
| EUR-B | <US\$ I/day | 6.3 | 32.4 | - | - | - | - | 9.3 | -5.6 | -13.6 | -9.1 |
|  | US\$ I-2/day | 4.0 | 25.9 | - | - | - | - | 9.6 | -8.2 | -17.3 | -6.5 |
|  | US\$ 2/day | 2.0 | 19.4 | - | - | - | - | 8.9 | -8.2 | -16.6 | -1.8 |
|  | >US\$ 2/day | 0.0 | 0.0 | - | - | - | - | 0.0 | 0.0 | 0.0 | 0.0 |
| Reference | >US\$ 2/day | 6.8 | 15.4 | - | - | - | - | 39.9 | 36.7 | 64.9 | 38.9 |


| EUR-C | <USS 1/day | 2.8 | 20.2 | - | -9.7 |  | -0.9 | ${ }^{6.3}$ | 2.7 | $-21.1$ | - |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Us\$ $1-2 /$ lay | ${ }_{17}^{2.2}$ | $\begin{array}{r}14.0 \\ 88 \\ \hline 8\end{array}$ |  | -8.6 |  | $-0.4$ | 8.3 | 1.0 0 | -15.8 -157 |  |
|  | Us\$ 2/day | 1.7 0.0 | 8.2 0.0 | - | -7.6 0.0 | - | -0.5 | 6.6 | -0.9 | -13.7 |  |
| Reference | > Ss $^{\text {2/day }}$ | 2.0 | 1.9 | - | 15.4 | - | 22.6 | 20.7 | 34.4 | 89.1 | - |
| SAAR-b |  |  |  |  |  |  |  |  |  |  |  |
|  | <USS I/day | 34.9 | 21.2 | - | - | - | - | - |  | - |  |
|  | Uss 1-2/day | 19.0 | 15.4 |  |  |  |  |  |  |  |  |
|  | Us\$2]day | 10.8 | 9.2 |  | - |  |  |  |  |  |  |
|  | > 3 S 2 day |  |  |  |  |  |  |  |  |  |  |
| Reference prevalence | >SS\$ 2 /day | 15.5 | 22.1 | - | - | - | - | - | - | - | - |
| SEAR-D | <USS I/day | 28.0 | 64.7 | - | - | - | - | 70.8 | - | - | -0.1 |
|  | Uss 1-2/day | 18.6 | 50.0 |  |  |  |  | 54.1 |  |  | -0.1 |
|  | Us $\$ 2$ /day | 7.0 | 17.0 |  | - | - | - | 22.5 |  | - | 0.0 |
|  | >US\$ / /day | 0.0 | 0.0 | - | - | - | - | 0.0 | - | - | 0.0 |
| Reference | > 3 \$ 2 /day | 26.1 | 16.0 | - | - | - | - | 29.2 | - | - | 0.2 |
| prevalence ${ }^{\text {e }}$ |  |  |  |  |  |  |  |  |  |  |  |
| WPR-B | <US\$ 1/day | 1.1 | 31.0 | - | - | - | - | 28.5 | 1.4 | -10.7 | 1.7 |
|  | Uss 1-2/day | 5.5 | 16.1 | - | - | - | - | 25.6 | -0.3 | -7.6 | -0.5 |
|  | Uss 2 /day | 4.0 | 2.3 | - | - |  | - | 18.4 | -0.7 | -4.4 | 3.0 |
|  | > S \$ $\$$ 2day | 0.0 | 0.0 | - | - | - | - | 0.0 | 0.0 | 0.0 | 0.0 |
|  | > 5 S 2 /day | 13.9 | 47.1 | - | - | - | - | 64.3 | 34.6 | 62.0 | 20.1 |
| prevalencee |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

Table 24.50 Absolute difference in risk factor prevalence by subregion by two-level poverty variable: reference group >US\$ I per

| Subregion |  | Underweight (low weight-for-age) | Unimproved water and/or sanitation | Non-marital sex (men) | Non-marital sex (women) | Condom use (men) | Condom use (women) | Indoor air pollution | Tobacco use | Alcohol use | Body weight (women) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | <US\$ 1/day | 17.3 | 53.1 | -18.9 | -10.4 | -19.0 | -17.0 | 5.1 | 2.4 | -5.9 | -10.1 |
|  | US\$ I/day | 7.5 | 17.4 | -6.8 | -2.6 | -4.7 | -7.6 | 4.7 | 1.5 | -4.9 | -4.7 |
|  | >US\$ 1/day | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Reference prevalence ${ }^{\text {a }}$ | >US\$ I/day | 22.6 | 16.4 | 49.2 | 20.4 | 44.7 | 26.5 | 70.6 | 14.2 | 41.6 | 20.1 |
| AFR-E | <US\$ 1/day | 19.2 | 41.7 | -6.2 | -5.4 | -27.0 | -21.4 | 19.5 | -6.0 | -77.3 | -18.4 |
|  | US\$ I/day | 13.6 | 32.6 | 3.7 | -3.8 | -15.4 | -16.1 | 49.6 | -5.3 | -80.2 | -13.8 |
|  | >US\$ 1/day | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Reference prevalence ${ }^{\text {a }}$ | >US\$ I/day | 25.8 | 46.2 | 37.5 | 17.1 | 44.4 | 24.7 | 80.5 | 24.6 | 50.5 | 33.1 |
| AMR-B | <US\$ 1/day | 4.8 | 79.7 | -11.9 | -2.1 | -18.8 | -24.8 | 62.1 | 0.9 | -34.8 | -5.1 |
|  | US\$ I/day | 3.3 | 59.0 | -9.4 | -4.1 | -15.8 | -20.7 | 49.7 | 3.9 | -26.6 | -2.9 |
|  | >US\$ I/day | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Reference prevalence ${ }^{\text {a }}$ | >US\$ I/day | 4.5 | 15.5 | 45.7 | 18.2 | 57.6 | 37.4 | 17.8 | 30.6 | 70.0 | 27.4 |
| AMR-D | <US\$ 1/day | 16.0 | 68.5 | -14.1 | -1.2 | -26.7 | -32.0 | 57.0 | -9.6 | -9.0 | -5.0 |
|  | US\$ I/day | 9.5 | 50.1 | -8.0 | -0.3 | -18.4 | -26.3 | 100.0 | -6.2 | -8.4 | -3.8 |
|  | >US\$ 1/day | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Reference prevalence ${ }^{\text {a }}$ | >US\$ I/day | 9.6 | 20.2 | 45.7 | 4.1 | 46.1 | 47.3 | 43.0 | 29.2 | 64.1 | 31.1 |







|।| | | | | | | | $\bar{\phi}$




$\stackrel{\propto}{\underset{\sim}{\infty}}$
Reference
prevalence ${ }^{\text {a }}$
$\stackrel{\stackrel{\alpha}{c}}{\underset{山}{\dot{\omega}}}$
Reference prevalence ${ }^{\text {a }}$
EUR-B

EUR-C
Reference
prevalence
$\infty$
$\stackrel{\alpha}{c}$
$\underset{\sim}{u}$
Reference
prevalence
Table 24.50 Absolute difference in risk factor prevalence by subregion by two-level poverty variable: reference group >US\$ I per day prevalence (continued)
Underweight Unimproved

|  |  |  |  |  |  |  |  |  |  |  |
| :--- | :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Subregion |  | (low weight- <br> for-age) | water and/or <br> sanitation | Non-marital <br> sex (men) | Non-marital <br> sex (women) | Condom <br> use (men) | Condom use <br> (women) | Indoor air <br> pollution | Tobacco use | Alcohol use |
| (women) |  |  |  |  |  |  |  |  |  |  |

[^109]underweight among children living on $>$ US $\$ 2$ per day in AFR-D was $17.9 \%$. Accordingly the prevalence of underweight among children living on <US\$ 1 per day can be deduced from the data in Table 24.49, i.e. $22.6+17.9=40.5 \%$.

While the relative risks across subregions for underweight were somewhat similar, the risk differences by poverty varied markedly given the different overall prevalences of underweight between subregions. There were also notable differences in the risk differences across subregions for unimproved water and sanitation and for indoor air pollution.

Finally, the negative risk differences correspond to those risk factors where the prevalence among the poor was less than that among the non-poor.

## 4. Systematic literature Reviews

### 4.1 Tobacco use

## Literature on developing countries

## Smoking associated with low socioeconomic status (SES)

The International Clinical Epidemiology Network (INCLEN) study collected data on risk factors for cardiovascular disease in men from 12 centres in seven developing countries (INCLEN 1994). It found significant SES trends (using education, occupation or income) for increased smoking among men of low SES in two centres in China and one in rural Thailand. No significant associations were obtained from the centres in Brazil, Chile, Colombia, Indonesia and the Philippines.

China. Marked gradients in smoking prevalence for men and women (increasing prevalence with lower levels of educational attainment) have been described in an urban Chinese population (Yu et al. 2000). Over the period 1989-1996 the gradient became less steep in men (owing largely to an increase in smoking by better educated men), whereas in women it changed little over the same period. This study also showed that less educated men smoked more cigarettes per day, but that this gap also decreased over the period 1989-1996. Another urban study revealed that those with lower and middle levels of education (particularly men but also women) had the highest prevalence of smoking (Koong et al. 1991). In contrast, a cohort study of urban male workers reported no association of SES with smoking prevalence, but daily cigarette consumption was significantly greater among men with low educational attainment (Siegrist et al. 1990). Surveys in three provinces of rural China also showed an increased prevalence of smoking among adults with low educational attainment ( Hu and Tsai 2000). Also, among urban adolescents, prevalence of smoking was found to be associated with lower parental SES (Zhu et al. 1996).

South Asia. Low educational attainment was found to be associated with smoking in men and women in urban populations in India (Narayan et al. 1996; Singh et al. 1998). One of these studies also reported that the smoking of beedi or chutta was associated with a low level of education and with manual occupations (Narayan et al. 1996). Another study in India also reported this finding for beedi smoking, and that smoking this type of cigarette (compared to other cigarettes) was independently associated with a higher frequency of respiratory symptoms and poorer lung function on clinical testing (Chhabra et al. 2001). In rural populations in India, the same pattern of increased smoking prevalence with lower SES was described for educational attainment in both men and women (Gupta et al. 1994) and for social class in men (Singh et al. 1997a) (with the smoking prevalence for women in this latter study being very low). Similarly, a national survey in Pakistan found that, among both males and females, illiteracy was strongly associated with higher smoking rates (Alam 1998). Also, among Bangladeshi male adolescents, a much higher rate of smoking was reported among slum dwellers with no formal education than among students (Ahsan et al. 1998).

Other Asian countries. In the Republic of Korea, smoking prevalence was found to be associated with low educational attainment in men but not in women (women had a very low smoking prevalence) (Chung et al. 1992). This was also the case for men in Cambodia (Smith et al. 1998). Studies of rural populations in Malaysia found the prevalence of smokeless tobacco use to be significantly higher among less educated women (Gan 1995, 1998). In the latter study, however, education was not associated with prevalence of tobacco smoking.

Middle East. In Saudi Arabian adults in three regions of the country, smoking prevalence was higher among uneducated people and among certain occupations of low SES such as manual workers (Jarallah et al. 1999). Similarly, in Bahrain (Hamadeh and Musaiger 2000) and Kuwait (Memon et al. 2000) low educational attainment was associated with increased smoking prevalence in adults.

Other countries. In urban Brazil, increased smoking prevalence was associated with low educational attainment in men and low social class in women (Duncan et al. 1993). Similarly, it was associated with more poor and uneducated adults in Nigeria (Obot 1990), men and women with low educational attainment in Tonga (Woodward et al. 1994) and men with low educational attainment on a Caribbean island (the results for women were not statistically significant) (Grol et al. 1997b). In the latter study, however, men and women of high SES who did smoke consumed significantly more cigarettes per day.

## Smoking associated with high SES

A gradient of increased smoking prevalence with higher SES was described in an urban study in Brazil for women of high social class (with no significant gradient among men and in contrast to the other Brazilian study described above) (Martins et al. 1995). This pattern was also described for adolescent smoking in the United Arab Emirates (Bener and al-Ketbi 1999) (based on paternal educational attainment and family income) and Ghana (for high SES homes, though no precise data were provided) (Amonoo-Lartson and Pappoe 1992).

## Smoking not associated with SES

As mentioned above, the INCLEN study (INCLEN 1994) found no significant associations between smoking prevalence and SES from the centres in Brazil, Chile, Colombia, Indonesia and the Philippines. Similar findings were obtained from surveys of villagers in India (Singh et al. 1997b) and of a town and village population in India (using level of educational attainment) (Chaturvedi et al. 1998). In this latter study, however, there was a higher rate of smoking among the unemployed than among the employed. Surveys of women in five Indian cities also found no significant gradient for smoking prevalence (Singh et al. 1999).

## LITERATURE ON DEVELOPED COUNTRIES

## Smoking associated with low SES

A review by Stellman and Resnicow (1997) found that in most developed countries, the prevalence of cigarette smoking is currently higher in low-SES groups. However, they note that in some of these countries smoking had been more prevalent among the higher social classes during the first half of the twentieth century. Studies examined in this review that reported an occupational class gradient by smoking prevalence (increased smoking among low-SES groups) come from Australia (with gradients for men and women); the United Kingdom of Great Britain and Northern Ireland (based on multiple surveys); and the United States of America (three studies cited with gradients for both men and women). The same pattern was seen for SES gradients determined by educational attainment in studies from France; Italy and Spain (for men under 65 years only); the United Kingdom; and the United States (seven studies cited). In the French study it was reported that less educated men much preferred nonfilter cigarettes manufactured with black tobacco. Similarly, there are survey data from the United States showing that better educated men tended to smoke cigarettes with lower tar yields.

The same pattern was seen when considering SES as determined by income in Australia (for men and women), the United Kingdom (based on housing data to reflect SES) (Stellman and Resnicow 1997) and the United States (owing principally to smoking cessation among the more affluent).

Similarly, the INTERSALT study involving 52 centres in 18 developed and 12 developing countries (ICRG 1988) found that smoking prevalence was related to low educational attainment (Stamler et al. 1992). Also, a review of smoking among adolescents identified four studies in which low SES was associated with increased smoking prevalence in developed countries (Canada, Finland, New Zealand and Norway) (Tyas and Pederson 1998).

Other data not detailed in the above-mentioned reviews also conform to this pattern of increased smoking prevalence by low-SES groups for:

- adults in the United Kingdom, based on occupational class for men and women (Acheson 1998);
- young people leaving school in the United Kingdom, for educational disadvantage independent of parental SES and parental smoking status (Glendinning et al. 1994; Green et al. 1991);
- students in the United States, for school performance and smoking prevalence and attempts to quit smoking (Hu et al. 1998);
- pregnant women in the United States, by educational attainment (Pamuk et al. 1998);
- adult New Zealanders, based on national data from the census (Borman et al. 1999; Crampton et al. 2000; Davis et al. 1999; Hay and Foster 1984), national health surveys (Howden-Chapman and Tobias 2000; Ministry of Health 1999) and other community surveys (Jackson et al. 1990; Klemp et al. 1998; Statistics New Zealand/ Ministry of Health 1993; Whitlock et al. 1997); and
- pregnant women in New Zealand (occupational class) (Fergusson et al. 1998).

Furthermore, low-SES adults who do not smoke were found to suffer from increased exposure to second-hand smoke (Whitlock et al. 1998).

## Tobacco-related disease and low SES

Stellman and Resnicow (1997) reported that "a large body of evidence confirms the inverse association of lung cancer and social class in many developed countries". These authors noted that in some situations it was "satisfactorily demonstrated" that such gradients were attributable to social class gradients in tobacco use. In general, however, the SES patterning of lung cancer cannot be fully explained by SES patterns of tobacco use (given some likely additional role for various dietary and occupational risk factors and access to health care). This review provided data for the following countries, all of which demonstrated increased tobacco-related disease rates among low-SES groups:

- Denmark, where there was a very strong occupational class gradient (lung and bladder cancer) in men and women;
- England and Wales, for cancer incidence and mortality;
- Italy, with a very strong relationship between educational attainment and various tobacco-related cancers (but not lung cancer); and
- New Zealand, where smoking patterns explained much of the increased mortality risk for social classes III and IV.

Other data from New Zealand also suggest that SES gradients for tobacco-related disease reflect SES gradients for smoking rates (higher smoking rates with lower SES) (Kawachi et al. 1991; Pearce 1997). Indeed, the latest analysis on smoking and inequality in New Zealand suggests that "at least one-third of the deprivation gradient in life expectancy at birth or older ages, one-quarter of the corresponding ethnic disparity, and one-fifth of the corresponding gender gap, is accounted for by tobacco consumption" (Ministry of Health 2001).

## Smoking associated with high SES

The review by Stellman and Resnicow (1997) reported that in the United Kingdom in the late 1940s and early to mid-1950s, smoking was more common among the higher social classes (based on data from multiple surveys). This review also reported on Spanish data indicating that smoking prevalence was relatively higher among college-educated men aged $>65$ years. Also, there was a very strong association with higher levels of education and smoking among younger Spanish women (in marked contrast to the data for men) (1987 data). More recent data from Spain suggest that the gap in smoking cessation rates by educational level has become wider over time (i.e. cessation rates are higher for those with higher levels of education) (Fernandez et al. 2001). This pattern was also apparent for women with higher educational levels. Similarly, other data from Spain suggest that the quitting rate is higher among women aged 25-44 years with non-manual occupations compared to those with manual occupations (Regidor et al. 2001). A review of smoking among adolescents identified studies in five developed countries (Iceland, Japan, New Zealand, Sweden and the United Kingdom) in which higher personal income was associated with higher levels of smoking (Tyas and Pederson 1998).

## Smoking not associated with SES

One review identified two studies that reported non-significant effects of parental education on adolescent smoking in Canada and the United States (Tyas and Pederson 1998). However, these studies examined maternal education only. In the United Kingdom there was little occupational class gradient in the proportion of children who had ever smoked (although some gradient existed for higher average consumption of cigarettes with lower occupational class) (Acheson 1998). It has been suggested that smoking in adolescence is an indicator of occupational
class of destination rather than occupational class of origin (Glendinning et al. 1994).

## DISCUSSION

## The relationship

In general in developing countries, lower SES was associated with higher prevalence of smoking. The association was generally stronger for men. There was also a much lower smoking prevalence among women in most developing countries. Nevertheless, some studies in the developing world actually show the opposite pattern, with high-SES adults and adolescents having higher smoking rates, while other studies show no association. This overall pattern contrasts with that for other cardiovascular disease risk factors in developing countries (cholesterol level, obesity, blood pressure and to some extent physical inactivity). This might reflect the low cost of tobacco products in many developing countries and the influence of advertising and the mass media in promoting their use. Nevertheless, there is a need for more historical data on smoking in developing countries to better interpret this pattern.

In developed countries the gradient of increased smoking with lower SES is even more predominant. There are still some exceptions, however, such as for adolescent smoking in some settings and for Spanish women. Furthermore, the SES gradient for tobacco-related disease and overall mortality is likely to be partly explained by the SES gradient for tobacco use.

In both developing and developed countries there is some evidence that the type of tobacco product smoked by low-SES adults may also be more hazardous (e.g. beedis in India) than that consumed by high-SES adults. Similarly, the number of cigarettes consumed daily may be higher among low-SES adults.

## Possible mechanism

Several factors are likely to be involved in the relatively increased prevalence of smoking by lower-SES groups.

- Cultural effects relating to health values and risk perception may be relevant. Similarly, neighbourhood effects may be important, such as the influence of disadvantaged neighbourhoods on smoking in men (though not women) described in the United States by Ross (2000).
- Parental influences on adolescent smoking may possibly be more important in low-SES populations (e.g. in countering the benefits of education about the risks of smoking).
- Access to cheap tobacco in tobacco-producing rural areas of the developing world may be relevant.
- Lower educational attainment may lead to poorer knowledge of the hazards of smoking and of how to obtain support for quitting.
- The relative lack of restrictions on smoking at the workplace for many blue-collar workers could reduce incentives to quit (compared to those in higher occupational classes).
- Lower rates of smoking cessation in low-SES groups might be related to a greater burden of psychosocial stress and the effect of others smoking.

Advertising by the tobacco industry is probably an important determinant of smoking trends (Stellman and Resnicow 1997), particularly the aggressive marketing in developing countries (Connolly 1992; Mackay 1992).

### 4.2 Hazardous alcohol use

## LITERATURE ON DEVELOPING COUNTRIES

## Hazardous alcohol use associated with low SES

In China, one cohort study of urban male workers reported that high levels of daily alcohol consumption were significantly more prevalent among men with low educational attainment (Siegrist et al. 1990). Similarly, a Brazilian study of urban residents (using a logistic regression model) found that both heavy drinking and alcohol dependence were associated with low educational attainment and low income (Moreira et al. 1996). Another study in a Brazilian city found that low social class was significantly associated with excessive alcohol consumption in both men and women (Duncan et al. 1993).

A study in a Caribbean island reported that men and women of low SES (based on educational attainment) drank more alcohol per week (Grol et al. 1997b). However, high-SES women were more likely to be regular drinkers than low-SES women. Of the two African studies identified, one in Ethiopia found that problem drinking (based on the CAGE screening instrument) was associated with a lower level of educational attainment among residents of Addis Ababa (Kebede and Alem 1999). Similarly, in a Nigerian study, alcohol consumption was higher in those with low SES (Bunker et al. 1992). In Brazil, the results of one study suggested that higher rates of oesophageal cancer among lower-SES groups are likely to be attributable to the higher use of sugar-cane spirit, black tobacco and mate in these groups (Bouchardy et al. 1993). In contrast, data from Colombia indicate that the incidence of alcohol-related cancers tended to show positive social class gradients (Cuello et al. 1982).

## Hazardous alcohol use associated with high SES

In an Ethiopian study, problem drinking (based on the CAGE screening instrument) was associated with higher income and education among men or women in a rural district (Alem et al. 1999) (in contrast to another Ethiopian study described above). In Thailand, a study reported that hazardous alcohol use (based on AUDIT scores) was associated with higher income among those seeking emergency treatment (Lapham et al. 1998). By contrast, the study found that those with a university degree were significantly less likely to have positive AUDIT scores. A study in the Republic of Korea found a trend towards greater alcohol consumption with increasing years of education among women, but no such trend among men (Chung et al. 1992). Alcohol intake among high-SES males explained a small but statistically significant part of the relationship between hypertension and high SES in villagers in North India (Singh et al. 1997b). This suggests that at least some members of this population were drinking fairly large quantities of alcohol.

## No association with SES

A study in a metropolitan area in Brazil found no statistically significant differences in alcohol use associated with SES, although there was a trend towards higher rates of drinking among low-SES men (Martins et al. 1995).

## Literature on developed countries

## Hazardous alcohol use associated with low SES

A major review of alcohol drinking, social class and cancer suggested a very likely role for alcohol drinking in the observed negative social class gradients for alcohol-related cancers in men in France, Italy and New Zealand (Moller and Tonnesen 1997). Evidence that was less strong, but still suggestive of such a role, was also reported for men in Denmark, Switzerland and the United Kingdom. This review identified studies showing associations between hazardous alcohol use and low SES in Denmark (for average number of drinks per week in men and women); Finland (for rates of alcohol intoxication and rates of hospital admission for acute alcohol-related conditions for men and women); France (for prevalence of "heavy drinkers" in men); Sweden (alcoholism and high alcohol consumption in young people); Switzerland (for alcoholism); the United Kingdom (for the proportion of heavy drinkers in manual vs non-manual occupations among men); and the United States (for heavy consumption-five or more drinks on one occasion in men and women).

The major review by Colhoun et al. (1998) on hypertension and SES reported that higher alcohol consumption by low-SES men explains part, though not all, of the association between SES and blood pressure in men in developed countries. This collectively considered the findings of
studies in Australia, Israel, the Netherlands, Norway, the United Kingdom (two studies) and the United States (two studies). Similarly, the INTERSALT study, involving 52 centres in 18 developed and 12 developing countries, found that adjusting for alcohol (along with BMI, smoking, and sodium and potassium excretion) halved the association between low SES and higher blood pressure in men so that it was no longer statistically significant (Stamler et al. 1992).

A combined analysis of population samples from France, Italy, Spain and Switzerland (Pequignot et al. 1988) found that daily alcohol consumption was higher in male manual workers than in male professionals, whereas in women there was no significant variation by occupational group. More recent data from the United Kingdom indicated that 10\% of low-SES men (social classes IV and V) were dependent on alcohol compared to $5 \%$ of high-SES men (social classes I and II) (Acheson 1998). Similarly, in the United States, heavy alcohol use was reported to be higher among those with poorer education among men and women in virtually all ethnic groups (Pamuk et al. 1998).

Data from surveys in 48 American states showed that hazardous consumption of alcohol was significantly more frequent in households with below-median income and in those with a lower educational level (for five or more drinks on one occasion at least once a week during the previous year) (Midanik and Clark 1994). For weekly drinking, however, the pattern was the opposite, being higher in those with above-median income and a higher level of education.

In a national survey in New Zealand based on AUDIT scores, lower social class was associated with an increased risk of a hazardous pattern of alcohol use (Howden-Chapman and Tobias 2000). There was some suggestion that lower educational attainment was associated with a more hazardous pattern of alcohol use, but this was not statistically significant. Earlier New Zealand work indicated that the pattern of drinking "high quantities but less often" was more common in the lower social classes, along with an increased prevalence of high daily consumption (Casswell and Gordon 1984).

## Hazardous alcohol use associated with high SES

The review by Moller and Tonnesen (1997) reported a French study indicating that high-SES women were more likely to be regular consumers of alcohol, and that in Sweden high alcohol consumption was associated with high SES among older people. The review by Colhoun et al. (1998) on hypertension and SES reported that higher alcohol consumption among high-SES women had been identified in studies in Australia and England (two studies) and in the INTERSALT study, where some of 52 centres collected data on alcohol.

A study in Wales revealed a tendency towards a higher prevalence of consumption of 22 drinks or more per week in the higher social classes (Farrow et al. 1988). A study of three regions in the United Kingdom
reported that weekly alcohol consumption increased with household income in both men and women (Crawford 1988), but that manual workers drank more alcohol than non-manual workers. In the United States, a combined analysis of 10 surveys showed some positive gradient for the frequency of alcohol drinking with SES (Knupfer 1989). However, there was a weakly inverse gradient for "frequent drunkenness". Among employees in Minneapolis-St Paul, the frequency of alcohol drinking was associated with high SES in women but not in men (Jeffery et al. 1991). A national survey in New Zealand (Statistics New Zealand/Ministry of Health 1993) reported a higher alcohol intake with greater educational attainment, in contrast to the other results from New Zealand detailed above.

## No association with SES

The review by Colhoun et al. (1998) reported on two British studies that showed no SES gradient with alcohol consumption. Similarly, there was no association with social class among a sample of attendees at a London health centre (King 1986). In New Zealand, a child cohort study found that family SES did not have a marked impact on drinking behaviour, but that those from low-SES families in this cohort drank more at age 15 years (Connolly 1992).

The international review of alcohol drinking, social class and cancer by Moller and Tonnesen (1997) found no evidence for such an association for Japan, except that women in the highest social classes had slightly elevated mortality from liver cirrhosis. There was a similar finding for Finland and for Sweden, except that pharyngeal cancer in Swedish women was more prevalent in low-SES women. Nevertheless, these two countries do show some evidence for SES gradients in hazardous alcohol use (as detailed above).

## DISCUSSION

## The relationship

The general pattern in developed countries is for lower-SES men to have a more hazardous pattern of alcohol consumption in terms of high or excessive intake of alcohol. This pattern is also evident in the distribution of alcohol-related disease, particularly alcohol-related cancers. The pattern for women can resemble that for men, although there are also many studies indicating excessive alcohol use among high-SES women. Different studies in the same countries sometimes show contrasting patterns that may reflect temporal trends, regional variation and different methodological approaches.

The pattern is also mixed in developing countries, although in this review more studies were identified that indicated that lower SES was associated with a more hazardous pattern of alcohol consumption. This pattern is similar to that found for tobacco use, but it does contrast with
the SES gradient seen for other disease risk factors where high SES is associated with increased risk (e.g. for cholesterol level, obesity, blood pressure and to some extent physical inactivity).

## Possible mechanisms

Some of the factors likely to be involved in the excessive use of alcohol by lower-SES groups include:

- cultural influences, including the presence of greater numbers of alcohol users in certain areas and the social acceptability of heavy drinking;
- the existence of psychological stressors associated with poverty and inequality from which alcohol users may seek to escape;
- access to cheap alcohol, especially in rural areas of the developing world and where alcohol is not taxed; and
- the effect of lower educational attainment on poorer knowledge of the hazards of excessive alcohol use.

High levels of alcohol use among higher-SES groups (especially women) may reflect in particular cultural patterns and the role of additional income in settings where alcohol is relatively expensive. The alcohol industry may also target advertising at those with the highest incomes. It is also plausible that alcohol is used as a coping mechanism by high-SES women who have heavy demands on their time.

Some of the studies in this review also reported higher levels of abstinence among lower-SES groups. This behavioural pattern may have health benefits in those aged $<50$ years (e.g. in terms of injury prevention), but abstinence could be considered a potential risk factor when considering the evidence for cardioprotective effects of moderate alcohol consumption in older populations (Rimm et al. 1996).

### 4.3 ILlicit DRUG USE

## LITERATURE ON DEVELOPING COUNTRIES

## Substance use associated with poverty

In an Indian city, substance abuse by adolescents (covering tobacco, alcohol and cannabis) was more common among those from Hindu families with low educational status and low family income (Kushwaha et al. 1992). Similarly, in Nepal, a small qualitative study suggested that urban poverty was a factor in substance use (Jutkowitz et al. 1997). In Brazil, youths living on the streets were significantly (eight times) more likely to use drugs (generally inhalants and marijuana) than those living at home (Pinto et al. 1994). Other work has documented relatively heavy drug use by street children in Brazilian cities, such as half of $7-8$-yearold street children in São Paulo (Carlini 1990). A study of military con-
scripts in Kuwait found that amphetamine use was associated with lower educational attainment and unemployment (Bilal et al. 1992).

## Substance use associated with high SES

A survey of students in an Indian city found that regular cannabis users "were mainly from professional colleges, hailing from metropolitan cities and with relatively higher amount of pocket money at their disposal" (Reddy et al. 1993). Also, a study of marijuana-related hospital admissions in Malawi found that the marijuana-abusing group had significantly more schooling than the control group (Carr et al. 1994).

## No relationshiplinadequate information

In Zimbabwe, there were no particularly pronounced differences in the use of inhalants and cannabis by students of different SES subgroups, in contrast to that of tobacco and alcohol (Eide and Acuda 1996). A review of the use of non-prescription psychoactive substances in Nigeria reported that: "as a result of insufficient information, it could not be established who uses the substances more among different age, religious, social class, educational, and occupational categories" (Omoluabi 1995).

## LIterature on developed countries

## Substance use associated with low SES

A United States study reported that people living in areas with high rates of poverty had relatively high rates of illicit drug use (marijuana or cocaine/crack) in the past year compared to a national sample (Ensminger et al. 1997). Also, frequent cocaine use in the United States was reported to be more prevalent in low-income urban areas than elsewhere (Brownsberger 1997). Another United States study found that social disadvantage was moderately associated with drug-related behaviour (Boardman et al. 2001). This work suggested that this was related to increased social stressors and higher levels of psychological distress among residents of disadvantaged neighbourhoods. Supporting evidence for such an association also comes from United States data on the early wave of injecting drug users with AIDS, who were characteristically poor or from ethnic minorities (Schrager et al. 1991).

Poverty and lower levels of education were associated with marijuana use in pregnancy in a United States review (Day et al. 1993). This review also reported work that found that cocaine use in pregnancy was associated with low educational attainment. In another United States review (Hans 1999), the studies reviewed suggested similar rates of substance use during pregnancy by women of different racial and social class categories (for all substances including alcohol). However, it did report that black women and poorer women were more likely to use illicit substances, particularly cocaine. Another United States review of
substance use in pregnancy also noted its association with poverty (Howell et al. 1999). A more recent study found that neighbourhood poverty was associated with higher rates of the use by pregnant women of cocaine, opiates, amphetamines and all illicit drugs (Finch et al. 2001).

A Swedish study suggested that low SES at the time of birth, relative to the general population, might be associated with dependence on amphetamines as an adult (Nyberg et al. 1992). However, this relationship did not exist for opiate dependence. In New Zealand, there is evidence that the Maori population has a significantly greater rate of cannabis use (Fergusson and Horwood 2000) and a relatively high rate of treatment for cannabis use (Adamson et al. 2000; Pomare et al. 1995). Similarly, Australian school survey data show that a relatively disadvantaged ethnic group (Aboriginal and Torres Strait Islander students) were more likely to have tried cannabis and other illicit substances (as well as cigarettes) than other students (Forero et al. 1999).

In terms of educational attainment and substance use, student survey data in the United States indicate that lower parental educational level was significantly associated with higher rates of substance use (Robinson et al. 1987), although "substance use" covered alcohol and tobacco as well as illicit drugs. A New Zealand cohort study found that adolescents who had less well educated mothers had significantly higher "deviant peer affiliations", i.e. associations with delinquent or substanceusing peers (Fergusson and Horwood 1999). These substances included alcohol, tobacco and other drugs.

## Substance use associated with high SES

Some United States research suggested higher rates of substance use among affluent, suburban or white youths than among poor, non-white, urban youths (Luthar and D'Avanzo 1999). Other work suggested minimal SES patterning for substance use: one United States study reported that illicit drug use was only 1.3 times more likely in disadvantaged neighbourhoods compared to the least disadvantaged ones (Saxe et al. 2001). However, visible drug sales were 6.3 times more likely to be reported in these disadvantaged neighbourhoods.

## DISCUSSION

## The relationship

The available literature in general suggests a relationship between SES disadvantage and increased prevalence of use of illicit substances. This relationship appears to exist in both developing and developed countries. Nevertheless, some studies suggest that, for some substances in some settings, the opposite relationship exists. A further complication is that not all the studies on this relationship appropriately adjust for relevant confounding variables such as ethnic status and religious affiliation. In India,
for example, caste was described as being a factor in determining substance use (Sharma 1996).

## Possible mechanisms

Many factors are likely to be significant in the relationship between poverty, low SES and drug use. For example, Finch et al. (2001) describe how neighbourhood poverty may increase substance use in a number of ways: ". . . the presence of greater numbers of substance users, access to substances, greater overall deviance, the social acceptability of substance use ...". These authors also consider that contextual effects might be involved: " $\ldots$. the increased stressors of poor housing conditions, a lack of health care and substance use services, and a general lack of social services within the community" (Finch et al. 2001). Specific United States data highlight the importance of exposure, with youths living in the most disadvantaged neighbourhoods being six times more likely to have been offered cocaine (Crum et al. 1996).

In drug production areas, it is likely that poor farmers and their neighbours will have relatively easy access to these substances at low cost. Evidence from Malawi suggests that the low cost of marijuana is a factor influencing its usage (Carr et al. 1994). Also, poverty may sustain illicit drug production in settings where other cash crops are less profitable.

On the demand side, it is plausible that lower educational attainment is associated with poorer knowledge of the hazards of illicit drug use. Also, the hardship of poverty may be associated with a greater demand for the temporary psychological escape achievable through drug use. Self-perceptions of low status may also be relevant, given that low social status may affect dopamine receptors in the brain that increase the addictiveness of some drugs (Morgan et al. 2002).

### 4.4 Hypertension/high blood pressure

## Literature on developing countries

A narrative systematic review covering literature published between 1966 and 1996 identified 13 studies in developing countries (Colhoun et al. 1998). Of these, a direct relationship between higher SES and higher blood pressure (BP) or hypertension was reported in five studies in India (three studies), Kenya and Nigeria. Another Indian study reported a non-significant direct relationship. The opposite pattern (of low SES associated with higher BP or hypertension) was reported in four studies in Brazil, India and Senegal and among South African Zulus. A further Indian study reported no relationship, while two studies in the Caribbean found conflicting results for males and females. Another review of cardiovascular disease risk factor data from 12 centres in seven developing countries also examined SES (three centres in Thailand, two each in Brazil, Chile and China and one each in

Colombia, Indonesia and the Philippines) (INCLEN 1994). For BP the associations with SES tended to be negative (i.e. similar to those from the developed world).

## High BP associated with high SES

A number of studies conducted in Asia and Africa showed this pattern, e.g. villagers in India (Singh et al. 1997b), urban residents in India (Singh et al. 1998), urban workers in Nigeria (Ekpo et al. 1992) and urban civil servants in two Nigerian cities (Markovic et al. 1998; Omokhodion et al. 2000). However, while another Nigerian study found this pattern for men it found the reverse for women, i.e. higher BP in those with less educational attainment (Kadiri et al. 1999). There was a positive association between BP and income among a rural population in Papua New Guinea, but this held for women only and did not reach statistical significance. Another Chinese study found that men paid according to a new, highly demanding salary system exhibited significant increases in systolic BP between the first and second screenings (Siegrist et al. 1990).

## High BP associated with low SES

Data from an urban population in China indicated that high BP and hypertension in men and women were associated with low SES (Yu et al. 2000). Over a seven-year period the SES gradient declined for men (systolic and diastolic BP). Over this period there was also an increase in systolic BP among better-educated women, while the reverse occurred in women with less education. A study in Saudi Arabia also found an association between high BP or hypertension and low SES (Wahid Saeed et al. 1996).

## Literature on developed countries

The narrative systematic review by Colhoun et al. (1998) identified 57 relevant studies in developed countries. The reviewers concluded that lower SES was associated with higher mean BP in almost all studies in developed countries. However, the magnitude of the association varied and was generally quite small (with age-adjusted mean systolic BP differences of about $2-3 \mathrm{mmHg}$ between the highest and lowest SES groups). Even so, this difference is still important in terms of its consequences for public health at the population level. This inverse gradient was both stronger and more consistently found in women than in men. An earlier review (the INTERSALT study) examined data on BP and educational attainment from 47 centres around the world (in 18 developed and 12 developing countries) (Stamler et al. 1992). It found an inverse association for men in 28 centres and for women in 38 centres. Ageadjusted systolic BP in men was significantly higher $(1.3 \mathrm{mmHg}$ higher for 10 fewer years of education), while for women the association was even more significant at 4.5 mmHg higher. Yet when the adjustment
included five lifestyle factors (24-hour sodium and potassium excretion, BMI, alcohol intake and smoking) these estimates were reduced by about $50 \%$. Furthermore, after this adjustment the inverse association was no longer significant for men. The results obtained for diastolic BP were similar to those for systolic BP. Of ten studies on children in developed countries, only four reported significant differences in BP associated with SES (Colhoun et al. 1998). All four showed that systolic BP was higher in children of low SES, although in two the association disappeared after adjustment, e.g. for weight or fitness.

## DISCUSSION

## The relationship

In developing countries there is a mixed picture, although slightly more studies have suggested a direct association between high SES and high BP. This pattern is consistent with that generally seen for the SES distribution of other cardiovascular disease risk factors in developing countries (obesity, hypertension, cholesterol level and to some extent physical inactivity). Nevertheless, in a number of developing countries the pattern is similar to the fairly consistent pattern in the developed world, with high BP associated with low SES. The pattern seen in developed countries is consistent with the evidence that mortality from hypertensionrelated diseases such as ischaemic heart disease, hypertensive heart disease, stroke and end-stage renal disease also shows an association with low SES (Colhoun et al. 1998).

## Possible mechanisms

Colhoun et al. (1998) suggested that the SES gradient in BP seen in the developed world was unlikely to be explained by differential treatment rates. Indeed, these authors considered that a substantial part of it was accounted for by the SES gradient in BMI (see also section 4.7). Also, alcohol consumption across SES groups accounted for part of the association in men (although this issue was examined in only a few studies). Furthermore, these reviewers considered there to be little evidence that adverse psychosocial factors associated with low SES caused chronic elevations in BP. Nevertheless, they noted that stronger relationships might exist between BP and more specific stressful aspects of low SES such as unemployment or job insecurity.

The INTERSALT study reported that those with less education had on average higher sodium excretion, lower potassium excretion, greater body mass and higher alcohol intake (Stamler et al. 1992). All of these are factors that tend to be associated with increased BP.

For some developing countries, the review by Colhoun et al. (1998) suggested that the direct association between higher SES and higher BP may reflect a higher prevalence of obesity and higher salt and alcohol intakes among those of higher SES. Other dietary factors (e.g. more
animal products) and less physical activity in these groups might also play a role. Furthermore, a Brazilian review concluded that the higher people's "cultural consonance" the lower their BP (with cultural consonance referring to how individuals are able to approximate, in their own behaviour, shared cultural models of life) (Dressler and Santos 2000).

In those developing countries where low SES is associated with high BP , poor maternal and perinatal nutrition may be important factors. Indeed, there is growing evidence from animal studies (Edwards et al. 1999; Ozanne 2001) and human studies (Roseboom et al. 2001) for the Barker hypothesis on the fetal origins of adult disease.

Regardless of the mechanisms involved, it is important to note that there is evidence that the intensive medical management of hypertension can abolish SES differences in hypertension-associated mortality (Anonymous 1977).

### 4.5 Serum cholesterol and lipid levels

Literature on developing countries

## Adverse lipid profile associated with high SES

Data from 12 centres in seven countries (the INCLEN study) found such a relationship for total cholesterol (TC) in six centres. These included centres in Brazil (for occupational class and income); Chile; China (for occupational class); the Philippines (for education, occupational class and income); and Thailand (one centre for income and one for education) (INCLEN 1994).

China. A study of an urban population reported that high-SES people had a more unfavourable serum lipid profile (for TC, HDL-c, LDL-c and triglycerides [TG]) than those in lower-SES groups (Yu et al. 2002). This significant association was especially apparent in men, and education seemed to be the most important predictor of serum lipids compared to the other SES indicators of occupation and income. A cohort study of urban male workers found that blood lipid profiles were less favourable in the better-educated groups (Siegrist et al. 1990). LDL-c and LDL-c/HDL-c ratio were significantly higher and HDL-c was significantly lower. Similar results were obtained using occupational class.

India. A study of an urban population found a significant SES gradient for both men and women for hypercholesterolaemia, TC, LDL-c, HDLc and TG (Singh et al. 1998). Similarly, surveys of women in five Indian cities found positive gradients by social class for TC and HDL-c (Singh et al. 1998). A study of an elderly urban population also found an association between hypercholesterolaemia and income (Singh et al. 1995). In rural India, one study found a significant SES gradient for hypercholesterolaemia, for TC and for TG in women (Singh et al. 1997a), but the
trends for higher LDL-c and HDL-c in both men and women were not significant.

Other countries. In a white rural South African population, the prevalence of hypercholesterolaemia increased with income for men but there was no association with education or between HDL-c and SES (Rossouw et al. 1990). Low HDL-c was more prevalent in women (by income and educational attainment) but, in contrast to the men, there was less hypercholesterolaemia with increasing level of income.

A study in Mauritius found that professional and skilled men had significantly higher mean TG, higher LDL-c and lower mean HDL-c than unskilled workers (Pereira et al. 1998). Unskilled and partly skilled women had lower TG compared with "homemakers". Also, in contrast to the men, professional and skilled women had higher HDL-c than unskilled women (with no significant association for LDL-c).

A study in Turkey found an association between income and TC in urban men and women and rural men, but not rural women (Onat et al. 1992). Another Turkish study found higher TC and HDL-c levels in high-SES than in low-SES children (Mahley et al. 2001). The high-SES group also consumed more saturated fat of animal origin and less carbohydrate.

## Lipid profile and development

Cholesterol levels in the population appear to increase with the modernization and economic development of a country. For example, a study of adult men in 13 countries in 1980 reported that mean TC levels were low in those from Africa, intermediate in those from Hungary, Pakistan, the Philippines, Poland, Suriname and the Mediterranean countries, and high in those from Finland and the Netherlands (Knuiman et al. 1982). HDL-c tended to be lower in men from Africa, Asia and Suriname than in those from Europe (with the highest values for both TC and HDL-c being found in Finnish men). The HDL-c:TC ratio was on average slightly higher in those from Africa than it was in those from Europe and from Asia and Suriname.

Another multi-country study found low TC levels in Amazonian Indians and Africans (traditional Bantu population) compared to Italian and Polish populations (Pavan et al. 1999). Other studies found relatively low TC levels in other traditional societies with diets based on complex carbohydrates and vegetables (Mancilha-Carvalho and Crews 1990). Even in developing countries, such a Zimbabwe, there appears to be an increase in TC levels in urban relative to rural populations (Allain et al. 1997).

## Adverse lipid profile associated with low SES

Some of the results for specific aspects of the lipid profile detailed in studies described above fit this pattern (e.g. low-SES white rural South

African women and hypercholesterolaemia, and unskilled women in Mauritius and HDL-c). Furthermore, data from the INCLEN study revealed such a relationship in one centre (Shanghai in China for low occupational class and raised TC) (INCLEN 1994). In India it was reported that TC and LDL-c were high in urban slum dwellers (Misra et al. 2001), but this study did not involve a comparison group.

## No association (lipid profile and SES)

Data from the INCLEN study found no significant relationship for TC in five centres in Brazil, Chile, Colombia, Indonesia and Thailand. In a rural population in western India there was no association of SES with the prevalence of hypercholesterolaemia (for TC, LDL-c, HDL-c and TG) in men and women by educational status. Similarly, SES was not related to HDL-c levels in adults, in contrast to the results for Turkish children detailed above (INCLEN 1994).

## Literature on developed countries

A review of cardiovascular disease risk factors and SES by Pickering (1999) reported that the SES gradient for TC was generally minimal or non-existent. He reviewed studies in England (two studies), Norway (one study) and the United States (four studies). However, further work has subsequently been published, some of which is described below.

## Adverse lipid profile associated with high SES

In a multi-country study (mainly in developed countries) the lipid profile was worse with increasing years of schooling for men in China, Poland and the Russian Federation (for raised TC, LDL-c and TG and lower HDL-c) (Perova et al. 2001). However, the findings were less consistent for women and for Israeli and American blacks of both sexes. Also, the opposite pattern was seen among white men from the United States.

A study in Scotland found that TC levels were higher in the nonmanual social class groups, for both men and women (Davey Smith et al. 1998). A study of working Japanese men showed that those in higher employment grades had lower levels of HDL-c (Martikainen et al. 2001).

A study of an economically depressed agricultural area in the United States revealed that those with low educational attainment and income below the poverty level had lower mean TC levels than the rest of the population in the area (Gold and Franks 1990), despite the fact that the low-SES groups in this population had a greater intake of dietary cholesterol.

## Adverse lipid profile associated with low SES

Perova et al. (2001) showed that white men in the United States with low educational attainment had an adverse lipid profile. Similarly, a national survey in the United States found that for whites, HDL-c levels
were highest for those in the highest category of earnings (although blacks generally had lower levels of HDL-c with increased earnings) (Linn et al. 1989). In a multivariate model, predictors of higher HDL-c included a higher frequency of alcohol intake and reported high physical activity, with smoking and high BMI being strongly negatively related to HDL-c levels. The authors of this national study suggested that the findings "support previous findings in selected populations in the United States".

In Finland, cholesterol levels were found to be higher in blue-collar than white-collar workers, for both men and women (Vartiainen et al. 1998). This study found that the decline in cholesterol levels between 1972 and 1987 was similar for men in both groups, but greater among white-collar women than blue-collar women. Similarly, in Norway, poverty during childhood was positively associated with age-adjusted levels of TC in adults (Arnesen and Forsdahl 1985). In Sweden, women of high SES were found to have a better lipid profile than low-SES women (lower levels of TC, LDL-c and TG and raised HDL-c), but there was no relation between occupational status and lipid levels for men.

In the Whitehall II longitudinal study of men and women in the United Kingdom, it was found that high occupational status was associated with lower cardiovascular disease risk factors, including HDL-c levels (Brunner et al. 1999). This study also reported that social position in childhood was associated with HDL-c level in women. Another study of working men in England found that higher occupational class was associated with a favourable lipid profile (for HDL-c and TC) (Martikainen et al. 2001).

In Greece, TC levels were found to be inversely associated with educational level in both sexes, and especially in women (Benetou et al. 2000). The authors report that this SES pattern contrasted with the previous pattern that existed two decades before in Greece. HDL-c levels were inversely associated with educational level in men, but the association was not clear for women.

## No association (lipid profile and SES)

In a German study, no poverty-related differences were found for prevalence of hypercholesterolaemia, despite a much higher prevalence of obesity in persons with an income below the poverty line (Helmert et al. 1997). In a national survey in New Zealand there was no overall SES gradient in TC levels (Howden-Chapman and Tobias 2000). In European men and Maori women, however, relatively higher TC levels were seen with high SES while, in contrast, higher TC levels were seen in lowSES European women (based on educational attainment).

TC levels for children in a disadvantaged inner-city population in the United States were not significantly different from those of other children from several large studies of North American populations (Wadowski et al. 1994).

## DISCUSSION

## The relationship

In the developing world, the predominant pattern is for the population as a whole to develop an adverse lipid profile along with industrialization and urbanization. Furthermore, within developing countries, highSES groups appear to have the most hazardous lipid profiles. This pattern is consistent with that generally seen for the SES distribution of other cardiovascular disease risk factors in developing countries (obesity, hypertension and to some extent physical inactivity).

The picture is more mixed in the developed world, but the predominant pattern is for low-SES groups to have the most adverse lipid profiles (particularly in the most developed countries in this group). Such a pattern is also consistent with that generally seen for the SES distribution of other cardiovascular disease risk factors in developed countries (smoking, obesity, hypertension and physical inactivity). It is also consistent with the relatively higher rates of cardiovascular disease among low-SES groups in many developed countries.

## Possible mechanisms

In the developing world, it is likely that dietary change is a key factor in the development of adverse lipid profiles associated with industrialization and urbanization. Higher total calories and increased intakes of animal products are probably important. Nevertheless, other factors may be relevant: a Brazilian study found significant interactions between elevated cholesterol levels and smoking, obesity and physical inactivity (Martins et al. 1995).

In developed countries, the generally poorer diet of low-SES groups is also likely to be a factor in generating adverse lipid profiles. For example, increased dietary intake of cholesterol was described among various United States populations, including poor elderly people (Prothro and Rosenbloom 1999), poor children (Casey et al. 2001) and children whose parents smoked (a crude marker for low SES) (Johnson et al. 1996). Similarly, the higher BMI of low-SES groups is likely to play a role: a 13country survey by Knuiman et al. (1982) found a direct relationship of BMI with TC and an inverse relationship with HDL-c and HDL-c:TC ratio.

Physical inactivity may also be relevant (given that it contributes to obesity) and there is some evidence that it is associated with a poorer lipid profile, although studies have reported conflicting results (Sowers et al. 1995). Regular consumption of modest quantities of alcohol may also play a role in improving the lipid profile of high-SES groups. For example, a recent French study found that wine drinkers had higher HDL-c levels than non-wine drinkers (Ruidavets et al. 2002). However, the authors reported that such differences became non-significant after adjustment for SES parameters.

Furthermore, there may be psychosocial pathways for determining lipid profiles in different populations. For example, a case-control study in England suggested that "lifestyle incongruity" is associated with higher cholesterol levels (Dressler et al. 1992). This variable is a type of status incongruence, involving the degree to which lifestyle (material consumption and status-enhancing behaviour) exceeds occupational status.

Low birth weight may play a role in influencing lipid metabolism in the adult. This is one component of the still controversial "Barker hypothesis" discussed in section 4.4.

In some developed countries there may also be SES patterning of cholesterol screening, awareness of the risks of elevated cholesterol levels, willingness to change one's diet to reduce cholesterol levels and access to cholesterol-lowering drugs.

### 4.6 Overweight/obesity

## Literature on developing countries

## Obesity associated with high SES

A review of 144 published studies on the relationship between SES and obesity was published by Sobal and Stunkard (1989). It revealed a strong direct relationship between SES and obesity in women in 10 studies in developing and non-Western societies (with another study showing no relationship and one showing the opposite relationship). For men, 12 studies found a relationship and two found no relationship. In children and adolescents, this pattern was seen in 14 studies in girls (with two other studies showing no relationship) and in 13 studies in boys (with two others showing no relationship). A subsequent review by Colhoun et al. (1998) reported that BMI and SES were measured in seven out of 13 studies on BP and SES in developing countries. In four of these studies-in rural India (two studies), rural Kenya and urban NigeriaBMI increased with higher SES. In Senegal, BMI did not vary with SES and in two studies the relationship of SES to BMI was not reported. These reviewers concluded that, overall, the SES differences in BMI "probably account for a substantial part, though not all, of the SES-BP association".

The INCLEN study collected cardiovascular disease risk factor data on men from 12 centres in seven developing countries (INCLEN 1994). It found significant trends for increasing BMI with higher SES in five out of the 12 centres. These were in urban China (by occupation and income), urban Indonesia (by education), urban Philippines (by education, occupation and income), rural Thailand (by education and income) and urban Thailand (by occupation and income). A recent analysis of obesity in women based on nationally representative surveys ( 39 surveys from 38 developing countries) (Martorell et al. 2000) reported that in developing countries, such as those in sub-Saharan Africa, obesity levels were higher among better educated women and urban women. Yet in
more developed countries, such as those in Latin America and central and eastern Europe, obesity levels were more equally distributed in the general population. At a national level, levels of obesity in countries increased sharply up to a GNP of US\$ 1500 per capita (1992 values) and then changed little thereafter.

Another multi-country study of developing and developed countries used data for children aged 6-18 years from nationwide surveys in China, the Russian Federation and the United States (NHANES III, 1988-1994) (Wang 2001). It found that higher-SES subjects were more likely to be obese in China and the Russian Federation, while in the United States low-SES groups were at higher risk. A review of preschool children by Martorell et al. (2000) examined 71 national nutrition surveys since 1986 from 50 developing countries. It found that in 22 countries (mainly those from Latin America, the Caribbean, the Middle East and North Africa) overweight was significantly more common in children of mothers with higher education. Overweight was also significantly more common in urban areas in 24 countries (and only more common in rural areas in two countries). At national level, the prevalence of overweight among preschool children tended to increase with increasing GNP ( $r=0.28, P=0.05$ ).

In China, a nationwide longitudinal survey conducted in 1989 and 1991 (Popkin et al. 1995a) found evidence that increased household income was significantly associated with increased BMI. Urban residence and higher income were associated with lower energy intake, higher fat intake and lower levels of physical activity compared to rural residence and other income categories. Diet was considered to be a particularly important determinant of body weight in this population (Paeratakul et al. 1998).

## Obesity associated with low SES

The INCLEN study (INCLEN 1994) found significant trends for increasing BMI with lower SES in two out of 12 centres. These were in urban Colombia (by education and occupation) and rural Thailand (by occupation). The review by Sobal and Stunkard (1989) found no studies reporting an association between obesity and low SES. Similarly, the review of overweight among preschool children from 50 developing countries found none in which overweight was significantly more common in the children of poorly educated women (Martorell et al. 2000).

A review of obesity in developing countries (Popkin et al. 1995b) cited evidence for increased obesity among low-SES groups from urban Brazil. Another recent review by Pena and Bacallao (2000) reported such a finding for urban Chile, for women in Uruguay and for urban Peru. A book on obesity in Latin America described the higher prevalence of obesity in middle-income compared to high-income women (Monteiro 2000).

A significant gradient was described for BMI and obesity prevalence by education level for urban women in China (i.e. greatest among those with the least education) (Yu et al. 2000). Over the period 1989-1996 this gradient declined for women, along with a decrease in BMI across all educational strata. The results for men suggest no significant gradient in BMI and obesity and no significant time trends.

## No association (obesity and SES)

The review by Colhoun et al. (1998) reported one country (Senegal) where there was no association found between BMI and SES. Also, the INCLEN study (INCLEN 1994) found no significant associations for BMI and SES in five out of 12 of the centres. These were in Brazil (two sites), Chile (two sites) and China. The review of overweight among preschool children from developing countries by Martorell et al. (2000) found no association with maternal education in 28 out of 50 countries. The review by Sobal and Stunkard (1989) found no association in one out of 11 studies of women. The equivalent figures for men, boys and girls, respectively, were $2 / 14,2 / 15$ and $2 / 16$. One multi-country study examined the prevalence of the coexisting underweight and overweight individuals in households using survey data from Brazil, China and the Russian Federation (Doak et al. 2000). It found no clear pattern in the prevalence of underweight/overweight in households by income, but this coexistence pattern was highest in the urban environment in all three countries.

## Literature on developed countries

## Obesity associated with low SES

The review by Colhoun et al. (1998) reported that adjustment for BMI in 26 studies (along with other variables) usually attenuated observed SES gradients in BP (increasing BP with lower SES). Furthermore, in five studies in Germany, the Netherlands, Sweden and the United States (two studies), in which adjustment was made for age and BMI only, this SES gradient was abolished completely. Some of the studies described showed stronger gradients for women compared to men for increased BMI and low SES (Shewry et al. 1992) and in some settings the association applied only to women. Similarly, the INTERSALT study found that increased BMI was significantly associated with lower SES, particularly for women (Stamler et al. 1992).

Surveys associated with WHO's MONICA (MONItoring of CArdiovascular diseases) Project collected data on BMI and years of schooling for 26 populations in the period 1979-1996 ( 25 developed countries plus China) (Molarius et al. 2000). An analysis of the combined data found that, for women, almost all populations ( 22 out of 26 ) showed a statistically significant association between lower educational level and increased BMI (the difference between the highest and the lowest edu-
cational tertiles ranged from -3.3 to $0.4 \mathrm{~kg} / \mathrm{m}^{2}$ ). For men, there was the same association in six out of 26 of the populations (with the equivalent range being from -1.5 to $2.2 \mathrm{~kg} / \mathrm{m}^{2}$ ). In about two thirds of the populations, the differences in BMI between the educational levels increased over the 10 -year period.

The earlier review of 144 published studies on SES and obesity by Sobal and Stunkard (1989) also revealed a strong inverse relationship among women in developed societies. There were 28 United States studies showing this relationship for women and 18 from other developed countries (with only seven studies showing no relationship). For men there were 12 United States studies and 22 studies from other developed countries that found this relationship (while 20 found the opposite relationship and 11 found no relationship). In children and adolescents this pattern was seen in 13 studies for girls (with eight showing the opposite relationship and 11 showing no relationship) and similarly in 11 studies in boys (with nine showing the opposite relationship and 14 showing no relationship).

More recently, a systematic review by Parsons et al. (1999) found that in developed countries there was a strong consistent relationship observed between low SES in early life and overweight in adulthood. These reviewers found that women who changed social class (social mobility) adopted the prevalence of obesity of the class they joined, while this association was not present in men. More recent survey data from various developed countries is consistent with the relationship described above of increased BMI being associated with low SES, for example New Zealand (e.g. for education) (Dryson et al. 1992; Howden-Chapman and Tobias 2000; Statistics New Zealand/Ministry of Health 1993); the United Kingdom (for occupational class among men and women) (Acheson 1998); and the United States (for education among men and women) (Pamuk et al. 1998).

## Obesity associated with high SES

The review by Sobal and Stunkard (1989) found this relationship for men in three United States studies and in eight studies from other developed countries. It was found in only one study of women (among migrants in Belgium). In the 26 MONICA Project populations this association was reported in two populations for men (in Poland and the Russian Federation) and none in women (Molarius et al. 2000). Among children, this relationship was reported for eight studies in girls and nine in boys.

## No association (obesity and SES)

The review by Sobal and Stunkard (1989) found no association between SES and obesity in men in 11 studies, in women in seven studies, in boys in 14 studies and in girls in 11 studies. Of the 26 MONICA Project surveys, 18 found no significant association between BMI and SES in
men and four found no such association for women (Molarius et al. 2000). Similarly, the lack of an association was described for many studies for men and children in the earlier review by Sobal and Stunkard (1989).

The systematic review by Parsons et al. (1999) found no clear relationship between SES in early life and overweight as a child, even though a relationship was reported for low SES in early life and overweight as an adult. These reviewers reported that very few of the relevant studies investigating SES considered confounding by parental obesity. Such findings are consistent with an examination of United States survey data on 12 states between 1980 and 1989, which found in general that children from low-income families did not have a greater prevalence of overweight than children from higher-income families (Yip et al. 1993). (In two states, however, low-income school-aged children and adolescents did show significant increases in body weight in relation to height.)

## DISCUSSION

## The relationship

A very large body of evidence from studies in developed countries indicates that the predominant gradient is that of lower SES being associated with increased body weight. This relationship appears very strong and consistent for women, but only six of the 26 MONICA Project surveys found such a relationship for men. Also, a systematic review found no clear relationship between SES in early life and childhood obesity, although a relationship was reported for low SES in early life and adult obesity.

The predominant pattern in developing countries is for higher-SES men and women to have higher body weight. Nevertheless, some studies have found no relationship and in some regions (e.g. Latin America) the opposite pattern appears to be emerging. Among preschool children there is also evidence that being overweight was significantly more common in those with mothers with higher education (in 22 out of 50 countries).

## Possible mechanisms

In developing countries the association between BMI and higher SES is probably related to a "nutrition transition" (Popkin 2001) to diets higher in fat. A trend towards poorer diets (e.g. increased dietary fat) with increasing income was described for China and the Philippines in a review by Popkin et al. (1995b). Societal attitudes to obesity and thinness, as described by Sobal and Stunkard (1989) may also be relevant in determining dietary restraint and participation in physical activity. For example, mild obesity appears to be equated with physical attractiveness in some countries. Fetal and infant nutrition may also be relevant in determining subsequent obesity, although a recent review of this evidence
in a developing country context suggests the relationship is not as clear for obesity as it is for hypertension (Dryson and Martorell 2000). Reduced physical activity was considered to be a relevant factor for obesity in high-SES groups (see also section 4.7), but reduced physical activity might also be an important factor for obesity in low-SES groups (Torun 2000).

In developed countries it is likely that nutrition also plays a key role in the association between low SES and increased BMI. A review by Potter (1997) reported evidence of poorer nutrition with low SES in a number of developed countries: Australia (three studies), Canada, Denmark, Finland and the United States (four studies). In all these studies the dietary pattern was suggestive of that associated with the greatest risk of excessive calorie intake (i.e. low in vegetables, fruit or wholegrain cereals and/or high in fat, meat, butter or whole milk). In particular, the relationship between dietary fat intake and obesity was fairly well established (Bray and Popkin 1998). Similarly a meta-analysis of 11 studies from seven European countries found lower vegetable and fruit intake in low-SES groups (using education and occupation) for both men and women (Irala-Estevez et al. 2000). Data from other studies provide further evidence for this pattern, for example for New Zealand (Howden-Chapman and Tobias 2000) and the United Kingdom (Acheson 1998). The SES gradient for breastfeeding, e.g. in the United Kingdom (Acheson 1998), might also be relevant in determining subsequent BMI in childhood.

In some developed countries, physical inactivity might well play a role in the SES-BMI gradient (see section 4.7). Other explanations include "a complex mix of poverty of information and skills, a distorted view of "status foods" and an approach to dietary priorities typical of the 1940 s, which was then much more focused on protein and calories" (Potter 1997). Furthermore, cultural attitudes to thinness and dietary restraint may be important (Sobal and Stunkard 1989).

When considering potential explanations for the relationship between SES of origin and adult obesity, Power and Parsons (2000) suggested that SES of origin may actually be confounded by parental body size. Alternative explanations include "(1) nutrition in infancy and childhood, either over- or undernutrition, followed subsequently by over nutrition; (2) psychological factors, possibly involving emotional deprivation in childhood; (3) cultural or social norms regarding dietary restraint and attitudes to fatness that may be acquired during childhood".

### 4.7 Physical inactivity

Literature on Developing countries

## Physical inactivity associated with low SES

In an urban Chinese population, low level of education was found to be associated with leisure-time physical inactivity in both men and women
(Yu et al. 2000). However, this gradient declined over the seven-year study period, particularly for men. Other Chinese survey data indicate that urban residents are far more sedentary than rural residents (Paeratakul et al. 1998). A study of a metropolitan area in Brazil found that high-SES women were less likely to be physically inactive (Martins et al. 1995). This was also the pattern for the skilled working class relative to the unskilled working class. A study in a Caribbean island reported that low-SES men and women (based on educational attainment) were reported to exercise less often (Grol et al. 1997b).

## Physical inactivity associated with high SES

In India, physical inactivity was reported to be more prevalent among those of higher SES in women in five cities (Singh et al. 1999), in men and women in villages (for SES based on occupation) (Singh et al. 1997b) and in men and women in a city (Singh et al. 1998). For SES based on educational attainment, the same pattern was seen in another study of men and women in Indian villages (Gupta et al. 1994). This pattern was also reported for men in a Brazilian city for overall physical activity, although there was no association for leisure-time activity (Duncan et al. 1993). In contrast, the higher-SES women in this Brazilian city had higher levels of physical activity. A study in Puerto Rico reported that urban men with more education had lower levels of physical activity (Sorlie and Garcia-Palmieri 1990). Also, in Mauritius, unskilled workers were reported to be significantly more physically active than members of other occupational groups (attributable largely to occupational rather than to leisure-time physical activity) (Pereira et al. 1998).

## Literature on developed countries

## Physical inactivity associated with low SES

In the United States, national survey data indicated increased leisure-time physical inactivity with lower SES (including people who were less educated, who lived below the poverty line and who lived in low-income households) (Crespo et al. 1999). Previous national survey data also showed this pattern (USDHHS 2001). However, these indicators still do not seem to explain the higher prevalence of leisure-time physical inactivity among specific populations (African Americans and Mexican Americans) (Crespo et al. 2000).

Consistent with these results are studies of specific United States population groups, for example for income and education in a disadvantaged urban community (Diez-Roux 1998); educational level and women among members of a health maintenance organization in California (Sternfeld et al. 1999); and educational level and income among Minneapolis-St Paul residents). The latter study reported time trends for leisure-time physical activity, indicating greater gains among men with
low SES and among less affluent women. There is also United States evidence that living in an poor area is independently associated with physical inactivity and in a decline in physical activity levels over time (Yen and Kaplan 1998).

Similarly, the Whitehall II longitudinal study in the United Kingdom found that low-status occupation was associated with leisure-time physical inactivity for both men and women (Brunner et al. 1999). Adjustment for earlier SES position (using father's social class and own education level simultaneously) did not weaken the effects of SES position as an adult. Other United Kingdom work showed non-participation in sport to be associated with low SES, and no significant changes in this gradient occurred between 1984 and 1993 (Bartley et al. 2000).

Survey data from Germany have also suggested that lack of regular exercise (for both sexes) was one of the most striking poverty-related differences in terms of cardiovascular disease risk factors (Helmert et al. 1997). Other German work has reported that high-SES men and women were, respectively, four and three times more likely to have an active leisure time than those with low SES (Mensink et al. 1997). Similarly, survey data on elderly people of low SES in Switzerland suggest that they get less physical exercise (Abelin and Schlettwein-Gsell 1986). In the Netherlands, a cohort study reported that lower levels of educational attainment were associated with physical inactivity (Schrijvers et al. 1999), as did a cross-sectional study (Droomers et al. 1998).

A Swedish cohort study (Lindstrom et al. 2001) reported increased levels of physical inactivity among skilled and unskilled male manual workers and unskilled female manual workers compared to high-level non-manual employees. However, it was found that adjusting for social participation almost completely erased the SES differences (social participation included involvement in political parties and organizations and was a strong predictor of increased physical activity). In Australia, it was reported that multicentre cross-sectional community survey data indicate a "clear socioeconomic gradient between leisure-time physical activity and education attainment" (Bennett 1995). Nevertheless, time-trend data indicate that walking for recreation or exercise has become more popular among older men of low education. Another Australian study found that a higher proportion of students participating in organized sport and health and fitness activities came from the higher-SES group (based on parental occupation) (Blanksby et al. 1996).

## Physical inactivity associated with high SES

Survey data from a Spanish city between 1983 and 1994 found that physical inactivity in men was always more prevalent in social classes I and II (for total daily activity including activity at work) (Borrell et al. 2000). This difference increased over time as more people of advantaged classes became physically inactive. Among men in the United Kingdom, those
in the manual classes had a higher level of physical activity than those in the non-manual classes, owing partly to work-related physical activity (Acheson 1998). However, the general pattern for the United Kingdom is for physical inactivity to be associated with low SES (as discussed above).

## No relationship

In New Zealand, the most recent national survey found no significant relationship between physical inactivity and social class or educational attainment (Howden-Chapman and Tobias 2000). This was in contrast to earlier New Zealand national survey data, which showed a gradient for lower levels of physical inactivity with higher educational attainment (for more than four hours of exercise in the previous seven days) (Statistics New Zealand/Ministry of Health 1993).

## DISCUSSION

## The relationship

In general, physical inactivity is associated with low SES in developed countries, with a fairly marked gradient. Nevertheless, there are some exceptions, such as the reverse pattern among men in Spain. The relationship is more mixed for developing countries, with opposing gradients seen for China and India (with the latter showing the pattern of the developed world).

For the developed countries, this relationship appears to have impacts on disease outcomes. A United States review suggested that SES gradients in cardiovascular disease may be partly attributable to SES gradients in physical activity (Pickering 1999). A Dutch cohort study also reported that physical inactivity partly explained the relationship between poor educational attainment and mortality rates (Schrijvers et al. 1999). In general terms, the pattern of low SES being associated with physical inactivity is consistent with the evidence from developed countries that low SES in men is associated with increased risk of developing cardiovascular disease (Gonzalez et al. 1998).

## Possible mechanisms

Physical inactivity and low SES. In the United Kingdom, barriers to exercise among low-income people include illness or disability, lack of money and lack of transport (Chinn et al. 1999). Fear of crime may also have an effect on walking and the use of parks in some neighbourhoods, along with access to such parks, recreational facilities and gymnasiums. Swedish data suggest that high-SES groups have increased social participation, and it may be that leisure-time physical activity is partly mediated by a higher extent of encouragement/peer pressure to participate in physical activity (Lindstrom et al. 2001). United States data also indicate that social support and confidence in one's ability to continue to
exercise when faced with other pressures and demands is important in determining levels of physical activity in women (Sternfeld et al. 1999).

Physical inactivity and high SES. In at least some parts of the developing world, the move to sedentary occupations among those with high SES appears to be an important factor in the SES patterning of physical inactivity (Singh et al. 1998). A decline in walking to work among members of this group may also be important (Duncan et al. 1993). Also, in some low-income neighbourhoods there may be "contagion effects" where a culture of being out on the streets and walking around prevails (Ross 2000), possibly combined with higher-density housing. In the United Kingdom, barriers to participation in physical activity among high-income people include lack of leisure-time and lack of motivation (Chinn et al. 1999).

### 4.8 Blood-lead levels

## Literature on developing countries

A review of lead poisoning in children in China found 17 relevant publications. Those children residing in industrial areas and areas with heavy traffic had average blood-lead levels of 21.8-67.9 $\mu \mathrm{g} / \mathrm{dl}$ (Shen et al. 1996). The percentages of blood-lead levels above $10 \mu \mathrm{~g} / \mathrm{dl}$, which defines lead poisoning in children, ranged from 64.9 to 99.5 . A more recent study in a Chinese city (Wuxi City) in children aged 1-5 years also found that blood-lead levels were significantly higher for those living in industrial areas (Gao et al. 2001).

In a survey in India of children aged from 6 months to 6 years, risk factors for elevated blood-lead levels were living in houses painted with lead-based paint, recent exposure to lead-based paint and the use of the eye cosmetic "ma". However, they also found that high caste was a risk factor. No association was found with traffic, parental occupational exposure or nutritional status.

In a study in a Mexican city, low family income was one of the most important factors related to raised blood-lead levels in children, along with family use of lead-glazed pottery and the use of animal fat in cooking (Azcona-Cruz et al. 2000). Another study examined blood-lead levels in pregnant women in Mexico City attending public hospital prenatal clinics (representing primarily women of low SES) and private hospitals (primarily women of high SES) (Farias et al. 1996). Overall, blood-lead levels were significantly higher in the public hospital group. Consumption of milk products significantly reduced lead levels in the higher-SES group, and taking calcium supplements lowered blood-lead levels in those women whose diets were deficient in calcium. The authors concluded that avoiding the use of lead-glazed ceramics, consuming diets rich in calcium and, if needed, taking calcium supplements would be
expected to result in a substantial lowering of blood-lead levels, especially in pregnant women of low SES.

A study of 4-12-year-old schoolchildren in Bangladesh found that elevated blood-lead levels were significantly associated with low parental education (odds ratio of $2.7,95 \%$ CI $2.0-3.8$ ) and living close to major roads (odds ratio of $2.3,95 \%$ CI 1.2-4.3) along with soil-eating and increasing age (Kaiser et al. 2001). The authors considered that living near major roads may be related to low SES in this population.

A study in South Africa (KwaZulu-Natal) examined children aged 3-5 and 8-10 years (Nriagu et al. 1997). Household factors that were significantly associated with blood-lead levels included distance from tarred roads, overcrowding, hygienic habits in the household and the burning of solid waste for heating or cooking, all of which are related to poverty and low SES. The lack of significant associations with risk behaviour in this study was attributed by the authors to "the over-riding influence of high levels of contaminated dusts both indoor and outdoor". Another study in South Africa found that low parental education and income were associated with raised blood-lead levels, along with dusty homes, homes in a poor state of repair and overcrowding (von Schirnding et al. 1991).

## LITERATURE ON DEVELOPED COUNTRIES

A recent review (Tong et al. 2000b) reported that in developed countries chronic exposure to low levels of lead is still a public health issue, "especially among some minorities and socioeconomically disadvantaged groups". Similarly, a review in the United States reported that poor and minority children have higher rates of elevated blood-lead levels (Powell and Stewart 2001).

National survey data on United States children aged 1-5 years found that independent risk factors for elevated blood-lead levels included: "being of minority race/ethnicity, living in an older home, residing in the northeast or midwest regions of the United States, being on Medicaid, having a head of household with $<12$ years of education, and having a history of anemia" (Kaufmann et al. 2000). Some of these factors are associated with SES disadvantage (living in older housing and low level of education). An earlier national survey also reported associations between blood-lead levels in children and low income and low parental educational attainment (Pirkle et al. 1994).

Another study in the United States found that lead concentrations in the tibia were significantly higher in men who did not graduate from high school than in men with $\geq 4$ years of college education (Elreedy et al. 1999). Living in an area with a poor level of education was associated with a significantly higher tibia lead level among those not graduating from high school, but not among college graduates. The authors concluded that the influence of individual SES on cumulative lead
exposure is modified by geographical area. This study also identified a significant relationship between low individual annual income and increased bone lead levels.

In Australia, a cohort study found that residential area and father's place of employment were the two variables most strongly predictive of a child's blood-lead concentration at the end of primary school (Baghurst et al. 1999). A poorer-quality home environment was also found to be an independent contributor to blood-lead concentrations. Other work on this cohort, however, did not find statistically significant interactions between lifetime average blood-lead concentration and parents' occupational status (used as a surrogate of SES) (Tong et al. 2000a).

In New Zealand, a study of a cohort at age 21 years found significant associations between higher blood-lead levels and high-risk occupational activities, living close to a main road, smoking and recreational exposure (Fawcett et al. 1996). Some of these risk factors have associations with SES (occupation and smoking). However, a study of this same cohort at age 11 years found no significant correlation between bloodlead levels and SES (Silva et al. 1986).

## DISCUSSION

## The relationship

The overall pattern in both developed and developing countries is that poverty and low SES are associated with elevated blood-lead levels in children and adults. Nevertheless, there is still a minority of studies in which such a relationship is not apparent (for example, high caste Indian children, who may have greater exposure to cosmetics and crayons containing lead).

## Possible mechanisms

The literature suggests a number of possible mechanisms for a relationship between poverty and elevated blood-lead levels.

- Air pollution from vehicle exhausts. People living near major roads are likely to have increased exposure to airborne lead in countries where lead is still added to petrol. In the United States, air pollution exposures and associated health risks appear to affect disproportionately populations that are poor and non-Caucasian (see section 3.5 for further details).
- Older and dilapidated housing. Older houses have an increased probability of having been painted with lead-based paint, and these tend to be concentrated in older urban centres where many disadvantaged groups live. In these settings, and where houses are not properly maintained, children are more likely to ingest flakes of lead-based paint.
- Housing in the developing world. It has been suggested that the type of construction of some low-income housing and its location close to roads may make it difficult to keep these houses free from lead in dust and soil (Kaiser et al. 2001).
- Emissions from industry. Poor people are more likely to be exposed to such emissions, since they more often live in or close to heavily industrialized areas. One United States review noted that: "minority and poor families disproportionately live in communities with landfills, hazardous waste facilities, incinerators, industrial plants, and old housing with poor indoor air quality and lead-based paint" (Powell and Stewart 2001).
- Occupational hazards. Certain occupations are associated with increased risks of high blood-lead levels (e.g. lead smelting and battery manufacture). Many of these blue-collar factory workers are in relatively disadvantaged groups. Also, these workers sometimes bring lead dust into the home environment and thereby expose family members. In developing countries poor people can be exposed to lead in cottage industries.
- Smoking. Smoking is a risk factor for high blood-lead levels and is generally more common in low-income groups.
- Home remedies. Some home remedies used in developing countries are known to contain high levels of lead (Azcona-Cruz et al. 2000). Poor people might be more likely to use such remedies than those who can afford to seek treatment from a health professional.
- Poor nutrition. The dangers of lead exposure are increased by several dietary conditions (Mahaffey 1990) that are more frequently present in economically disadvantaged children.
- Knowledge of risks and preventive measures. Poorly educated parents may be less likely to know of the risks of lead exposure and to make use of preventive measures (such as frequent hand-washing with soap and water; thoroughly cleaning fruit and vegetables; regularly dusting and sweeping the house; and preventing children sucking fingers, toys or pencils or eating soil).
- Overcrowding. In a South African study (von Schirnding et al. 1991) it was suggested that "the over-crowded nature of the homes could have a direct bearing on the quality of the care-giving environment, providing opportunity for children's activities to go unsupervised. This could lead young children to be more exposed to accessible sources of lead associated with poor housing conditions".

Poverty and low SES might also actually accentuate the effect of lead poisoning. An Australian cohort study found that, after adjustment for
a wide range of covariates, children from socially disadvantaged backgrounds were more sensitive to the effects of lead on their IQ relative to those of a higher SES (Tong et al. 2000a). This finding is generally consistent with the results of some (but not all) other studies of children. It is also consistent with the results of a recent animal study (Schneider et al. 2001).

## 5. Discussion

We estimated associations between risk factors and absolute poverty across the developing subregions of the world, and calculated attributable risks based on these, but without adjustment for confounding. Many of the causes of ill-health that we studied tend to weigh most heavily on the poorest groups in each of the 11 subregions. However, the socioeconomic gradient is not uniform across all the risk factors, and in some instances the associations vary between subregions. For some risk factors, variation between subregions in overall risk factor prevalence appears to be more important than variation by individual-level poverty within subregions. Our findings are based on a combination of quantitative analyses of existing survey data, and literature reviews for selected risk factors with some limitations.

### 5.1 Limitations

## CONCEPTUAL LIMITATIONS

We examined the individual-level association of poverty with risk factors, whereas previous work had examined the global differences in health status and poverty using country-level analyses (e.g. Gwatkin et al. 1999). Our presentation of individual-level analyses is arguably preferable to country- or subregional-level analyses of poverty and health. However, it is important not to let our analyses distract from two issues: societies tend to adopt health behaviours collectively; and the greatest income inequalities in the world are between countries rather than within countries. Regarding the former point, it is interesting to note that the between-region variation in the overall prevalence of risk factors such as tobacco use, indoor air pollution and unsafe water and sanitation was more strongly associated with individual risk than income poverty.

There are several other limitations to our analyses of the association of poverty with risk factors, including the inability to make causal inferences, lack of consideration of the role of time-lags, exclusion of residual income heterogeneity within level of absolute poverty, and exclusion of the heterogeneity of the association of poverty with risk factors within and between countries in each subregion. Consequently, the results must be interpreted as a mapping of risk factors by income poverty. They should not be interpreted as a determination of the causal association
of poverty and risk factors or as a determination of the causal impact fractions of poverty for various risk factors.

## Quantitative analyses

The major limitation was data availability. First, both poverty and risk factor data were available for only some countries within each subregion, requiring extrapolations from those countries with data. Second, some subregions had no data at all for some risk factors (e.g. unsafe sex, alcohol), limiting the number of subregions for which we could conduct analyses. Third, all results were based on survey data with their own random and systematic errors.

It was necessary to make methodological assumptions in order to map risk factors by absolute poverty. The most important of these assumptions was that we could use the association of risk factors with a composite measure of socioeconomic position (e.g. asset scores in DHS data) to estimate the (unobserved) association with income poverty.

The definition and measurement of risk factors also need to be considered when interpreting the findings. In some instances, the variables used are indirect measures. Tobacco use, for example, was estimated in some subregions from household expenditure on tobacco (which would probably skew the distribution of use by socioeconomic group towards the better off). Other variables may not fully capture all aspects of risk. For example, alcohol use was also based on expenditure data.

### 5.2 Key findings

Having accepted the limitations described above, many striking results appeared that are highly unlikely to be spurious. For several important risk factors the effect of poverty appears to be very strong. Lack of improved water and sanitation, for instance, was 10-15 times more common in households living on <US\$ 1 per day than in those living on more than $>$ US $\$ 2$ per day. Indeed, lack of adequate water and sanitation is widely accepted as being so closely aligned to poverty that it is used as a variable in asset scores to measure poverty. (Note that we did not use water or sanitation in the estimation of asset scores.) Worldwide, we estimated that $36-51 \%$ of instances of inadequate water and/or sanitation could be prevented if the prevalence of inadequate water and/or sanitation among people living on $<$ US $\$ 2$ per day was the same as that of people living on $>$ US $\$ 2$ per day (assuming a causal relationship).

We found a consistently strong association of absolute poverty with underweight children in all subregions except WPR-B, which is dominated by the China analyses (see Table 24.19). Previous research also found the same strong association, including for China (Chen 1996; Li et al. 1999; Popkin et al. 1993). Thus, our estimate of there being little association of poverty with underweight children in China is probably due to the small sample size in that survey. Putting China aside, the con-
sistency of the relative association of absolute poverty with underweight children across other subregions was notable. Put another way, regardless of where children live in the world the relative increase in underweight children with increasing socioeconomic deprivation is remarkably consistent. Across all developing countries, we estimated that 23-37\% of cases of underweight could be prevented if the prevalence of underweight among children living on $<$ US $\$ 2$ per day was changed to that of children living on $>$ US $\$ 2$ per day. Assuming a causal relationship, poverty appears to have a strong and important association with child malnutrition.

The third risk factor in our quantitative analyses that demonstrated a strong association with poverty was indoor air pollution-although with some interesting heterogeneity between subregions (see Table 24.35). Regardless of the level of poverty in AFR-E, exposure to indoor air pollution appeared to be very high. In other subregions (often based on analyses for just one country per subregion), the associations appeared somewhat variable with regard to overall level of exposure to indoor air pollution and the relative risk association. Nevertheless, there was no exception to the general rule that those living in absolute poverty are exposed more frequently to indoor air pollution. Using the population impact fraction estimates (US\$2 per day cut-off), we estimate that a third to a half of the exposure to indoor air pollution in developing countries could be averted if the poor had the exposure prevalence of the nonpoor (assuming a causal relationship).

Beyond these three risk factors (underweight, water and sanitation and indoor air pollution), the associations we found in our quantitative analyses tended to be weaker and more variable between subregions. Data were limited in our quantitative analyses on unsafe sex, and the most reliable overall conclusion was of no consistent pattern of unsafe sex by poverty within subregions. However, behind the summary variable of "unsafe sex", it was interesting that "sex with a non-marital partner" and "condom use among those having sex with a non-marital partner" were less common among poor males and females. Also, there was some indication of variation in the association by sex. In AFR-E for instance, unsafe sex was associated with poverty among men but there was no apparent association for women (see Tables 24.26 and 24.27).

We carried out quantitative analyses for outdoor air, although these were not conducted in the same manner as for other risk factors. The results for outdoor air pollution show lower exposures to fine particles in those living on <US $\$ 1$ per day in all subregions studied (see Table 24.37). Our method assigned people in urban or rural areas the same outdoor air pollution exposure regardless of income. Because rural areas have lower air pollution, and poor people tend to live in rural areas, this means that the poorest (on average) are exposed to lower levels of outdoor air pollution. However, this result is likely to obscure some
reverse socioeconomic patterns within urban centres, where it is the poorest people who live in the more polluted areas. This within-urban socioeconomic difference is likely to partially offset the urban-rural effect that gives rise to our results, but by how much is unclear. Without doubt, however, our method overstated the degree to which the poorest people are protected from outdoor air pollution relative to those who are less poor.

The quantitative analyses for overweight were based on DHS data (probably superior to LSMS data) supplemented by CHNS data, but only analyses for maternal overweight were possible. Nevertheless, a clear pattern emerged. Among the poorest subregions of the world (Africa and South East Asia) poor women were only half as likely to be overweight as the non-poor, whereas among subregions without quite the same depth of poverty (south-eastern Europe, central Asia, Central and South America) poor women were only about $20 \%$ less likely to be overweight. The relationship in WPR-B (dominated by China) was flat. We are uncertain about the accuracy of the data for China, but Yu et al. (2000) observed a higher prevalence of obesity among the least educated women in urban China. Our literature review tended to support the pattern of obesity becoming a greater problem among lower socioeconomic groups in South American countries. Popkin et al. (1995b) commented that the socioeconomic gradient in developing countries is shifting towards a developed world pattern of increased obesity among groups of low SES. Indeed, this is already occurring in some urbanized parts of Latin America. It is interesting to speculate whether the patterns we found were simply due to individuals in more poor subregions living in deeper poverty or whether this pattern is due to the socioeconomic gradient starting to "tip the other way" in countries with increasing affluence.

The final two risk factors for which we were able to conduct quantitative analyses-tobacco and alcohol consumption-showed weak and variable associations between subregions. The data used in the analyses were not of high quality, being often based on expenditure data and income. Summarizing the analyses as a (rather crude) estimate, there was little "net" association: the "total" population impact fractions for subregions with data combined were less than $10 \%$. Nevertheless, there were some indications of between-subregion variability. For example, it is reasonable to conclude that if poor people in Africa (AFR-D and AFRE ) adopted the lifestyle of the non-poor, the prevalence of alcohol consumption might increase by approximately $20-40 \%$.

In terms of tobacco and alcohol consumption, we may expect a variable association between subregions. It has often been the experience of the developed world that higher socioeconomic groups initially adopt adverse behavioural patterns (such as smoking) then discard them on learning of the health consequences. For various reasons, possibly asso-
ciated with economics or marketing, lower socioeconomic groups take up such behavioural patterns later. It is not unreasonable to expect the current pattern among developing countries to reflect historical patterns in the developed world. Moreover, it seems likely that such transitions might be at different stages in different subregions. Our quantitative analyses are not inconsistent with this conjecture, but neither can they offer strong support for it. For example, one might expect to see a stronger association of poverty with tobacco in subregions with relatively high prevalence of tobacco use, assuming that prevalence is related to the stage of the tobacco "epidemic". This was not apparent in our data.

Regarding the international literature on tobacco, the general pattern in developing countries is for a higher prevalence of smoking in those of lower SES, with a generally stronger association for men. Nevertheless, some studies in the poorer countries actually show the opposite pattern, with adults and adolescents of high SES having higher smoking rates, while some studies show no association. Our quantitative analyses mostly demonstrated a null association of poverty with prevalence of tobacco use except in EMR-D (i.e. Pakistan), where there seemed to be higher prevalence of tobacco use among more poor people (see Table 24.39). It should be noted that our quantitative analyses may have underestimated tobacco use among poor people, owing to our having to rely on expenditure data. Conversely, it is possible that the published literature has tended to arise from less poor areas of the developing world.

Our literature review demonstrated a tendency towards more hazardous alcohol consumption among lower socioeconomic groups in the developing world. However, the pattern varied between countries and studies. Our quantitative analyses were based on the prevalence of alcohol use, as indicated by expenditure on alcohol. This approach might have underestimated an adverse socioeconomic gradient for hazardous alcohol consumption.

The local effects of poverty are important, but it must be noted that variations in prevalence between subregions are frequently greater than the variations in risk factors by income within subregions. Tobacco use and exposure to indoor air pollution are two examples. The effect of poverty on smoking rates might be considered as a (not insignificant) ripple on a much larger wave, with "peak" subregions having smoking prevalence figures four times greater than those of the subregions in the "trough". Global tobacco control must deal with both the forces that generate disadvantage for lower income groups within populations and those forces (mostly on the supply side) operating at the level of countries and subregions.

We have included detailed literature reviews for risk factors for which we were unable to conduct quantitative analyses. For cholesterol and
lipids, hypertension and physical inactivity, the general patterns in the developing world were of either a mixed pattern or of more adverse profiles among higher socioeconomic groups. However, just as with obesity and overweight, it seems likely that in the future the average profile for these risk factors will worsen in the developing world, and become more concentrated among lower socioeconomic groups.

## Acknowledgements

We thank Cara Marshall for assisting with the retrieval of journal articles; Emmanuela Gakidou for advice on factor analysis; and Majid Ezzati, Alan Lopez and Anthony Rodgers for ongoing guidance and advice. We are also indebted to the other risk factor teams, who forwarded data and reports to assist in the analysis.

Funding for this work was provided by WHO.

## Notes

1 See preface for an explanation of this term.
2 For the 11 subregions with DHS data.
3 Referred to as scenarios in chapter 16.

## References

Abelin T, Schlettwein-Gsell D (1986) [Handicaps and needs of the elderly. A multifactor epidemiologic study under urban conditions]. Schweizerische Medizinische Wochenschrift. Journal Suisse de Médecine, 116:1524-1542.
Acheson D (1998) Independent inquiry into inequalities in health report. The Stationery Office, London.
Adamson SJ, Sellman JD, Futterman-Collier A et al. (2000) A profile of alcohol and drug clients in New Zealand: results from the 1998 national telephone survey. New Zealand Medical Journal, 113:414-416.
Agarwal AK, Yunus M, Khan A, Ahmad J (1994) A clinical-epidemiological study of hypertension in rural population of Jawan Block, Distt, Aligarh (UP) India. Journal of the Royal Society of Health, 114:17-19.
Ahsan H, Underwood P, Atkinson D (1998) Smoking among male teenagers in Dhaka, Bangladesh. Preventive Medicine, 27:70-76.
Alam SE (1998) Prevalence and pattern of smoking in Pakistan. JPMA—Journal of the Pakistan Medical Association, 48:64-66.
Alem A, Kebede D, Kullgren G (1999) The epidemiology of problem drinking in Butajira, Ethiopia. Acta Psychiatrica Scandinavica, Supplementum, 397:77-83.
Allain TJ, Wilson AO, Gomo ZA, Adamchak DJ, Matenga JA (1997) Diet and nutritional status in elderly Zimbabweans. Age and Ageing, 26:463-470.
al-Mannai A, Dickerson JW, Morgan JB, Khalfan H (1996) Obesity in Bahraini adults. Journal of the Royal Society of Health, 116:30-2, 37-40.
al-Nuaim AA, Bamgboye EA, al-Rubeaan KA, al-Mazrou Y (1997) Overweight and obesity in Saudi Arabian adult population, role of socio-demographic variables. Journal of Community Health, 22:211-223.
al-Nuaim AR, al-Rubeaan K, al-Mazrou Y, al-Attas O, al-Daghari N, Khoja T (1996) High prevalence of overweight and obesity in Saudi Arabia. International Journal of Obesity and Related Metabolic Disorders, 20:547-552.
Amonoo-Lartson R, Pappoe ME (1992) Prevalence of smoking in secondary schools in the greater Accra region of Ghana. Social Science and Medicine, 34:1291-1293.

Anonymous (1977) Race, education and prevalence of hypertension. American Journal of Epidemiology, 106:351-361.

Arnesen E, Forsdahl A (1985) The Tromso Heart Study: coronary risk factors and their association with living conditions during childhood. Journal of Epidemiology and Community Health, 39:210-214.

Azcona-Cruz MI, Rothenberg SJ, Schnaas L, Zamora-Munoz JS, RomeroPlaceres M (2000) Lead-glazed ceramic ware and blood lead levels of children in the city of Oaxaca, Mexico. Archives of Environmental Health, 55:217-222.

Baghurst PA, Tong S, Sawyer MG, Burns J, McMichael AJ (1999) Sociodemographic and behavioural determinants of blood lead concentrations in children aged 11-13 years. The Port Pirie Cohort Study. Medical Journal of Australia, 170:63-67.

Bartley M, Fitzpatrick R, Firth D, Marmot M (2000) Social distribution of cardiovascular disease risk factors: change among men in England 1984-1993. Journal of Epidemiology and Community Health, 54:806-814.
Bener A, al-Ketbi LM (1999) Cigarette smoking habits among high school boys in a developing country. Journal of the Royal Society of Health, 119: 166-169.

Benetou V, Chloptsios Y, Zavitsanos X, Karalis D, Naska A, Trichopoulou A (2000) Total cholesterol and HDL-cholesterol in relation to socioeconomic status in a sample of 11645 Greek adults: the EPIC study in Greece. European prospective investigation into nutrition and cancer. Scandinavian Journal of Public Health, 28:260-265.

Bennett S (1995) Cardiovascular risk factors in Australia: trends in socioeconomic inequalities. Journal of Epidemiology and Community Health, 49: 363-372.

Bilal AM, Khattar MA, Hassan KI, Berry D (1992) Psychosocial and toxicological profile of drug misuse in male army conscripts in Kuwait. Acta Psychiatrica Scandinavica, 86:104-107.
Blakely T (2002) The New Zealand census-mortality study: socioeconomic inequalities and adult mortality 1991-94. Ministry of Health, Wellington.

Blakely T, Pearce N (2002) Socio-economic position is more than just NZDep. New Zealand Medical Journal, 115:109-111.

Blakely T, Woodward A, Pearce N, Salmond C, Kiro C, Davis P (2002) Socioeconomic factors and mortality among 25-64 year olds followed from 1991 to 1994: The New Zealand Census-Mortality Study. New Zealand Medical Journal, 115:93-97.
Blanksby BA, Anderson MJ, Douglas GA (1996) Recreational patterns, body composition and socioeconomic status of western Australian secondary school students. Annals of Human Biology, 23:101-112.
Blum A (1997) Cancer prevention: Preventing tobacco-related cancers. In: Cancer: principles and practice of oncology. DeVita V, Hellman S, Rosenburg S, eds. Lippincott-Raven, Philadelphia, PA.
Boardman JD, Finch BK, Ellison CG, Williams DR, Jackson JS (2001) Neighborhood disadvantage, stress, and drug use among adults. Journal of Health and Social Behavior, 42:151-165.

Borman B, Wilson N, Mailing C (1999) Socio-demographic characteristics of New Zealand smokers: results from the 1996 census. New Zealand Medical Journal, 112:460-463.
Borrell C, Rue M, Pasarin MI, Rohlfs I, Ferrando J, Fernandez E (2000) Trends in social class inequalities in health status, health-related behaviors, and health services utilization in a southern European urban area (1983-1994). Preventive Medicine, 31:691-701.
Bouchardy C, Parkin DM, Khlat M et al. (1993) Education and mortality from cancer in Sao Paulo, Brazil. Annals of Epidemiology, 3:64-70.

Bray GA, Popkin BM (1998) Dietary fat intake does affect obesity! American Journal of Clinical Nutrition, 68:1157-1173.

Brownsberger W (1997) Prevalence of frequent cocaine use in urban poverty areas. Contemporary Drug Problems, 24:349-371.

Brunner E, Shipley MJ, Blane D, Smith GD, Marmot MG (1999) When does cardiovascular risk start? Past and present socioeconomic circumstances and risk factors in adulthood. Journal of Epidemiology and Community Health, 53:757-764.
Bunker CH, Ukoli FA, Nwankwo MU et al. (1992) Factors associated with hypertension in Nigerian civil servants. Preventive Medicine, 21:710-722.
Carlini EA (1990) Research is badly needed to improve programmes for the prevention and treatment of drug abuse and drug dependence in Brazil. Drug and Alcohol Dependence, 25:169-173.

Carr S, Ager A, Nyando C, Moyo K, Titeca A, Wilkinson M (1994) A comparison of chamba (marijuana) abusers and general psychiatric admissions in Malawi. Social Science and Medicine, 39:401-406.

Casey PH, Szeto K, Lensing S, Bogle M, Weber J (2001) Children in food-insufficient, low-income families: prevalence, health, and nutrition status. Archives of Pediatrics and Adolescent Medicine, 155:508-514.
Casswell S, Gordon A (1984) Drinking and occupational status in New Zealand men. Journal of Studies on Alcohol, 45:144-148.

Chaturvedi HK, Phukan RK, Zoramtharga K, Hazarika NC, Mahanta J (1998) Tobacco use in Mizoram, India: Sociodemographic differences in pattern. Southeast Asian Journal of Tropical Medicine and Public Health, 29:66-70.

Chen CM (1996) Nutrition status of the Chinese people. Biomedical and Environmental Sciences, 9:81-92.
Chen S, Ravallion M (2000) How did the world's poorest fare in the 1990s? World Bank, Washington, DC. Accessed 29/11/01 at http://www.worldbank.org/ research/povmon itor/pdfs/methodology.pdf.
Chhabra SK, Rajpal S, Gupta R (2001) Patterns of smoking in Delhi and comparison of chronic respiratory morbidity among beedi and cigarette smokers. Indian Journal of Chest Diseases and Allied Sciences, 43:19-26.

Chinn DJ, White M, Harland J, Drinkwater C, Raybould S (1999) Barriers to physical activity and socioeconomic position: implications for health promotion. Journal of Epidemiology and Community Health, 53:191-192.

Chung MH, Chung KK, Chung CS, Raymond JS (1992) Health-related behaviors in Korea: Smoking, drinking, and perinatal care. Asia-Pacific Journal of Public Health, 6:10-15.
Colhoun H, Hemingway H, Poulter NR (1998) Socioeconomic status and blood pressure: an overview analysis. Journal of Human Hypertension, 12:91-110.
Commission on Macroeconomics and Health (2001) Investing in health for economic development. World Health Organization, Geneva.
Connolly GN (1992) Worldwide expansion of transnational tobacco industry. Journal of the National Cancer Institute. Monographs, 12:29-35.
Crampton P, Salmond C, Woodward A, Reid P (2000) Socioeconomic deprivation and ethnicity are both important for anti-tobacco health promotion. Health Education and Behavior, 27:317-327.

Crawford A (1988) Self-reported alcohol consumption among population subgroups in three areas of Britain. Drug and Alcohol Dependence, 21:161-167.

Crespo CJ, Ainsworth BE, Keteyian SJ, Heath GW, Smit E (1999) Prevalence of physical inactivity and its relation to social class in U.S. adults: results from the third national health and nutrition examination survey, 1988-1994. Medicine and Science in Sports and Exercise, 31:1821-1827.
Crespo CJ, Smit E, Andersen RE, Carter-Pokras O, Ainsworth BE (2000) Race/ethnicity, social class and their relation to physical inactivity during leisure time: results from the third national health and nutrition examination survey, 1988-1994. American Journal of Preventive Medicine, 18:46-53.

Crum RM, Lillie-Blanton M, Anthony JC (1996) Neighborhood environment and opportunity to use cocaine and other drugs in late childhood and early adolescence. Drug and Alcohol Dependence, 43:155-161.

Cuello C, Correa P, Haenszel W (1982) Socioeconomic class differences in cancer incidence in Cali, Colombia. International Journal of Cancer, 29:637-643.

Davey Smith G, Hart C, Watt G, Hole D, Hawthornem V (1998) Individual social class, area-based deprivation, cardiovascular disease risk factors, and
mortality: the Renfrew and Paisley Study. Journal of Epidemiology and Community Health, 52:399-405.

Davis P, McLeod K, Ransom M, Ongley P, Pearce N, Howden-Chapman P (1999) The New Zealand socioeconomic index: developing and validating an occupationally-derived indicator of socioeconomic status. Australian and New Zealand Journal of Public Health, 23:27-33.
Day NL, Cottreau CM, Richardson GA (1993) The epidemiology of alcohol, marijuana, and cocaine use among women of childbearing age and pregnant women. Clinical Obstetrics and Gynecology, 36:232-245.
Delpeuch F, Cornu A, Massamba JP, Traissac P, Maire B (1994) Is body mass index sensitively related to socioeconomic status and to economic adjustment? A case study from the Congo. European Journal of Clinical Nutrition, 48:S141-147.

Delpeuch F, Traissac P, Martin-Prevel Y, Massamba JP, Maire B (2000) Economic crisis and malnutrition: socioeconomic determinants of anthropometric status of preschool children and their mothers in an African urban area. Public Health Nutrition, 3:39-47.
Dhurandhar NV, Kulkarni PR (1992) Prevalence of obesity in Bombay. International Journal of Obesity and Related Metabolic Disorders, 16:367-375.
Diez-Roux A (1998) Bringing context back into epidemiology: variables and fallacies in multilevel analysis. American Journal of Public Health, 88:216-222.
Doak CM, Adair LS, Monteiro C, Popkin BM (2000) Overweight and underweight coexist within households in Brazil, China and Russia. Journal of Nutrition, 130:2965-2971.

Dressler WW, Evans P, Gray DJ (1992) Status incongruence and serum cholesterol in an English general practice. Social Science and Medicine, 34:757-762.

Dressler WW, Santos JE (2000) Social and cultural dimensions of hypertension in Brazil: a review. Cadernos de Saude Publica, 16:303-315.
Drever F, Whitehead M, eds. (1997) Health inequalities. Office for National Statistics, London.
Droomers M, Schrijvers CT, van de Mheen H, Mackenbach JP (1998) Educational differences in leisure-time physical inactivity: a descriptive and explanatory study. Social Science and Medicine, 47:1665-1676.
Dryson E, Martorell R (2000) Poor fetal and child growth and later obesity and chronic disease: relevance for Latin America. In: Obesity and poverty: a new public health challenge. Pena M, Bacallao J, eds. Pan American Health Organization, Washington, DC.

Dryson E, Metcalf P, Baker J, Scragg R (1992) The relationship between body mass index and socioeconomic status in New Zealand: ethnic and occupational factors. New Zealand Medical Journal, 105:233-235.
Duncan BB, Schmidt MI, Achutti AC, Polanczyk CA, Benia LR, Maia AA (1993) Socioeconomic distribution of noncommunicable disease risk factors in urban Brazil: the case of Porto Alegre. Bulletin of the Pan American Health Organization, 27:337-349.

Edwards R, Brown JS, Hodgson P, Kyle D, Reed D, Wallace B (1999) An action plan for tobacco control at regional level. Public Health, 113:165-170.

Eide AH, Acuda SW (1996) Adolescents’ drug use in Zimbabwe-comparing two recent studies. Central African Journal of Medicine, 42:128-135.

Ekpo EB, Udofia O, Eshiet NF, Andy JJ (1992) Demographic, life style and anthropometric correlates of blood pressure of Nigerian urban civil servants, factory and plantation workers. Journal of Human Hypertension, 6:275-280.
Elreedy S, Krieger N, Ryan PB, Sparrow D, Weiss ST, Hu H (1999) Relations between individual and neighborhood-based measures of socioeconomic position and bone lead concentrations among community-exposed men: the normative aging study. American Journal of Epidemiology, 150:129-141.

Ensminger ME, Anthony JC, McCord J (1997) The inner city and drug use: initial findings from an epidemiological study. Drug and Alcohol Dependence, 48:175-184.

Evans T, Whitehead M, Diderichsen F, Bhuiya A, Wirth M (2001) Challenging inequalities in health: from ethics to action, Oxford University Press, Oxford.

Farias P, Borja-Aburto VH, Rios C, Hertz-Picciotto I, Rojas-Lopez M, ChavezAyala R (1996) Blood lead levels in pregnant women of high and low socioeconomic status in Mexico City. Environmental Health Perspectives, 104: 1070-1074.
Farrow SC, Charny MC, Lewis PC (1988) A community survey of alcohol consumption. Alcohol and Alcoholism, 23:315-322.
Fawcett JP, Williams SM, Heydon JL, Walmsley TA, Menkes DB (1996) Distribution of blood lead levels in a birth cohort of New Zealanders at age 21. Environmental Health Perspectives, 104:1332-1335.

Fergusson DM, Horwood LJ (1999) Prospective childhood predictors of deviant peer affiliations in adolescence. Journal of Child Psychology and Psychiatry and Allied Disciplines, 40:581-592.
Fergusson DM, Horwood LJ (2000) Cannabis use and dependence in a New Zealand birth cohort. New Zealand Medical Journal, 113:156-158.

Fergusson DM, Woodward LJ, Horwood LJ (1998) Maternal smoking during pregnancy and psychiatric adjustment in late adolescence. Archives of General Psychiatry, 55:721-727.
Fernandez E, Schiaffino A, Garcia M, Borras JM (2001) Widening social inequalities in smoking cessation in Spain, 1987-1997. Journal of Epidemiology and Coттииity Health, 55:729-730.

Filmer D, Pritchett L (1988) Estimating wealth effects without expenditure, data or tears. Development Economics Research Group, Washington, DC.

Finch BK, Vega WA, Kolody B (2001) Substance use during pregnancy in the state of California, USA. Social Science and Medicine, 52:571-583.

Forero R, Bauman A, Chen JX, Flaherty B (1999) Substance use and sociodemographic factors among aboriginal and Torres Strait Islander school students in New South Wales. Australian and New Zealand Journal of Public Health, 23:295-300.

Gan CY (1995) Smokeless tobacco use among rural Kadazan women in Sabah, Malaysia. Southeast Asian Journal of Tropical Medicine and Public Health, 26:291-296.

Gan CY (1998) Tobacco usage among rural Bajaus in Sabah, Malaysia. Southeast Asian Journal of Tropical Medicine and Public Health, 29:643-648.
Gao W, Li Z, Kaufmann RB et al. (2001) Blood lead levels among children aged 1 to 5 years in Wuxi City, China. Environmental Research, 87:11-19.
Ge K, Weisell R, Guo X et al. (1994) The body mass index of Chinese adults in the 1980s. European Journal of Clinical Nutrition, 48:S148-154.
Glanville J, Lefebvre C (2000) Identifying systematic reviews: key resources. Evidence-Based Mental Health, 3:68-69.
Glendinning A, Shucksmith J, Hendry L (1994) Social class and adolescent smoking behaviour. Social Science and Medicine, 38:1449-1460.
Gold MR, Franks P (1990) The social origin of cardiovascular risk: an investigation in a rural community. International Journal of Health Services, 20: 405-416.

Gonzalez MA, Rodriguez Artalejo F, Calero JR (1998) Relationship between socioeconomic status and ischaemic heart disease in cohort and casecontrol studies: 1960-1993. International Journal of Epidemiology, 27: 350-358.

Green G, Macintyre S, West P, Ecob R (1991) Like parent like child? Associations between drinking and smoking behaviour of parents and their children. British Journal of Addiction, 86:745-758.
Grol ME, Eimers JM, Alberts JF et al. (1997a) Alarmingly high prevalence of obesity in Curacao: data from an interview survey stratified for socioeconomic status. International Journal of Obesity and Related Metabolic Disorders, 21:1002-1009.

Grol ME, Halabi YT, Gerstenbluth I, Alberts JF, O’Niel J (1997b) Lifestyle in Curacao. Smoking, alcohol consumption, eating habits and exercise. West Indian Medical Journal, 46:8-14.
Gupta R, Gupta VP, Ahluwalia NS (1994) Educational status, coronary heart disease, and coronary risk factor prevalence in a rural population of India. British Medical Journal, 309:1332-1336.
Gwatkin D, Guillot M, Heuveline P (1999) The burden of disease among the global poor. The Lancet, 354:586-589.
Hadi A (2000) A participatory approach to sanitation: experience of Bangladeshi NGOs. Health Policy and Planning, 15:332-337.

Hamadeh RR, Musaiger AO (2000) Lifestyle patterns in smokers and nonsmokers in the state of Bahrain. Nicotine and Tobacco Research, 2:65-69.

Hans SL (1999) Demographic and psychosocial characteristics of substanceabusing pregnant women. Clinics in Perinatology, 26:55-74.

Hay DR, Foster FH (1984) Intercensal trends in cigarette smoking in New Zealand 2: social and occupational factors. New Zealand Medical Journal, 97:395-398.

Helmert U, Mielck A, Shea S (1997) Poverty, health, and nutrition in Germany. Reviews on Environmental Health, 12:159-170.

Hoa DP, Thanh HT, Hojer B, Persson LA (1995) Young child feeding in a rural area in the Red River Delta, Vietnam. Acta Paediatrica, 84:1045-1049.

Hodge AM, Dowse GK, Toelupe P, Collins VR, Imo T, Zimmet PZ (1994) Dramatic increase in the prevalence of obesity in Western Samoa over the 13 year period 1978-1991. International Journal of Obesity and Related Metabolic Disorders, 18:419-428.
Howden-Chapman P, Tobias M, eds. (2000) Social inequalities in health: New Zealand 1999, Ministry of Health, Wellington.
Howell EM, Heiser N, Harrington M (1999) A review of recent findings on substance abuse treatment for pregnant women. Journal of Substance Abuse Treatment, 16:195-219.

Hu TW, Lin Z, Keeler TE (1998) Teenage smoking, attempts to quit, and school performance. American Journal of Public Health, 88:940-943.
Hu TW, Tsai YW (2000) Cigarette consumption in rural China: survey results from 3 provinces. American Journal of Public Health, 90:1785-1787.

ICRG (1988) Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. INTERSALT Cooperative Research Group. British Medical Journal, 297: 319-328.

INCLEN (1994) Socioeconomic status and risk factors for cardiovascular disease: a multicentre collaborative study in the international clinical epidemiology network (INCLEN). The INCLEN multicentre collaborative group. Journal of Clinical Epidemiology, 47:1401-1409.

Irala-Estevez JD, Groth M, Johansson L, Oltersdorf U, Prattala R, MartinezGonzalez MA (2000) A systematic review of socioeconomic differences in food habits in Europe: consumption of fruit and vegetables. European Journal of Clinical Nutrition, 54:706-714.
Jackson R, Beaglehole R, Yee RL, Small C, Scragg R (1990) Trends in cardiovascular risk factors in Auckland, 1982 to 1987. New Zealand Medical Journal, 103:363-365.
Jarallah JS, al-Rubeaan KA, al-Nuaim AR, al-Ruhaily AA, Kalantan KA (1999) Prevalence and determinants of smoking in three regions of Saudi Arabia. Tobacco Control, 8:53-56.

Jeffery RW, French SA, Forster JL, Spry VM (1991) Socioeconomic status differences in health behaviors related to obesity: the healthy worker project. International Journal of Obesity, 15:689-696.
Johnson RK, Wang MQ, Smith MJ, Connolly G (1996) The association between parental smoking and the diet quality of low-income children. Pediatrics, 97:312-317.

Jutkowitz JM, Spielmann H, Koehler U, Lohani J, Pande A (1997) Drug use in Nepal: the view from the street. Substance Use and Misuse, 32:987-1004.

Kadiri S, Walker O, Salako BL, Akinkugbe O (1999) Blood pressure, hypertension and correlates in urbanised workers in Ibadan, Nigeria: a revisit. Journal of Human Hypertension, 13:23-27.
Kaiser R, Henderson AK, Daley WR et al. (2001) Blood lead levels of primary school children in Dhaka, Bangladesh. Environmental Health Perspectives, 109:563-566.
Kaufman J, Cooper R (1999) Seeking causal explanations in social epidemiology. American Journal of Epidemiology, 150:113-120.
Kaufmann RB, Clouse TL, Olson DR, Matte TD (2000) Elevated blood lead levels and blood lead screening among us children aged one to five years: 1988-1994. Pediatrics, 106:E79.

Kawachi I, Marshall S, Pearce N (1991) Social class inequalities in the decline of coronary heart disease among New Zealand men, 1975-1977 to 1985-1987. International Journal of Epidemiology, 20:393-398.

Kebede D, Alem A (1999) The epidemiology of alcohol dependence and problem drinking in Addis Ababa, Ethiopia. Acta Psychiatrica Scandinavica, Supplementum, 397:30-34.

Kikafunda JK, Walker AF, Collett D, Tumwine JK (1998) Risk factors for early childhood malnutrition in Uganda. Pediatrics, 102:E45.
King M (1986) At risk drinking among general practice attenders: Prevalence, characteristics and alcohol-related problems. British Journal of Psychiatry, 148:533-540.

Klemp P, Robertson MC, Stansfield S, Klemp JA, Harding E (1998) Factors associated with smoking and the reasons for stopping in Maori and Europeans: a comparative study. New Zealand Medical Journal, 111:148-501.

Knuiman JT, West CE, Burema J (1982) Serum total and high density lipoprotein cholesterol concentrations and body mass index in adult men from 13 countries. American Journal of Epidemiology, 116:631-642.
Knupfer G. (1989) The prevalence in various social groups of eight different drinking patterns, from abstaining to frequent drunkenness: analysis of 10 US surveys combined. British Journal of Addiction, 11:1305-1318.
Koong SL, Serdula MK, Williamson DF, Malison MD, Davis RM (1991) Smoking prevalence in the United States and Taipei City, Taiwan. American Journal of Preventive Medicine, 7:161-165.
Krieger N (2001) A glossary for social epidemiology. Journal of Epidemiology and Community Health, 56:693-700.
Kunst A, Groenhof F, Mackenbach J, The EU Working Group on Socioeconomic Inequalities in Health (1998a) Mortality by occupational class among men 30-64 years in 11 European countries. Social Science and Medicine, 46:1459-1476.

Kunst A, Groenhof F, Mackenbach J, The EU Working Group on Socioeconomic Inequalities in Health (1998b) Occupational class and cause specific mortality in middle-aged men in 11 European countries: comparison of population based studies. British Medical Journal, 316:1636-1642.

Kushwaha KP, Singh YD, Rathi AK, Singh KP, Rastogi CK (1992) Prevalence and abuse of psychoactive substances in children and adolescents. Indian Journal of Pediatrics, 59:261-268.

Lapham SC, Skipper BJ, Brown P, Chadbunchachai W, Suriyawongpaisal P, Paisarnsilp S (1998) Prevalence of alcohol problems among emergency room patients in Thailand. Addiction, 93:1231-1239.
Leon D, Walt G, eds. (2001) Poverty, inequality and health. Oxford University Press, Oxford.
Li Y, Guo G, Shi A, Anme T, Ushijima H (1999) Prevalence and correlates of malnutrition among children in rural minority areas of China. Pediatrics International, 41:549-556.

Lindstrom M, Hanson BS, Ostergren PO (2001) Socioeconomic differences in leisure-time physical activity: the role of social participation and social capital in shaping health related behaviour. Social Science and Medicine, 52:441-451.

Linn S, Fulwood R, Rifkind B et al. (1989) High density lipoprotein cholesterol levels among us adults by selected demographic and socioeconomic variables. The Second National Health and Nutrition Examination Survey 1976-1980. American Journal of Epidemiology, 129:281-294.
Luthar SS, D'Avanzo K (1999) Contextual factors in substance use: a study of suburban and inner-city adolescents. Development and Psychopathology, 11:845-867.

Mackay J (1992) US tobacco export to Third World: Third World War. Journal of the National Cancer Institute. Monographs, 12:25-28.

Mackenbach J, Kunst A (1997) Measuring the magnitude of socioeconomic inequalities in health: an overview of available measures illustrated with two examples from Europe. Social Science and Medicine, 44:757-771.

Mackenbach J, Kunst A, Cavelaars A, Groenhof F, Geurts J (1997) Socioeconomic inequalities in morbidity and mortality in western Europe. The Lancet, 349:1655-1659.

Mahaffey KR (1990) Environmental lead toxicity: nutrition as a component of intervention. Environmental Health Perspectives, 89:75-78.
Mahley RW, Arslan P, Pekcan G et al. (2001) Plasma lipids in Turkish children: impact of puberty, socioeconomic status, and nutrition on plasma cholesterol and HDL. Journal of Lipid Research, 42:1996-2006.
Mancilha-Carvalho JJ, Crews DE (1990) Lipid profiles of Yanomamo Indians of Brazil. Preventive Medicine, 19:66-75.

Markovic N, Bunker CH, Ukoli FA, Kuller LH (1998) John Henryism and blood pressure among Nigerian civil servants. Journal of Epidemiology and Community Health, 52:186-190.

Martikainen P, Ishizaki M, Marmot MG, Nakagawa H, Kagamimori S (2001) Socioeconomic differences in behavioural and biological risk factors: a comparison of a Japanese and an English cohort of employed men. International Journal of Epidemiology, 30:833-838.

Martins IS, Coelho LT, Casajus MI, Okani ET (1995) Smoking, consumption of alcohol and sedentary life style in population grouping and their relationships with lipemic disorders. Revista de Saude Publica, 29:38-45.

Martorell R, Kettel Khan L, Hughes ML, Grummer-Strawn LM (2000) Overweight and obesity in preschool children from developing countries. International Journal of Obesity and Related Metabolic Disorders, 24:959-967.
Martorell R, Khan LK, Hughes ML, Grummer-Strawn LM (1998) Obesity in Latin American women and children. Journal of Nutrition, 128:1464-1473.
Memon A, Moody PM, Sugathan TN et al. (2000) Epidemiology of smoking among Kuwaiti adults: prevalence, characteristics, and attitudes. Bulletin of the World Health Organization, 78:1306-1315.

Mensink GB, Loose N, Oomen CM (1997) Physical activity and its association with other lifestyle factors. European Journal of Epidemiology, 13:771-778.

Midanik LT, Clark WB (1994) The demographic distribution of US drinking patterns in 1990: Description and trends from 1984. American Journal of Public Health, 84:1218-1222.

Misra A, Pandey RM, Devi JR, Sharma R, Vikram NK, Khanna N (2001) High prevalence of diabetes, obesity and dyslipidaemia in urban slum population in northern India. International Journal of Obesity, 25:1722-1729.
Ministry of Health (1999) Taking the pulse: The 1996/97 New Zealand Health Survey. Ministry of Health, Wellington.
Ministry of Health (2001) Inhaling inequality: tobacco's contribution to health inequality in New Zealand. Ministry of Health, Wellington.
Molarius A, Seidell JC, Sans S, Tuomilehto J, Kuulasmaa K (2000) Educational level, relative body weight, and changes in their association over 10 years: an international perspective from the WHO MONICA Project. American Journal of Public Health, 90:1260-1268.

Moller H, Tonnesen H (1997) Alcohol drinking, social class and cancer. IARC Scientific Publications, 138:251-263.

Monteiro C (2000) The epidemiologic transition in Brazil. In: Obesity and poverty: a new public health challenge. Pena M, Bacallao J, eds. Pan American Health Organization, Washington, DC.
Monteiro CA, Mondini L, de Souza AL, Popkin BM (1995) The nutrition transition in Brazil. European Journal of Clinical Nutrition, 49:105-113.
Moreira LB, Fuchs FD, Moraes RS et al. (1996) Alcoholic beverage consumption and associated factors in Porto Alegre, a southern Brazilian city: a population-based survey. Journal of Studies on Alcohol, 57:253-259.

Morgan D, Grant KA, Gage HD et al. (2002) Social dominance in monkeys: dopamine D2 receptors and cocaine self-administration. Nature Neuroscience, 5:169-174.

Naidu AN, Rao NP (1994) Body mass index: a measure of the nutritional status in Indian populations. European Journal of Clinical Nutrition, 48:S131-140.

Narayan KM, Chadha SL, Hanson RL et al. (1996) Prevalence and patterns of smoking in Delhi: cross sectional study. British Medical Journal, 312: 1576-1579.

Nriagu J, Jinabhai CC, Naidoo R, Coutsoudis A (1997) Lead poisoning of children in Africa, II. Kwazulu/Natal, South Africa. Science of the Total Environment, 197:1-11.

Nyberg K, Allebeck P, Eklund G, Jacobson B (1992) Socioeconomic versus obstetric risk factors for drug addiction in offspring. British Journal of Addiction, 87:1669-1676.

Obot IS (1990) The use of tobacco products among Nigerian adults: a general population survey. Drug and Alcohol Dependence, 26:203-208.
Omokhodion FO, Umar US, Ogunnowo BE (2000) Prevalence of low back pain among staff in a rural hospital in Nigeria. Occupational Medicine (Oxford), 50:107-110.

Omoluabi PF (1995) A review of the incidence of nonprescription psychoactive substance use/misuse in Nigeria. International Journal of the Addictions, 30:445-458.

Onat A, Surdum-Avci G, Senocak M, Ornek E, Gozukara Y (1992) Plasma lipids and their interrelationship in Turkish adults. Journal of Epidemiology and Community Health, 46:470-476.
Ozanne SE (2001) Metabolic programming in animals. British Medical Bulletin, 60:143-152.

Paeratakul S, Popkin BM, Keyou G, Adair LS, Stevens J (1998) Changes in diet and physical activity affect the body mass index of Chinese adults. International Journal of Obesity and Related Metabolic Disorders, 22:424-431.

Pamuk E, Makuc D, Heck k, Reuben C, Lochner K (1998) Socioeconomic status and health chartbook. Health, United States, 1998. National Center for Health Statistics, Hyattsville, MD.

Parsons TJ, Power C, Logan S, Summerbell CD (1999) Childhood predictors of adult obesity: a systematic review. International Journal of Obesity and Related Metabolic Disorders, 23:S1-107.

Pavan L, Casiglia E, Braga LM et al. (1999) Effects of a traditional lifestyle on the cardiovascular risk profile: the Amondava population of the Brazilian Amazon. Comparison with matched African, Italian and Polish populations. Journal of Hypertension, 17:749-756.
Pearce N (1997) Why study socioeconomic factors and cancer? IARC Scientific Publications, 138:17-23.

Pena M, Bacallao J (2000) Obesity among the poor: an emerging problem in Latin America and the Caribbean. In: Obesity and poverty: a new public health challenge. Pena M, Bacallao J, eds. Pan American Health Organization, Washington, DC.

Pequignot G, Crosignani P, Terracini B et al. (1988) A comparative study of smoking, drinking and dietary habits in population samples in France, Italy, Spain and Switzerland. III. Consumption of alcohol. Revue d'Épidémiologie et de Santé Publique, 36:177-185.
Pereira MA, Kriska AM, Collins VR et al. (1998) Occupational status and cardiovascular disease risk factors in the rapidly developing, high-risk population of Mauritius. American Journal of Epidemiology, 148:148-159.

Perova NV, Davis CE, Tao S et al. (2001) Multi-country comparison of plasma lipid relationship to years of schooling in men and women. International Journal of Epidemiology, 30:371-379.
Pickering T (1999) Cardiovascular pathways: socioeconomic status and stress effects on hypertension and cardiovascular function. Annals of the New York Academy of Sciences, 896:262-277.
Pinto JA, Ruff AJ, Paiva JV et al. (1994) HIV risk behavior and medical status of underprivileged youths in Belo Horizonte, Brazil. Journal of Adolescent Health, 15:179-185.
Pirkle JL, Brody DJ, Gunter EW et al. (1994) The decline in blood lead levels in the United States. The National Health and Nutrition Examination Surveys (NHANES). Journal of the American Medical Association, 272: 284-291.

Pomare E, Keefe-Ormsby V, Ormsby C et al. (1995) Hauora: Maori standards of health III. Te Ropu Rangahau Hauora a Eru Pomare; Wellington School of Medicine, Wellington.
Popkin BM (2001) The nutrition transition and obesity in the developing world. Journal of Nutrition, 131:S871-873.
Popkin BM, Keyou G, Zhai F, Guo X, Ma H, Zohoori N (1993) The nutrition transition in China: a cross-sectional analysis. European Journal of Clinical Nutrition, 47:333-346.
Popkin BM, Paeratakul S, Zhai F, Ge K (1995a) Dietary and environmental correlates of obesity in a population study in China. Obesity Research, 3:S135-143.

Popkin BM, Paeratakul S, Zhai F, Ge K (1995b) A review of dietary and environmental correlates of obesity with emphasis on developing countries. Obesity Research, 3:S145-153.
Potter JD (1997) Diet and cancer: Possible explanations for the higher risk of cancer in the poor. IARC Scientific Publications, 138:S265-283.
Powell DL, Stewart V (2001) Children. The unwitting target of environmental injustices. Pediatric Clinics of North America, 48:1291-1305.
Power C, Parsons T (2000) Nutritional and other influences in childhood as predictors of adult obesity. Proceedings of the Nutrition Society, 59:267-272.
Prothro JW, Rosenbloom CA (1999) Description of a mixed ethnic, elderly population. I. Demography, nutrient/energy intakes, and income status. Journals of Gerontology, Series A—Biological Sciences and Medical Sciences, 54:M315-324.

Prüss A, Kay D, Fewtrell L, Bartam J (2001) Estimating the burden of disease due to water, sanitation, and hygiene at the global level. World Health Organization, Geneva.
Radebe BZ, Brady P, Siziya S, Todd H (1996) Maternal risk factors for childhood malnutrition in the Mazowe district of Zimbabwe. Central African Journal of Medicine, 42:240-244.

Reddy DC, Singh SP, Tiwari IC, Shukla KP, Srivastava MK (1993) An epidemiological study of cannabis abuse among college students of Varanasi. Indian Journal of Public Health, 37:10-15.

Regidor E, Gutierrez-Fisac JL, Calle ME, Navarro P, Dominguez V (2001) Trends in cigarette smoking in Spain by social class. Preventive Medicine, 33: 241-248.

Ricci JA, Becker S (1996) Risk factors for wasting and stunting among children in Metro Cebu, Philippines. American Journal of Clinical Nutrition, 63: 966-975.

Rimm EB, Klatsky A, Grobbee D, Stampfer MJ (1996) Review of moderate alcohol consumption and reduced risk of coronary heart disease: is the effect due to beer, wine, or spirits. British Medical Journal, 312:731-736.

Robinson TN, Killen JD, Taylor CB et al. (1987) Perspectives on adolescent substance use. A defined population study. Journal of the American Medical Association, 258:2072-2076.

Roseboom TJ, van der Meulen JH, Osmond C, Barker DJ, Ravelli AC, Bleker OP (2001) Adult survival after prenatal exposure to the Dutch famine 1944-45. Paediatric and Perinatal Epidemiology, 15:220-225.
Ross CE (2000) Walking, exercising, and smoking: does neighborhood matter? Social Science and Medicine, 51:265-274.
Rossouw JE, Jooste PL, Steenkamp HJ, Thompson ML, Jordaan PC, Swanepoel AS (1990) Socioeconomic status, risk factors and coronary heart disease. The CORIS baseline study. South African Medical Journal, 78:82-85.

Ruidavets JB, Ducimetiere P, Arveiler D et al. (2002) Types of alcoholic beverages and blood lipids in a French population. Journal of Epidemiology and Community Health, 56:24-28.

Sakamoto N, Wansorn S, Tontisirin K, Marui E (2001) A social epidemiologic study of obesity among preschool children in Thailand. International Journal of Obesity and Related Metabolic Disorders, 25:389-394.
Saxe L, Kadushin C, Beveridge A et al. (2001) The visibility of illicit drugs: implications for community-based drug control strategies. American Journal of Public Health, 91:1987-1994.
Schneider JS, Lee MH, Anderson DW, Zuck L, Lidsky TI (2001) Enriched environment during development is protective against lead-induced neurotoxicity. Brain Research, 896:48-55.

Schrager L, Friedland G, Feiner C, Kahl P (1991) Demographic characteristics, drug use, and sexual behavior of I.V. drug users with AIDS in Bronx, New York. Public Health Reports, 106:78-84.
Schrijvers CT, Stronks K, van de Mheen HD, Mackenbach JP (1999) Explaining educational differences in mortality: the role of behavioral and material factors. American Journal of Public Health, 89:535-540.
Sen A (1999) Development as freedom. Alfred A. Knopf, New York.
Sharma HK (1996) Sociocultural perspective of substance use in India. Substance Use and Misuse, 31:1689-1714.

Shen X, Rosen JF, Guo D, Wu S (1996) Childhood lead poisoning in China. Science of the Total Environment, 181:101-109.

Shewry MC, Smith WC, Woodward M, Tunstall-Pedoe H (1992) Variation in coronary risk factors by social status: results from the Scottish Heart Health Study. British Journal of General Practice, 42:406-410.
Siegrist J, Bernhardt R, Feng ZC, Schettler G (1990) Socioeconomic differences in cardiovascular risk factors in China. International Journal of Epidemiology, 19:905-910.
Silva PA, Hughes P, Faed JM (1986) Blood lead levels in Dunedin 11 year old children. New Zealand Medical Journal, 99:179-183.
Singh RB, Beegom R, Mehta AS et al. (1999) Social class, coronary risk factors and undernutrition, a double burden of diseases, in women during transition, in five Indian cities. International Journal of Cardiology, 69:139-147.
Singh RB, Niaz MA, Ghosh S et al. (1995) Epidemiological study of coronary artery disease and its risk factors in an elderly urban population of north India. Journal of the American College of Nutrition, 14:628-634.

Singh RB, Niaz MA, Thakur AS, Janus ED, Moshiri M (1998) Social class and coronary artery disease in a urban population of north India in the Indian Lifestyle and Heart Study. International Journal of Cardiology, 64:195-203.

Singh RB, Sharma JP, Rastogi V et al. (1997a) Social class and coronary disease in rural population of north India. The Indian Social Class and Heart Survey. European Heart Journal, 18:588-595.

Singh RB, Sharma JP, Rastogi V, Niaz MA, Singh NK (1997b) Prevalence and determinants of hypertension in the Indian Social Class and Heart Survey. Journal of Human Hypertension, 11:51-56.

Smith M, Umenai T, Radford C (1998) Prevalence of smoking in Cambodia. Journal of Epidemiology, 8:85-89.

Statistics New Zealand/Ministry of Health (1993) A picture of health. Statistics New Zealand and Ministry of Health, Wellington.
Sobal J, Stunkard AJ (1989) Socioeconomic status and obesity: a review of the literature. Psychological Bulletin, 105:260-275.
Sorlie P, Backlund E, Keller J (1995) US mortality by economic, demographic, and social characteristics: the National Longitudinal Mortality Study. American Journal of Public Health, 85:949-956.
Sorlie PD, Garcia-Palmieri MR (1990) Educational status and coronary heart disease in Puerto Rico: the Puerto Rico Heart Health Program. International Journal of Epidemiology, 19:59-65.
Sowers M, Gonzalez Villalpando C, Stern MP, Fox C, Mitchell BD (1995) Relationships between physical activity, insulin levels and lipids in non-diabetic low income residents of Mexico City: The Mexico City Diabetes Study. Archives of Medical Research, 26:133-140.
Stamler R, Shipley M, Elliott P, Dyer A, Sans S, Stamler J (1992) Higher blood pressure in adults with less education. Some explanations from INTERSALT. Hypertension, 19:237-241.

Stellman SD, Resnicow K (1997) Tobacco smoking, cancer and social class. IARC Scientific Publications, 138:229-250.

Sternfeld B, Ainsworth BE, Quesenberry CP (1999) Physical activity patterns in a diverse population of women. Preventive Medicine, 28:313-323.
Taylor R, Cumming R, Woodward A, Black M (2001) Passive smoking and lung cancer: a cumulative meta-analysis. Australian and New Zealand Journal of Public Health, 25:203-211.

Tong S, McMichael AJ, Baghurst PA (2000a) Interactions between environmental lead exposure and sociodemographic factors on cognitive development. Archives of Environmental Health, 55:330-335.
Tong S, von Schirnding YE, Prapamontol T (2000b) Environmental lead exposure: a public health problem of global dimensions. Bulletin of the World Health Organization, 78:1068-1077.

Torun B (2000) Physical activity patterns in Central America. In: Obesity and poverty: a new public health challenge. Pena M, Bacallao J, eds. Pan American Health Organization, Washington, DC.

Tyas SL, Pederson LL (1998) Psychosocial factors related to adolescent smoking: a critical review of the literature. Tobacco Control, 7:409-420.

Ueshima H, Zhang XH, Choudhury SR (2000) Epidemiology of hypertension in China and Japan. Journal of Human Hypertension, 14:765-769.
UNDP (2000) Human development report 2000: human rights and human development. United Nations Development Programme, Oxford University Press, New York.

USDHHS (2001) The Surgeon General's call to action to prevent and decrease overweight and obesity. U.S. Department of Health and Human Services. Public Health Service, Office of the Surgeon General, Rockville, MD.

Vartiainen E, Pekkanen J, Koskinen S, Jousilahti P, Salomaa V, Puska P (1998) Do changes in cardiovascular risk factors explain the increasing socioeconomic difference in mortality from ischaemic heart disease in Finland? Journal of Epidemiology and Community Health, 52:416-419.
von Schirnding YE, Fuggle RF, Bradshaw D (1991) Factors associated with elevated blood lead levels in inner city Cape Town children. South African Medical Journal, 79:454-456.
Wadowski SJ, Karp RJ, Murray-Bachmann R, Senft C (1994) Family history of coronary artery disease and cholesterol: screening children in a disadvantaged inner-city population. Pediatrics, 93:109-113.

Wagstaff A (2000) Socioeconomic inequalities in child mortality: comparisons across nine developing countries. Bulletin of the World Health Organization, 78:19-29.

Wagstaff A, Watanabe N (2001) Socioeconomic inequalities in child malnutrition in the developing world. World Bank, Washington, DC.

Wahid Saeed AA, al Shammary FJ, Khoja TA, Hashim TJ, Anokute CC, Khan SB (1996) Prevalence of hypertension and sociodemographic characteristics of adult hypertensives in Riyadh City, Saudi Arabia. Journal of Human Hypertension, 10:583-587.

Wang Y (2001) Cross-national comparison of childhood obesity: the epidemic and the relationship between obesity and socioeconomic status. International Journal of Epidemiology, 30:1129-1136.

Whitlock G, MacMahon S, Vander Hoorn S, Davis P, Jackson R, Norton R (1997) Socioeconomic distribution of smoking in a population of 10529 New Zealanders. New Zealand Medical Journal, 110:327-330.
Whitlock G, MacMahon S, Vander Hoorn S, Davis P, Jackson R, Norton R (1998) Association of environmental tobacco smoke exposure with socioeconomic status in a population of 7725 New Zealanders. Tobacco Control, 7:276-280.

Woodward A, Newland H, Kinahoi M (1994) Smoking in the Kingdom of Tonga: report from a national survey. Tobacco Control, 3:41-45.

World Bank (2001a) World development report 2000/2001: attacking poverty, Oxford University Press, New York.

World Bank (2001b) World development indicators 2001. CD-ROM. World Bank. Washington, DC.

Yen IH, Kaplan GA (1998) Poverty area residence and changes in physical activity level: Evidence from the Alameda County Study. American Journal of Public Health, 88:1709-1712.

Yip R, Scanlon K, Trowbridge F (1993) Trends and patterns in height and weight status of low-income US children. Critical Reviews in Food Science and Nutrition, 33:409-421.

Yu Z, Nissinen A, Vartiainen E, Hu G, Tian H, Guo Z (2002) Socioeconomic status and serum lipids: a cross-sectional study in a Chinese urban population. Journal of Clinical Epidemiology, 55:143-149.

Yu Z, Nissinen A, Vartiainen E, Song G, Guo Z, Tian H (2000) Changes in cardiovascular risk factors in different socioeconomic groups: seven year trends in a Chinese urban population. Journal of Epidemiology and Community Health, 54:692-696.
Zhu BP, Liu M, Shelton D, Liu S, Giovino GA (1996) Cigarette smoking and its risk factors among elementary school students in Beijing. American Journal of Public Health, 86:368-75.

## Appendix A: DHS asset scores

Table A.I Distribution of DHS respondents by subregion and country, by values of the variables comprising the DHS asset score

| Subregion | Country | Variable value | Electricity availability | Urban-rural status | Educational level | Floor material |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | Benin |  |  |  |  |  |
|  |  | 0 | 4648 | 547 | 4026 |  |
|  |  | I | 806 | 317 | 1018 | 2591 |
|  |  | 2 |  | 955 | 425 | 17 |
|  |  | 3 |  | 3672 | 22 | 2883 |
| AFR-D | Burkina Faso |  |  |  |  |  |
|  |  | 0 | 5503 | 963 | 5240 |  |
|  |  | I | 738 | 337 | 643 | 3965 |
|  |  | 2 |  | 351 | 535 |  |
|  |  | 3 |  | 4794 | 27 | 2480 |
| AFR-D | Cameroon |  |  |  |  |  |
|  |  | 0 | 2309 | 1379 | 1329 |  |
|  |  | I | 2854 | 340 | 1992 | 2223 |
|  |  | 2 |  | 990 | 2057 | 12 |
|  |  | 3 |  | 2792 | 123 | 3266 |
| AFR-D | Chad |  |  |  |  |  |
|  |  | 0 | 7081 | 1355 | 5365 |  |
|  |  | 1 | 364 | 467 | 1620 | 6879 |
|  |  | 2 |  | 1191 | 454 |  |
|  |  | 3 |  | 4441 | 15 | 564 |
| AFR-D | Comoros |  |  |  |  |  |
|  |  | 0 | 2011 | 214 | 1635 |  |
|  |  | I | 1031 |  | 788 | 1382 |
|  |  | 2 |  | 689 | 600 | 42 |
|  |  | 3 |  | 2147 | 27 | 1626 |
| AFR-D | Ghana |  |  |  |  |  |
|  |  | 0 | 2921 | 635 | 1737 |  |
|  |  | I | 1919 | 232 | 813 | 869 |
|  |  | 2 |  | 718 | 2188 | 6 |
|  |  | 3 |  | 3258 | 105 | 3968 |
| AFR-D | Guinea |  |  |  |  |  |
|  |  | 0 | 5248 | \| 351 | 5361 |  |
|  |  | 1 | 1442 | 284 | 721 | 3361 |
|  |  | 2 |  | 790 | 530 | 74 |
|  |  | 3 |  | 4328 | 141 | 3318 |
| AFR-D | Liberia |  |  |  |  |  |
|  |  | 0 | 2266 |  | 3347 |  |
|  |  | I | 2969 | 933 | 1027 | 806 |
|  |  | 2 |  | 1011 | 808 | 4199 |
|  |  | 3 |  | 3295 | 57 | 234 |
| AFR-D | Madagascar |  |  |  |  |  |
|  |  | 0 | 5655 | 1286 | 1465 |  |
|  |  | 1 | 1398 | 351 | 3439 | 485 |
|  |  | 2 |  | 739 | 1972 | 4861 |
|  |  | 3 |  | 4684 | 182 | 1714 |

Table A.I Distribution of DHS respondents by subregion and country, by values of the variables comprising the DHS asset score (continued)

| Subregion | Country | Variable value | Electricity availability | Urban-rural status | Educational level | Floor material |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-D | Mali |  |  |  |  |  |
|  |  | 0 | 8679 | 1265 | 7773 |  |
|  |  | I | 976 | 1708 | 1218 | 7310 |
|  |  | 2 |  | 536 | 685 | 5 |
|  |  | 3 |  | 6195 | 28 | 2389 |
| AFR-D | Niger |  |  |  |  |  |
|  |  | 0 | 6573 | 1048 | 6066 |  |
|  |  | I | 963 | 1315 | 934 | 5721 |
|  |  | 2 |  |  | 549 |  |
|  |  | 3 |  | 5214 | 28 | 1856 |
| AFR-D | Nigeria |  |  |  |  |  |
|  |  | 0 | 5033 | 661 | 3706 |  |
|  |  | 1 | 4682 | 1750 | 2692 | 3508 |
|  |  | 2 |  | 807 | 2891 | 24 |
|  |  | 3 |  | 6592 | 521 | 6278 |
| AFR-D | Senegal |  |  |  |  |  |
|  |  | 0 | 5724 | 1017 | 6020 |  |
|  |  | 1 | 2474 | 1083 | 1705 | 3430 |
|  |  | 2 |  | 1122 | 795 |  |
|  |  | 3 |  | 5371 | 73 | 5163 |
| AFR-D | Togo |  |  |  |  |  |
|  |  | 0 | 6877 | 1417 | 4423 |  |
|  |  | I | 1671 | 763 | 2800 | 2074 |
|  |  | 2 |  | 869 | 1302 | 9 |
|  |  | 3 |  | 5520 | 44 | 6486 |
| AFR-E | Burundi |  |  |  |  |  |
|  |  | 0 | 2955 |  | 2924 |  |
|  |  | 1 | 1015 | 619 | 814 | 3450 |
|  |  | 2 |  | 31 | 200 | 23 |
|  |  | 3 |  | 3320 | 31 | 497 |
| AFR-E | Central African |  |  |  |  |  |
|  | Republic | 0 | 5570 | I 207 | 3083 |  |
|  |  | 1 | 298 |  | 2039 | 4813 |
|  |  | 2 |  | 1267 | 730 | 20 |
|  |  | 3 |  | 3410 | 32 | 1051 |
| AFR-E | Côte d'lvoire |  |  |  |  |  |
|  |  | 0 | 4448 | 1264 | 4909 |  |
|  |  | 1 | 3625 | 1296 | 2032 | 1619 |
|  |  | 2 |  | 1292 | 1112 | 2 |
|  |  | 3 |  | 4247 | 46 | 6478 |
| AFR-E | Ethiopia |  |  |  |  |  |
|  |  | 0 | 10811 | 2015 | 10586 |  |
|  |  | I | 3902 | 1507 | 2530 | 11759 |
|  |  | 2 |  | 1021 | 2092 | 158 |
|  |  | 3 |  | 10824 | 159 | 3450 |

Table A.I Distribution of DHS respondents by subregion and country, by values of the variables comprising the DHS asset score (continued)

| Subregion | Country | Variable value | Electricity availability | Urban-rural status | Educational level | Floor material |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AFR-E | Kenya |  |  |  |  |  |
|  |  | 0 | 6830 | 884 | 1010 |  |
|  |  | I | 981 | 341 | 4719 | 5134 |
|  |  | 2 |  | 241 | 2004 | 63 |
|  |  | 3 |  | 6415 | 148 | 2684 |
| AFR-E | Malawi |  |  |  |  |  |
|  |  | 0 | 4485 | 336 | 1834 |  |
|  |  | I | 356 | 568 | 2633 | 3720 |
|  |  | 2 |  | 412 | 369 |  |
|  |  | 3 |  | 3533 | 13 | 1129 |
| AFR-E | Mozambique |  |  |  |  |  |
|  |  | 0 | 7522 | 2507 | 3434 |  |
|  |  | 1 | 1166 | 510 | 4844 | 5795 |
|  |  | 2 |  | 1204 | 486 | 368 |
|  |  | 3 |  | 4558 | 15 | 2616 |
| AFR-E | Namibia |  |  |  |  |  |
|  |  | 0 | 4227 | 666 | 799 |  |
|  |  | I | 1169 |  | 2674 | 3391 |
|  |  | 2 |  | 1225 | 1859 | 92 |
|  |  | 3 |  | 3530 | 89 | 1938 |
| AFR-E | Rwanda |  |  |  |  |  |
|  |  | 0 | 5851 | 722 | 2342 |  |
|  |  | I | 430 | 436 | 3492 | 5086 |
|  |  | 2 |  |  | 693 |  |
|  |  | 3 |  | 5393 | 24 | 1465 |
| AFR-E | United Republic of Tanzania | 0 | 3104 | 335 | 1026 |  |
|  |  | I | 642 | 506 | 2461 | 2526 |
|  |  | 2 |  | 577 | 540 | 3 |
|  |  | 3 |  | 2611 | 2 | 1500 |
| AFR-E | Zambia |  |  |  |  |  |
|  |  | 0 | 6418 | 943 | 1168 |  |
|  |  | I | 1544 | 1219 | 4833 | 4757 |
|  |  | 2 |  | 839 | 1828 | 15 |
|  |  | 3 |  | 5020 | 191 | 3249 |
| AFR-E | Zimbabwe |  |  |  |  |  |
|  |  | 0 | 3894 | 435 | 437 |  |
|  |  | I | 2011 | 966 | 2518 | 1959 |
|  |  | 2 |  | 408 | 2803 | 11 |
|  |  | 3 |  | 4098 | 149 | 3937 |
| AMR-B | Brazil |  |  |  |  |  |
|  |  | 0 | 795 | 4580 | 769 |  |
|  |  | I | 11813 | 3002 | 4254 | 733 |
|  |  | 2 |  | 2672 | 6839 | 859 |
|  |  | 3 |  | 2358 | 747 | 11020 |

Table A.I Distribution of DHS respondents by subregion and country, by values of the variables comprising the DHS asset score (continued)

| Subregion | Country | Variable value | Electricity availability | Urban-rural status | Educational level | Floor material |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AMR-B | Colombia |  |  |  |  |  |
|  |  | 0 | 417 | 3821 | 404 |  |
|  |  | I | II 154 | 3280 | 3804 | 928 |
|  |  | 2 |  | 1698 | 5819 | 422 |
|  |  | 3 |  | 2786 | I 558 | 10230 |
| AMR-B | Dominican |  |  |  |  |  |
|  | Republic | 0 | 2871 | 1337 | 793 |  |
|  |  | I | 5529 | 2878 | 4336 | 864 |
|  |  | 2 |  | 911 | 2356 |  |
|  |  | 3 |  | 3296 | 935 | 7558 |
| AMR-B | Mexico |  |  |  |  |  |
|  |  | 0 | I 081 | 2411 | 813 |  |
|  |  | I | 8228 |  | 4352 | 1500 |
|  |  | 2 |  | \| 391 | 3504 | 4046 |
|  |  | 3 |  | 5508 | 639 | 3764 |
| AMR-B | Paraguay |  |  |  |  |  |
|  |  | 0 | 2751 | 1561 | 177 |  |
|  |  | I | 3075 | 359 | 3744 | 2044 |
|  |  | 2 |  | 981 | 1591 | 232 |
|  |  | 3 |  | 2926 | 313 | 3551 |
| AMR-B | Trinidad and |  |  |  |  |  |
|  | Tobago | 0 | 244 |  | 32 |  |
|  |  | I | 3558 | 189 | 1725 | 471 |
|  |  | 2 |  | 1501 | 1972 | 2058 |
|  |  | 3 |  | 2116 | 76 | 1277 |
| AMR-D | Bolivia |  |  |  |  |  |
|  |  | 0 | 2976 | 4846 | 1072 |  |
|  |  |  | 8189 | 1769 | 4023 | 3383 |
|  |  | 2 |  | 807 | 4434 | 1584 |
|  |  | 3 |  | 3765 | 1658 | 6220 |
| AMR-D | Ecuador |  |  |  |  |  |
|  |  | 0 | 908 |  | 368 |  |
|  |  | I | 3805 | 1920 | 2238 | 524 |
|  |  | 2 |  | 870 | 1673 | 2065 |
|  |  | 3 |  | 1923 | 434 | 2124 |
| AMR-D | Nicaragua |  |  |  |  |  |
|  |  | 0 | 4293 | 1616 | 2436 |  |
|  |  | 1 | 9323 | 2961 | 5818 | 6233 |
|  |  | 2 |  | 3267 | 4637 | 616 |
|  |  | 3 |  | 5790 | 743 | 6785 |
| AMR-D | Peru |  |  |  |  |  |
|  |  | 0 | 9538 | 11320 | 2127 |  |
|  |  | I | 19373 | 4105 | 9620 | 12792 |
|  |  | 2 |  | 3709 | 11387 | 2012 |
|  |  | 3 |  | 9817 | 5817 | 13528 |

Table A.I Distribution of DHS respondents by subregion and country, by values of the variables comprising the DHS asset score (continued)

| Subregion | Country | Variable value | Electricity availability | Urban-rural status | Educational level | Floor material |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| EMR-B | Tunisia |  |  |  |  |  |
|  |  | 0 | 1035 |  | 2372 |  |
|  |  | I | 3149 | 1218 | 1302 | 10 |
|  |  | 2 |  | 1244 | 440 | 359 |
|  |  | 3 |  | 1722 | 70 | 3815 |
| EMR-D | Egypt |  |  |  |  |  |
|  |  | 0 | 880 | 1916 | 6793 |  |
|  |  | I | 13897 | 2628 | 3539 | 5463 |
|  |  | 2 |  | 1735 | 3629 |  |
|  |  | 3 |  | 8500 | 818 | 9316 |
| EMR-D | Morocco |  |  |  |  |  |
|  |  | 0 | 4508 | 1799 | 5866 |  |
|  |  | I | 4745 | 1142 | 1561 | 2487 |
|  |  | 2 |  | 1609 | 1629 | 34 |
|  |  | 3 |  | 4706 | 200 | 6735 |
| EMR-D | Pakistan |  |  |  |  |  |
|  |  | 0 | 1753 |  | 5055 |  |
|  |  | I | 4838 | 1820 | 600 | 3304 |
|  |  | 2 |  | 1564 | 842 | 478 |
|  |  | 3 |  | 3227 | 114 | 2783 |
| EMR-D | Sudan |  |  |  |  |  |
|  |  | 0 | 3933 |  | 3425 |  |
|  |  | 1 | 1926 | 1503 | 1543 | 5493 |
|  |  | 2 |  | 673 | 821 | 367 |
|  |  | 3 |  | 3684 | 71 |  |
| EMR-D | Yemen |  |  |  |  |  |
|  |  | 0 | 2692 | 431 | 5124 |  |
|  |  | I | 3304 | 463 | 669 | 2998 |
|  |  | 2 |  | 636 | 153 | 6 |
|  |  | 3 |  | 4480 | 59 | 3006 |
| EUR-B | Kyrgyzstan |  |  |  |  |  |
|  |  | 0 | 8 | 893 | 3 |  |
|  |  | I | 3833 | 477 | 10 | 150 |
|  |  | 2 |  | 115 | 3097 | 3154 |
|  |  | 3 |  | 2363 | 738 | 544 |
| EUR-B | Turkey |  |  |  |  |  |
|  |  | 0 | * | 1376 | 1590 |  |
|  |  | 1 |  | 2854 | 4455 | 535 |
|  |  | 2 |  | 1472 | 2005 | 1469 |
|  |  | 3 |  | 2874 | 526 | 6572 |
| EUR-B | Uzbekistan |  |  |  |  |  |
|  |  | 0 | 12 | 828 | 3 |  |
|  |  | I | 4401 | 711 | 11 | 463 |
|  |  | 2 |  | 767 | 3815 | 3376 |
|  |  | 3 |  | 2109 | 586 | 576 |


| Table A | Dist by (contis |  | S respond riables com | nts by sub prising the | gion and DHS asset | ountry, core |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Subregion | Country | Variable value | Electricity availability | Urban-rural status | Educational level | Floor materia |
| EUR-C | Kazakhstan |  |  |  |  |  |
|  |  | 0 | 143 | 1850 | 15 |  |
|  |  | 1 | 4561 | 434 | 11 | 50 |
|  |  | 2 |  | 643 | 3706 | 2949 |
|  |  | 3 |  | 1873 | 1068 | 1801 |
| SEAR-B | Indonesia |  |  |  |  |  |
|  |  | 0 | 8086 | 2497 | 3866 |  |
|  |  | , | 20724 | 2759 | 15285 | 4966 |
|  |  | 2 |  | 3416 | 8646 | 7748 |
|  |  | 3 |  | 20138 | 1013 | 16096 |
| SEAR-B | Sri Lanka |  |  |  |  |  |
|  |  | 0 | 4495 | * | 734 |  |
|  |  | 1 | 1368 |  | 1777 | 2457 |
|  |  | 2 |  |  | 2062 | 23 |
|  |  | 3 |  |  | 1289 | 3385 |
| SEAR-B | Thailand |  |  |  |  |  |
|  |  | 0 | 1265 |  | 599 |  |
|  |  | , | 5509 | 1248 | 4984 | 312 |
|  |  | 2 |  | 1175 | 777 | 6057 |
|  |  | 3 |  | 4352 | 415 | 406 |
| SEAR-D | Bangladesh |  |  |  |  |  |
|  |  | 0 | 6795 | 462 | 4899 |  |
|  |  | I | 2323 | 65 | 2530 | 7887 |
|  |  | 2 |  | 922 | 1403 | 42 |
|  |  | 3 |  | 7678 | 295 | 1198 |
| SEAR-D | India |  |  |  |  |  |
|  |  | 0 | 36333 | 7825 | 50823 |  |
|  |  | , | 53444 | 8581 | 15546 | 18196 |
|  |  | 2 |  | 11128 | 19281 | 26507 |
|  |  | 3 |  | 62243 | 3855 | 45074 |
| SEAR-D | Nepal |  |  |  |  |  |
|  |  | 0 | 6855 | 203 | 6736 |  |
|  |  |  | 1566 | 751 | 893 | 7575 |
|  |  | 2 |  | 7475 | 688 | 193 |
|  |  | 3 |  |  | 112 | 661 |
| WPR-B | Philippines |  |  |  |  |  |
|  |  | 0 | 3981 | 954 | 366 |  |
|  |  | I | 9705 | 4037 | 4010 | 1022 |
|  |  | 2 |  | 1739 | 5718 | 5107 |
|  |  | 3 |  | 7253 | 3889 | 7854 |

[^110]Table A. 2 Numbers of DHS observations by DHS asset score, by subregion

| Asset score | Subregion |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | AFR-D | AFR-E | AMR-B | AMR-D | EMR-B | EMR-D | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-B |
| -1.11788 | 30876 | 22999 | 609 | 3117 | 6 | 8165 |  |  | 961 | 11202 | 14 |
| -0.9019455 | 1795 | 1446 | 58 | 198 |  | 396 |  |  | 13 | 6376 | 2 |
| -0.8412057 | 4676 | 19249 | 2496 | 6912 |  | 888 |  |  | 1798 | 2586 | 275 |
| -0.8398404 | 2209 | 194 | 68 | 141 | 161 | 101 |  | 1 | 838 | 8931 | 220 |
| -0.6860113 | 1451 | 379 | 15 | 70 |  | 281 |  |  | 7 | 503 | 1 |
| -0.6252716 | 645 | 1653 | 165 | 478 |  | 105 |  |  | 28 | 764 | 11 |
| -0.6239063 | 337 | 48 | 14 | 6 | 8 | 8 |  |  | 24 | 653 | 24 |
| -0.5645318 | 999 | 3029 | 208 | 1479 |  | 198 | 15 | 18 | 324 | 818 | 158 |
| -0.5631665 | 2497 | 188 | 410 | 746 | 13 | 19 |  |  | 2726 | 1275 | 1333 |
| -0.5618012 | 8206 | 2169 | 222 | 360 | 398 | 2338 |  |  | 335 | 9518 | 14 |
| -0.4700771 | 833 | 544 | 6 | 47 |  | 16 |  |  |  | 266 |  |
| -0.4305602 | 508 | 525 | 253 | 405 | 3 | 5015 | 1 |  | 338 | 3462 | 1 |
| -0.4093374 | 491 | 735 | 45 | 248 |  | 100 |  |  | 12 | 103 | 5 |
| -0.4079721 | 121 | 1 |  | 2 | 1 | 22 |  |  | 6 | 200 | 11 |
| -0.3485976 | 184 | 370 | 31 | 212 |  | 30 |  |  | 15 | 331 | 14 |
| -0.3472323 | 270 | 37 | 52 | 33 | 6 | 2 |  |  | 79 | 159 | 125 |
| -0.345867 | 1362 | 448 | 36 | 58 | 133 | 161 |  |  | 8 | 491 | 1 |
| -0.2878579 | 16 | 45 | 5 | 155 |  | 2 |  | 2 | 5 | 45 | 20 |
| -0.2864926 | 562 | 26 | 119 | 227 | 1 | 8 | 2 | 98 | 394 | 549 | 727 |
| -0.2851273 | 3196 | 3447 | 1526 | 998 | 75 | 280 |  |  | 1231 | 2233 | 188 |
| -0.214626 | 93 | 166 | 93 | 190 |  | 870 |  |  | 20 | 1165 | 1 |
| -0.1934032 | 410 | 1003 | 25 | 206 |  | 6 |  |  | I | 82 |  |
| -0.1920379 | 8 | 6 | 2 | 4 |  |  |  |  |  | 93 |  |
| -0.1538863 | 324 | 248 | 1035 | 1449 |  | 1310 | 2 |  | 1221 | 1001 | 111 |
| -0.152521 | 805 | 10 | 218 | 67 | 127 | 72 | 2 | 3 | 672 | 5242 | 42 |
| -0.1326635 | 165 | 99 | 3 | 161 |  | 37 |  |  | 1 | 50 | 13 |
| -0.1312982 | 79 | 5 | 1 | 10 |  | 3 |  |  | 26 | 72 | 97 |
| -0.1299329 | 1235 | 354 | 39 | 17 | 85 | 61 |  |  | 5 | 144 | 3 |
| -0.0719237 |  | 2 | 1 | 20 |  | 1 |  |  |  | 14 | 4 |
| -0.0705584 | 127 | 2 | 19 | 18 |  | 2 |  | 5 | 11 | 82 | 90 |
| -0.0691931 | 737 | 870 | 280 | 125 | 68 | 32 |  |  | 47 | 209 | 21 |
| -0.0098187 | 3 |  |  | 29 |  |  |  | 3 | 10 | 5 | 114 |
| -0.0084534 | 1989 | 1886 | 324 | 404 |  | 77 | 1 | 6 | 354 | 1504 | 208 |
| 0.0013082 | 153 | 204 | 24 | 166 |  | 560 |  |  | 18 | 610 |  |
| 0.0620479 | 73 | 137 | 330 | 719 |  | 267 |  |  | 103 | 379 | 33 |
| 0.0634132 | 273 | 2 | 77 | 35 | 22 | 61 | 2 | 1 | 72 | 974 | 2 |
| 0.0832707 | 199 | 210 | 2 | 244 |  | 1 |  |  |  | 48 |  |
| 0.084636 | 37 | 13 | 6 | 35 |  |  |  |  |  | 30 | 2 |
| 0.0860013 | 1517 | 537 | 15 | 11 |  | 157 |  |  | 1 | 41 |  |
| 0.1227876 | 165 | 39 | 306 | 723 |  | 454 | 541 | 27 | 285 | 708 | 161 |
| 0.1241529 | 290 | 8 | 2030 | 468 | 11 | 22 | 9 | 6 | 4584 | 1608 | 343 |
| 0.1255182 | 1065 | 515 | 437 | 199 | 681 | 3111 |  | 1 | 751 | 7685 | 9 |
| 0.1440104 | 3 | 1 |  | 24 |  | 3 |  |  |  | 1 | 2 |

Table A. 2 Numbers of DHS observations by DHS asset score, by subregion (continued)

| Asset score | Subregion |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | AFR-D | AFR-E | AMR-B | AMR-D | EMR-B | EMR-D | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-B |
| 0.1453757 | 54 |  |  | 21 |  | 6 | 2 | 3 | 5 | 33 | 82 |
| 0.146741 | 632 | 814 | 419 | 49 | 52 | 25 |  |  | 29 | 76 | 14 |
| 0.2061155 |  |  |  | 7 |  |  |  |  |  | 1 | 14 |
| 0.2074808 | 506 | 488 | 126 | 86 | 12 | 27 |  |  | 20 | 148 | 15 |
| 0.2172423 | 75 | 117 | 18 | 165 |  | 58 |  |  | 1 | 958 |  |
| 0.2682205 | 57 | 98 | 42 | 68 |  | 3 |  |  | 10 | 53 | 93 |
| 0.2779821 | 48 | 164 | 129 | 791 | 1 | 362 |  |  | 36 | 281 | 38 |
| 0.2793474 | 266 | 2 | 8 | 76 | 4 | 208 | 1 |  | 98 | 613 | 6 |
| 0.3387218 | 66 | 136 | 303 | 713 |  | 155 | 28 | 1 | 38 | 376 | 32 |
| 0.3400871 | 151 | 8 | 765 | 266 | 3 | 27 | 3 |  | 785 | 476 | 67 |
| 0.3414524 | 1018 | 568 | 245 | 179 | 423 | 1819 |  |  | 113 | 1323 | 3 |
| 0.3599446 | 2 | 5 |  | 26 |  |  |  |  |  |  |  |
| 0.3613099 | 30 | 4 | 4 | 32 |  |  |  | 5 |  | 25 | 4 |
| 0.3626752 | 942 | 1659 | 181 | 71 |  | 109 |  |  | 5 | 32 | 2 |
| 0.3994615 | 8 | 3 | 8 | 95 |  | 10 | 18 |  | 11 | 54 | 55 |
| 0.4008268 | 148 | 2 | 1403 | 275 |  | 12 | 3421 | 1277 | 1157 | 1329 | 524 |
| 0.4021921 | 1165 | 638 | 3871 | 1295 | 222 | 1206 | 2 |  | 3867 | 3953 | 461 |
| 0.4220496 |  |  |  | 3 |  |  |  |  |  | 3 | 13 |
| 0.4234149 | 398 | 402 | 286 | 60 | 15 | 10 |  | 2 | 15 | 63 | 24 |
| 0.4841547 | 14 | 12 | 19 | 13 | 1 | 1 |  |  |  | 9 | 12 |
| 0.4939162 | 96 | 162 | 149 | 991 |  | 17 | 1 |  | 12 | 409 | 3 |
| 0.4952815 | 3 | 25 | 42 | 20 |  | 1 |  | 6 | 6 | 603 |  |
| 0.554656 | 73 | 163 | 101 | 825 |  | 351 | 5 |  | 21 | 353 | 34 |
| 0.5560213 | 161 | 4 | 113 | 443 | 1 | 129 | 1 |  | 913 | 410 | 143 |
| 0.5573866 | 1085 | 770 | 245 | 169 | 320 | 1593 |  | 1 | 84 | 750 | 6 |
| 0.6153957 | 4 | 7 | 18 | 150 |  | 9 | 1 |  | 3 | 53 | 13 |
| 0.616761 | 213 | 16 | 740 | 488 | 1 | 27 | 713 | 351 | 453 | 762 | 160 |
| 0.6181263 | 987 | 751 | 2155 | 1250 | 407 | 824 |  |  | 1079 | 1207 | 146 |
| 0.6379838 | 1 |  |  | 6 |  |  |  |  |  | 2 |  |
| 0.6393491 | 615 | 610 | 130 | 86 |  | 65 |  |  | 6 | 17 |  |
| 0.6775007 | 1 |  | 87 | 69 |  | 1 | 345 | 182 | 151 | 77 | 193 |
| 0.678866 | 1594 | 555 | 2834 | 981 | 20 | 1137 | 88 | 151 | 2247 | 5373 | 1012 |
| 0.7000889 | 14 | 10 | 86 | 7 |  | 1 |  |  | 1 | 6 | 11 |
| 0.7705901 | 114 | 145 | 82 | 1526 |  | 6 | 1 | 1 | 4 | 494 | 1 |
| 0.7719554 | 47 | 41 | 467 | 222 |  | 2 | 2 | 1 | 50 | 357 | 22 |
| 0.7733208 | 1492 | 872 | 239 | 248 |  | 1132 |  | 2 | 94 | 644 | 4 |
| 0.8313299 | 1 | 3 | 7 | 157 |  | 49 |  | 1 | 1 | 51 | 16 |
| 0.8326952 | 282 | 9 | 221 | 746 |  | 136 | 804 | 161 | 460 | 842 | 332 |
| 0.8340605 | 1174 | 1297 | 2379 | 1511 | 443 | 1216 | 1 |  | 694 | 752 | 391 |
| 0.8934349 | 18 | 1 | 39 | 169 |  | 2 | 91 | 51 | 145 | 131 | 79 |
| 0.8948002 | 1529 | 1109 | 3148 | 2336 | 139 | 1193 | 34 | 185 | 1295 | 2624 | 401 |

Table A. 2 Numbers of DHS observations by DHS asset score, by subregion (continued)

|  | Subregion |  |  |  |  |  |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| Asset score | AFR-D | AFR-E | AMR-B | AMR-D | EMR-B | EMR-D | EUR-B | EUR-C | SEAR-B | SEAR-D | WPR-B |
| 0.916023 | 25 | 16 | 61 | 28 |  | 5 |  |  |  |  |  |
| 0.9555399 | 163 | 63 | 455 | 231 | 4 | 111 | 19 | 32 | 220 | 513 | 812 |
| 1.047264 | 5 | 4 | 5 | 261 |  | 4 |  |  |  | 38 | 3 |
| 1.048629 | 173 | 86 | 597 | 744 |  | 4 | 670 | 566 | 43 | 736 | 69 |
| 1.049995 | 1983 | 1914 | 3166 | 2044 |  | 954 |  | 3 | 942 | 512 | 104 |
| 1.109369 | 28 | 2 | 32 | 304 |  | 10 | 178 | 52 | 94 | 195 | 226 |
| 1.110734 | 1651 | 2216 | 4359 | 3300 | 252 | 1899 | 141 | 154 | 1157 | 2286 | 1198 |
| 1.171474 | 177 | 94 | 436 | 845 | 21 | 189 | 9 | 41 | 240 | 775 | 437 |
| 1.325303 | 49 | 15 | 81 | 564 |  |  | 284 | 177 | 3 | 187 | 37 |
| 1.326669 | 3716 | 2824 | 6718 | 6240 |  | 1235 | 438 | 628 | 1118 | 1819 | 331 |
| 1.387408 | 219 | 176 | 1182 | 1550 | 44 | 487 | 53 | 56 | 323 | 974 | 1293 |
| 1.603342 | 554 | 323 | 1698 | 3793 |  | 374 | 325 | 443 | 211 | 1074 | 350 |
| Total | 94811 | 83252 | 51504 | 57811 | 4184 | 42443 | 8254 | 4704 | 35584 | 107044 | 13686 |
|  |  |  |  |  |  |  |  |  |  |  |  |

## Appendix B: Loess plots for China data

Locally linear kernel regression smooth plots of the various risk factors according to normalized equivalized income ranking are shown in Figure B.1. Each sub-figure plots the proportion of individuals or households with the risk factor in question (y-axis ranging from 0 to 1 ) by normalized equivalized income rank for China (x-axis ranging from 0 [poorest] to 1 [richest in China]).

Figure B.I Loess plots of the prevalence of various risk factors ( $y$-axis) by normalized income rank (x-axis)


Appendix C: Literature review summaries

## UNDERWEIGHT

A total of 17 journal articles on malnutrition were summarized. These studies used several measures of nutritional status, including underweight, wasting and stunting. Many studies did not report results for underweight, but the pattern of the results was comparable across the three different measures and between studies. There were negative associations between malnutrition and income, maternal education, urban residence, availability of water or sanitation, and household goods such as a radio or television.

Wagstaff and Watanabe (2001) reviewed data on malnutrition in relation to equivalized household consumption for 20 developing countries. They found that malnutrition tended to decline monotonically with rising living standards. Inequalities in underweight tended to be larger than for wasting and stunting. Between-country differences in underweight were not statistically significant.

| Authorlyear | Anonymous (2000) |
| :---: | :---: |
| Subregion (country) | AMR-D (Bolivia) |
| Study type/design | Cross-sectional survey (DHS)/random sampling |
| Study population | Nationally representative sample of women aged 15-49 years ( $n=1$ I I87), their children aged 0-5 years and men aged 15-64 years $(n=3780)$ |
| Survey year(s) | 1998 |
| Risk factor measure(s) | Weight-for-height: $\leq-2$ SD below median of CDC/WHO reference population defined as moderate or severely wasted, $\leq-3$ SD below median defined as severely wasted Height-for-age: $\leq-2$ SD below median of CDC/WHO reference population defined as moderate or severely stunted, $\leq-3$ SD below median defined as severely stunted |
| Socioeconomic measure(s) | Area of residence (rural/urban) <br> Mother's education level: $\mathrm{a}=$ none, $\mathrm{b}=$ basic, $\mathrm{c}=$ intermediate, $\mathrm{d}=$ secondary or higher |
| Measure(s) of association | Prevalence of malnutrition by socioeconomic measure |
| Bivariate summary of results (including controlling for sex and age) | Weight-for-height <br> MODERATE/SEVERE WASTING (Total = $1.8 \%$ ): rural $=2.4$, <br> urban $=1.3$ <br> Educational level: $\mathrm{a}=2.9, \mathrm{~b}=1.9, \mathrm{c}=2.1, \mathrm{~d}=1.1$ <br> SEVERE WASTING (Total $=0.5 \%$ ): rural $=0.6$, urban $=0.4$ <br> Educational level: $a=0.3, b=0.5, c=I . I, d=0.2$ <br> Height-for-age <br> MODERATE/SEVERE STUNTING (Total $=25.6 \%$ ) rural $=35.6$, <br> urban $=18.3$ <br> Educational level: $\mathrm{a}=44.3, \mathrm{~b}=33.6, \mathrm{c}=19.1, \mathrm{~d}=12.6$ <br> SEVERE STUNTING (Total $=8.9 \%$ ): rural $=14.0$, urban $=5.1$ <br> Educational level: $\mathrm{a}=19.1, \mathrm{~b}=12.1, \mathrm{c}=6.3, \mathrm{~d}=2.9$ |
| Multivariate summary of results (including controlling for rurality, ethnicity, etc.) | Not applicable |
| Comments | Weight-for-age (underweight) results not provided No further definitions of area of residence or mother's education provided |


| Authorlyear | Anonymous (1998) |
| :---: | :---: |
| Subregion (country) | AMR-B (Dominican Republic) |
| Study type/design | Cross-sectional survey ( DHS )/random sampling |
| Study population | Nationally representative sample of women aged 15-49 years ( $n=8422$ ), their children aged $0-5$ years and men aged 15-64 years ( $n=2279$ ) |
| Survey year(s) | 1996 |
| Risk factor measure(s) | Weight-for-height: $\leq-2$ SD below median of CDC/WHO reference population defined as moderate or severely wasted, $\leq-3$ SD below median defined as severely wasted. Height-for-age: $\leq-2$ SD below median of CDC/WHO reference population defined as moderate or severely stunted, $\leq-3$ SD below median defined as severely stunted |
| Socioeconomic measure(s) | Area of residence (rural/urban) <br> Mother's educational level: $\mathrm{a}=$ none, $\mathrm{b}=$ primary ( $\mathrm{I}-4$ years), $\mathrm{c}=$ primary ( $5-8$ years), $\mathrm{d}=$ secondary, $\mathrm{e}=$ superior |
| Measure(s) of association | Prevalence of malnutrition by socioeconomic measure |
| Bivariate summary of results | Weight-for-height <br> SEVERE WASTING (Total $=0.2 \%$ ): rural $=0.2$, urban $=-$ <br> Educational level: $\mathrm{a}=0.4, \mathrm{~b}=0.3, \mathrm{c}=0.3, \mathrm{~d}=0.0, \mathrm{e}=0.0$ <br> MODERATE/SEVERE WASTING (Total $=1.2 \%$ ) rural $=1.2$, <br> urban = - <br> Educational level: $\mathrm{a}=1.7, \mathrm{~b}=1.9, \mathrm{c}=1.0, \mathrm{~d}=1.2, \mathrm{e}=0.4$ <br> Height-for-age <br> SEVERE STUNTING (Total $=2.8 \%$ ): rural $=4.4$, urban $=1.6$ <br> Educational level: $\mathrm{a}=7.7, \mathrm{~b}=5.5, \mathrm{c}=2.4, \mathrm{~d}=0.8, \mathrm{e}=0.0$ <br> MODERATE/SEVERE STUNTING (Total = $10.7 \%$ ):rural $=15.2$, urban $=7.3$ <br> Educational level: $\mathrm{a}=23 . \mathrm{I}, \mathrm{b}=16.3, \mathrm{c}=\mathrm{II} . \mathrm{I}, \mathrm{d}=5.6, \mathrm{e}=1.6$ |
| Multivariate summary of results | Not applicable |
| Comments | No data on underweight (weight-for-age). No further definitions of area of residence or mother's education provided |


| Authorlyear | Anonymous (1992) |
| :---: | :---: |
| Subregion (country) | EMR-D (Pakistan) |
| Study type/design | Cross-sectional survey ( DHS )/random sampling |
| Study population | Nationally representative sample of women aged 15-49 years ( $n=66 \mathrm{II}$ ), their children aged $0-5$ years and subset of their husbands ( $n=1354$ ) |
| Survey year(s) | 1990/1991 |
| Risk factor measure(s) | Weight-for-age (underweight), height-for-age (stunting) and weight-for-height (wasting) below -2 SD and -3 SD |
| Socioeconomic measure(s) | Rural/urban <br> Mother's educational level: $\mathrm{a}=$ no education, $\mathrm{b}=$ primary, $c=$ middle, $d=$ secondary or higher |
| Measure(s) of association | Prevalence of underweight, stunting and wasting below -2 SD and -3 SD by socioeconomic measure |
| Bivariate summary of results | BELOW -2 SD <br> Weight-for-age (underweight) <br> (Total $=40.4 \%$ ): rural $=44.6$, urban $=32.5$ <br> Educational level: $\mathrm{a}=44.9, \mathrm{~b}=37 . \mathrm{I}, \mathrm{c}=25.8, \mathrm{~d}=13.0$ |
|  | Height-for-age (stunting) <br> (Total $=50.0 \%$ ): rural $=54.9$, urban $=40.7$ <br> Educational level: $\mathrm{a}=55.5, \mathrm{~b}=43.8, \mathrm{c}=33.2, \mathrm{~d}=18.2$ |
|  | Weight-for-height (wasting) <br> (Total $=9.2 \%$ ): rural $=9.8$, urban $=8.1$ <br> Educational level: $\mathrm{a}=10.3, \mathrm{~b}=7.5, \mathrm{c}=5.3, \mathrm{~d}=3.6$ |
|  | A similar pattern was observed for -3 SD |
| Multivariate summary of results | Not applicable |
| Comments | No detailed definition provided for residence, maternal education and measures of malnutrition |


| Author/year | Anonymous (1991) |
| :---: | :---: |
| Subregion (country) | AFR-E (Zimbabwe) |
| Study type/design | Cross-sectional survey (DHS)/random sampling |
| Study population | Nationally representative sample of women aged 15-49 years ( $n=420 \mathrm{I}$ ) and their children aged $3-60$ months |
| Survey year(s) | 1988 |
| Risk factor measure(s) | Weight-for-height: $\leq-2$ SD below median of CDC/WHO reference population defined as moderate or severely wasted, - I to -I.99 SD below median defined as mildly wasted Height-for-age: $\leq-2$ SD below median of CDC/WHO reference population defined as moderate or severely stunted, -I to -1.99 SD below median defined as mildly stunted |
| Socioeconomic measure(s) | Area of residence (rural/urban) <br> Mother's educational level: $\mathrm{a}=$ none, $\mathrm{b}=$ primary, $\mathrm{c}=$ secondary/higher |
| Measure(s) of association | Prevalence of malnutrition by socioeconomic measure |
| Bivariate summary of results | ```Weight-for-height SEVERE/MODERATE WASTING (Total = I.3%): rural = I.3, urban = I.5 Educational level: a = I.9, b=1.4, c=0.5``` |
|  | MILD WASTING (Total $=9.6 \%$ ): rural $=10.1$, urban $=8.1$ <br> Educational level: $\mathrm{a}=12.6, \mathrm{~b}=9.0, \mathrm{c}=8.7$ |
|  | Height-for-age <br> SEVERE/MODERATE STUNTING (Total = 29.0\%): rural $=33.6$, <br> urban $=14.3$ <br> Educational level: $\mathrm{a}=37.0, \mathrm{~b}=30.4, \mathrm{c}=15.6$ |
|  | MILD STUNTING (Total $=35.5 \%$ ): rural $=36.5$, urban $=32.0$ <br> Educational level: $\mathrm{a}=35.6, \mathrm{~b}=36.2, \mathrm{c}=32.7$ |
| Multivariate summary of results | Not applicable |
| Comments | No data on underweight (weight-for-age) |


| Authorlyear | Chen (1996) |
| :---: | :---: |
| Subregion (country) | WPR-B (China) |
| Study type/design | Meta-analysis using survey data from 1987 Child Survey (9 provinces), 1992 National Child Survey, 1990 Nutrition Surveillance—State Statistic Bureau (7 provinces including Beijing), 1992 Third National Nutritional Survey (all provinces), 1990 nutrition surveillance data |
| Study population | Nationally representative samples |
| Survey year(s) | 1987, 1992, 1990, 1992, 1990, respectively |
| Risk factor measure(s) | Prevalence of underweight and stunting (neither is defined) |
| Socioeconomic measure(s) | Area of residence (urban, rural) <br> Household income (groups I-6 based on percentile groupings or 100 yuan/person per year increase); <br> Safe drinking-water (percentage piped drinking-water coverage) <br> Mother or father illiterate <br> Poverty (no definition provided) <br> Results also given according to geographical area: <br> $\mathrm{I}=$ Beijing, Tianjin, Liaoning <br> II = Jilin, Heilongjiang, Jiangsu, Shandong <br> III = Shanghai, Zejiang, Guangdong <br> IV = Hebei, Inner Mongolia, Shanxi, Anhui, Henan <br> V = Fujian, Jiangxi, Hubei, Hunan <br> VI = Hainan, Guizhou, Shaanxi, Gansu, Qinghai, Ningxia <br> VII = Guangxi, Sichuan, Yunnan, Xinjiang |
| Measure(s) of association | Prevalence of malnutrition by area of residence Attributable risk of factors relating to stunting in rural China |
| Bivariate summary of results | The national prevalence of moderate underweight in children aged $0-6$ years was $21.3 \%$ in 1987 and $17.9 \%$ in 1992 (figures for urban and rural areas are provided by province only-on average rural prevalences are markedly higher than urban) |
|  | In 1992 for children aged 0-5 years, national prevalence of underweight (moderate plus severe) was $17.9 \%$ and that of stunting was $34.0 \%$ <br> Prevalence of stunting in urban areas in this group was II.4\% for boys and II.5\% for girls, and in rural areas $39.3 \%$ and $40.5 \%$, respectively |
| Multivariate summary of results | Attributable risk of factors related to stunting in rural China by geographical area: <br> Poverty: $\mathrm{IV}=5.7, \mathrm{~V}=50.0, \mathrm{VI}=6.5, \mathrm{VII}=32.4$ <br> Safe drinking-water: $\mathrm{II}=23.7, \mathrm{III}=35 . \mathrm{I}, \mathrm{IV}=\mathrm{I} 5.3, \mathrm{~V}=24.2$, <br> $\mathrm{VII}=26.5$ <br> Illiterate mother or father: $\mathrm{I}=3 \mathrm{I} .0, \mathrm{II}=20.0, \mathrm{IV}=9.9, \mathrm{~V}=26.5$, $\mathrm{VI}=18.7$ <br> Income (IOO yuan/person per year increase): $I=-I, I I=-2$, $\mathrm{III}=-2, \mathrm{IV}=-3, \mathrm{~V}=-\mathrm{I}, \mathrm{VI}=-3, \mathrm{VII}=-\mathrm{I}$ |
| Comments | Difficult to interpret, broken up into regions, no raw data Few definitions given (none of stunting, underweight or poverty) |


| Authorlyear | Delpeuch et al. (2000) |
| :--- | :--- |
| Subregion (country) | AFR-E (Congo) |
| Study type/design | Cross-sectional survey/cluster sampling |
| Study population | Urban (Brazzaville) children aged 0-5 years ( $n=2373$ ) |
| Survey year(s) | I986, I991 |
| Risk factor measure(s) | Low weight-for-height: -I SD below mean weight-for-height of |
|  | CDC/WHO reference population (prevalence of wasting, defined |
|  | as weight-for-height <-2 SD, was only 4.2\%) |
|  | Stunting: -2 SD below mean height-for-age of CDC/WHO |
|  | reference population |
|  | BMI (kg/m2): thin <l8.5 kg/m², overweight $\geq 25$ kg/m |



|  | Firewood only fuel | Yes $=25.0, \mathrm{No}=20.0$ | 0.412 |
| :---: | :---: | :---: | :---: |
|  | Charcoal main fuel | Yes $=20.8, \mathrm{No}=25.6$ | 0.372 |
|  | Paraffin main fuel | Yes $=13.2, \mathrm{No}=27.5$ | 0.018 |
|  | Cleanliness | Clean $=15.2$ | 0.290 |
|  |  | Moderately clean $=23.4$ |  |
|  |  | Dirty/very dirty $=29.5$ |  |
|  | Maternal education | None $=29.4$ | 0.885 |
|  |  | Primary (incomplete) $=24.0$ |  |
|  |  | Primary (complete) $=25.0$ |  |
|  |  | Secondary and above $=20.9$ |  |
|  | When age of child was mothers had less stunt | controlled for, however, bette children ( $P=0.045$ ) | ucated |
| Multivariate summary | Underweight by economic | status of family by odds ratio |  |
| of results | Economic status | Odds ratio (95\% CI) |  |
|  | Mid-upper (reference) | 1.00 |  |
|  | Lower | 2.57 (0.68-9.80) |  |
|  | Low/very low | 2.62 (0.63-11.03) |  |
| Comments | Not documented how | economic status measured |  |
|  | Positive association bet been shown in other st in this study once age | ween maternal education and udies in Uganda; only positive controlled for | ting has ciation |
|  | No table showing numb | er of mothers in each educati | group |
|  | Proportion underweigh from 1988/I989 Ugand | in this study (24\%) similar to DHS (23\%) | ults |
|  | No record of how many malnutrition, although children showed greate and $P=0.0 I 2$, respecti | children had more than one tated that both stunted and incidence of underweight ( $P$ ely) | of d 001 |
|  | Neither age, sex nor b underweight, although older than younger chil | rth order had any effect on in here was higher incidence of dren ( $P \leq 0.000$ I) | ce of ing in |


| Authorlyear | Li et al. (1999) |
| :---: | :---: |
| Subregion (country) | WPR-B (China) |
| Study type/design | Cross-sectional survey/random sampling |
| Study population | Children aged 0-7 years from four poor rural counties in Yunnan Province ( $n=2019$ ). Sample included four minority ethnic groups (Hani, Yi, Hui, Miao) and representatives of the dominant ethnic group in the area (Han) |
| Survey year(s) | Not recorded |
| Risk factor measure(s) | Prevalence of stunting (height-for-age), underweight (weight-forage) and wasting (weight-for-height) from -2 to -2.99 SD (moderate) or $\leq-3$ (severe) SD below median of NCHS/WHO reference population |
| Socioeconomic measure(s) | Family income per capita in previous year: low $=<200$ yuan; medium $=200-499$ yuan; high $=\geq 500$ yuan <br> Ethnicity: Hani, Yi, Hui, Miao and Han (dominant in area) <br> Drinking-water: tap water, pump water, well or spring water, rain or snow water, lake, river or pool water <br> Lavatory: public toilet, indoor flush toilet, private pit, no toilet |
| Measure(s) of association | Odds ratio <br> Prevalence of malnutrition by socioeconomic measure |
| Bivariate summary of results | Overall prevalences <br> Moderate and severe underweight: $18.9 \%$ <br> The prevalence of being underweight increased with age, peaking at 12-17 months then decreased steadily; boys were more likely to be underweight than girls $\left(\chi^{2}=15.58, P<0.00 \mathrm{I}\right)$; most underweight children (87.1\%) were also stunted <br> Moderate and severe stunting: 51.0\% <br> The prevalence of being underweight increased with increasing age, peaking at 3 years but maintained fairly high prevalence; boys were more likely to be stunted than girls $\left(\chi^{2}=7.36\right.$, P<0.0I) <br> Moderate and severe wasting: $1.4 \%$ |
|  | Moderate and severe stunting adjusted for sex and age <br> Family income: low: $\mathrm{OR}=5.2(2.8-9.9)$; medium: $\mathrm{OR}=2.2$ <br> (I.3-3.6); high: $O R=I$ (reference) |
|  | Drinking-water: well and spring: $O R=1.5(0.9-2.5)$; lake, river, pool: $\mathrm{OR}=4.5$ (2.6-7.9); tap: $\mathrm{OR}=\mathrm{I}$ (reference) |
|  | Lavatory: private pit: $\mathrm{OR}=2.3$ (I.0-5.3); no toilet: $\mathrm{OR}=3.0$ (I.2-7.4); public toilet: $O R=I$ (reference) |
|  | Certain ethnic groups (in particular Miao) had high odds ratios for moderate/severe stunting compared with Han |
| Multivariate summary of results | Moderate and severe stunting adjusted for sex and age: Family income: low: $\mathrm{OR}=4.1$ (1.9-8.5); medium: $\mathrm{OR}=1.2$ (0.7-2.3); high: $O R=1.0$ (reference) |
|  | Drinking water: well and spring: $\mathrm{OR}=1.2$ (0.7-2.2); lake, river, pool: $O R=2.3$ (I.0-5.I); tap: $O R=I$ (reference) |
| Comments | Most of the analysis was performed on stunting only |


| Author/year | Madzingira (1995) |
| :--- | :--- |
| Subregion (country) | AFR-E (Zimbabwe) |
| Study type/design | Cross-sectional survey (DHS)/random sampling |
| Study population | Nationally representative sample of women aged I5-49 years <br> $(n=420$ I) and children aged 0-5 years |
| Survey year(s) | 1988 |
| Risk factor measure(s) | Height-for-age, weight-for-age, weight-for-height |
| Socioeconomic | Area of residence (urban/rural) <br> measure(s) |
| Maternal education |  |
| Measure(s) of | Odds ratio <br> association |
| Prevalence of malnutrition by area of residence |  |


| Author/year | Matulessy et al. (1992) |
| :--- | :--- |
| Subregion (country) | SEAR-B (Jakarta, Indonesia) |
| Study type/design | Analysis of survey data/nationally representative routinely <br> collected data by Central Bureau of Statistics of Indonesia and <br> Directorate of Nutrition (I986); also National Survey of Vitamin |
| A (I978) and other nationally collected data |  |


| Author/year | Monteiro et al. (1992) |
| :---: | :---: |
| Subregion (country) | AMR-B (Brazil) |
| Study type/design | Analysis of two national nutrition surveys (both multistage cluster stratified randomized sampling): Estudo Nacional da Despesa Familiar (1975) and Pesquisa Nacional sobre Saude e Nutricao (1989) |
| Study population | Children aged 0-5 years for whom weight, age and sex information was provided ( $n=36407$ in 1975, $n=3571$ in 1989) |
| Survey year(s) | 1975, 1989 (only results for 1989 reported here) |
| Risk factor measure(s) | Prevalence of malnutrition (weight-for-age indices $<-2 z$-scores below reference average weight expected for age and sex) |
| Socioeconomic measure(s) | Area of residence Quartile of monthly per capita family income |
| Measure(s) of association | Prevalence of malnutrition by socioeconomic measure in 1989 |
| Bivariate summary of results | Measure Prevalence |
|  | AREA OF RESIDENCE |
|  | Urban 5.6\% |
|  | Rural 10.6\% |
|  | INCOME QUARTILE |
|  | Q1 (lowest) 13.6\% |
|  | Q2 $9.5 \%$ |
|  | Q3 $4.8 \%$ |
|  | Q4 $1.4 \%$ |
| Multivariate summary of results | Not applicable |
| Comments | Nationally representative. Over time (1975-1989) the prevalence of weight-for-age malnutrition decreased, although this decrease was lowest in most vulnerable groups, i.e. those in the lowest income quartile <br> Also geographical pattern of malnutrition |


| Authorlyear | Monteiro et al. (1995) |
| :---: | :---: |
| Subregion (country) | AMR-B (Brazil) |
| Study type/design | Cross-sectional anthropometric surveys/random sampling |
| Study population | Nationally representative sample, excluding pregnant women (in 1989: 5969 children and 23544 adults) |
| Survey year(s) | 1974-1975, 1989 (only results for 1989 reported here) |
| Risk factor measure(s) | Weight-for-height indices compared with CDC/WHO reference population used to measure children; <br> Weight-for-age indices compared with CDC/WHO reference population used to measure children: <-2 z-scores classified as underweight; $>2 \mathrm{z}$-scores classified as overweight BMI (underweight $<18.5 \mathrm{~kg} / \mathrm{m} 2$, obese $>30.0 \mathrm{~kg} / \mathrm{m}^{2}$ ) used to classify adults |
| Socioeconomic measure(s) | Per capita family income (obtained through the standardized questionnaire of Instituto Brasiliero de Geografia e Estatística): poorest $=$ lowest $30 \%$ of sample (approx. $<$ US $\$ 3$ I per month); middle $=$ next $40 \%$ (approx. US\$ 30-9I per month); highest $=$ top 30\% (approx. >US\$ 90) |
| Measure(s) of association | Prevalence of underweight (weight-for-age $<-2$ z-scores for children) by socioeconomic measure |
| Bivariate summary of results | Total prevalence of underweight in children (aged I-4 years in $1989=7.6 \%$ |
|  | Prevalence of underweight in children (aged I-4 years) by socioeconomic measure, 1974 |
|  | Poorest 26.5\% (SE 0.49) |
|  | Intermediate II.6\% (SE 0.49) |
|  | Richest 3.9\% (SE 0.44) |
|  | Prevalence of underweight in children (aged I-4 years) by socioeconomic measure, 1989 |
|  | Poorest I2.2\% (SE 0.83) |
|  | Intermediate 3.8\% (SE 0.63) |
|  | Richest $\quad 1.4 \%$ (SE 0.59) |
|  | Also results for adults by sex and BMI (sex difference among adults) |
| Multivariate summary of results | Not applicable |
| Comments | Decreasing prevalence of underweight children over time, but still a problem among the poorest |


| Author/year | Nube et al. (1998) |
| :---: | :---: |
| Subregion (country) | AFR-D (Ghana) |
| Study type/design | Analysis of household survey data (LSMS) |
| Study population | Adults aged 20-65 years and their children aged $<5$ years, excluding individuals from households with expenditure $>1$ million cedis/year, pregnant and lactating women and those with BMI $<10$ or $>40$ ( $n=4228: 2114$ males and 2114 females $)$ |
| Survey year(s) | 1988-1989 |
| Risk factor measure(s) | Height-for-age $z$-scores (number of SDs below mean of reference data [children aged $<5$ years only]) |
| Socioeconomic measure(s) | Area/type of residence (residence code): <br> Rursavfarm = rural savannah farm household <br> Rurforfarm = rural forest farm household <br> Rurcoafarm = rural coastal farm household <br> Rursavnonf = rural savannah non-farm household <br> Rurfornonf = rural forest non-farm household <br> Rurcoanonf = rural coastal non-farm household <br> Urbsavunsk = urban savannah unskilled household <br> Urbforunsk = urban forest unskilled household <br> Urbcoasunsk = urban coastal unskilled household <br> Urbsavskil = urban savannah skilled household <br> Urbforskil = urban forest skilled household <br> Urbcoaskil = urban coastal skilled household |
|  | Note: Correlation of these codes with mean per capita expenditure, mean years of schooling of head of household, mean number of households with electricity and building material of house also provided (Urbcoaskil highest and Rursavfarm lowest scores) |
| Measure(s) of association | Mean height-for-age z-score |
| Bivariate summary of results | Mean height-for-age $z$-score by residence code |
|  | Residence code Mean height-for-age z-score |
|  | Rursavfarm -I.46 |
|  | Rurforfarm -1.65 |
|  | Rurcoafarm -1.17 |
|  | Rursavnonf -1.08 |
|  | Rurfornonf -1.15 |
|  | Rurcoanonf -0.89 |
|  | Urbsavunsk -I.69 |
|  | Urbforunsk -1.19 |
|  | Urbcoasunsk $\quad-0.79$ |
|  | Urbsavskil -0.26 |
|  | Urbforskil -I.15 |
|  | Urbcoaskil -0.71 |
| Multivariate summary of results | Not applicable |
| Comments | Article focuses on how strongly BMI or other nutritional measures are correlated with various measures of SES |


| Author/year | Quinn et al. (1995) |
| :---: | :---: |
| Subregion (country) | AFR-E (Malawi) |
| Study type/design | Meta-analysis of survey data/three cross-sectional surveys carried out during similar time periods using same protocol for anthropometric data collection |
| Study population | Children aged over 24 months. Characteristics of three surveys are (i) preschool children attending institutions known to cater to high-income families in urban centres (Blantyre, Lilongwe, Zomba) ( $n=350$ ); (ii) rural survey of villages conducted in Ntchisi district ( $n=667$ ); (iii) survey of low-income urban neighbourhoods conducted in Blantyre and Lilongwe ( $n=225$ ) |
| Survey year(s) | 1990-1991, 1989, 1991, respectively |
| Risk factor measure(s) | Prevalence of stunting or wasting (height-for-age and weight-forheight, respectively, below -2 z-score) compared with mean of CDC/WHO reference population |
| Socioeconomic measure(s) | As per survey sample: I = high income urban area; 2 = low income rural area; $3=$ low income urban area |
| Measure(s) of association | Mean height and weight by survey |
| Bivariate summary of results | Prevalence of stunting or wasting (height-for-age and weight-for-height respectively below -2 z-score), all ages (24-59 months) combined |
|  | I. Urban rich 2. Rural poor 3. Urban poor |
|  | $\begin{array}{llll}\text { Stunting } & 7.8 & 83.2 & 69.3\end{array}$ |
|  | $\begin{array}{llll}\text { Wasting } & 0.6 & 0.4 & 4.0\end{array}$ |
| Multivariate summary of results | Not applicable |
| Comments | No measures of underweight |





| Father's education (years) | - | 0.92 | 0.95 |
| :---: | :---: | :---: | :---: |
| Mother's education (years) | 0.91 | - | 0.96 |
| RURAL |  |  |  |
|  | Age group (months) |  |  |
| Variable | 0-5 | 6-11 | 12-29 |
| TV/radio |  |  |  |
| Radio only | 0.59 | 0.74 | 0.86 |
| Television | 0.61 | 0.56 | 0.77 |
| Neither (reference) | I | I | I |
| Water source |  |  |  |
| Purchased | - | - | 0.92 |
| Pump/pipe | - | - | 0.80 |
| Well (reference) | - | - | I |
| Toilet type |  |  |  |
| Pit | - | 0.82 | 0.83 |
| Sealed | - | 0.66 | 0.71 |
| None (reference) | - | I | I |
| Flooring material |  |  |  |
| Wood | - | - | 1.01* |
| Cement | - | - | 0.81 |
| Bamboo/all else (reference) | - | - | I |
| Cooking fuel |  |  |  |
| Gas or oil | - | - | 0.70 |
| Wood (reference) | - | - | 1 |
| Father's education (years) | 0.93 | 0.94 | 0.93 |
| Mother's education (years) | - | - | 0.95 |

(all $P<0.05$ except * not significant)
Wasting
URBAN

|  | Age group (months) |  |  |
| :--- | :--- | :--- | :--- |
| Variable | $0-5$ | $6-11$ | $12-29$ |
| TV/radio |  |  |  |
| Radio only | - | 1.05 | - |
| Television | - | 0.66 | - |
| Neither (reference) | - | 1 | - |
| Wall material | - |  |  |
| Cement | - | 0.80 | - |
| All else | - | 1 | - |
| Wood (reference) |  |  |  |


| Mother's employment |  |  |  |
| :--- | :--- | :--- | :--- |
| At home | - | - | 0.94 |
| Away from home | - | - | 0.77 |
| None (reference) | - | - | 1 |
| Father's education (years) | - | - | 0.95 |
| Mother's education (years) | - | - | 0.96 |

RURAL

|  | Age group (months) |  |  |
| :--- | :--- | :--- | :--- |
| Variable | $0-5$ | $6-11$ | $12-29$ |
| TV/radio | - |  |  |
| Radio only | - | - | 0.90 |


| Television | - | - | 0.58 |
| :--- | :--- | :--- | :--- |
| Neither (reference) | - | - | 1 |
| Water source | - | 1.59 | - |
| Purchased | - | 0.95 | - |
| Pump/pipe | - | 1 | - |
| Well (reference) | - | - | 1.33 |
| Mother's employment | - | - | 0.91 |
| At home | - | - | 1 |
| Away from home |  |  |  |
| None (reference) |  |  |  |

(all $P<0.05$ )

## Multivariate summary of results

## Comments

Model applied
Similar results as for bivariate analysis
Results not given in paper
No indication of how representative this study is for the Philippines as a whole. Weight-for-age given only at descriptive level, as investigators did not think it differentiated stunted from wasted children.
Investigators showed a clear socioeconomic difference between urban and rural households (in terms of housing characteristics such as water source, cooking fuel, ownership of television and radio, floor and wall materials, as well as parental education) but similar household size and density

## Water and sanitation

Good quality water and adequate sanitation are commonly regarded as basic socioeconomic prerequisites for health, and thus their lack is associated with low SES. It is difficult to find individual-level data on water supply, as most research on water quality examines the issue at aggregate rather than individual level.

As water quality and sanitation facilities are regarded as indicators of SES, and the relationship between poverty and poor water quality and sanitation is widely accepted, it was difficult to find appropriate articles as the relationship between these two factors is rarely examined. More often, water quality and/or sanitation are used as part of socioeconomic indices. As a result, only nine journal articles were revealed by the literature review search, of which three were retrieved. On examination, two of these articles were deemed unsuitable for our purposes and were discarded.

The remaining article reported on a study carried out in rural Bangladesh. However, some of the subjects of the study were recipients of support from nongovernmental organizations (NGOs) and thus were not necessarily typical in terms of education and access to water and sanitation. In general, however, there was a strong positive relationship between SES and safe sanitation and a weaker positive relationship between SES and safe waste disposal. A number of measures for SES were used, including educational level, housing level, occupation and land ownership. A similar pattern was seen across all these measures.

| Authorlyear | Hadi (2000) |
| :---: | :---: |
| Region (country) | SEAR-D (Bangladesh) |
| Study type/design | Survey/random sample of proportionally representative households selected from those covered by Watch, the demographic and health surveillance system that covers 70 villages located in 10 regions of Bangladesh ( $n=1556$ households) |
| Study population | A total of 1556 households representative of those covered by Watch Rural households only |
| Survey year(s) | 1995 |
| Risk factor measure(s) | Safe sanitation behaviour ( $\mathrm{Yes} / \mathrm{No}$ ) <br> Safe disposal of solid waste (Yes/No) |
| Socioeconomic measure(s) | In Bangladesh, households that own <50 decimals of land and in which the principal worker had to sell at least 100 days of labour over the past year in order to subsist are eligible for aid from NGOs <br> Survey sample was divided into: participants (those households eligible for and receiving NGO aid); non-participants (those households eligible for but not receiving NGO aid); and not eligible (those households not eligible for NGO aid) Comparison of mean years of schooling, percentage literate, mean amount of land (in decimals), occupation and exposure to the media showed that non-participants were the most disadvantaged, and the not eligible group were the least disadvantaged <br> Education: none, I-5 years, $\geq 6$ years <br> Land ownership: none, I-I99 decimals, $\geq 200$ decimals <br> Occupation: labourer, agricultural worker, service/business sector <br> Housing conditions: poor, good <br> Exposure to the media: poor, better |
| Measure(s) of association | Prevalence of safe sanitation behaviour and safe disposal of solid waste by socioeconomic measure; log odds ratio (looking at variables outside of CRA) |
| Bivariate summary of results (including controlling for sex and age) | Prevalence of safe sanitation behaviour (i.e. sanitary latrine use) and safe disposal of solid waste by socioeconomic measure |
|  | Safe sanitation Safe disposal of |
|  | NGO eligibility status |
|  | $\begin{array}{lll}\text { Non-participants } & 6.7\end{array}$ |
|  | Participants 23.047 .0 |
|  | $\begin{array}{lll}\text { Not eligible } & 35.3 & 45.9\end{array}$ |
|  | Education |
|  | None 15.7 44.7 |
|  | $\begin{array}{ll}\text { I-5 years } & 30.3\end{array}$ |
|  | $\begin{array}{lll}\geq 6 \text { years } & 48.8 & 50.7\end{array}$ |
|  | Land ownership |
|  | Landless 17.2 48.5 |
|  | $\begin{array}{ll}\text { I-I99 decimals } & 26.6\end{array}$ |
|  | $\geq 200$ decimals 48.5 |



## Unsafe sex

Two journal articles on unsafe sex were retrieved of the 30 revealed by the literature search. One was unsuitable for our purposes and thus discarded. The remaining article was the result of a study on knowledge of sexually transmitted infections (particularly syphilis and gonorrhoea) among married women in rural Bangladesh. The results showed that increasing knowledge of these sexually transmitted infections correlated with increasing level of education among these women. This knowledge was also more common among women whose husbands were professionals or service-holders.

| Authorlyear | Khan et al. (1997) |
| :---: | :---: |
| Subregion (country) | SEAR-D (Bangladesh) |
| Study type/design | Survey |
| Study population Sample characteristics: region, urban/rural, socioeconomic strata, random, sample size | Married women of reproductive age in four thanas (Abhoynagar, Bagherpara, Keshopur, Sirajganj) of Bangladesh ( $n=8674$ ) |
| Survey year(s) | 1994 |
| Risk factor measure(s) | Knowledge of which diseases are spread through sexual intercourse <br> Knowledge of how syphilis and gonorrhoea are spread if they have heard of them <br> Knowledge of how to prevent sexually transmitted diseases |
| Socioeconomic measure(s) | Subject's educational level: none, primary, above primary Husband's educational level: none, primary, above primary Husband's occupation: farmer or self-employed, labourer, trader/businessman, service-holder, professional |
| Measure(s) of association | Prevalence of awareness of syphilis and gonorrhoea Odds ratio |
| Bivariate summary of results (including controlling for sex and age) | Prevalence of awareness of syphilis and gonorrhoea and how they are spread, by socioeconomic measure |
|  | Socioeconomic variable Awareness of syphilis/gonorrhoea Woman's education |
|  | None 21.5\% |
|  | Primary 27.8\% |
|  | Above primary 40.7\% |
|  | $\chi^{2}$ ( 173.82 |
|  | $P$ value $<0.001$ |
|  | Husband's education |
|  | None 21.5\% |
|  | Primary 25.5\% |
|  | Above primary 38.6\% |
|  | $\chi^{2}$ I86.47 |
|  | $P$ value $<0.001$ |
|  | Husband's occupation |
|  | Farmer/self employed 26.2 |
|  | Labourer 23.4 |
|  | Trader/business 30.2 |
|  | Service-holder 45.4 |
|  | Professional 53.4 |
|  | $\chi^{2}$ 123.03 |
|  | $P$ value $<0.001$ |


| Multivariate summary of results (including controlling for rurality, ethnicity, etc.) | Odds ratios measuring association between socioeconomic measures and awareness of syphilis and gonorrhoea |  |  |
| :---: | :---: | :---: | :---: |
|  | Socioeconomic variable Woman's education | OR | 95\% Cl |
|  | None | 1.00 |  |
|  | Primary | 1.55 | 1.34-1.78 |
|  | Above primary | 2.93 | 2.38-3.60 |
|  | Husband's education |  |  |
|  | None | 1.00 |  |
|  | Primary | 1.27 | 1.10-1.46 |
|  | Above primary | 1.68 | 1.42-1.98 |
|  | Husband's occupation |  |  |
|  | Farmer/self employed | 1.00 |  |
|  | Labourer | 1.04 | 0.91-1.19 |
|  | Trader/business | 1.12 | 0.94-1.36 |
|  | Service-holder | 1.36 | 1.02-1.81 |
|  | Professional | 1.72 | 1.19-2.49 |

## Appendix references

Anonymous (1991) Zimbabwe 1988: results from the demographic and health survey. Studies in Family Planning, 22:395-399.
Anonymous (1992) Pakistan 1990/91: results from the demographic and health survey. Studies in Family Planning, 23:274-278.
Anonymous (1998) Dominican Republic 1996: results from the demographic and health survey. Studies in Family Planning, 29:423-427.
Anonymous (2000) Bolivia 1998: Results from the demographic and health survey. Studies in Family Planning, 31:257-261.

Chen CM (I996) Nutrition status of the Chinese people. Biomedical and Environmental Sciences, 9:81-92.
Delpeuch F, Traissac P, Martin-Prevel Y, Massamba JP, Maire B (2000) Economic crisis and malnutrition: socioeconomic determinants of anthropometric status of preschool children and their mothers in an African urban area. Public Health Nutrition, 3:39-47.

Hadi A (2000) A participatory approach to sanitation: experience of Bangladeshi NGOs. Health Policy and Planning, I5:332-337.
Khan MA, Rahman M, Khanam PA, Barkat-e-Khuda, Kane TT, Ashraf A (I997) Awareness of sexually transmitted disease among women and service providers in rural Bangladesh. International Journal of STD and AIDS, 8:688-696.
Kikafunda JK, Walker AF, Collett D, Tumwine JK (1998) Risk factors for early childhood malnutrition in Uganda. Pediatrics, 102:E45.
Li Y, Guo G, Shi A, Anme T, Ushijima H (1999) Prevalence and correlates of malnutrition among children in rural minority areas of China. Pediatrics International, 41:549-556.

Madzingira N (1995) Malnutrition in children under five in Zimbabwe: effect of socioeconomic factors and disease. Social Biology, 42:239-246.

Matulessy PF, Asumi R, Thamrin MH, Husaini, Angeles TI, Kariadi D (1992) Nutrition situation in Metropolitan Jakarta. Southeast Asian Journal ofTropical Medicine and Public Health, 23:S3-16.

Monteiro CA, Benicio MH, lunes R, Gouveia NC, Taddei JA, Cardoso MA (1992) Nutritional status of Brazilian children: trends from 1975 to 1989. Bulletin of the World Health Organization, 70:657-666.
Monteiro CA, Mondini L, de Souza AL, Popkin BM (1995) The nutrition transition in Brazil. European Journal of Clinical Nutrition, 49:105-1I3.
Nube M, Asenso-Okyere WK, van den Boom GJ (1998) Body mass index as indicator of standard of living in developing countries. European Journal of Clinical Nutrition, 52:136-144.

Popkin BM, Keyou G, Zhai F, Guo X, Ma H, Zohoori N (1993) The nutrition transition in China: a cross-sectional analysis. European Journal of Clinical Nutrition, 47:333-346.

Quinn VJ, Chiligo-Mpoma MO, Simler K, Milner J (1995) The growth of Malawian preschool children from different socioeconomic groups. European Journal of Clinical Nutrition, 49:66-72.

Radebe BZ, Brady P, Siziya S, Todd H (1996) Maternal risk factors for childhood malnutrition in the Mazowe district of Zimbabwe. Central African Journal of Medicine, 42:240-244.

Ricci JA, Becker S (I996) Risk factors for wasting and stunting among children in Metro Cebu, Philippines. American Journal of Clinical Nutrition, 63:966-975.

Wagstaff A, Watanabe N (200I) Socioeconomic inequalities in child malnutrition in the developing world. World Bank, Washington, DC.

## Chapter 25

# Estimating attributable burden OF DISEASE FROM EXPOSURE AND HAZARD DATA 

Stephen Vander Hoorn, Majid Ezzati, Anthony Rodgers, Alan D. Lopez and Christopher J.L. Murray

## 1. Estimating population attributable FRACTIONS

As described in earlier chapters, the contribution of a risk factor to disease burden (expressed as the fraction of disease or death attributable to the risk factor in a population) is given by the generalized "potential impact fraction" (PIF) in Equation 1a (Drescher and Becher 1997; Eide and Heuch 2001; Walter 1980).

$$
\begin{equation*}
\text { PIF }=\frac{\int_{x=0}^{m} R R(x) P(x) d x-\int_{x=0}^{m} R R(x) P^{\prime}(x) d x}{\int_{x=0}^{m} R R(x) P(x) d x} \tag{1a}
\end{equation*}
$$

$R R(\mathrm{x})$ : relative risk at exposure level x
$P(x)$ : population distribution of exposure
$P^{\prime}(x)$ : counterfactual distribution of exposure, and
$m$ : maximum exposure level
The first and second terms in the numerator of Equation 1a represent the total exposure-weighted risk of disease or mortality in the population under current and counterfactual exposure distributions. The corresponding relationship when exposure is described as a discrete variable with $n$ levels is given by:

$$
\begin{equation*}
P I F=\frac{\sum_{i=1}^{n} P_{i} R R_{i}-\sum_{i=1}^{n} P_{i}^{\prime} R R_{i}}{\sum_{i=1}^{n} P_{i} R R_{i}} \tag{1b}
\end{equation*}
$$

The PIF equation can be used to estimate the population attributable fraction (PAF), defined as the proportional reduction in disease or death that would occur if exposure to the risk factor were reduced to the counterfactual exposure distribution. The remainder of this chapter outlines how data on exposure, hazard and disease burden were combined to derive estimates of attributable disease burden, with estimation of the population attributable fraction as the intermediate step. The application of Equations 1 a and 1 b in the context of the comparative risk assessment (CRA) project is discussed and several issues regarding its implementation are detailed.

## 2. Estimating attributable mortality and BURDEN OF DISEASE

For each risk factor-disease outcome pair, PAFs for each of the 224 age, sex, subregion ${ }^{1}$ groups were calculated using the relationships in Equation 1, separately for mortality $\left(P A F_{M}\right)$ and incidence $\left(P A F_{I}\right)$ when the relative risks for mortality and incidence were different. For each of these 224 groups, the estimates of mortality $\left(A M_{i j}\right)$ and burden of disease $\left(A B_{i j}\right)$ from disease $j$ attributable to risk factor $i$ were calculated as below. Burden of disease, reported annually in the annexes of the World Health Report, was expressed in disability-adjusted life years (DALYs), with methods and assumptions described elsewhere (Murray and Lopez 1996). Specifically:

$$
\begin{aligned}
A M_{i j} & =P A F_{M-i j} \times M_{j} \\
A Y L L_{i j} & =P A F_{M-i j} \times Y L L_{j} \\
A Y L D_{i j} & =P A F_{I-i j} \times Y L D_{j} \\
A B_{i j} & =A Y L L_{i j}+A Y L D_{i j}
\end{aligned}
$$

Where "A" indicates "attributable" and
YLL: years of life lost to premature mortality
YLD: years of life lived with disability due to disease incidence
For those risk factors with insufficient data to estimate a relative risk model (e.g. occupational or alcohol-caused injuries or the effects of lead exposure on blood pressure), disease burden or mortality was estimated using existing registers or corresponding hazard relationships. Estimates were then aggregated across age groups to obtain subregional estimates,
and across subregions to obtain global estimates. The details of this aggregation are described later in this chapter.

## 3. Counterfactual exposure distribution

The estimates of burden of disease and injuries due to risk factors in the CRA project are based on a counterfactual of theoretical-minimum-risk exposure distribution, defined in chapter 1 and described in individual risk factor chapters. By using the theoretical-minimum-risk exposure distribution, which by definition has a relative risk of 1.0, as the counterfactual exposure distribution or level, Equations 1a and 1b are reduced to:

$$
\begin{equation*}
A F=\frac{\int_{x=0}^{m} R R(x) P(x) d x-1}{\int_{x=0}^{m} R R(x) P(x) d x} \tag{2a}
\end{equation*}
$$

and

$$
\begin{equation*}
A F=\frac{\sum_{i=1}^{n} P_{i}\left(R R_{i}-1\right)}{\sum_{i=1}^{n} P_{i}\left(R R_{i}-1\right)+1} \tag{2b}
\end{equation*}
$$

### 3.1 Theoretical-MINIMUM-RISK EXPOSURE DISTRIBUTION FOR CONTINUOUS EXPOSURE VARIABLES

The theoretical-minimum-risk exposure distribution for continuous risk factors is itself often a distribution of exposure levels, vs a constant baseline. Figure 25.1, for example, illustrates a scenario for systolic blood pressure (SBP) with typical exposure levels in an older population (mean: $150 \mathrm{mmHg} ; \mathrm{SD}: 9 \mathrm{mmHg}$ ) compared with the theoretical-minimum-risk exposure distribution (mean: 115 mmHg ; SD: 6 mmHg ). The non-zero standard deviation of the theoretical-minimum-risk distribution reflects the reality that there always is some inter-person variability within any given population, even after hypothetical reductions such as that shown in Figure 25.1.

The optimal exposure distribution for a population would overlap precisely with the theoretical-minimum-risk exposure distribution. By definition of theoretical-minimum risk, such a population would be collectively without any increased risk and therefore with zero attributable burden due to the risk factor of interest. Any population containing individuals outside this distribution will then have a population attributable fraction greater than zero and exposure distributions converging on the

Figure 25.1 Theoretical-minimum-risk exposure distribution for continuous risk factors using systolic blood pressure (SBP) as an example


Note: Each point represents a hypothetical individual or small group of individuals in the population. The solid straight lines represent the increasing relative risk, on a log scale, for ischaemic heart disease with increasing SBP.
theoretical minimum will have attributable burden tending towards zero. The risk for any individual (or groups of individuals in a narrow range of exposure) in the population would be determined by the difference between her/his current exposure (SBP level) and the SBP level that s/he would have when the population distribution overlaps with the theoret-ical-minimum-risk exposure distribution.

Estimating the total hazard at the population level can be achieved using a micro-simulation approach in which individuals are drawn randomly from current and theoretical-minimum-risk exposure distributions. For most risk factors, however, individual exposures "track" over relatively long periods of time (Lauer and Clarke 1988; Voors et al. 1979; Wilsgaard et al. 2001). In other words, those with higher/lower exposure levels of a particular risk factor are expected to have higher/lower exposure levels within the theoretical-minimum-risk exposure distribution (see the hypothetical individuals in Figure 25.1). Random (uncorrelated) draws of individuals from current and theoretical-minimum-risk exposure distributions would be inconsistent with the empirical evidence on tracking. Consistent with this evidence, we assumed that the ordering of individuals in the exposure distribution remains unchanged (i.e.
the rank-order correlation of individual exposures equals 1 ) in the transition to the theoretical-minimum-risk distribution in estimating the PAF.

With correlated rank-ordering of individuals in current and theoretical-minimum-risk exposure distributions, if hazards were a linear function of exposure, then for those risks with symmetric distributions, shifting the population to the theoretical minimum distribution would be computationally equivalent to shifting everyone to the mean exposure of the theoretical-minimum-risk exposure distribution (i.e. the standard deviation of the theoretical-minimum-risk exposure distribution would not change the total hazard). This is because, with a linear hazard function, the changes in hazards for individuals above and below the mean, as a result of changing the standard deviation of the theoretical-minimum-risk exposure distribution, would fully compensate each other.

Risk is, however, an exponential function of exposure in most epidemiological models. With an exponential hazard function, when the baseline is the mean of theoretical-minimum-risk exposure distribution, the integrated risk is larger for those above the mean than those below the mean (Figure 25.2), compared to the case of treating theoretical-minimum-risk exposure as a distribution with non-zero standard deviation. The net difference would depend on both the steepness of the risk curve (i.e. increased risk per unit increase in exposure) and the standard deviations of the current and theoretical-minimum-risk exposure distributions.

For computational reasons, we estimated PAFs for continuous risk factors relative to the mean of the theoretical-minimum-risk exposure distribution. In these calculations, the relative risk for any individual in the population with an exposure below the mean of the theoretical-minimum-risk exposure distribution was set to 1.0 (e.g. in Figures 25.1 and 25.2, the lower tail of current blood pressure distribution is inside the theoretical minimum distribution with some individuals already at a level below 115 mmHg . These individuals were assigned a relative risk of 1.0). Sensitivity analysis showed that in the scenarios analysed in the CRA project, global PAFs estimated by integrating risk relative to the mean of the theoretical-minimum-risk exposure distribution were up to $2 \%$ larger than those estimated by integrating risk relative to the full distribution. As described above, this is because of the non-linear shape of most hazard functions.

## 4. Aggregation of attributable burden across age, sex and subregion

Within each of the 14 subregions, all-age-sex population attributable fractions $\left(P A F_{\text {subregion }}\right)$ were calculated by aggregating attributable burden estimates across the 16 age-sex-specific estimates within the subregion

Figure 25.2 Effect of a non-linear hazard function and choice of baseline on total population risk


Note: With an exponential hazard function, when theoretical-minimum-risk exposure is a distribution with a non-zero standard deviation, those falling above the mean of the current distribution (e.g. 155 mmHg for SBP) contribute more to total population hazard than those below it, relative to the case when the baseline is a constant level ( 115 mmHg for SBP). In the figure, the solid lines represent the hazard when the theoretical-minimum-risk exposure is a distribution, and the dotted line when a constant baseline is considered. The difference between the two relative risks on the right $\left(R R_{2}\right)$ is larger than those on the left $\left(R R_{1}\right)$. As a result of this imbalanced contribution to hazard, using the mean of theoretical-minimum-risk exposure distribution as baseline in estimating total population hazard would result in slightly larger PAFs than using the complete theoretical minimum distribution.
and then dividing by the total subregional disease burden using the relationship in Equation 3 (the estimates could similarly be aggregated across ages separately for males and females).

$$
\begin{equation*}
P A F_{\text {subregion }}=\frac{\sum_{\text {age,sex }} A B_{\text {subregion,age,sex }}}{\sum_{\text {age,sex }} B_{\text {subregion,age,sex }}} \tag{3}
\end{equation*}
$$

Similarly, for each age-sex group, world attributable fractions ( $P A F_{\text {age,sex }}$ ) were calculated by aggregating attributable burden estimates across all the 14 subregion-specific estimates and then dividing by the disease burden for that age-sex group using the relationship in Equation 4.

$$
\begin{equation*}
P A F_{\text {age,sex }}=\frac{\sum_{\text {subregion }=1}^{14} A B_{\text {subregion,age,sex }}}{\sum_{\text {subregion }=1}^{14} B_{\text {subregion,age,sex }}} \tag{4}
\end{equation*}
$$

This is shown in Tables 25.1-25.3 for the case of SBP and ischaemic heart disease (IHD). The non-italic numbers in Table 25.1 are the sub-region-age-sex specific PAFs estimated using Equation 2. Next, these fractions were applied to the Global Burden of Disease (GBD) 2000 estimates of disease burden for IHD, shown in Table 25.2, producing the estimates of IHD disease burden attributable to SBP in Table 25.3 (similar estimates could be made for mortality or YLL). Dividing the total attributable burden in any subregion (e.g. 1.548 million DALYs for AMR-A in the highlighted cell) by the total IHD burden for the subregion in the GBD database ( 3.506 million DALYs for AMR-A in the highlighted cell) gives the all-age-sex subregional PAFs ( $44 \%$ for AMR-A in the highlighted cell, obtained by dividing 1.548 by 3.506 ).

The all-age-sex PAF estimates for the remaining 13 subregions are also shown in Table 25.1 in italics. Similarly, world PAFs were calculated within each age and sex group, as well as overall, using Equation 4, and are shown in the bottom row of Table 25.1. For example, the world PAF for $60-69$ years old males was obtained by dividing the total attributable burden in that age-sex group ( 4.71 million DALYs) by the total world IHD burden in the GBD database ( 9.015 million DALYs), giving a PAF of $52 \%$.

Computationally, aggregate PAFs (whether aggregated across age-sex groups or subregions) are equivalent to weighted averages of the subre-gion-age-sex specific estimates, with weights being the same as the total number of events (i.e. deaths, YLL, or DALYs) for each subregion-agesex group. In the above example, total subregion-age-sex specific DALYs are the weighting factor. As a result, subregion and world PAFs are weighted more towards the ages and/or subregions that have higher DALYs (rather than those with larger populations). For each risk factor, these separate aggregate PAFs were estimated for deaths, YLL, and DALYs with the corresponding GBD estimates used as the denominator (or weighting factor). As a result, even when the age-sex-subregionspecific PAFs are the same for the three measures, the aggregate ones may differ.

For example, although the subregion-age-sex-specific PAFs are the same for deaths and DALYs in the case of elevated SBP and IHD, dividing the IHD deaths attributable to this risk factor in AMR-A (203000) by the total IHD deaths in this region (618000), gives a regional PAF for mortality of $33 \%$ in AMR-A. In fact, the smaller subregional PAF for mortality compared to that of DALYs highlights the higher weighting towards the older age PAFs (which are smaller) in the case of
Table 25.I PAFs for IHD attributable to increased SBP (\%), by age, sex and subregion

Table 25.2 GBD 2000 estimates of total disease burden for IHD (000s of DALYs), by age, sex and subregion

|  | 0-4 (years) |  | 5-14 (years) |  | 15-29 (years) |  | 30-44 (years) |  | 45-59 (years) |  | 60-69 (years) |  | 70-79 (years) |  | $\geq 80$ (years) |  | Total |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |  |
| Subregion |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| AFR-D | 3 | 4 | 6 | 5 | 31 | 48 | 116 | 88 | 242 | 281 | 190 | 189 | 122 | 166 | 37 | 48 | 1576 |
| AFR-E | 4 | 5 | 7 | 7 | 49 | 66 | 149 | 88 | 273 | 278 | 200 | 188 | 110 | 148 | 29 | 50 | 1653 |
| AMR-A | 1 | 1 | 1 | 0 | 15 | 7 | 214 | 74 | 709 | 260 | 514 | 289 | 470 | 391 | 219 | 340 | 3506 |
| AMR-B | 2 | 2 | 3 | 2 | 46 | 21 | 218 | 107 | 594 | 301 | 412 | 279 | 255 | 223 | 69 | 96 | 2631 |
| AMR-D | 1 | 2 | 1 | I | 10 | 7 | 25 | 11 | 54 | 33 | 42 | 30 | 31 | 24 | 9 | 11 | 294 |
| EMR-B | 3 | 5 | 6 | 4 | 32 | 17 | 179 | 59 | 411 | 143 | 219 | 128 | 114 | 99 | 26 | 29 | 1474 |
| EMR-D | 24 | 24 | 21 | 12 | 90 | 84 | 322 | 170 | 771 | 475 | 522 | 469 | 292 | 328 | 69 | 73 | 3746 |
| EUR-A | I | 0 | 1 | 7 | 15 | 15 | 181 | 40 | 657 | 146 | 717 | 284 | 671 | 518 | 239 | 389 | 3882 |
| EUR-B | 1 | 0 | 1 | 1 | 35 | 15 | 266 | 78 | 734 | 258 | 647 | 440 | 431 | 471 | 93 | 176 | 3647 |
| EUR-C | 0 | 0 | I | 1 | 78 | 15 | 693 | 127 | 1708 | 477 | 1496 | 920 | 847 | 1223 | 171 | 563 | 8319 |
| SEAR-B | 7 | 3 | 4 | 3 | 98 | 49 | 222 | 123 | 394 | 279 | 318 | 286 | 188 | 188 | 43 | 54 | 2259 |
| SEAR-D | 70 | 52 | 99 | 58 | 283 | 622 | 1105 | 900 | 3688 | 2119 | 2657 | 2361 | 1412 | 1484 | 278 | 291 | 17480 |
| WPR-A | 0 | 0 | 3 | 3 | 10 | 6 | 40 | 10 | 151 | 40 | 135 | 59 | 112 | 85 | 46 | 66 | 765 |
| WPR-B | 8 | 7 | 14 | 7 | 155 | 77 | 495 | 278 | 1099 | 665 | 945 | 783 | 682 | 776 | 182 | 340 | 6513 |
| World | 125 | 107 | 168 | 111 | 946 | 1049 | 4225 | 2154 | 11484 | 5755 | 9015 | 6704 | 5737 | 6127 | 1510 | 2526 | 57743 |

Table 25.3 Burden of IHD attributable to increased SBP (000s of DALYs), by age, sex and subregion

mortality, because greater numbers of IHD deaths occur in these age groups. On the other hand, because deaths at younger ages contribute to larger loss of life (YLL), when DALYs are considered, the contribution of PAFs at younger ages to the all-age-sex PAF becomes greater.

## 5. Exceptions to the general ESTIMATION PROCEDURE

The following list briefly describes the major departures from the standard analysis framework which were required so that all risk factors could be adequately assessed within the project. Further details are provided in the individual risk factor chapters.

- Theoretical minima varied by age, sex and subregion for iron deficiency, since this was the haemoglobin distribution that would be observed if iron deficiency were eliminated from each population.
- Theoretical minima varied by age, sex and subregion for lack of contraception as a risk factor to reflect different fertility preferences across populations.
- Fruit and vegetable intakes in any population were truncated at zero. In other words, all individuals falling below zero in the distribution were allocated a value of zero. Sensitivity analyses were also performed to assess the effects of possible skewness in the distribution of fruit and vegetable intake.
- The burden of cardiovascular diseases attributable to lead exposure was estimated by assessing different scenarios of elevated blood pressure due to lead and then estimating the total mediated effect through blood pressure.


## 6. Other methodological issues

It has been shown that the attributable fraction estimates using the PIF relationship in Equation 1 lead to biased estimates when the relative risk has been adjusted for confounding (Greenland 1984; Greenland and Robins 1988). This bias is in fact a result of the correlation among multiple risks (the risk factor of interest and other risk factors that act as confounders), as well as the diseases affected by them (Ezzati et al. 2003). Accounting for this correlation in the estimation of attributable burden, however, would require the availability of exposure and disease data stratified by the confounding variable(s). In general, such stratified data are not available and therefore reliance on the formula with direct use of the adjusted relative risk estimates was necessary. In the case of positive correlation among risk factors, this would generally result in an underestimation of population attributable fraction.

## Note

1 See preface for an explanation of this term.

## References

Drescher K, Becher B (1997) Estimating the generalized impact fraction from case-control data. Biometrics, 53:1170-1176.
Eide GE, Heuch I (2001) Attributable fractions: fundamental concepts and their visualization. Statistical Methods in Medical Research, 10:159-193.
Ezzati M, Vander Hoorn S, Rodgers A, Lopez AD, Mathers CD, Murray CJ, Comparative Risk Assessment Collaborating Group (2003) Estimates of global and regional potential health gains from reducing multiple major risk factors. The Lancet, 362:271-280.

Greenland S (1984) Bias in methods for deriving standardized morbidity ratio and attributable fraction estimates. Statistics in Medicine, 3:131-141.
Greenland S, Robins JM (1988) Conceptual problems in the definition and interpretation of attributable fractions. American Journal of Epidemiology, 128: 1185-1197.
Lauer R, Clarke W (1988) A longitudinal view of blood pressure during childhood: the Muscatine Study. Statistics in Medicine, 7:47-57.
Murray CJL, Lopez AD, eds. (1996) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Global Burden of Disease and Injury, Vol 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.
Voors A, Webber L, Berenson G (1979) Time course studies of blood pressure in children-the Bogalusa Heart Study. American Journal of Epidemiology, 109:320-334.

Walter SD (1980) Prevention of multifactorial disease. American Journal of Epidemiology, 112:409-416.
Wilsgaard T, Jacobsen BK, Schirmer H et al. (2001) Tracking of cardiovascular risk factors: the Tromso Study, 1979-1995. American Journal of Epidemiology, 154:418-426.

## Chapter 26

# Mortality and burden of disease ATTRIBUTABLE TO INDIVIDUAL 

 RISK FACTORSMajid Ezzati, Anthony Rodgers, Alan D. Lopez, Stephen Vander Hoorn and Christopher J.L. Murray

Population attributable fractions (PAFs) for mortality and burden of disease attributable to individual risk factors were calculated, as described in chapter 25 , using risk factor exposure and hazard estimates provided in risk factor chapters. Mortality and burden of disease attributable to individual risk factors were then calculated by multiplying the PAFs by the estimates of total mortality and burden of disease from the Global Burden of Disease (GBD) databases in each of the 224 subregion-age-sex groups, as described in chapter 25 . These results are presented in the Annex Tables for each risk factor and summarized here across risks.

## 1. Aggregate disease burden attributable to INDIVIDUAL RISK FACTORS

All-cause mortality and burden of disease estimates for females and males attributable to CRA risk factors in the 14 subregions ${ }^{1}$ are presented in Table 26.1. Figure 26.1 shows the contribution of the 20 leading global risk factors to mortality and burden of disease in the world and three broad combinations of subregions-demographically and economically developed (AMR-A, EUR and WPR-A), low-mortality developing (AMR-B, EMR-B, SEAR-B and WPR-B) and high-mortality developing (AFR, AMR-D, EMR-D and SEAR-D). Figure 26.2 presents the burden of disease due to the leading 10 risk factors for each subregional grouping, also showing the cause composition, divided into broad groups of diseases and injuries. The different ordering of risk factors in their contributions to mortality and disease burden reflects the age profile of mortality (e.g. under-five mortality for underweight has larger

Table 26.1(a) Mortality for females and males due to selected risk factors in 14 subregions

|  | AFRICA <br> Mortality stratum |  | THE AMERICAS <br> Mortality stratum |  |  | EASTERN MEDITERRANEAN <br> Mortality stratum |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | High child, high adult | High child, very high adult | Very low child, very low adult | Low child, low adult | High child, high adult | Low child, low adult | High child, high adult |
|  | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female |
| Total population (000s) | 147133/146945 | $171600 / 173915$ | 160494/164689 | $213309 / 217623$ | $35471 / 35759$ | $72156 / 66903$ | 174275/168301 |
| Total mortality (000s) | 2206/2050 | $3154 / 3001$ | 1 342/1 392 | $1459 / 1120$ | 290/237 | 409/287 | $1750 / 1602$ |
| Childhood and maternal undernutrition |  |  |  |  |  |  |  |
| Childhood and maternal underweight | $438 / 402$ | 487/441 | 0/0 | 14/11 | 14/11 | 8/8 | 223/229 |
| Iron deficiency anaemia | $59 / 67$ | 65/80 | 2/3 | 13/13 | 3/4 | 3/4 | 36/44 |
| Vitamin A deficiency | 90/112 | 120/15 | 0/0 | 2/3 | 2/2 | 0/0 | 34/53 |
| Zinc deficiency | 74/68 | 128/116 | 0/0 | $3 / 2$ | 5/4 | 2/2 | 44/45 |
| Other nutrition-related risk factors and physical inactivity |  |  |  |  |  |  |  |
| High blood pressure | 87/128 | 79/116 | 179/191 | 170/162 | 20/20 | 76/57 | 164/171 |
| High cholesterol | 34/52 | 36/53 | 161/189 | 88/79 | 10/9 | 51/31 | 114/101 |
| Overweight and obesity (high BMI) | 14/19 | 21/35 | 135/137 | 117/144 | 15/18 | 36/28 | 58/67 |
| Low fruit and vegetable consumption | $21 / 31$ | 33/41 | 92/79 | 81/58 | 7/7 | 27/15 | 51/48 |
| Physical inactivity | 20/25 | 21/27 | 74/81 | 52/55 | 6/6 | 21/13 | 47/43 |
| Addictive substances |  |  |  |  |  |  |  |
| Smoking and oral tobacco use | 43/7 | 84/26 | 352/294 | 163/58 | 5/1 | 43/10 | 114/19 |
| Alcohol use | 53/15 | 125/30 | 27/-22 | 207/39 | 22/6 | 6/1 | 8/1 |
| Illicit drug use | 5/1 | 1/0 | 10/7 | 7/4 | 1/0 | 5/I | 18/4 |
| Sexual and reproductive health |  |  |  |  |  |  |  |
| Unsafe sex | 198/234 | 805/923 | 8/8 | 22/27 | 17/11 | 0/4 | 33/39 |
| Non-use and use of ineffective methods of contraception | NA/16 | NA/33 | NA/0 | NA/5 | NA/4 | NA/I | NA/23 |
| Environmental risk factors |  |  |  |  |  |  |  |
| Unsafe water, sanitation and hygiene | 129/103 | 207/169 | 0/1 | 16/15 | 13/10 | 9/9 | 117/135 |
| Urban air pollution | 11/11 | 5/5 | 14/14 | 16/14 | 3/2 | 5/3 | 28/23 |
| Indoor air pollution from household use of solid fuels | 93/80 | 118/101 | 0/0 | 7/9 | 5/5 | 1/1 | 56/60 |
| Lead exposure | 5/4 | 4/3 | 2/1 | 14/7 | 2/1 | 5/2 | $12 / 6$ |
| Global climate change | 9/9 | 18/18 | 0/0 | 0/0 | 0/0 | 0/0 | 10/11 |
| Occupational risk factors |  |  |  |  |  |  |  |
| Risk factors for injury | 14/1 | 18/1 | 3/0 | 17/1 | 2/0 | 8/0 | 27/2 |
| Carcinogens | I/0 | 1/0 | 7/2 | 4/1 | 0/0 | 1/0 | 1/0 |
| Airborne particulates | 5/2 | 7/3 | $12 / 2$ | $9 / 1$ | 1/0 | 1/0 | $9 / 2$ |
| Ergonomic stressors | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 |
| Noise | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 |
| Other selected risks factors |  |  |  |  |  |  |  |
| Contaminated injections in health care settings | 10/7 | 27/23 | 0/0 | 1/0 | I/I | 0/0 | 24/20 |
| Child sexual abuse | 0/0 | 2/1 | 1/1 | 1/0 | 0/0 | 0/0 | 1/1 |


| Very low child, very low adult | EUROPE <br> Mortality stratum Low child, low adult | Low child, high adult | SOUTH-EAST ASIA <br> Mortality stratum |  | WESTERN PACIFIC <br> Mortality stratum |  | WORLD |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Low child, low adult | High child, high adult | Very low child, very low adult | Low child, low adult |  |  |
| Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Total |
| 201514/210376 | 108182/110277 | 114051/129133 | $147173 / 146646$ | 639087/602719 | 75796/78558 | 785 055/747878 | 3045 295/2999722 | 6045017 |
| 2020/2054 | 1034/916 | $1878 / 1721$ | I 234/I 022 | 6358/5 764 | 616/519 | 5483/4944 | 29232/26629 | 55861 |
| 0/0 | 9/8 | 0/0 | 40/29 | 573/614 | 0/0 | 95/94 | $1900 / 1848$ | 3748 |
| 2/3 | 3/3 | 2/2 | 15/19 | 139/185 | 0/0 | 34/39 | 375/466 | 841 |
| 0/0 | 0/0 | 0/0 | 10/13 | 68/101 | 0/0 | 7/9 | 333/445 | 778 |
| 0/0 | 2/2 | 0/0 | 5/4 | 132/14\| | 0/0 | 6/6 | 400/389 | 789 |
| 325/354 | 281/289 | 514/671 | 133/139 | 668/519 | 85/76 | 711/758 | 3491/3649 | 7141 |
| 265/282 | 144/136 | 387/518 | 72/40 | 488/507 | 39/39 | 222/265 | $2112 / 2303$ | 4415 |
| 183/197 | 117/141 | 202/265 | 44/58 | 42/110 | 21/20 | 163/184 | $1168 / 1423$ | 2591 |
| 95/75 | 80/67 | 234/247 | 55/48 | 378/311 | 26/19 | 269/232 | 1 449/1 277 | 2726 |
| 103/103 | 64/62 | 147/175 | 34/34 | 218/185 | 23/19 | 132/134 | 961/961 | 1922 |
| 531/145 | 255/53 | 548/73 | 181/12 | 785/132 | 128/49 | 661/137 | 3893/1014 | 4907 |
| 65/-85 | 100/25 | 338/88 | 51/9 | 148/21 | 23/-28 | 465/66 | $1638 / 166$ | 1804 |
| $11 / 6$ | 3/1 | 18/5 | 13/1 | 40/8 | 2/1 | 28/2 | 163/41 | 204 |
| 3/9 | 1/8 | 3/13 | 30/25 | 231/177 | 0/3 | 18/36 | $1370 / 1516$ | 2886 |
| NA/O | NA/O | NA/O | NA/7 | NA/56 | NA/O | NA/3 | NA/I49 | 149 |
| 0/1 | 8/7 | 1/1 | 25/21 | 326/327 | 0/0 | 42/35 | 895/835 | 1730 |
| 12/11 | 20/18 | 22/24 | 17/15 | 72/60 | 10/8 | 176/179 | 411/388 | 799 |
| 0/0 | 8/9 | 1/3 | 15/22 | 218/304 | 0/0 | 137/366 | 658/961 | 1619 |
| 4/2 | 15/8 | 26/13 | 6/3 | 38/19 | 0/0 | 21/10 | 155/79 | 234 |
| $0 / 0$ | 0/0 | $0 / 0$ | 1/0 | 35/38 | 0/0 | 2/1 | 76/78 | 154 |
| 4/0 | 5/0 | 15/1 | 19/1 | 79/5 | 2/0 | 78/5 | 291/19 | 310 |
| $12 / 2$ | 6/1 | $13 / 2$ | 3/0 | 1 1/1 | 4/1 | 28/8 | 92/17 | 109 |
| $17 / 2$ | 7/2 | $15 / 3$ | $10 / 3$ | 54/17 | 4/1 | 113/54 | 264/92 | 356 |
| 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0 |
| 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0 |
| 0/0 | 1/0 | 6/4 | 19/9 | 92/62 | 0/0 | 137/58 | 317/184 | 501 |
| I/I | 1/1 | 3/2 | 1/0 | 16/18 | 1/1 | 10/14 | 38/41 | 79 |

Table 26.l(b) Burden of disease for females and males due to selected risk factors in 14 subregions

|  | AFRICA <br> Mortality stratum |  | THE AMERICAS <br> Mortality stratum |  |  | EASTERN MEDITERRANEAN <br> Mortality stratum |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | High child, high adult | High child, very high adult | Very low child, very low adult | Low child, low adult | High child, high adult | Low child, low adult | High child, high adult |
|  | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female |
| Total population (000s) | $147133 / 146945$ | $171600 / 173915$ | 160494/164689 | $213309 / 217623$ | $35471 / 35759$ | $72156 / 66903$ | 174275/168301 |
| Total DALYs (000s) | $73650 / 70695$ | $103191 / 101977$ | 24480/21 804 | 45 372/35 065 | $9158 / 7895$ | 12590/10 131 | $55790 / 54140$ |
| Childhood and maternal undernutrition |  |  |  |  |  |  |  |
| Childhood and maternal underweight | I 5530/14375 | $17189 / 15710$ | 12/11 | 570/498 | 512/410 | 324/312 | 8203/8407 |
| Iron deficiency anaemia | 2263/2521 | 2451/2905 | 223/255 | 446/465 | 121/217 | 239/277 | $1449 / 1746$ |
| Vitamin A deficiency | $3178 / 3856$ | 4208/5 167 | 0/0 | 79/103 | 53/68 | 9/8 | I 159/I 758 |
| Zinc deficiency | 2625/2414 | $4563 / 4150$ | 1/1 | $115 / 99$ | 174/138 | 66/63 | I 547/I 574 |
| Other nutrition-related risk factors and physical inactivity |  |  |  |  |  |  |  |
| High blood pressure | 980/1 295 | 984/I 177 | $1642 / 1141$ | $1807 / 1438$ | 208/178 | 840/570 | $1781 / 1698$ |
| High cholesterol | 395/563 | 456/578 | $1451 / 1012$ | $1070 / 803$ | 109/87 | 605/320 | \| 273/I 051 |
| Overweight and obesity (high BMI) | 246/318 | 341/546 | I 825/I 654 | $1505 / 1918$ | 189/234 | 534/456 | 882/I 027 |
| Low fruit and vegetable consumption | 253/354 | 434/471 | 833/536 | 896/581 | 72/67 | 322/172 | 607/550 |
| Physical inactivity | 225/280 | 262/309 | 691/576 | 582/585 | 61/68 | 265/164 | 559/492 |
| Addictive substances |  |  |  |  |  |  |  |
| Smoking and oral tobacco use | 591/97 | \| 3| 1/367 | $3567 / 2606$ | $2190 / 813$ | 51/14 | 593/197 | $1780 / 379$ |
| Alcohol use | 1441/393 | $3621 / 785$ | 2925/702 | 7854/1 443 | 789/170 | 162/22 | 328/36 |
| Illicit drug use | 543/156 | 495/163 | 808/379 | 791/310 | 200/71 | 449/78 | 620/153 |
| Sexual and reproductive health |  |  |  |  |  |  |  |
| Unsafe sex | 6205/7753 | 24059/29 664 | 281/235 | 843/912 | $521 / 310$ | 30/162 | $1125 / 1508$ |
| Non-use and use of ineffective methods of contraception | NA/997 | NA/I 732 | NA/2 | NA/375 | NA/203 | NA/II9 | NA/I 210 |
| Environmental risk factors |  |  |  |  |  |  |  |
| Unsafe water, sanitation and hygiene | 3797/3119 | 6365/5 355 | 31/30 | 686/603 | 436/320 | $314 / 315$ | $3797 / 4506$ |
| Urban air pollution | 153/132 | 80/67 | 87/65 | 133/99 | 24/20 | 47/30 | 305/253 |
| Indoor air pollution from household use of solid fuels | 3036/2358 | 3865/3059 | 2/4 | 193/25 1 | 175/154 | 32/32 | $1817 / 1691$ |
| Lead exposure | 512/488 | 460/433 | 68/49 | 907/789 | 140/125 | 238/187 | 606/504 |
| Global climate change | $321 / 305$ | 631/636 | 1/2 | 35/36 | 13/10 | 10/10 | 357/391 |
| Occupational risk factors |  |  |  |  |  |  |  |
| Risk factors for injury | 486/39 | 583/46 | 82/6 | 606/51 | 80/6 | 253/18 | 961/68 |
| Carcinogens | 9/2 | $13 / 4$ | 56/16 | 38/8 | 3/1 | 12/1 | 18/2 |
| Airborne particulates | 106/37 | 141/69 | 184/36 | 213/44 | 21/4 | 37/4 | 148/39 |
| Ergonomic stressors | 21/16 | 25/20 | 17/10 | 32/15 | 4/2 | 9/3 | 25/16 |
| Noise | 109/49 | $127 / 60$ | 92/31 | 122/43 | $15 / 6$ | 60/21 | 142/88 |
| Other selected risks factors |  |  |  |  |  |  |  |
| Contaminated injections in health care settings | 244/187 | 804/742 | 0/0 | 13/5 | 20/12 | 0/0 | 437/390 |
| Child sexual abuse | 49/102 | 167/238 | 98/320 | 147/118 | 46/27 | 41/83 | 85/225 |


| Very low child, very low adult | EUROPE <br> Mortality stratum Low child, low adult | Low child, high adult | SOUTH-EAST ASIA <br> Mortality stratum |  | WESTERN PACIFIC <br> Mortality stratum |  | WORLD |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Low child, low adult | High child, high adult | Very low child, very low adult | Low child, low adult |  |  |
| Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Male/Female | Total |
| 201514/210376 | 108182/110277 | 114051/129133 | $147173 / 146646$ | 639087/602719 | 75796/78558 | 785 055/747878 | 3045 295/2999722 | 6045017 |
| 28006/25314 | $21304 / 17689$ | 35099/24 144 | 33 585/29 302 | 178923/177345 | 8780/7 591 | $131634 / 110818$ | $761562 / 693911$ | 1455473 |
| 10/9 | 367/324 | 32/29 | 1634/1239 | 21 297/22766 | 6/6 | 4048/3972 | 69733/68067 | 137801 |
| 87/211 | 166/271 | 110/161 | 681/847 | 5614/6883 | $31 / 81$ | $1876 / 2462$ | 15756/19301 | 35057 |
| 0/0 | 1/1 | 0/0 | 347/406 | 2321/3 368 | 0/0 | 241/306 | 11596/15042 | 26638 |
| 0/0 | 65/56 | 5/4 | 197/152 | 4635/4961 | 0/0 | 208/219 | 14201/13833 | 28034 |
| 2624/1828 | 2699/2180 | $5386 / 4632$ | $1394 / 1402$ | $7010 / 5316$ | 781/451 | 6783/6044 | 34 920/29350 | 64270 |
| 2062/1317 | $1461 / 996$ | $4109 / 3211$ | 828/412 | $5562 / 5528$ | 380/227 | 2376/2 195 | 22 136/18301 | 40437 |
| 1922/I 735 | 1 420/l 445 | 2578/2684 | 650/818 | 686/I 939 | 334/295 | 2430/2804 | 15543/17872 | 33415 |
| 785/413 | 777/511 | 2431/I 684 | 614/524 | $4139 / 3521$ | 237/118 | 2718/2042 | $15117 / 11544$ | 26662 |
| 852/654 | 636/494 | \| 46|/| 236 | 414/409 | 2489/2 186 | 228/160 | 1436/1318 | 10159/8933 | 19092 |
| 4991// 464 | $3381 / 715$ | 7230/832 | 2712/180 | 10474/I 621 | 994/325 | 8313/1296 | 48 177/10904 | 59081 |
| $3103 / 416$ | $2183 / 446$ | 7543/1570 | $1793 / 284$ | $4927 / 675$ | 708/43 | 12020/1941 | $49397 / 8926$ | 58323 |
| 786/344 | 181/81 | 762/223 | 406/121 | $1386 / 282$ | 231/101 | $1110 / 259$ | 8769/2719 | 11488 |
| 114/202 | 50/240 | 134/295 | $1009 / 925$ | 7413/6004 | 12/65 | 804/995 | 42600/49 269 | 91869 |
| NA/3 | NA/83 | NA/47 | NA/397 | NA/3 354 | NA/I | NA/290 | NA/8814 | 8814 |
| 33/33 | 287/262 | 64/57 | 734/506 | 8762/9725 | 14/13 | $2112 / 1879$ | 27432/26726 | 54158 |
| 73/44 | 170/118 | 191/129 | 154/128 | 718/594 | 53/31 | $1343 / 1161$ | $3533 / 2871$ | 6404 |
| 0/0 | 233/244 | 18/49 | 458/532 | 6641/7596 | 0/0 | 2569/3 528 | 19040/19499 | 38539 |
| 75/43 | 304/189 | 424/211 | 379/337 | $1489 / 1198$ | 15/10 | $1496 / 1251$ | $7112 / 5814$ | 12926 |
| 1/2 | 5/5 | $2 / 2$ | 19/15 | 1213/1325 | $0 / 1$ | 92/77 | 2700/2816 | 5517 |
| 130/12 | 203/15 | 410/31 | 577/39 | 2857/184 | 56/5 | 2495/199 | 9779/718 | 10496 |
| 95/13 | 63/7 | 129/16 | 35/5 | 119/12 | 24/4 | 227/87 | 891/179 | 1070 |
| 216/43 | 105/32 | 167/43 | 135/47 | 862/315 | 68/18 | $1726 / 493$ | $4130 / 1224$ | 5354 |
| 21/11 | 18/12 | 21/14 | 26/19 | 111/78 | $9 / 5$ | 146/110 | 485/333 | 818 |
| $117 / 47$ | 92/50 | 136/92 | 219/185 | 799/303 | 26/22 | 735/365 | 2788/1 362 | 4151 |
| 0/0 | 8/5 | 106/59 | 356/156 | 2341/1759 | 0/0 | 2028/791 | 6356/4 105 | 10461 |
| 61/175 | 72/158 | 132/205 | 42/56 | $1079 / 2340$ | 29/96 | 888/1 158 | 2934/5 302 | 8235 |

[^111]Figure 26.1 (a) Mortality and (b) burden of disease due to leading global risk factors

응

Note: High-mortality developing: AFR, AMR-D, EMR-D and SEAR-D subregions; low-mortality developing: AMR-B, EMR-B, SEAR-B and WPR-B; developed: AMR-A, EUR and WPR-A. The figure shows the estimated mortality and disease burden for each risk factor considered individually. These risks act in part through other risks and act jointly with other risks. Consequently, the burden due to groups of risk factors will usually be less than the sum of individual risks (see chapter 27).
ぜ
Figure 26.2 Burden of disease due to leading regional risk factors divided by disease type in (a) highmortality developing, (b) low-mortality developing and (c) developed subregions

$\square$ Infectious and parasitic
$\square$ Maternal and perinatal
$\square$ Nutritional deficiency
$\square$ Vascular
$\square$ Cancer
$\square$ Chronic respiratory
$\square$ Neuro-psychiatric
س Other noncommunicable
$\square$ Intentional injury
$\square$ Unintentional injury

contribution to disease burden) and the non-fatal effects (e.g. neuropsychological outcomes of alcohol).

Despite disaggregation into underweight and micronutrient deficiency (which are not additive; see chapter 27) and methodological changes, undernutrition has remained the single leading global cause of health loss with comparable contributions in 1990 ( 220 million DALYs, 16\%, for malnutrition) (Murray and Lopez 1997) and 2000 ( 140 million DALYs, $9.5 \%$, for underweight; $2.4 \%, 1.8 \%, 1.9 \%$ for iron, vitamin A and zinc deficiency respectively; $0.1 \%$ for iodine deficiency disorders). This is because while prevalence of underweight has decreased in most regions of the world in the past decade, it has increased in sub-Saharan Africa (de Onis et al. 2000) where its effects are disproportionately large due to simultaneous exposure to other childhood disease risk factors. A substantial part of the decrease in the burden of disease due to poor water, sanitation and hygiene (from $6.8 \%$ in 1990 to $3.7 \%$ in 2000) is due to a decline in global diarrhoeal disease mortality (from 2.9 million deaths in 1990 to 2.1 million in 2000), and partly a result of improved case management interventions, particularly oral rehydration therapy.

Leading causes of burden of disease in all high-mortality developing subregions were childhood and maternal undernutrition-including underweight ( $14.9 \%$ ) and micronutrient deficiencies ( $3.1 \%$ for iron deficiency, $3.0 \%$ for vitamin A deficiency and $3.2 \%$ for zinc deficiency)unsafe sex $(10.2 \%)$, poor water, sanitation and hygiene (5.5\%) and indoor smoke from solid fuels ( $3.6 \%$ ). The relative contribution of unsafe sex was disproportionately larger (26\%) in AFR-E, where HIV/AIDS prevalence is the highest, making it the leading cause of burden of disease in this subregion. The outcomes of these risk factors were mostly communicable, maternal, perinatal and nutritional conditions (Figure 26.2) which dominate the disease burden in high-mortality developing subregions. Despite the very large contribution of these diseases and their underlying risk factors, tobacco, blood pressure and cholesterol already resulted in significant loss of healthy life years in these subregions. For example, in SEAR-D (dominated by India in terms of population) the burden of disease attributable to tobacco, blood pressure and cholesterol was already of comparable magnitude to micronutrient deficiencies and is only marginally smaller than indoor smoke from solid fuels and poor water, sanitation and hygiene. In addition to their relative magnitude, the absolute size of the loss of healthy life years attributed to risk factors in high-mortality developing subregions was substantial. Childhood and maternal underweight and unsafe sex in these subregions alone (with $38 \%$ of global population) contributed as much ( $>200$ million DALYs) to loss of healthy life as all diseases and injuries in developed countries (with $22 \%$ of global population).

Across developed subregions, tobacco ( $12.2 \%$ ), high blood pressure ( $10.9 \%$ ), alcohol ( $9.2 \%$ ), high cholesterol ( $7.6 \%$ ) and high BMI ( $7.4 \%$ ) were consistently the leading causes of loss of healthy life, contributing
mainly to noncommunicable diseases and injuries. Tobacco was the leading cause of disease burden in all developed subregions, except EURC (dominated by Russia) where high blood pressure and alcohol resulted in slightly larger loss of healthy life. The increase in the disease burden due to blood pressure compared to 1990 (Murray and Lopez 1997) (from $3.9 \%$ in the established market economies and $5.9 \%$ in the formerly socialist economies) mainly reflects new evidence on hazard size after correction for regression dilution bias (MacMahon et al. 1990). The contributions of these risk factors are consistently larger than those of leading diseases of the developed subregions (i.e. ischaemic heart disease [ $9.4 \%$ ], unipolar depressive disorders [7.2\%], cerebrovascular disease [ $6.0 \%$ ], etc.), which emphasizes the potential health gains from reducing risk factors.

The low-mortality developing subregions present possibly the most striking mixture of leading risk factors. The leading risk factors in these subregions ( $40 \%$ of global population) include those from both developed and high-mortality developing subregions with comparable magnitudes (e.g. underweight [3.1\%] and high BMI [2.7\%] had comparable contributions to the burden of disease. See also Monteiro et al. 2002). In addition, the decline in the share of burden of disease due to the risk factors in low-mortality developing subregions was less marked than that in high-mortality developing and developed subregions (e.g. the ratio of 1st to 10 th leading risk factors was smaller). This lower clustering of risk factor burden further emphasizes the role of a more extended and mixed group of risk factors in low-mortality developing subregions. Alcohol was the leading cause of burden of disease in low-mortality developing subregions as a whole ( $6.2 \%$ ) and in AMR-B and WPR-B, but made a relatively low contribution to the burden of disease in EMRB. In general, AMR-B and EMR-B had risk factor profiles similar to the developed subregions (tobacco, blood pressure, cholesterol, BMI and alcohol), while SEAR-B and WPR-B had a more mixed risk factor profile (with the leading five risks being underweight, blood pressure, tobacco, unsafe sex and alcohol in SEAR-B; alcohol, blood pressure, tobacco, underweight and indoor smoke from solid fuels in WPR-B).

An important finding of this analysis is the key role of nutrition in health worldwide. Approximately $13 \%$ of the global disease burden can be attributed to the joint effects of childhood and maternal underweight or micronutrient deficiencies. In addition, almost as much as $7 \%(16 \%$ for those aged 30 years and above) can be attributed to risk factors that have substantial dietary determinants-high blood pressure, high cholesterol, high BMI and low fruit and vegetable intake. These patterns are not uniform within subregions, however, and in some countries the transition has been healthier than in others (Lee et al. 2000; Popkin et al. 2001). Further, the major nutritional risk factors show interregional heterogeneity (e.g. the relative contributions of blood pressure, cholesterol and BMI were different in AMR-A, SEAR-D and WPR-B). This het-
erogeneity further illustrates the importance of concurrent and comparable quantification of distal and proximal risk factors to provide a more complete picture of the role of various distal and proximal risk factors in reducing disease.

This analysis also provides the first quantitative evidence of the public health consequences of a number of risk factors including indoor smoke from solid fuels ( $2.6 \%$ of global disease burden), high BMI ( $2.3 \%$ ) and zinc deficiency $(1.9 \%)$. On the other hand, the burden of disease due to some risks (e.g. physical inactivity) was lower than expected if the methodology and results from the limited number of industrialized countries had been extrapolated (Powell and Blair 1994). This is partially because of difficulties in measuring exposure to this risk factor. A categorical exposure variable with a conservative baseline of "sufficient" (vs vigorous) activity was used. In part, it also reflects the inclusion of occupational and transportation domains of activity (that are common among rural populations of developing countries) in this analysis, above and beyond leisure-time activity which is more relevant to developed countries and urban populations (Jacobs et al. 1993; Levine et al. 2001).

## 2. Distributions of risk factor-attributable disease burden

An important feature of risk assessment, with implications for broad prevention policies and specific interventions and programmes, is the distribution of disease burden among population subgroups. These subgroups may be defined by factors such as age, sex, socioeconomic status or the current level of exposure to a risk factor, if exposures are defined in multiple categories or continuously. For example, reducing the large disease burden due to road traffic accidents among young adult males, largely associated with binge alcohol consumption, would require designing interventions that focus on this population subgroup and their specific drinking behaviours. On the other hand, the majority of effects from risk factors such as blood pressure have been found to occur among those at moderately elevated levels, suggesting the need for interventions beyond those intended for clinical hypertension (Cook et al. 1995; Murray et al. 2003; Rodgers et al. 2000). While the distribution of health effects by age and by exposure level has been studied in specific cohorts and for specific risk factors (Peto et al. 1992; Rodgers and MacMahon 1999; Rose 1992), there are no such estimates at the global level and for multiple risks.

The distributions of mortality and disease burden attributable to the risk factors included in this book by age and sex is shown in Table 26.2. The estimated disease burden from childhood and maternal undernutrition, unsafe water, sanitation and hygiene, and global climate change (much of whose estimated effects are mediated through nutritional and water variables) was almost exclusively among children aged <5 years.
The distribution of risk factor-attributable mortality and burden of disease by age and sex

|  | Mortality (\%) |  |  |  |  |  | Disease burden (\%) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0-4 | 5-14 | 15-59 | $\geq 60$ | Males | Females | 0-4 | 5-14 | 15-59 | $\geq 60$ | Males | Females |
| Childhood and maternal undernutrition |  |  |  |  |  |  |  |  |  |  |  |  |
| Childhood and maternal underweight | 100 | 0 | 0 | 0 | 51 | 49 | 100 | 0 | 0 | 0 | 51 | 49 |
| Iron deficiency anaemia | 72 | I | 22 | 4 | 45 | 55 | 62 | 6 | 30 | 2 | 45 | 55 |
| Vitamin A deficiency | 85 | 1 | 14 | 0 | 43 | 57 | 86 | 1 | 12 | 0 | 44 | 56 |
| Zinc deficiency | 100 | 0 | 0 | 0 | 51 | 49 | 100 | 0 | 0 | 0 | 51 | 49 |
| Other nutrition-related risk factors and physical inactivity |  |  |  |  |  |  |  |  |  |  |  |  |
| High blood pressure | 0 | 0 | 19 | 81 | 49 | 51 | 0 | 0 | 43 | 57 | 54 | 46 |
| High cholesterol | 0 | 0 | 22 | 78 | 48 | 52 | 0 | 0 | 50 | 50 | 55 | 45 |
| Overweight and obesity (high BMI) | 0 | 0 | 26 | 74 | 45 | 55 | 0 | 0 | 57 | 43 | 47 | 53 |
| Low fruit and vegetable consumption | 0 | 0 | 23 | 77 | 53 | 47 | 0 | 0 | 49 | 51 | 57 | 43 |
| Physical inactivity | 0 | 0 | 21 | 79 | 50 | 50 | 0 | 0 | 48 | 52 | 53 | 47 |
| Addictive substances |  |  |  |  |  |  |  |  |  |  |  |  |
| Smoking and oral tobacco use | 0 | 0 | 30 | 70 | 79 | 21 | 0 | 0 | 61 | 39 | 82 | 18 |
| Alcohol use | I | 1 | 65 | 33 | 91 | 9 | 1 | 3 | 87 | 9 | 85 | 15 |
| Illicit drugs use | 0 | 0 | 100 | 0 | 80 | 20 | 0 | 2 | 98 | 0 | 77 | 23 |
| Sexual and reproductive health |  |  |  |  |  |  |  |  |  |  |  |  |
| Unsafe sex | 16 | 1 | 77 | 6 | 47 | 53 | 18 | 1 | 79 | 2 | 46 | 54 |
| Non-use and use of ineffective methods of contraception | 0 | 0 | 100 | 0 | 0 | 100 | 0 | 0 | 100 | 0 | 0 | 100 |
| Environmental risk factors |  |  |  |  |  |  |  |  |  |  |  |  |
| Unsafe water, sanitation and hygiene | 68 | 5 | 13 | 14 | 52 | 48 | 77 | 8 | 13 | 3 | 51 | 49 |
| Urban air pollution | 3 | 0 | 16 | 81 | 51 | 49 | 12 | 0 | 40 | 49 | 56 | 44 |
| Indoor air pollution from household use of solid fuels | 56 | 0 | 5 | 38 | 41 | 59 | 83 | 0 | 8 | 9 | 49 | 51 |
| Lead exposure | 0 | 0 | 41 | 57 | 66 | 34 | 75 | 0 | 16 | 8 | 55 | 45 |
| Global climate change | 86 | 3 | 6 | 5 | 49 | 51 | 88 | 5 | 6 | 1 | 49 | 51 |
| Selected occupational risk factors |  |  |  |  |  |  |  |  |  |  |  |  |
| Risk factors for injuries | 0 | 0 | 85 | 14 | 94 | 6 | 0 | 0 | 95 | 5 | 93 | 7 |
| Carcinogens | 0 | 0 | 28 | 72 | 85 | 15 | 0 | 0 | 51 | 49 | 83 | 17 |
| Airborne particulates | 0 | 0 | 17 | 83 | 74 | 26 | 0 | 0 | 65 | 35 | 77 | 23 |
| Ergonomic stressors | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 95 | 5 | 59 | 41 |
| Noise | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 89 | 11 | 67 | 33 |
| Other selected risk factors |  |  |  |  |  |  |  |  |  |  |  |  |
| Contaminated injections in health care settings | 10 | 2 | 53 | 35 | 63 | 37 | 16 | 3 | 67 | 13 | 61 | 39 |
| Child sexual abuse | 0 | 0 | 81 | 22 | 48 | 52 | 0 | 0 | 96 | 4 | 36 | 64 |

For these risks, more than $85 \%$ of the total attributable burden occurred in this age group, with the exception of iron deficiency where $30 \%$ of burden was borne by women of childbearing age. The disease burden from other diet-related risks, tobacco and occupational risks (except injuries and back pain) was almost equally distributed among adults above and below the age of 60 years. For example, $43 \%$ and $61 \%$ of disease burden due to high blood pressure and tobacco respectively, occurred from adverse events in the 15-59-year age group.

More than $90 \%$ of disease burden attributable to lack of contraception, illicit drugs, occupational ergonomic stressors and risk factors for injury and child sexual abuse occurred in adults below the age of 60 years. About three-quarters (77-80\%) of disease burden for alcohol, unsafe sex and contaminated injections in health care settings occurred between the ages of 15 and 59 years. Most of the risks whose burden is concentrated in younger adults are those with outcomes that include HIV/AIDS, maternal conditions, neuropsychiatric diseases and injuries. Moreover, with the exception of alcohol, which has a global presence, the majority of disease burden from these risks is concentrated in developing countries (Figures 26.1 and 26.2). This illustrates the large, and at times neglected disease burden from risks that affect young adults in developing countries, with important consequences for economic development.

Only a small fraction of disease burden from the risk factors considered occurred among 5-14-year olds. This was because some of the leading causes of ill-health of this age group (e.g. motor vehicle accidents and other injuries, depression) have complex causes that could not easily be included in the current risk-based framework. For other leading diseases at these ages (e.g. diarrhoea and lower respiratory infections), most epidemiological studies have focused on children aged $<5$ years and do not provide hazard estimates for older children.

The disease burden attributable to underweight and micronutrient deficiencies in children was equally distributed among males and females, but the total all-age disease burden from iron and vitamin A deficiencies was slightly greater in females due to effects on maternal conditions. Other diet-related risks, physical inactivity, environmental risks and unsafe sex contributed almost equally to disease burden in males and females. Approximately $80 \%$ of disease burden from addictive substances and $60-90 \%$ from various occupational risks occurred among men. The former reflects the social, behavioural and economic forces that have so far made addictive substances more widely used by men, especially in developing countries. The latter was partially due to the inclusion of formal employment only and partially because men tend to make up most of the workforce engaged in heavy industrial jobs and formal agriculture. Women suffered an estimated two-thirds of disease burden from childhood sexual abuse and the entire burden caused by non-use and use of ineffective methods of contraception, as defined in chapter 15.

The distributions of disease burden attributable to risk factors by exposure levels are shown in Table 26.3 for those risks quantified using categorical variables, and in Figure 26.3 for those with continuous variables. For most of these risks a substantial proportion of attributable burden occurred among those with modest elevation of risk. For example, only $35 \%$ of the disease burden from underweight, the leading global risk, occurred in severely underweight children ( $<-3$ SD from referent group median); the rest was among those in the 1-3 SD below the median range. The large majority of the burden of disease from unsafe water, sanitation and hygiene was approximately equally distributed among three of the five exposure scenarios. This reflects the fact that the exposure categories were defined as the presence of water and sanitation technology-based interventions, and during decades of water and sanitation projects, many countries have "clustered" in a limited number of technology groups. However, there is likely to be large heterogeneity of exposure within each scenario (Curtis et al. 2000).

Figure 26.3(a) shows the distribution of the estimated cardiovascular (CVD) burden of disease (in DALYs) attributable to four major continuous risk factors, by exposure levels. Half the attributable burden occurs to the left of the solid vertical line and half occurs to the right. The dashed vertical lines indicate commonly used thresholds-140 or 160 mmHg for hypertension, $6.5 \mathrm{mmol} / \mathrm{l}$ for hypercholesterolaemia, and $30 \mathrm{~kg} / \mathrm{m}^{2}$ for obesity (the cut-off for hypertension is shown at 150 mmHg , between the two commonly used values of 140 and 160 mmHg ). Figure 26.3 (b) shows the cumulative percentage of attributable burden by exposure levels. In reality, a modest rightward skew (not modelled here) of distributions would lead to slightly more events occurring in those who were hypertensive, hypercholesterolaemic or obese.

Figure 26.3 shows that a substantial proportion of the disease burden attributable to high blood pressure, cholesterol and body mass index (BMI) and inadequate fruit and vegetable intake occurred in the "midrange" exposures. For example, the 2nd and 3rd quartiles (i.e. half of attributable burden) occurred between SBP of approximately 130 and 150 mmHg , cholesterol of 5.0 and $6.1 \mathrm{mmol} / 1 \mathrm{and}$ BMI of $25-32 \mathrm{~kg} / \mathrm{m}^{2}$ and fruit and vegetable intake of $150-300 \mathrm{~g} / \mathrm{day}$. This was similar to or greater than the amount of burden occurring among individuals with risk factor levels above the commonly used (but arbitrary) thresholds of hypertension, hypercholesterolaemia and obesity, as shown in Figure 26.3.

The distribution by levels of demographic and economic development suggest that the burden of disease due to risks such as undernutrition and unsafe water, sanitation and hygiene occurred virtually entirely in the high-mortality developing subregions of the world, whereas other risks, such as tobacco, alcohol and dietary risks had global effects (Figure 26.1). Categorization based on economic and demographic development was also a key modifier of the age distribution patterns. Most notably,
Table 26.3 Distribution by exposure level of attributable burden due to selected categorical risk factors

| Risk factor | Referent category | Exposure categories |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Childhood and maternal underweight | Same fraction of children <-I SD weight-for-age as the international reference group | $<-1$ to $<-2$ SD below the international reference group median | $<-2$ to $<-3$ SD below the international reference group median | <-3 SD below the international reference group median |  |  |
| Proportion of total attributable disease burden | 0 | 0.20 | 0.46 | 0.35 |  |  |
| Physical inactivity | All having at least 2.5 hours per week of moderate-intensity activity or equivalent ( $400 \mathrm{~kJ} /$ week) | Some but less than 2.5 hours per week of moderate-intensity activity | Little or no physical activity |  |  |  |
| Proportion of total attributable disease burden | 0 | 0.49 | 0.51 |  |  |  |
| Unsafe water, sanitation and hygiene | Absence of transmission of diarrhoeal disease through water, sanitation and hygiene | Regulated water supply and full sanitation coverage, with partial treatment for sewage | Improved water supply, basic sanitation, improved access to drinking water, improved personal hygiene and water disinfected at point of use | Improved water supply and basic sanitation | Basic sanitation but no improved water supply | No improved water supply and no basic sanitation |
| Proportion of total attributable disease burden | 0 | 0 | 0.39 | 0.03 | 0.28 | 0.30 |
| Child sexual abuse | No sexual abuse | Non-contact abuse | Contact abuse | Intercourse |  |  |
| Proportion of total attributable disease burden | 0 | 0.08 | 0.44 | 0.48 |  |  |

Figure 26.3 Distribution by exposure level of cardiovascular disease (CVD) burden attributable to selected continuous risk factors


Note: For blood pressure and cholesterol, the plots represent the estimated usual levels (MacMahon et al. 1990), which tend to be closer to population means than levels based on one-off measurements commonly used in population surveys. For example, the distribution of usual blood pressure is approximately half as wide as the distribution of one-off blood pressure measures and so less people would be classified as hypertensive if classifications were based on usual rather than one-off blood pressure. Thus, with a population mean systolic blood pressure (SBP) of 134 mmHg , the SD of one-off measures might be 17 mmHg (with about $18 \%$ of the population having one-off SBP over 150 mmHg ) and the SD of usual SBP 9 mmHg (hence about $5 \%$ of the population would have usual SBP over 150 mmHg ).
disease burden due to many major risks for chronic diseases occurred in younger ages in developing regions compared to developed regions. For example, in high-mortality developing subregions, $69 \%$ of disease burden attributable to tobacco occurred in people aged 15-59 years, whereas this share was $63 \%$ for low-mortality developing subregions and $55 \%$ for developed subregions. The different age structures across major world regions, together with exposure differences, resulted in different distributions of attributable burden by region. For example, Figure 26.4 shows that disease burden attributable to elevated blood pressure, cholesterol and BMI occurred at lower levels in developing regions compared to developed regions, mainly because of lower age-specific exposure levels in those populations (see chapters 6-8).

Figure 26.4 Distribution by exposure level of attributable cardiovascular disease (CVD) burden due to selected risk factors, by age and subregional grouping (see Figure 26.3 for details)


Figure 26.4 also shows that skewness of the distribution of disease burden was not substantially different across different age groups for BMI. This is because the comparatively larger relative risk per unitBMI at younger ages (which leads to more right-hand skew) is counterbalanced by the comparatively lower BMI at younger ages (which leads to left-hand skew). This is in contrast to blood pressure, for which disease burden in younger age groups occurred at lower exposures because the age patterns of exposure and relative risk do not entirely compensate.

## 3. Sources of uncertainty

Broad sources of uncertainty in risk assessment were discussed in chapter 1 of this book. Uncertainty about disease causation (Evans 1978; Hill 1965) in practice was secondary to uncertainty about hazard size, because when causality was uncertain, estimates of hazard needed for risk assessment were also unknown or uncertain. For example, while there is uncertainty about whether climate change would increase incidence of certain diseases, or whether the relationships between occupational factors or physical inactivity and lower back pain are causal, in each case quantitative risk assessment would also require estimates of hazard magnitude. The collectivity of scientific knowledge from disciplines such as behavioural science, vector biology, physiology, biomechanics and epidemiology would confirm the possibility of a causal relationship in the above cases, but would shift the debate to hazard size. As a result, for some risk factors, only the contribution to a subset of disease outcomes could be quantified because epidemiological studies did not provide enough information for hazard quantification for all risk factor-disease pairs, even when the causal relationships were believed or suspected.

Estimates of hazard size in individual studies were as far as possible adjusted for confounding. Extrapolation of hazard from a limited number of studies to other populations on the other hand has received less attention. While the robustness of relative risk measures has been confirmed for more proximal factors in studies across populations (Eastern Stroke and Coronary Heart Disease Collaborative Research Group 1998; Horton 2000; Law et al. 1994), hazard extrapolation is an important source of uncertainty for more distal risks (e.g. child sexual abuse) or those whose effects are heterogeneous (e.g. alcohol and injuries vs alcohol and cancer).

Direct exposure data for many risk factors were limited due to difficulties both in their measurement and under-investment in risk factor surveillance, especially in developing countries. To allow maximum use of available data, such risk factors were represented using indirect or aggregate indicators (e.g. smoking impact ratio (SIR) for accumulated hazards of smoking, weight-for-age for childhood undernutrition and use of solid fuels for indoor air pollution). For some risks, multiple data sources allowed limiting the possible range of exposure estimates. For example, in the absence of alcohol surveys, total alcohol production, trade and unrecorded consumption provided upper bounds on the fraction of population that would be in the highest consumption category. Finally, some of the risk factors in this analysis were represented using continuous exposure variables (e.g. high blood pressure). Others have used categorical variables (e.g. indoor smoke from solid fuels, underweight and physical inactivity) even though the health effects occur along a continuum. This choice reflected the availability of exposure data and
hazard estimates for categories. In such cases, the contribution to disease within the categories may be under-estimated.

The findings of this work should, therefore, be considered within the context of limited available data and subject to uncertainty. This uncertainty varies across risk factors and geographical regions. Further discussion of sources and quantification of uncertainty has been provided in individual risk factor chapters.

## 4. Discussion

Despite inherent uncertainties, the quantification of the burden of disease attributable to selected risk factors illustrates that the loss of health in the world is dominated by those risk factors that affect the poorest regions and populations, such as undernutrition, poor water, sanitation and hygiene and indoor smoke from solid fuels. Coupled with these are hazards such as alcohol, tobacco, high blood pressure and high cholesterol that in the year 2000-even compared to a decade earlier (Murray and Lopez 1997)—are widespread or are estimated to have large health impacts. Nowhere is this picture more apparent than in the lowmortality developing regions, which account for $48 \%$ of global population, and are affected by both groups of risk factors.

Comparing the burden attributable to risk factors across the three groupings of countries in this work (Figures 26.1 and 26.2) provides a cross-sectional picture of a "risk factor transition" in which the relative contribution of adult or noncommunicable disease risk factors increases as childhood and communicable disease risk factors decrease with economic development. Analysis of previous development-based transitions, such as changes in inequality or environment with economic development, has demonstrated the role of policy in inducing or delaying, and shaping the dynamics of the transition (Bowman 1997). Examples in public health include rapid control of vector-borne diseases (Chitsulo et al. 2000), high maternal mortality where contraception and abortions are not accessible for non-economic reasons, and potential HIV epidemics in some developed countries (MacLehose et al. 2002). At the same time, at least some risk factor transitions are confirmed by the increasing role of hazards such as tobacco and obesity over time (Ebbeling et al. 2002; Pelletier 1998; WHO 1997). The increase in the global burden of disease due to tobacco from $2.6 \%$ in 1990 to $4.1 \%$ in 2000, while partially due to new evidence on hazard size after correction for confounding (Thun et al. 2000), mostly reflects the increased accumulated hazards, and is most noticeable in developing countries. The cross-sectional comparison demonstrates that risk factors such as alcohol and high blood pressure and cholesterol, if not increasing in absolute terms (Popkin 2002; Reddy and Yusuf 1998), are important contributors to loss of health in all regions.

The large remaining burden from childhood disease and mortality risks such as undernutrition, poor water and sanitation, and indoor smoke from solid fuels shows the continued need for developing and delivering effective interventions. At the same time, four of the five leading causes of lost healthy life affect adults (Figure 26.1). Risk factors for both adult communicable and noncommunicable diseases already make substantial contributions even in regions with low income and high infant mortality. It is imperative therefore that health programmes and policy continually reassess the appropriate balance between interventions addressing childhood disease risk factors and those that affect adult health. Dynamic and systematic policy responses can mitigate the spread of such risk factors and their more distal causes to a large extent throughout the development process, such as a healthier nutritional or environmental transitions (Arrow et al. 1995; Lee et al. 2000). Also, as illustrated by the persistence of diseases such as malaria or the large increase in the disease burden due to HIV/AIDS and its risk factors since 1990 (e.g. unsafe sex from $3.5 \%$ to $6.3 \%$ ), as well as the potential for generalized HIV/AIDS epidemics in some eastern European countries (MacLehose et al. 2002) or China (Kaufman and Jing 2002), important communicable disease risk factors also need dynamic monitoring and policy responses.

There are a number of reasons why risk factors that were not among the leading global causes of disease burden should not be neglected. Most obviously, this analysis could be expanded with other risk factors that are both prevalent and hazardous. Second, although smaller than other risks, many such risk factors make non-negligible contributions to burden of disease in specific populations. For example in WPR-B (dominated by China in terms of population), where there is considerable industrial activity based on coal, ambient air pollution and lead exposure have health effects comparable to poor water, sanitation and hygiene and some micronutrient deficiencies. Similarly, lack of contraception was among the 10 leading risk factors for female burden of disease in a number of subregions.

Some risk factors with comparatively low global disease burden are highly concentrated among sectors of society (e.g. occupational exposures among mine workers) and have implications for health inequalities. This concentration may also imply that risks can be targeted more easily. For other risk factors, such as child sexual abuse, ethical considerations may outweigh direct contributions to disease burden in policy debate. Finally, while the burden of disease due to a risk factor may be comparatively small, effective or cost-effective interventions may be available. Examples include reducing the number of unnecessary medical injections coupled with the use of sterile syringes and reduction in exposure to lead or ambient air pollution in industrialized countries in the second half of the 20th century which often also led to benefits such as energy saving.

Beyond their total magnitude, this study has also provided a picture of the distribution of risk factor-attributable disease burden by age, sex and exposure. Analysis in multiple age and exposure categories, or along a continuum of exposures, suggests that globally a considerable proportion of the disease burden attributable to many major risk factors occurred among those with only moderately raised levels, not the extremes (e.g. hypertension, obesity or severe malnutrition).

For acute exposures and outcomes, the underlying relationship is more complex. For example, while in many societies the majority of alcoholattributable injury (e.g. traffic accidents) arises among people who on average drink moderately (Kreitman 1986), these people would be at the more extreme end of the distribution in a different dimension: volume of drinking before the injury. This finding suggests that the shapes of both exposure distributions and risk relationships are important determinants of the distribution of disease burden. If exposure to risk factors is clustered or the risk relationship does not follow a linear pattern, high exposure groups may indeed play a disproportionately important role (Lemmens 2001; Skog 1999). Further implications of these findings for research, and for policies and programmes aimed at improving population health, are discussed in chapter 29.

## Note

1 See preface for an explanation of this term.

## References

Arrow K, Bolin B, Costanza R et al. (1995) Economic growth, carrying capacity, and the environment. Science, 168:520-521.
Bowman KS (1997) Should the Kuznets effect be relied on to induce equalizing growth: evidence from post-1950 development. World Development, 25:127-143.

Chitsulo L, Engels D, Montresor A, Savioli L (2000) The global status of schistosomiasis and its control. Acta Tropica, 77:41-51.

Cook NR, Cohen J, Hebert P, Taylor JO, Hennekens CH (1995) Implications of small reductions in diastolic blood pressure for primary prevention. Archives of Internal Medicine, 155:701-709.

Curtis V, Cairncross S, Yonli R (2000) Domestic hygiene and diarrhoea-pinpointing the problem. Tropical Medicine and International Health, 5:22-32.
de Onis M, Frongillo E, Blössner M (2000) Is malnutrition declining? An analysis of changes in levels of child malnutrition since 1980. Bulletin of the World Health Organization, 78:1222-1233.
Eastern Stroke and Coronary Heart Disease Collaborative Research Group (1998) Blood pressure, cholesterol, and stroke in Eastern Asia. The Lancet, 352:1801-1807.

Ebbeling CB, Pawlak DB, Ludwig DS (2002) Childhood obesity: public health crisis, common sense cure. The Lancet, 360:473-482.

Evans AS (1978) Causation and disease: a chronological journey. American Journal of Epidemiology, 108:249-258.

Hill AB (1965) The environment and disease: association or causation? Proceedings of the Royal Society of Medicine, 58:295-300.
Horton R (2000) Common sense and figures: the rhetoric of validity in medicine (Bradford Hill Memorial Lecture 1999). Statistics in Medicine, 19: 3149-3164.

Jacobs DR Jr, Ainsworth BE, Hartman TJ, Leon AS (1993) A simultaneous evaluation of 10 commonly used physical activity questionnaires. Medicine and Science in Sports and Exercise, 25:81-91.
Kaufman J, Jing J (2002) China and AIDS—the time to act is now. Science, 296:2339-2340.

Kreitman N (1986) Alcohol consumption and the preventive paradox. British Journal of Addiction, 81:353-363.

Law MR, Wald NJ, Thompson SG (1994) By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? British Medical Journal, 308:367-373.

Lee M-J, Popkin BM, Kim S (2000) The unique aspects of the nutrition transition in South Korea: the retention of healthful elements in their traditional diet. Public Health Nutrition, 5:197-203.
Lemmens P (2001) Relationship of alcohol consumption and alcohol problems at the population level. In: International handbook of alcohol dependence and problems, Heather N, Peters TJ, Stockwell T, eds. John Wiley and Sons Ltd, Chichester.

Levine JA, Weisell R, Chevassus S, Martinez CD, Burlingame B, Coward WA (2001) The work burden of women. Science, 294:812.

MacLehose L, McKee M, Weinberg J (2002) Responding to the challenge of communicable disease in Europe. Science, 295:2047-2050.

MacMahon S, Peto R, Cutler J et al. (1990) Blood pressure, stroke, and coronary heart disease. Part I. Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. The Lancet, 335:765-774.
Monteiro CA, Conde WL, Popkin BM (2002) Is obesity replacing or adding to undernutrition? Evidence from different social classes in Brazil. Public Health Nutrition, 5:105-112.
Murray CJL, Lopez AD (1997) Global mortality, disability, and the contribution of risk factors: global burden of disease study. The Lancet, 349:1436-1442.

Murray CJL, Lauer JA, Hutubessy RC et al. (2003) Effectiveness and costs of interventions to lower systolic blood pressure and cholesterol: a global and regional analysis on reduction of cardiovascular-disease risk. The Lancet, 361:717-725.

Pelletier DL, Rahn M (1998) Trends in body mass index in developing countries. Food and Nutrition Bulletin, 19:223-239.

Peto R, Lopez AD, Boreham J, Thun M, Heath Jr C (1992) Mortality from tobacco in developed countries. The Lancet, 339:1268-1278.

Popkin BM (2002) An overview on the nutrition transition and its health implications: the Bellagio meeting. Public Health Nutrition, 5:93-103.

Popkin BM, Horton S, Kim S, Mahal A, Shuigao J (2001) Trends in diet, nutritional status and diet-related noncommunicable diseases in China and India: the economic costs of the nutrition transition. Nutrition Reviews, 59: 379-390.

Powell K, Blair S (1994) The public health burdens of sedentary living habits: theoretical but realistic estimates. Medicine and Science in Sports and Exercise, 26:851-856.

Reddy KS, Yusuf S (1998) Emerging epidemic of cardiovascular disease in developing countries. Circulation, 97:596-601.

Rodgers A, Lawes C, MacMahon S (2000) The global burden of cardiovascular disease conferred by raised blood pressure. Benefits of reversal of blood pres-sure-related cardiovascular risk in Eastern Asia. Journal of Hypertension, 18(Suppl.):S3-S5.
Rodgers A, MacMahon S (1999) Blood pressure and the global burden of cardiovascular disease. Clinical and Experimental Hypertension, 21:543-552.
Rose G (1992) The strategy of preventive medicine. Oxford University Press, Oxford.

Skog OJ (1999) Prevention paradox revisited. Addiction, 94:751-757.
Thun MJ, Apicella LF, Henley SJ (2000) Smoking vs other risk factors as the cause of smoking-attributable mortality: confounding in the courtroom. Journal of American Medical Association, 284:706-712.

WHO (1997) Tobacco or health: a global status report. World Health Organization, Geneva.

## Chapter 27

# Potential health gains from REDUCING MULTIPLE RISK FACTORS 

Majid Ezzati, Stephen Vander Hoorn, Anthony Rodgers, Alan D. Lopez, Colin D. Mathers and Christopher J.L. Murray

## 1. Introduction

Estimates of the burden of disease attributable to selected individual risk factors were presented in chapter 26. Diseases and injuries are, however, almost always caused by multiple risk factors (Rothman 1976; Walter 1980), motivating analysis of the health benefits of simultaneous reductions in multiple risks. Estimating the joint effects of multiple distal and proximal risks is particularly important because many factors act through other, intermediate, factors (Murray and Lopez 1999; Yerushalmy and Palmer 1959), or in combination with other factors, as we described in chapter 1 of this book. For example, education, occupation and income may affect smoking, physical activity and diet, which are risk factors for cardiovascular diseases, both directly and through further layers of intermediate factors such as body mass index (BMI), blood pressure and cholesterol. Multi-causality also means that a range of interventions can be used for disease prevention, with the specific choice determined by factors such as cost, technology availability, infrastructure and preferences.

A number of works have estimated the joint effects of two or more risk factors in specific cohorts (Hirayama 1990; Neaton and Wentworth 1992; Rothman and Keller 1972; Stampfer et al. 2000; Willet 2002), or for specific groups of diseases and risks (Doll and Peto 1981; Smith et al. 1999). Innovative models and methods have also been developed to quantify the complexity of multiple risk factor effects, especially as they interact over time (Manton et al. 1993; Robins 1999). Estimating joint risk factor effects beyond specific diseases or cohorts, however, remains relatively unexplored in epidemiology and population health. Using comprehensive reviews of data on selected major risk factors in various levels of causality, this chapter is an attempt to do so.

[^112]We further used the joint effects of multiple risk factors to estimate the potential gain in healthy life expectancy (HALE) from reducing these risks. Analysis of multiple risk factors, with heterogeneous contributions to disease burden in different populations, would also allow estimating how much of the cross-population health differentials (e.g. differences in HALE) are due to the selected risk factors. By estimating gains in HALE based on causes of disease, this work also contributes in a systematic way to the continued debate on the potential limits to life expectancy (Oeppen and Vaupel 2002; Riley 2001).

## 2. Methods

### 2.1 Estimating joint population attributable fractions

Methods and data sources for estimating the burden of disease attributable to individual risk factors were described in chapter 25 . The contribution of a risk factor to disease or mortality relative to some alternative exposure scenario (i.e. population attributable fraction, PAF, defined as the proportional reduction in population disease or mortality that would occur if exposure to the risk factor were reduced to an alternative exposure scenario [Eide and Heuch 2001; Miettinen 1974]) is given by the generalized "potential impact fraction" in Equation 1.

$$
\begin{equation*}
P I F=\frac{\int_{x=0}^{m} R R(x) P(x) d x-\int_{x=0}^{m} R R(x) P^{\prime}(x) d x}{\int_{x=0}^{m} R R(x) P(x) d x} \tag{1}
\end{equation*}
$$

where
$R R(x)$ : relative risk at exposure level $x$
$P(x)$ : population distribution of exposure
$P^{\prime}(x)$ : alternative or counterfactual distribution of exposure, and $m$ : maximum exposure level
In equation $1, R R, P$, and $P^{\prime}$ may represent joint relative risks and exposure distributions for multiple risk factors (i.e. $x$ may be a vector of risk factors), with $R R$ for each risk factor estimated at the appropriate level of the remaining ones (Eide and Heuch 2001). Alternatively, for $n$ biologically independent and uncorrelated risk factors, the joint PAF is given by equation 2 (Miettinen 1974; Walter 1976). If risk factors are independent and uncorrelated, the proportion of the remaining disease which is attributed to the $i$ th additional risk factor equals $P A F_{i}$ (and hence $1-P A F_{i}$ not attributable to this factor). Therefore, the second term in the right hand side of equation 2 (i.e. the product of all
[1 - $P A F_{i}$ ] terms) is the fraction of disease not attributable to any of the $n$ risk factors. One minus this term is the fraction attributable to the combined effects of the $n$ risk factors:

$$
\begin{equation*}
P A F=1-\prod_{i=1}^{n}\left(1-P A F_{i}\right) \tag{2}
\end{equation*}
$$

where $P A F_{i}$ is the PAF of individual risk factors
Estimating the joint effects of multiple risk factors is in practice complex for several reasons. First, some of the effects of the more distal factors (e.g. physical inactivity) are mediated through intermediate factors (e.g. high BMI itself through blood pressure) (Figure 27.1). Estimating the joint effects of distal and intermediate factors requires knowledge of independent hazards of the distal ones (vs individual risk factor effects, which are based on total hazard) (Figure 27.1). Second, the hazard due to a risk factor may depend on the presence of other risk factors (effect modification) (Koopman 1981; Rothman and Greenland 1998). Third, there may be correlation between exposure to various risk factors, because they are affected by the same distal factors and policies.

Figure 27.I Mediated and direct effects


For example, undernutrition, poor water and sanitation and the use of solid fuels are more common among poor rural households in developing countries, or smokers generally have higher and more harmful patterns of alcohol consumption and worse diet than non-smokers.

While the current literature refers to scenarios 1 and 2 as biological interaction and to scenario 3 as statistical interaction (Miettinen 1974; Rothman and Greenland 1998; Rothman et al. 1980), this distinction is somewhat arbitrary and the three scenarios may occur simultaneously. For example, zinc deficiency affects mortality from diarrhoea directly as well as through lowering growth (weight-for-age) (scenario 1) (Brown et al. 2002; Zinc Investigators' Collaborative Group 1999), and may also be correlated with underweight, other micronutrient deficiencies, and poor water and sanitation (scenario 3). Similarly alcohol and smoking may not only be correlated (scenario 3), but also affect each other's hazard for some diseases (scenario 2) (Rothman and Keller 1972). Although the epidemiological literature has placed much emphasis on removing or minimizing the effects of confounding covariates, mediated and stratified hazards have received disproportionately little empirical attention. Therefore, we used reviews of extant literature and re-analysed existing cohort data to strengthen the empirical basis for considering interactions in sensitivity analyses.

In one set of estimates (referred to as the unadjusted scenario), we assumed no mediated effects or interactions among risk factors. We then included the mediated effects and interactions described above in a second scenario (referred to as the adjusted scenario).

## JOINT EFFECTS OF CARDIOVASCULAR DISEASE RISK FACTORS

Epidemiological studies on the effects of high BMI, physical inactivity, and low fruit and vegetable intake on cardiovascular disease risk have illustrated some attenuation of the effects after adjustment for intermediate factors (e.g. blood pressure or cholesterol) (Berlin and Colditz 1990; Blair et al. 2001; Eaton 1992; Gaziano et al. 1995; Jarrett et al. 1982; Jousilahti et al. 1999; Khaw and Barrett-Connor 1987; Liu et al. 2001, 2000; Manson et al. 1990, 2002; Rosengren et al. 1999; Tate et al. 1998). This attenuation confirms that some of the hazard of the more distal factors is mediated through the intermediate ones (Figure 27.1). The attenuation has varied among studies but has consistently been less that one half of the excess risk of the distal factors. We used an upper bound of $50 \%$ as the proportion of the excess risk from these risk factors mediated through intermediate factors that are themselves among the selected risks.

To include effect modification, deviations from the multiplicative model of $10 \%$ for ischaemic heart disease (IHD) and $30 \%$ for ischaemic stroke were used based on existing studies (both sub-multiplicative) (Eastern Stroke and Coronary Heart Disease Collaborative Research Group 1998; Neaton and Wentworth 1992).

## Joint effects of smoking and other risk factors

Liu et al. (1998) found that in China, the relative risks for mortality from lung and other cancers, respiratory diseases and cardiovascular diseases were approximately constant in different cities whose non-smoker mortality rates from these diseases varied by a factor of 4-10 (see Figures 4 and 6 in Liu et al. 1998). This finding has also been confirmed in studies which stratified hazards for serum cholesterol (Jee et al. 1999).

## JOINT EFFECTS OF CHILDHOOD UNDERNUTRITION FOR INFECTIOUS DISEASES

Zinc affects child growth (Brown et al. 2002) and some of its effects on infectious diseases may be mediated through growth (e.g. underweight). As no published source for these mediated effects existed, data from some of the available zinc trials (Zinc Investigators’ Collaborative Group 1999) were re-analysed and an upper bound of $50 \%$ on the proportion of zinc deficiency risk mediated through underweight was used. Vitamin A deficiency, which affects some of the same diseases as underweight and zinc deficiency, has been found not to change the hazard size for the other two risk factors based on stratified results from clinical trials and recent reviews of micronutrient deficiency literature (Christian and West Jr. 1998; Ramakrishnan and Martorell 1998; Ramakrishnan et al. 1995; West et al. 1991).

## Joint effects of undernutrition and environmental risk factors in CHILDHOOD DISEASES

Anthropometric (growth) indicators of childhood nutrition (e.g. weight-for-age) are aggregate measures of multiple factors which include nutrition (e.g. protein-energy intake) and previous infection (Pelletier et al. 1993; Scrimshaw et al. 1968; UNICEF 1990). Therefore, some of the risks for indoor smoke from solid fuels and poor water, sanitation and hygiene (which result in acute lower respiratory infections [ALRI] and diarrhoea, respectively) may be mediated through underweight. In a review of existing literature, Briend (1990) concluded that attempts to disentangle direct and mediated contributions, especially over long time periods needed to affect population-level anthropometry, have not established diarrhoea as a significant cause of underweight. Other works, however, have found evidence that infection (especially diarrhoea) could result in reduced growth and increased the prevalence of underweight (Black 1991; Guerrant et al. 1992; Lutter et al. 1989, 1992; Martorell et al. 1975a, 1975b; Stephensen 1999). To account for potential mediated effects, we chose an upper bound of $50 \%$ for the proportion of the excess risks for indoor smoke from solid fuels and for poor water, sanitation and hygiene mediated through underweight in subregions ${ }^{1}$ where underweight was a cause of disease burden.

RISK FACTOR CORRELATION
To estimate the joint effects of risk factors with a continuous exposure variable (e.g. blood pressure and cholesterol), each integral in the PIF re-
lationship may be replaced with $\int_{x_{1}=0}^{m_{1}} \int_{x_{2}=0}^{m_{2}} R R_{1}\left(x_{1}\right) R R_{2}\left(x_{2}\right) P\left(x_{1}, x_{2}\right) d x_{1} d x_{2}$,
where subscripts 1 and 2 denote the two risk factors and $P$ is the joint distribution of the two exposures. If the joint $R R$ were a linear function of exposure levels ( $x_{1}$ and $x_{2}$ ), then correlation between the two risk factors would not affect total hazard. Because individual $R R s$ are nonlinear functions of exposure (e.g. in a logistic or Cox proportional hazard model) and joint RRs are the product of such terms, positive correlation between risk factors would, in general, imply a larger PAF than zero correlation, which in turn would be larger than negative correlation (submultiplicative effect modification could result in smaller PAF even with positive correlation for some $R R$ values). Similarly, for categorical risk factors, positive correlation would in general result in larger PAF (see also Greenland 1984).

For the range of exposures and relative risks observed here, this secondary effect of risk factor correlation would be considerably smaller than the joint attributable fraction, which may be confirmed by microsimulation of exposure distributions and relative risks. This is because the PAF relationship is an increasing concave function of individual or joint $R R$ (i.e. rate of increase declines with increasing $R R$ or prevalence) (Figure 27.2). Because the risk factors considered in this analysis individually accounted for large fractions of the diseases affected by them in populations where these diseases are important components of disease burden, the joint effects approached $100 \%$ asymptotically, limiting the overestimation potential (e.g. individually, underweight accounted for $60-70 \%$ of under-five diarrhoea in AFR-D, AFR-E and SEAR-D; poor water, sanitation and hygiene for approximately $90 \%$; vitamin A deficiency for $20 \%-30 \%$; and zinc deficiency for $10-17 \%$; similarly, in various developed subregions, individually, high blood pressure accounted for 44-64\% of IHD; high cholesterol for 51-68\%; high BMI for $17-36 \%$; low fruit and vegetable intake for $19-35 \%$; and physical inactivity for 15-16\%).

### 2.2 Gains in healthy life expectancy (HALE)

The incidence of many conditions (e.g. neuropsychiatric conditions or long-term effects of injuries) may cause ill health but not death. It is therefore important to capture both fatal and non-fatal health outcomes in describing population health. Healthy life expectancy or healthadjusted life expectancy (HALE) reduces total life expectancy into equivalent years of "full health" by taking into account the distribution and severity of health states in the population (Mathers 2002). Inputs to the calculation of HALE include the period life table (or age-sex-specific

Figure 27.2 PAF relationship as a function of prevalence and relative risk


Note: PAF relationship is an increasing concave function of both prevalence (seen in the shape of each curve) and RR (seen in the declining distance between each adjacent pair of curves). As a result, for joint risk factor PAF, errors due to deviations from a simple uncorrelated multiplicative model (due to risk factor correlation or effect modification) are secondary to the joint PAF. For example, for two risk factors with $R R=2$ and $R R=3$, a multiplicative model would result in a joint $R R$ of 6 . The PAF would be approximately correct even if the true joint RR were 5 or 7 , as the PAF curves are close for these RR values. This phenomenon becomes increasingly dominant with increasing number of risk factors (i.e. the curves for $R R=I I, I 2$, and 13 are even closer than those for $R R=$ 5,6 and 7). Further, the flattening of PAF curves at high exposures limits the error due to risk factor correlation.
mortality rates) and prevalences of health states (resulting from diseases, their sequelae, and their combinations) for each country. Methods for estimating HALE have been described in detail elsewhere (Mathers et al. 2001). Unlike estimates of the burden of disease which compare current mortality and disability to a normative survivorship function (Murray and Lopez 1996), estimates of the gain in HALE account for competing risks.

The estimates in this chapter show the improvements in HALE for the year 2000, that would have been observed if exposure to the selected risk factors had been reduced to the theoretical-minimum-risk counterfactual distribution, as described in each of the risk factor chapters. In each of the 14 subregions, joint disease-specific PAFs were estimated for all diseases affected by the 20 leading global risk factors (chapter 26; see also individual risk factor chapters), for all age and sex groups.

Mortality and incidence in the counterfactual scenario are ( $1-P A F_{M}$ ) and $\left(1-P A F_{I}\right)$ times their original values where $P A F_{M}$ and $P A F_{I}$ are the PAF of mortality and incidence attributable to the joint effects of the risk
factors. The age-sex-specific mortality and age-sex-cause-specific prevalence of diseases and their sequelae were adjusted to these levels to estimate the HALE gain as a result of multiple risk factor removal. Reduction in prevalence was obtained from reduction in incidence under equilibrium conditions (Kruijshaar et al. 2002). Cause-specific estimates were necessary for non-fatal conditions because of different disability weights (Murray and Lopez 1996) but not for fatal conditions.

## 3. Results

Table 27.1 shows the individual and joint contributions of the 20 selected risk factors for the 10 leading diseases in the world and in three broad combinations of subregions-high-mortality developing ( $38 \%$ of global population), lower-mortality developing ( $40 \%$ of global population) and demographically and economically developed ( $22 \%$ of global population). ${ }^{2}$ For most diseases, the joint effects of these risk factors were substantially less than the crude sum of the individual effects (e.g. globally four separate risk factors were each responsible for $10 \%, 18 \%, 45 \%$ and $88 \%$ of diarrhoeal disease, but with a joint PAF of $92-94 \%$ ), confirming that a large number of cases are caused by the joint actions of more than one of these risk factors acting as sufficient causes (Rothman 1976), or through other factors.

Globally, large fractions of diarrhoea (92-94\%), ALRI (55-62\%), lung cancer ( $72 \%$ ), upper aerodigestive cancer ( $60 \%$ ), chronic obstructive pulmonary disease (COPD) ( $60 \%$ ), IHD ( $83-89 \%$ ) and stroke ( $70-76 \%$ ) were attributable to the joint effects of the risk factors considered here (see Willet 2002 and Stampfer et al. 2000 for consistent vascular disease examples from specific cohorts). The joint PAFs for cancers other than lung and upper aerodigestive ( $23 \%$ ), perinatal conditions ( $23 \%$ ), maternal conditions ( $42 \%$ ), and intentional ( $29 \%$ ) and unintentional ( $20 \%$ ) injuries, which have more diverse risk factors, were smaller but non-negligible. Although the fraction of total malaria burden attributable to childhood undernutrition was relatively large ( $56-59 \%$ ), this was because of the contribution of mortality at younger ages to disease burden. No adult malaria was attributed to the above risk factors because the epidemiological literature has focused on quantifying increased risk of malaria as a result of childhood undernutrition only. Finally, with the exception of alcohol and drug dependence, which were fully attributable to their specific risk factors, very small fractions or none of neuropsychiatric conditions, tuberculosis, congenital anomalies, and a number of other diseases were attributed to the risk factors considered in this book.

Figure 27.3 shows the individual and joint contributions (including overlap) of selected major risk factors to each of the following disease categories in the three subregional groups described above: I. communicable, maternal, perinatal and nutritional conditions;
Table 27.I Individual and joint contributions of 20 selected risks to 10 leading diseases and total burden of disease in the world and different subregional groupings ${ }^{\text {a }}$

| (a) World |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Disease/condition | \% global disease burden (total I. 46 billion DALYs) | \% global mortality (total 55.9 million deaths) | Contributing risk factors (individual PAF for disease burden) | Joint PAF ${ }^{\text {b }}$ <br> (disease burden) | Joint PAF ${ }^{\text {b }}$ <br> (mortality) |
| Lower respiratory infections | 6.1 | 6.8 | Underweight (childhood) (40\%); zinc deficiency (16\%); indoor smoke from solid fuels (36\%); tobacco (2\%) ${ }^{\text {c }}$ | 55-62\% | 40-45\% |
| HIV/AIDS | 5.5 | 4.6 | Unsafe sex (94\%); unsafe health care injections (5\%); illicit drugs (3\%) | 96\% | 96\% |
| Unipolar depressive disorders | 4.5 | 0.0 | Alcohol (2\%); childhood sexual abuse (6\%) | 7\% | $N A^{\text {d }}$ |
| Diarrhoeal diseases | 4.2 | 3.5 | Underweight (childhood) (45\%); vitamin A deficiency (18\%); zinc deficiency (10\%); unsafe water, sanitation and hygiene (88\%) | 92-94\% | 92-94\% |
| Ischaemic heart disease | 4.0 | 12.6 | High blood pressure (49\%); high cholesterol (56\%); high BMI (21\%); low fruit and vegetable intake (3I\%); physical inactivity (22\%); tobacco (I2\%); alcohol (2\%) | 83-89\% | 78-85\% |
| Low birth weight | 3.5 | 2.5 | Underweight (maternal) (10\%); iron deficiency (19\%); alcohol (0.2\%) | 29\% | 31\% |
| Stroke | 3.1 | 9.6 | High blood pressure (62\%); high cholesterol (18\%); high BMI (I3\%); low fruit and vegetable intake (1I\%); physical inactivity (7\%); tobacco (I2\%); alcohol (4\%) | 70-76\% | 65-73\% |

Table 27．I Individual and joint contributions of 20 selected risks to 10 leading diseases and total burden of disease in the world
Individual and joint contributions of 20 selected risks to 10 leading diseases and total burden of disease in the world
and different subregional groupings ${ }^{\text {a }}$（continued）

| （a）World（continued） |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Disease／condition | \％global disease burden（total I． 46 billion DALYs） | \％global mortality （total 55.9 million deaths） | Contributing risk factors（individual PAF for disease burden） | Joint PAF ${ }^{\text {b }}$ <br> （disease burden） | Joint PAF ${ }^{\text {b }}$ <br> （mortality） |

60－62\％
29\％
12\％
50－51\％
49－52\％
$25 \%$
$47-49 \%$
56－59\％
28\％
10\％
49－50\％
35－36\％
సे ஃ우
（a）World（continued）
Underweight（childhood）（45\％）；vitamin A deficiency（16\％）；zinc deficiency（I8\％）
Alcohol（20\％）；illicit drugs（2\％）；occupational risk factors for injuries
（6\％）
Tobacco（10\％）${ }^{\text {c }}$
Tobacco（10\％）
$32.6 \quad$ Multiple risks（see chapter 26）
$\begin{array}{ll}58.3 & \text { Multiple risks（see chapter 26）} \\ 9.1 & \text { Multiple risks（see chapter 26）} \\ & \end{array}$
$\begin{array}{ll}58.3 & \text { Multiple risks（see chapter 26）} \\ 9.1 & \text { Multiple risks（see chapter 26）} \\ & \end{array}$

\％global disease \％global mortality
billion DALYs）deaths）
2.0
でて
$2.9 \quad$ Tobacco（10\％）${ }^{\text {c }}$
$32.6 \quad$ Multiple risks（see chapter 26）
$\begin{array}{ll}58.3 & \text { Multiple risks（see chapter 26）} \\ 9.1 & \text { Multiple risks（see chapter 26）} \\ & \end{array}$
$\begin{array}{ll}58.3 & \text { Multiple risks（see chapter 26）} \\ 9.1 & \text { Multiple risks（see chapter 26）} \\ & \end{array}$
으
Disease／condition
2.9
2.6
2.5
42.0
45.7
12.3
100

て
N

| (b) High-mortality developing subregions |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Disease/condition | \% regional disease burden (total 830 million DALYs) | \% regional mortality (total 26.4 million deaths) | Contributing risk factors (individual PAF for disease burden) | Joint PAF ${ }^{b}$ <br> (disease burden) | Joint PAF ${ }^{\text {b }}$ <br> (mortality) |
| HIV/AIDS | 9.0 | 9.2 | Unsafe sex (97\%); unsafe health care injections (5\%); illicit drugs (0.3\%) | 97\% | 97\% |
| Lower respiratory infections | 8.2 | 9.8 | Underweight (childhood) (46\%); zinc deficiency (19\%); indoor smoke from solid fuels (4I\%); tobacco (I\%) ${ }^{\text {c }}$ | 62-69\% | 49-54\% |
| Diarrhoeal diseases | 6.3 | 6.6 | Underweight (childhood) (49\%); vitamin A deficiency (19\%); zinc deficiency (11\%); unsafe water, sanitation and hygiene (88\%) | 93-95\% | 93-94\% |
| Low birth weight | 5.0 | 4.4 | Underweight (maternal) (I2\%); iron deficiency (22\%); alcohol (0.2\%) | 32\% | 34\% |
| Malaria | 4.9 | 4.2 | Underweight (childhood) (45\%); vitamin A deficiency (17\%); zinc deficiency (19\%) | 57-60\% | 60-63\% |
| Unipolar depressive disorders | 3.1 | 0.0 | Alcohol (1\%); childhood sexual abuse (8\%) | 9\% | $N A^{\text {d }}$ |
| Measles | 3.0 | 2.7 | Underweight (childhood) (34\%); vitamin A deficiency (15\%) | 42\% | 43\% |
| Ischaemic heart disease | 3.0 | 9.1 | High blood pressure (44\%); high cholesterol (54\%); high BMI (II\%); low fruit and vegetable intake (33\%); physical inactivity (2I\%); tobacco (8\%); alcohol (4\%) | 80-87\% | 77-84\% |
| Tuberculosis | 2.9 | 3.8 | Tobacco (8\%) ${ }^{\text {c }}$ | 8\% | 10\% |
| Birth asphyxia and birth trauma | 2.7 | 1.9 | Iron deficiency (20\%) | 20\% | 27\% |

Table 27.I
ndividual and joint contributions of 20 selected risks to 10 leading diseases and total burden of disease in the world and different subregional groupings ${ }^{\text {a }}$ (continued)

| (b) High-mortality developing subregions (continued) |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Disease/condition | \% regional disease burden (total 830 million DALYs) | \% regional mortality (total 26.4 million deaths) | Contributing risk factors (individual PAF for disease burden) | Joint PAF ${ }^{\text {b }}$ (disease burden) | Joint PAF ${ }^{\text {b }}$ <br> (mortality) |
| Communicable, maternal, perinatal, and nutritional conditions | 58.9 | 54.5 | Multiple risks (see chapter 26) | 54-56\% | 56-57\% |
| Noncommunicable diseases | 30.5 | 37.0 | Multiple risks (see chapter 26) | 33-34\% | 47-50\% |
| Injuries | 10.6 | 8.4 | Multiple risks (see chapter 26) | 17\% | 19\% |
| All causes | 100 | 100 | All 20 selected risks (see chapter 26 ) | 44-45\% | 50-51\% |


| (c) Low-mortality developing subregions |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Disease/condition | \% regional disease burden (total 408 million DALYs) | \% regional mortality (total 16.0 million deaths) | Contributing risk factors (individual PAF for disease burden) | Joint PAF ${ }^{\text {b }}$ <br> (disease burden) | Joint PAF ${ }^{D}$ <br> (mortality) |
| Unipolar depressive disorders | 5.9 | 0.0 | Alcohol (1\%); childhood sexual abuse (4\%) | 5\% | $N A^{\text {d }}$ |
| Stroke | 4.7 | 13.8 | High blood pressure (58\%); high cholesterol (13\%); high BMI (II\%); low fruit and vegetable intake (10\%); physical inactivity (5\%); tobacco (8\%); alcohol (7\%) | 67-74\% | 62-69\% |


| Lower respiratory infections | 4.1 | 4.6 | Underweight (childhood) (24\%); zinc deficiency (5\%); indoor smoke from solid fuels (20\%); tobacco (3\%) ${ }^{\text {c }}$ | 35-42\% | 25-29\% |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Road traffic accidents | 4.1 | 3.4 | Alcohol (20\%); illicit drugs (I\%); occupational risk factors for injuries (6\%) | 27\% | 27\% |
| Chronic obstructive pulmonary disease | 3.8 | 9.2 | Indoor smoke from solid fuels (26\%); tobacco (26\%) | 52\% | 55\% |
| Ischaemic heart disease | 3.2 | 9.3 | High blood pressure (45\%); high cholesterol (48\%); high BMI (22\%); low fruit and vegetable intake (31\%); physical inactivity (22\%); tobacco (8\%); alcohol (3\%) | 79-87\% | 73-82\% |
| Birth asphyxia and birth trauma | 2.6 | 1.1 | Iron deficiency (10\%) | 10\% | 17\% |
| Tuberculosis | 2.4 | 3.3 | Tobacco (10\%) | 12\% | 13\% |
| Alcohol use disorders | 2.3 | 0.2 | Alcohol (100\%); childhood sexual abuse (5\%) | 100\% | 100\% |
| Hearing loss | 2.2 | 0.0 | Tobacco (5\%) ${ }^{\text {c }}$ | 5\% | $N A^{\text {d }}$ |
| Communicable, maternal, perinatal, and nutritional conditions | 25.0 | 18.1 | Multiple risks (see chapter 26) | 28-29\% | 27-28\% |
| Noncommunicable diseases | 59.8 | 70.5 | Multiple risks (see chapter 26) | 33-34\% | 46-48\% |
| Injuries | 15.3 | 11.5 | Multiple risks (see chapter 26) | 24\% | 25\% |
| All causes | 100 | 100 | All 20 selected risks (see chapter 26 ) | 30-31\% | 40-42\% |

Table 27.I Individual and joint contributions of 20 selected risks to 10 leading diseases and total burden of disease in the world and different subregional groupings ${ }^{\text {a }}$ (continued)

| (d) Developed subregions |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Disease/condition | \% regional disease burden (total 214 million DALYs) | \% regional mortality (total 13.5 million deaths) | Contributing risk factors (individual PAF for disease burden) | Joint PAF ${ }^{\text {b }}$ <br> (disease burden) | Joint PAF ${ }^{\text {b }}$ <br> (mortality) |

- (mortaly
$87 \%$

$N A^{d}$
$71-79 \%$
$100 \%$
$N A^{d}$
$74 \%$
$71 \%$
$00 \%$
$0 \%$
$10 \%$ .

| Road traffic accidents |  | 2.5 | 1.4 | Alcohol (38\%); illicit drugs (4\%); occupational risk factors for injuries (4\%) | 45\% | 44\% |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Osteoarthritis |  | 2.5 | 0.0 | High BMI (21\%); Tobacco (10\%) ${ }^{\text {c }}$ | 28\% | $N A^{\text {d }}$ |
| Trachea, bronchus and lung cancers |  | 2.4 | 4.5 | Indoor smoke from solid fuels (coal only) (0\%); tobacco (85\%); low fruit and vegetable intake (11\%) | 86\% | 87\% |
| Communicable, maternal, perinatal, and nutritional conditions |  | 9.0 | 6.7 | Multiple risks (see chapter 26) | 24-25\% | 20-21\% |
| Noncommunicable diseases |  | 78.2 | 85.7 | Multiple risks (see chapter 26) | 41-42\% | 54-57\% |
| Injuries |  | 12.8 | 7.6 | Multiple risks (see chapter 26) | 36\% | 37\% |
| All causes |  | 100 | 100 | All 20 selected risks (see chapter 26) | 39-40\% | 51-53\% |
| NA Not applicable. |  |  |  |  |  |  |
| a The risk factors also contribute to other diseases in each subregion which are not among the leading 10. |  |  |  |  |  |  |
|  | The first number is the PAF for adjusted scenario and the second for the unadjusted scenario in cases where adjustment for mediated effects and methods). |  |  |  |  |  |
|  | Affected by tobacco in the category "other respiratory diseases" or "selected other medical causes" (Peto et al. 1992). The PAF has large uncertainty. |  |  |  |  |  |
|  | The number of deaths coded to "hearing loss", "unipolar depressive disorders", "osteoarthritis", and "alzheimer and other dementias" is zero or making the mortality PAF for these diseases undefined or unstable. |  |  |  |  |  |

Figure 27.3 Individual and joint contributions (adjusted scenario as described in methods) of selected risk factors to different disease groups
(a) High-mortality developing subregions (38\% of global population)

(c) Developed subregions ( $22 \%$ of global population)

$\square$ Total group I 19M DALYs
$\square$ Unsafe sex 6\%
$\square$ Underweight 4\%
$\square$ Unsafe WSH 4\%
$\square$ Indoor smoke 2\%
Joint PAF 14\%

$\square$ Total group II 168M DALYs

- High BP 14\%
$\square$ High cholesterol 10\%
$\square$ Tobacco 15\%
Alcohol non-vascular 7\% Joint PAF 34\%

$\square$ Total group III 27M DALYs - Occupational injury risks 4\% $\square$ Alcohol 29\%
$\square$ Illicit drugs 3\%
Joint PAF 36\%


## Notes for Figure 27.3

Key: High-mortality developing subregions: AFR, AMR-D, EMR-D and SEAR-D. Low-mortality developing subregions: AMR-B, EMR-B, SEAR-B and WPR-B. Developed subregions: AMR-A, EUR and WPR-A. Group I: communicable, maternal, perinatal and nutritional conditions; Group II: noncommunicable diseases; Group III: injuries; WSH, water, sanitation and hygiene; BP, blood pressure.
Note: The size of each circle shows the absolute size of the burden (in millions of DALYs). Numbers for individual risk factors show the total burden including those overlapping with the remaining factors shown in lined pattern. Note that each risk factor may also have contributions to other disease groups (e.g. indoor smoke also causes COPD, which is in Group II and alcohol also causes injuries, which are in Group III). In reality, there is a small overlap between underweight and unsafe sex since underweight children with HIV/AIDS are likely to survive for a shorter period (not estimated in this work).
II. noncommunicable diseases; III. injuries. Communicable, maternal, perinatal and nutritional conditions, and their underlying risk factors, in high-mortality developing subregions contributed disproportionately to global loss of healthy life (i.e. large Group I disease burden and large fraction (54-56\%) attributable to the selected risk factors). Noncommunicable diseases and their risk factors (54-57\% attributable to the selected risk factors) dominated the burden of disease in developed subregions, although, by comparison, this was considerably smaller than total disease burden in high-mortality developing subregions. Both disease groups and their risk factors, with intermediate levels, affected low-mortality developing subregions.

Table 27.1 and Figure 27.3 also show that the selected risk factors for each disease group exhibited heterogeneous contributions to disease burden across clusters of countries and subregions. For Group I diseases, unsafe sex had the lowest proportional contribution in low-mortality developing countries but made a very large contribution in highmortality developing countries. For Group II diseases, the relative contributions of high cholesterol and tobacco varied across the three subregional groupings, with their relative contributions reversing from one to another. Similarly, for Group III, the relative contributions of alcohol and occupational factors showed considerable heterogeneity across subregional groups.

Gains in HALE from removing these 20 selected risk factors are shown in Figure 27.4. Globally, in the year 2000, an estimated $47 \%$ of mortality and $39 \%$ of disease burden were attributable to the joint effects of 20 selected risk factors. Global HALE would increase from 56.2 to 65.5 years in the absence of these risks (adjusted scenario). The corresponding results for the unadjusted scenario were nearly identical with HALE increasing to 66.0 years.

Figure 27.4 shows that the removal of major risk factors would not only have resulted in improvements in each subregion, but also, in general, reduced the health differentials across subregions (i.e. larger gains in subregions with lower HALE) with the largest gain in health in

Figure 27.4 The joint effects of leading 20 global risk factors on HALE in year 2000, by subregion (adjusted scenario)


Note: The numbers show the fraction of the total burden of disease in each subregion attributable to the selected risk factors.

AFR-E (16.1 years) and the smallest in WPR-A (4.4 years). Important exceptions to the monotonic decreasing relationship between HALE gain and initial HALE were EUR-C and EUR-B (mainly consisting of the countries of eastern and central Europe and the former Soviet Union). In these subregions, the leading global risk factors jointly account for a disproportionately larger share of disease burden (49\% in EUR-C and $37 \%$ in EUR-B) and led to substantial loss of healthy life years (10.7 in EUR-C and 8.3 in EUR-B), emphasizing the concentration of disease burden among a few important risk factors (alcohol, tobacco, high blood pressure and high cholesterol) in these two subregions.

## 4. Discussion

The estimates of the joint contributions of 20 selected leading global risk factors showed that these risks together were responsible for a considerable loss of healthy life in different regions of the world. In particular, for some of the leading global diseases (e.g. ALRI, diarrhoea, lung cancer, IHD and stroke), substantial proportions were attributable to these selected risk factors. Removing these 20 risk factors would not only have resulted in a 9.3 -year ( $17 \%$ ) gain in global HALE, but also would have
accounted for some of the interregional HALE differences. In fact, the analysis showed that even populations with currently high HALE (e.g. developed regions of the western Pacific and Europe) could further benefit from risk reduction. These results provide a guide to the potential gains in (healthy) life expectancy (estimated statistically from past trends [Oeppen and Vaupel 2002; Riley 2001]) through disease prevention by reducing known risks. Similar analyses for the leading 10 selected global risks suggest a gain of 8.1 years in HALE (vs 9.3 years for the leading 20). This concentration of disease burden further emphasizes the contribution of leading risks such as undernutrition, unsafe sex, high blood pressure, tobacco and alcohol to global loss of healthy life.

At the same time, the estimated joint contributions of these risk factors left an important part of the global disease burden unexplained and did not fully explain interregional HALE differentials. This was because only a small fraction of some important diseases was attributable to the selected risk factors considered here. These include diseases whose determinants: i) are diffuse among environmental and behavioural factors (e.g. some cancers, perinatal conditions, and neuropsychiatric diseases) (see Doll and Peto 1981 for examples from cancers); ii) have more complex, multi-factor etiology and often heterogeneous determinants in different populations and therefore difficult to quantify without data at very small scale (e.g. tuberculosis and injuries); iii) involve long delays; or iv) have limited quantitative research at the population level (e.g. neuropsychiatric diseases), often as a result of the above three factors as well as difficulties in measuring exposure or outcome (Evans 1978). Mitigation of many such diseases (e.g. malaria, tuberculosis or injuries) may be better guided by analyses of the effects of interventions tailored to individual settings than by risk factor analysis.

The results of this analysis changed little with plausible assumptions about mediated risks or effect modification among risk factors. An important reason for this is the concave shape of the PAF relationship (Figure 27.2). Because risk factors considered in the analysis individually accounted for large fractions of the diseases affected by them (e.g. diarrhoea and IHD), the joint affects approached $100 \%$ asymptotically limiting the sensitivity of results to assumptions about interaction. We emphasize that this does not include the considerably larger uncertainty in each of the individual PAF estimates discussed in detail in chapters 1 and 26 of this book. At the same time, since for many of the important causes of global disease burden (e.g. childhood infectious and vascular diseases), multiple important risk factors were included, the joint effects would likely remain large regardless of uncertainties in the individual PAF.

An additional important source of uncertainty, affecting both individual and joint risk factor estimates, is the concentration of both risks and diseases in specific subgroups (vs correlations of risks alone, discussed above). For many risk factors and diseases, exposure and outcome
are simultaneously higher in some groups (e.g. higher malnutrition, unsafe water, sanitation and hygiene, and indoor smoke in poor rural households in developing countries; unhealthy diet, and higher smoking and BMI in some groups in developed countries). In these circumstances, PAFs based on population averages would in general underestimate the effects compared to group-specific analysis, even if the relative risks are constant across groups (Greenland 1984) (also confirmed by microsimulation). Higher concentration of disease and mortality (e.g. childhood mortality and vascular diseases) in the same groups due to factors such as limited access to health services would magnify this effect, becoming an important contributor to underestimation of the benefits of risk reduction when population level exposure and mortality data are used. In addition to risk factor analysis, estimates of HALE include large uncertainty, especially in countries with poor mortality and disease registration systems as estimated and discussed elsewhere (Mathers et al. 2001). Further implications of these findings for research, and for policies and programmes aimed at improving population health, are discussed in chapter 29.

## Acknowledgements

We thank T. Armstrong, R.E. Black, F. Bull, G. Colditz (with E. Rimm and M. Stampfer), C. Lawes, K. Lock, V. Parag, J. Powles, A.J. Rice, K.P. West Jr., G. Whitlock, W. Willet and M. Woodward for discussion and references on independent and mediated effects.

## Notes

1 See preface for an explanation of this term.
2 High-mortality developing subregions: AFR, AMR-D, EMR-D and SEAR-D. Low-mortality developing subregions: AMR-B, EMR-B, SEAR-B and WPRB. Developed subregions: AMR-A, EUR and WPR-A.

## References

Berlin JA, Colditz GA (1990) A meta-analysis of physical activity in the prevention of coronary heart disease. American Journal of Epidemiology, 132:612-628.

Black RE (1991) Would control of childhood infectious diseases reduce malnutrition? Acta Paediatrica Scandinavica Supplement, 374:133-140.

Blair SN, Cheng Y, Holder JS (2001) Is physical activity or physical fitness more important in defining health benefits? Medicine and Science in Sports and Exercise, 33:S379-399.

Briend A (1990) Is diarrhoea a major cause of malnutrition among the underfives in developing countries? A review of available evidence. European Journal of Clinical Nutrition, 44:611-628.

Brown KH, Peerson JM, Rivera J, Allen LH (2002) Effect of supplemental zinc on the growth and serum zinc concentrations of prepubertal children: a meta-analysis of randomized controlled trials. American Journal of Clinical Nutrition, 75:1062-1071.
Christian P, West Jr KP (1998) Interactions between zinc and vitamin A: an update. American Journal of Clinical Nutrition, 68:5435-441.
Doll R, Peto R (1981) The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. Journal of the National Cancer Institute, 66:191-308.
Eastern Stroke and Coronary Heart Disease Collaborative Research Group (1998) Blood pressure, cholesterol, and stroke in Eastern Asia. The Lancet, 352:1801-1807.

Eaton CB (1992) Relation of physical activity and cardiovascular fitness to coronary heart disease, Part I: a meta-analysis of the independent relation of physical activity and coronary heart disease. Journal of the American Board of Family Practice, 5:31-42.
Eide GE, Heuch I (2001) Attributable fractions: fundamental concepts and their visualization. Statistical Methods in Medical Research, 10:159-193.
Evans AS (1978) Causation and disease: a chronological journey. American Journal of Epidemiology, 108:249-258.
Gaziano JM, Manson JE, Branch LG, Colditz GA, Willett WC, Buring JE (1995) A prospective study of consumption of carotenoids in fruits and vegetables and decreased cardiovascular mortality in the elderly. Annals of Epidemiology, 5:255-260.

Greenland S (1984) Bias in methods for deriving standardized morbidity ratio and attributable fraction estimates. Statistics in Medicine, 3:131-141.

Greenland S (1987) Quantitative methods in the review of epidemiologic literature. Epidemiologic Reviews, 9:1-30.
Guerrant RL, Schorling JB, McAuliffe JF, de Souza MA (1992) Diarrhea as a cause and an effect of malnutrition: diarrhea prevents catch-up growth and malnutrition increases diarrhea frequency and duration. American Journal of Tropical Medicine and Hygiene, 47:28-35.
Hirayama T (1990) Life-style and mortality: a large-scale census-based cohort study in Japan. Karger, Tokyo.
Jarrett RJ, Shipley MJ, Rose G (1982) Weight and mortality in the Whitehall study. British Medical Journal, 285:535-537.
Jee SH, Suh I, Kim IS, Appel LJ (1999) Smoking and atherosclerotic cardiovascular disease in men with low levels of serum cholesterol: the Korea Medical Insurance Corporation study. Journal of the American Medical Association, 282:2149-2155.

Jousilahti P, Vartiainen E, Tuomilehto J, Puska P (1999) Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14786 middle-aged men and women in Finland. Circulation, 99:1165-1172.

Khaw KT, Barrett-Connor E (1987) Dietary fiber and reduced ischemic heart disease mortality rates in men and women: a 12 -year prospective study. American Journal of Epidemiology, 126:1093-1102.
Koopman JS (1981) Interaction between discrete causes. American Journal of Epidemiology 113:716-724.
Kruijshaar ME, Barendregt JJ, Hoeymans N (2002) The use of models in the estimation of disease epidemiology. Bulletin of the World Health Organization, 80:622-628.
Liu BQ, Peto R, Chen ZM et al. (1998) Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. British Medical Journal, 317:1411-1422.

Liu S, Lee IM, Ajani U, Cole SR, Buring JE, Manson JE (2001) Intake of vegetables rich in carotenoids and risk of coronary heart disease in men: the Physicians' Health Study. International Journal of Epidemiology, 30: 130-135.

Liu S, Manson JE, Lee IM, Cole SR, Hennekens CH, Willett WC, Buring JE (2000) Fruit and vegetable intake and risk of cardiovascular disease: the Women's Health Study. American Journal of Clinical Nutrition, 72: 922-928.

Lutter CK, Habicht JP, Rivera JA, Martorell R (1992) The relationship between energy intake and diarrhoeal disease in their effects on child growth: biological model, evidence, and implications for public health policy. Food and Nutrition Bulletin, 14:36-42.

Lutter CK, Mora JO, Habicht JP et al. (1989) Nutritional supplementation: effects on child stunting because of diarrhea. American Journal of Clinical Nutrition, 50:1-8.
Manson JE, Colditz GA, Stampfer MJ et al. (1990) A prospective study of obesity and risk of coronary heart disease in women. New England Journal of Medicine, 322:882-889.
Manson JE, Greenland P, LaCroix AZ et al. (2002) Walking compared with vigorous exercise for the prevention of cardiovascular events in women. New England Journal of Medicine, 347:755-756.
Manton KG, Singer BH, Suzman RM (1993) Forecasting the health of elderly populations. Springer-Verlag, New York.

Martorell R, Habicht JP, Yarbrough C, Lechtig A, Klein RE, Western KA (1975a) Acute morbidity and physical growth in rural Guatemalan children. American Journal of Diseases of Children, 129:1296-1301.
Martorell R, Yarbrough C, Lechtig A, Habicht JP, Klein RE (1975b) Diarrheal diseases and growth retardation in preschool Guatemalan children. American Journal of Physical Anthropology, 43:341-346.
Mathers CD (2002) Health expectancies: an overview and critical appraisal. In: Summary measures of population health: concepts, ethics, measurement and applications. Murray CJL, Salomon JA, Mathers CD, Lopez AD, eds. World Health Organization, Geneva.

Mathers CD, Sadana R, Salomon JA, Murray CJL, Lopez AD (2001) Healthy life expectancy in 191 countries, 1999. The Lancet, 357:1685-1691.
Miettinen OS (1974) Proportion of disease caused or prevented by a given exposure, trait or intervention. American Journal of Epidemiology, 99:325-332.

Murray CJL, Lopez AD, eds. (1996) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990. The Global Burden of Disease and Injury, Vol. 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.
Murray CJL, Lopez AD (1999) On the comparable quantification of health risks: lessons from the Global Burden of Disease. Epidemiology, 10:594-605.
Neaton JD, Wentworth D (1992) Serum cholesterol, blood pressure, cigarette smoking, and death from coronary heart disease. Overall findings and differences by age for 316099 white men. Multiple risk factor intervention Trial Research Group. Archives of Internal Medicine, 152:56-64.

Oeppen J, Vaupel JW (2002) Broken limits to life expectancy. Science, 296: 1029-1030.

Pelletier DL, Frongillo EA Jr, Habicht JP (1993) Epidemiologic evidence for a potentiating effect of malnutrition on child mortality. American Journal of Public Health, 83:1130-1133.
Peto R, Lopez AD, Boreham J, Thun M, Heath Jr C (1992) Mortality from tobacco in developed countries. The Lancet, 339:1268-1278.
Ramakrishnan U, Latham MC, Abel R (1995) Vitamin A supplementation does not improve growth of preschool children: a randomized, double-blind field trial in South India. Journal of Nutrition, 125:202-211.
Ramakrishnan U, Martorell R (1998) The role of vitamin A in reducing child mortality and morbidity and improving growth. Salud Publica de Mexico, 40:189-198.
Riley JC (2001) Rising life expectancy: a global history. Cambridge University Press, Cambridge.
Robins JM (1999) Marginal structural models versus structural nested models as tools for causal inference. In: Statistical models in epidemiology. Halloran E, ed. Springer Verlag, New York.
Rosengren A, Wedel H, Wilhelmsen L (1999) Body weight and weight gain during adult life in men in relation to coronary heart disease and mortality: a prospective population. European Health Journal, 20:269-277.
Rothman KJ (1976) Causes. American Journal of Epidemiology, 104:587-592.
Rothman KJ, Greenland S (1998) Modern epidemiology. Lippincott-Raven, Philadelphia, PA.
Rothman KJ, Greenland S, Walker AM (1980) Concepts of interaction. American Journal of Epidemiology, 112:467-470.

Rothman KJ, Keller A (1972) The effect of joint exposure to alcohol and tobacco on the risk of cancer of the mouth and pharynx. Journal of Chronic Disease, 25:711-716.

Scrimshaw NS, Taylor CE, Gordon JE (1968) Interactions of nutrition and infection. (WHO Monograph Series No. 57.) World Health Organization, Geneva.
Smith KR, Corvalan CF, Kjellstrom T (1999) How much global ill health is attributable to environmental factors. Epidemiology, 10:573-584.

Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC (2000) Primary prevention of coronary heart disease in women through diet and lifestyle. New England Journal of Medicine, 343:16-22.
Stephensen CB (1999) Burden of infection on growth failure. Journal of Nutrition, 129:S534-538.
Tate RB, Manfreda J, Cuddy TE (1998) The effect of age on risk factors for ischemic heart disease: the Manitoba Follow-up Study, 1948-1993. Annals of Epidemiology, 8:415-421.
UNICEF (1990) Strategy to improve nutrition of children and women in developing countries: a UNICEF policy review. United Nations Children's Fund, New York.

Walter SD (1976) The estimation and interpretation of attributable risk in health research. Biometrics, 32:829-849.

Walter SD (1980) Prevention for multifactorial diseases. American Journal of Epidemiology, 112:409-416.
West KP Jr, Pokhrel RP, Katz J et al. (1991) Efficacy of vitamin A in reducing preschool child mortality in Nepal. The Lancet, 338:67-71.
Willet WC (2002) Balancing life-style and genomics research for disease prevention. Science, 296:695-698.
Yerushalmy J, Palmer CE (1959) On the methodology of investigations of etiologic factors in chronic diseases. Journal of Chronic Disease, 108:27-40.
Zinc Investigators' Collaborative Group (1999) Prevention of diarrhea and pneumonia by zinc supplementation in children in developing countries: pooled analysis of randomized controlled trials. Journal of Pediatrics, 135:689-697.

## Chapter 28

# Effects of multiple interventions 

James Robins, Miguel Hernan and Uwe Siebert

## 1. Introduction

The purpose of this chapter is (i) to describe some currently available analytical methods for using individual level epidemiological data to estimate the impact of multiple risk factor interventions on health and (ii) to carefully review the conditions under which these methods deliver unbiased estimates of impact. The chapter is organized as follows. In sections 2 and 3, we discuss estimation of effects of short-term, timeindependent interventions. Specifically, we discuss estimating the effect of a single risk factor intervention on life expectancy or quality-adjusted life expectancy over a specified period of follow-up in a single population, when essentially ideal epidemiological data are available. That is, we assume a random sample of the population is randomly assigned to different levels of the risk factor and followed prospectively for a fixed time period. Second, we consider the same study design, except now we are interested in the joint effect of interventions on several risk factors. Third, we consider the problem of extrapolation of the results to longer periods of follow-up and to other populations for which no primary epidemiological data are available. Sections 2 and 3 serve to indicate the possibilities and limitations of even ideal epidemiological data for estimating the effects of multiple time-independent risk factor interventions. In sections 4 and 5 we turn to the main topic of this chapter: the estimation of the effect of multiple time-dependent interventions from observational data, possibly plagued by confounding, selection bias, measurement error, information bias and ill-defined interventions. In sections 6 to 8 , we illustrate our methods by attempting to estimate the effects of various time-varying interventions on subjects entered in the Framingham Offspring cohort study. Finally, in section 9 we offer some conclusions.

## 2. Time-independent interventions

### 2.1 An ideal intervention

We suppose we have data from a randomized trial of the effect of a onetime short-term intervention (say, an anti-smoking intervention involving one week of intense study of literature on the health consequences of smoking) initiated at calendar time, say 1983. We suppose the subjects in the trial constitute a random sample of the population of a given country, say the United States of America, and these subjects are followed for twenty years. Let $S_{0}(t)$ and $S_{1}(t)$ be the survival curves of the exposed and unexposed respectively at $t$ years from the beginning of the trial for $t \leq 20$ years. Thus $S_{0}(t)$ and $S_{1}(t)$ are both equal to one at time $t=0$ and decrease as $t$ increases. Suppose there is no sharing of information in the sense subjects who receive the anti-smoking intervention do not pass on (i.e. infect or contaminate) others in the trial or in the general population with their acquired knowledge. Then it is well known that the area between the survival curves is equal to the expected years of life saved over the first twenty years of the trial due to the intervention. If we are interested in quality-adjusted years of life lost, we use subject-yearspecific health and interview data to associate with each year a subject lives a quality measure (taking values between 0 and 1 where 1 indicates optimal quality) and compare expected quality-adjusted years of life lost.

### 2.2 MUlTiple ideal interventions

Turn now to estimating the effect of simultaneous interventions on $k$ time-independent risk factors $A=\left(A_{1}, \ldots, A_{k}\right)$, such as smoking, alcohol, blood pressure, cholesterol, etc., where for the moment we assume all interventions were randomly assigned at the same moment in time and only once. Here $A$ is the $k$-vector of all risk factors. Each risk factor $A_{m}$ has some number $\left|A_{m}\right|$ of possible levels $a_{m 1}, \ldots, a_{m\left|A_{m}\right|}$. Then $A$ has $|A|=\left|A_{1}\right| \times\left|A_{2}\right| \times \ldots \times\left|A_{k}\right|$ possible joint levels. Let the set $\mathcal{A}=$ $\{a\}$ of size $|A|$ denote the set of possible values $a$ of the vector $A$. Let $S_{a}(t)$ be the survival curve for the random sample of subjects assigned to joint level $a$ of the various risk factors. Let $S(t)$ denote the survival curve for a random subset of the population on which no interventions were made. Then the expected years of life gained by intervening and setting each subject's level of the $k$ risk factors to $a$ compared to no intervention is precisely the area between the survival curves $S_{a}(t)$ and $S(t)$. The optimal intervention $a^{*}$ is the value of $a$ for which the area under $S_{a}(t)$ is largest. Further, the loss in life expectancy under intervention a compared to the optimal is simply the difference in area under $S_{a^{*}}(t)$ and $S_{a}(t)$. In principle, we need make no assumptions concerning necessary and sufficient causes, multiplicative or additive interactions, or the fraction of deaths attributable to any particular cause to order the health benefits of various joint interventions on life expectancy from such ideal
epidemiological data. All we need is a way to accurately estimate $S_{a}(t)$ for each joint intervention $a$ in $\mathcal{A}$. Due to random sampling variability, in order to obtain reliable estimates of each of the $|A|$ curves $S_{a}(t)$ would require an inconceivably large randomized trial if $|A|$ was at all large, since a large number of subjects would need to be randomly assigned to each of the $|A|$ possible levels of $A$. In practice such large trials are infeasible. As a consequence, we can randomize subjects to only a subset of the possible interventions. In that case we would need to make modelling assumptions as to the nature of the interactions between the risk factors on survival (e.g. by assuming no interaction between risk factors on the mortality rate on a particular scale such as multiplicative or additive) both in order to obtain estimates of the $S_{a}(t)$ that are not too variable and to extrapolate to values of $a$ outside the range of the data (i.e. to values of $a$ to which no one was randomized). In the final analysis it is the area under the curves $S_{a}(t)$ that remains of interest. If our models are misspecified (e.g. we assume no interaction on an additive scale when in fact such interaction is present), the resulting estimates of $S_{a}(t)$ will be biased. Thus we would like to avoid use of models as much as possible. However the use of models cannot be done away with because of our inability to conduct a sufficiently large study.

In the randomized experiment of the previous paragraph, we can also estimate the effects of conditional (or dynamic) interventions. Let $L=$ $\left(L^{(1)}, \ldots L^{(p)}\right)$ denote a $p$-vector of measured baseline (pretreatment) covariates such as age, sex and measures of baseline health status. Let $d(l)$ be a function that assigns to each value of the vector $L$ a value of $a$ in the set $\mathcal{A}$ of possible joint risk factor interventions. If a regime $d$ assigns the same value $a$ to each $L$, we refer to the regime $d$ as nondynamic. Otherwise, we refer to $d$ as a conditional or dynamic treatment regime, strategy or plan as it individualizes the treatment (i.e. joint risk factor intervention) a subject receives based on the subject's value of $L$. A wise choice of $d$ should allow us to optimize therapy for individuals and thus should be a better strategy than even the optimal non-dynamic intervention $a^{*}$ discussed above. Let $S_{d}(t)$ be the survival curve that would result if we randomly assigned individuals to plan $d$. For subjects with a given value $l$ of $L$, the conditional survival curve $S_{d}(t \mid L=l)$ given $L=l$ under regime $d$ equals $S_{a}(t \mid L=l)$ for the value $a=d(l)$ that they receive under the plan. Thus for the population as a whole $S_{d}(t)=\Sigma_{l} S_{a}(t \mid L=l) \operatorname{pr}(L=l)$ is weighted average of $S_{a}(t \mid L=l)$ with $a=d(l)$ and weights proportional to the fraction $\operatorname{pr}(L=l)$ of the population with $L=l$. Thus the survival curve $S_{d}(t)$ of a dynamic regime can be estimated from the data in the randomized trial wherein each subject is randomized to a non-dynamic regime. Define $d_{o p}(l)$ to be the treatment plan that minimizes the area under $S_{d}(t)$ over all possible dynamic and nondynamic treatment plans $d$. The loss in life expectancy under intervention $d$ compared to the optimal is the difference in area under $\mathrm{S}_{d_{o p}}(t)$ and $S_{d}(t)$. In an ideal world, we would estimate $d_{o p}(l)$ from a large ideal
randomized study and, after analysing the trial, treat a new individual with $L=l$ with treatments $d_{o p}(l)$.

## 3. Some limits of ideal data

In this section, we return to the simple randomized trial of section 2.1.

### 3.1 Extrapolation

Although we have a precise estimate of the effect of this intervention on twenty year mortality of United States citizens in the calendar period 1983-2003, the trial provides no direct evidence concerning the following more relevant policy questions. What would be the continuing effect of this intervention on the United States population through 2013 or 2023? What would be the effect of this same intervention, if it began now in 2003 rather than in 1983? What would be the effect of a similar intervention on a population that differs from the United States population on both measured and unmeasured determinants of mortality including smoking, age, cholesterol, high blood pressure, lifestyle pattern, access to state-of-the-art health care, etc? Obviously any of these questions can only be addressed by assuming a model that extrapolates beyond the observed data. A common simple approach to extrapolation is to first statistically test from the available data whether the relative risk (equivalently, mortality ratio, hazard ratio, rate ratio) in the exposed compared to the non-exposed remains nearly constant both over the 20 years of follow-up and within levels of measured covariates such as age, ethnicity, socioeconomic status and smoking. If the test accepts the constant relative risk hypothesis then we extrapolate by assuming the same will be true if follow-up was continued past 2003, if the intervention was in 2003, and if the intervention was applied to a population with different smoking and risk factor distribution than the United States. In most studies, however, the power to detect deviations from a constant rate ratio is fairly small and there is rarely any strong biological reason to believe that rate ratios rather than other effect measures (such as rate differences) should be constant over time and location. Further, we have no way to test whether the effect of an intervention on a rate ratio scale is the same across groups that differ in unmeasured factors. Finally, even if we assume the relative risk to be constant, nevertheless, to estimate the effect of intervention on life expectancy, we still require a method to estimate covariate-calendar yearspecific baseline rates in various populations in future years, since the effect on life expectancy depends both on the relative risk and on these baseline rates.

### 3.2 Contamination

In the discussion so far, we have assumed that the exposed do not "infect" or "contaminate" the unexposed with their exposure. This
assumption would not hold if subjects given, say, the aforementioned anti-smoking intervention distribute their anti-smoking materials to the control group. In that case, the beneficial effect of the intervention will be underestimated by the difference in the exposed and unexposed survival curves because a number of the unexposed will actually have been exposed. The difficulty is that the trial suffered from noncompliance in the sense that, contrary to our wishes, some of those assigned to no treatment actually received treatment. Certain analytic methods, referred to as instrumental variable methods, can partially adjust for this type of noncompliance if a randomly selected (possibly stratified) subsample of the trial population is interviewed to determine how much treatment they received as measured, say, by the fraction of the anti-smoking articles provided to the treatment group the controls actually read. This approach to correction for noncompliance is useful when the mechanism by which treatment exerts its effects is directly through access to the anti-smoking materials (Robins and Greenland 1996).

However, there are types of contamination that operate in quite different ways. For example, suppose that an effect of the treatment was that one of the treated subjects became so motivated that she started a public health campaign that resulted in the legal prohibition of smoking in public, leading to additional decreases in the cigarette consumption in both the exposed and unexposed. Then the difference in the treatmentarm specific survival curves would underestimate the total effect of the intervention (which should include the indirect effect through the legal prohibition) and correction by instrumental variable methods would not be possible. In this case a different design would be necessary. For example, one could use a cluster randomization design wherein different cities, counties or states are the experimental units and are randomly assigned as a whole to either treatment or control. The goal of the design is to have the experimental units sufficiently isolated from one another that one can be reasonably certain that between-unit contamination of treatment will not occur. If data from an appropriate cluster randomized design are not available, other different, and less reliable approaches to estimating the effect of the intervention on life expectancy must be used. Examples of such approaches include the following: (i) assume death rates would have remained unchanged had the intervention not occurred; (ii) specify and fit complex stochastic models for the mechanism by which the intervention reduced deaths, say, by assuming any decrease in the population exposure to cigarettes was wholly due to the intervention and modelling the effect on mortality of the observed decrease in smoking based on past or current data on changes in smoking and mortality; and (iii) create the observational equivalent of a cluster randomized design by assuming mortality and/or smoking data from other communities can be used to estimate what would have happened in the study population had no intervention taken place. A well known example of the type of contamination we are considering in this paragraph occurs in random-
ized trials studying the effect of vaccines on infectious disease and is the basis for the so-called phenomenon of "herd immunity" wherein an epidemic can be prevented in the unvaccinated (untreated) by vaccinating a sufficiently large fraction of the entire community.

The considerable difficulties caused by contamination and the need for extrapolation will not be considered further in this chapter due to space limitations and to the fact that the main goal of this chapter lies elsewhere. Manton et al. (1992) describe some models that may be used for extrapolation. We now turn to observational settings in which ideal epidemiologic data with which to estimate the effect of various treatments or risk factors even on the population under study may not be available.

## 4. Observational data and time-Independent and time-dependent interventions

For financial, ethical or logistical reasons, randomized trial evidence concerning the effectiveness of many if not most interventions is lacking and data from observational studies must be utilized. In this section we will use a hypothetical observational study of the effect of antiretroviral therapy on the progression of HIV-related disease as a specific example. It is well understood that causal effects can be estimated from observational studies only when data on all relevant time-independent and timedependent confounding factors have been obtained. What is less well known is that for time-varying treatments, standard approaches to confounder control can be biased, even when the causal null hypothesis of no treatment effect is true and there are no unmeasured confounding factors. Specifically, the standard approach to the estimation of the causal effect of a time-varying treatment on survival has been to model the hazard of failure at $t$ as a function of treatment history with a timedependent proportional hazards model. Robins et al. (1986) have shown that, even in the absence of unmeasured confounding factors or model misspecification, the usual approach may be biased even under the causal null hypothesis, whether or not one further adjusts for the past history of measured covariates in the analysis, when (i) there exists a timedependent risk factor (say CD4 lymphocyte count and/or Pneumocystis carinii pneumonia [PCP] history) for survival that also predicts subsequent treatment, and (ii) past treatment history predicts subsequent risk factor level. Specifically, condition (i) implies that the analysis that does not adjust for covariates is biased due to confounding by time-dependent risk factors such as CD4 count and/or PCP. Condition (ii) implies that the analysis that includes current CD4 count and/or PCP history as a regressor is biased since it adjusts for a variable (CD4 count and/or PCP history) affected by past treatment. In contrast to standard methods, estimation methods based on marginal structural models (MSMs), the parametric g-computation formula, and structural nested models provide
consistent estimates of causal effects whenever unmeasured confounding and model misspecification are absent (see the article, discussion and rejoinder in Robins et al. 1999, for additional details). To describe difficulties with standard approaches and the basis for the effectiveness of these three alternatives, it will be useful to informally introduce causal directed acyclic graphs (DAGs) as discussed by Spirtes et al. (1993), Pearl (1995), and Greenland et al. (1999).

A causal graph is a DAG in which the vertices (nodes) of the graph represent variables, the directed edges (arrows) represent direct causal relations between variables, and there are no directed cycles, since no variable can cause itself. For a DAG to be causal, the variables represented on the graph must include the variables and additional unmeasured variables, such that if any two measured variables on the graph have a cause in common, that common cause is itself included as a variable on the graph.

A variable is a cause of a second variable if there is a directed path from the first variable to the second consisting solely of arrows pointing towards the second variable. As an example consider a follow-up study of AIDS patients. Let $A(t)=A_{t}$ be the dose of the treatment or exposure of interest, say an antitetroviral therapy, at $t$ with time measured in months since start of follow-up. The time $t$ at which a treatment occurs will either be placed in parentheses or subscripted, depending on the context. Let $Y$ be a dichotomous outcome of interest (e.g. $Y=1$ if HIV RNA is not detectable in the blood and is 0 otherwise) measured at end of follow-up at time $K+1$. Our goal is to estimate the causal effect of the time-dependent treatment $A(t)$ on the outcome $Y$.

Figure 28.1 is a causal graph that represents our study with $K=1$. In Figure 28.1, $L(t)=L_{t}$ represents the value at time $t$ of the vector of all measured causal risk factors for the outcome, such as CD4 count, white blood count (WBC), red blood count (RBC), the presence or absence of various opportunistic infections such as PCP, age and weight. We assume that $L(t)$ is temporally prior to $A(t)$ since physicians commonly obtained data recorded in $L(t)$ such as CD4 count before deciding on a treatment $A(t)$ to be given in month $t$. Similarly, $U(t)=U_{t}$ represents the value at time $t$ of all unmeasured causal risk factors for Y. Figure 28.1(b) differs from Figure 28.1(a) only in that the arrows from the unmeasured risk factors into the treatment variables have been removed. When, as in Figure 28.1(b), there are no arrows from unmeasured risk factors directly into treatment variables, we say that there are no unmeasured confounders given data on the measured confounders $L(t)$. Figure 28.1(c) differs from Figures 28.1(a) and 28.1(b) in that none of the causal risk factors for $Y$ (measured or unmeasured) has arrows into any treatment variable. Note, however, that earlier treatment $A(0)$ can causally affect later treatment $A(1)$. When, as in Figure 28.1(c), there are no arrows from any (non-treatment) risk factor into any treatment variable, there

Figure 28.I Causal graphs for a time-dependent exposure
(a)

(b)

(c)

is no confounding by either measured or unmeasured factors, in which case we say that treatment is unconfounded or, equivalently, causally exogenous.

The distinctions drawn above apply equally to more familiar point treatment studies where the treatment is not time-dependent. As indicated in Figure 28.2, a point treatment study is a special case of the general set-up in which $K=0$. Figures 28.2(a)-28.2(c) are the analogues of Figures 28.1(a)-28.1(c) for a point treatment study.

Our causal DAGs would be of no use without an assumption linking the causal structure represented by the DAG to the statistical data obtained in an epidemiological study. Recall that if a set of variables $X$ is statistically independent of (i.e. unassociated with) another set of

Figure 28.2 Causal graphs for a point exposure $A_{0}$
(a)

(b)

(c)

variables $Y$ conditional on a third set of variables $Z$, then within joint strata defined by the variables in $Z$, any variable in $X$ is unassociated with any variable in $Y$. For example, suppose all variables are dichotomous and the set $Z$ consists of the two variables $Z_{1}$ and $Z_{2}$. Then conditional independence implies that the population odds ratio and population risk ratio between any variable in $X$ and any variable in $Y$ is 1 within each of the $4=2^{2}$ strata of $Z:\left(Z_{1}, Z_{2}\right)=(0,0),\left(Z_{1}, Z_{2}\right)=(0$, 1), $\left(Z_{1}, Z_{2}\right)=(1,0)$ and $\left(Z_{1}, Z_{2}\right)=(1,1)$. We use the symbol $\amalg$ to indicate statistical independence, e.g. $X \amalg Y \mid Z$ means $X$ is conditionally independent of $Y$ given $Z$. The following so-called causal Markov assumption links the causal structure of the DAG with the various statistical independencies.

### 4.1 Causal Markov assumption (CMA)

On a causal graph any variable that is not caused by a given variable $V$ will be independent of (i.e. unassociated with) $V$ conditional on (i.e. within joint strata defined by) V's direct causes. Thus, in Figure 28.1(c), $A(t)$ being causally exogenous implies that $A(0) \coprod L(0), U(0)$ and $A(1) \coprod\{L(0), L(1), U(0), U(1)\} \mid A(0)$.

As in any observational study, we cannot determine from the observed data whether there is confounding by unmeasured risk factors. We can only hope that whatever residual confounding there may be due to the $U_{1}<y$ is small. However, as discussed further below, under the untestable assumption that there is no unmeasured confounding given the $L_{k}$, we
can empirically test from the data whether treatment is causally exogenous. Specifically, a sufficient condition for treatment to be unconfounded is that, at each time $k$, among subjects with the same past treatment history $A_{0}, \ldots, A_{k-1}$, the treatment $A_{k}$ is unassociated with (i.e. statistically independent of) the past history of measured covariates $L_{0}, \ldots, L_{k}$. In particular, in our point treatment study, treatment will be causally exogenous if $A_{0}$ is unassociated with $L_{0}$.

### 4.2 Inverse probability of treatment weighted estimation

In this section, our goal is to estimate using MSMs the causal effect of the time-dependent treatment $A(t)$ on the mean of $Y$, which for a dichotomous $(0,1)$ response is just the probability that $Y=1$. In this section, we assume that there is no loss to follow-up so $Y$ is observed on each study subject. For any time-dependent variable we use overbars to denote the history of that variable up to and including $t$. For example, $\bar{L}(t)=$ $[L(0), L(1), L(2), \ldots, L(t)]$ is the covariate process through $t$. Consider first the association (i.e. regression) model that states that the mean of Y, given treatment history $\bar{A} \equiv \bar{A}(K)$, is a linear logistic function of a subject's duration of antiretroviral therapy. That is,

$$
E[Y \mid \bar{A}]=g(\bar{A} ; \gamma)
$$

where

$$
\begin{equation*}
g(\bar{A} ; \gamma)=\frac{\exp \left[\gamma_{0}+\gamma_{1} \operatorname{Dur}(\bar{A})\right]}{1+\exp \left[\gamma_{0}+\gamma_{1} \operatorname{Dur}(\bar{A})\right]} \tag{1}
\end{equation*}
$$

$\operatorname{Dur}(\bar{A})=\sum_{k=0}^{K} A(k)$ is the subject's duration of treatment in months, and $A(k)$ equals 1 if the subject is on treatment in month $k$ and is 0 otherwise. That is, we are assuming a linear logistic regression model

$$
\begin{equation*}
\text { logit } \operatorname{pr}(Y=1 \mid \bar{A})=\gamma_{0}+\gamma_{1} \operatorname{Dur}(\bar{A}) \tag{2}
\end{equation*}
$$

The logistic regression maximum likelihood estimator (MLE) of $\gamma=\left(\gamma_{0}\right.$, $\gamma_{1}$ ) that is computed by all standard packages maximizes $\prod_{i=1}^{n} L i k_{i}(\gamma)$ with $\operatorname{Lik}_{i}(\gamma)=g\left(\bar{A}_{i} ; \gamma\right)^{\gamma_{i}}\left[1-g\left(\bar{A}_{i} ; \gamma\right)\right]^{1-\gamma_{i}}$ being the likelihood contribution for a single subject and $n$ the sample size.

Assuming the association model (1) is correct, when does $\gamma_{1}$ and $\gamma_{2}$ have a causal interpretation? To answer this question, imagine that the decision to administer treatment at each time $t$ was made totally at random by the treating physician. In that hypothetical randomized trial, treatment at time $t$ is not expected to be associated with the history up
to $t$ of any measured or unmeasured prognostic factors (i.e. there is no confounding). In the absence of confounding, association implies causation and we would expect $\gamma_{1}$ to represent the effect of antiretroviral therapy on the mean of Y. More generally, the marginal association between treatment and response represents causation whenever the treatment is causally exogenous, that is, the conditional probability of receiving a treatment $A(t)$ at time $t$ given past treatment and prognostic factor history for $Y$ (measured and unmeasured) depends only on past history of treatment $\bar{A}(t-1)$ as in Figures 28.1(c) and 28.2(c). A more technical definition is provided below after we define counterfactual outcomes. It is well-recognized in the social sciences, econometrics, epidemiological, and biostatistical literature that the treatment parameters of a correctly specified association model will have a causal interpretation if treatment is causally exogenous.

To help assess whether antiretroviral therapy may be causally exogenous, we introduce the concept of "statistical exogeneity". We say that treatment $A(t)$ is a "statistically exogenous or ancillary" process if the probability of receiving treatment at time $t$ does not depend on the history of measured time-dependent prognostic factors up to $t$, conditional on treatment history prior to $t$, i.e.

$$
\bar{L}(t) \mathrm{C} A(t) \mid \bar{A}(t-1)
$$

Note that a nearly necessary condition for $A(t)$ to be "causally exogenous" is for it to be "statistically exogenous". However, that a process is "statistically exogenous" does not imply it is "causally exogenous", because there may be unmeasured prognostic factors $\bar{U}(t)$ for the outcome (i.e. confounders) that predict the probability of treatment $A(t)$ at time $t$ given past treatment history. Thus we can test from the data whether $A(t)$ is statistically exogenous as it is a relationship between observed variables, but are unable to test whether a statistically exogenous process is causally exogenous. However, as mentioned above and discussed further below, if we make the untestable assumption that there are no unmeasured confounders, then statistical exogeneity will imply causal exogeneity.

Suppose $A(t)$ is discrete and we can correctly model both the probability $f[a(t) \mid \bar{l}(t), \bar{a}(t-1)]$ of taking treatment $a(t)$ on day $t$ as a function
of past treatment $\bar{a}(t-1)$ and measured prognostic factor history $\bar{l}(t)$, and the probability $f[a(t) \mid \bar{a}(t-1)]$ of taking treatment $a(t)$ in month $t$ as a function only of past treatment $\bar{a}(t-1)$ history. Here we use the convention that random variables (i.e. variables whose values can differ from subject to subject) are denoted by upper case letters. Lower case letters denote possible values of the corresponding random variables. Thus, for example, $f[a(t) \mid \bar{a}(t-1)]$ is the proportion of subjects in the target population with treatment $A(t)$ equal to $a(t)$ among subjects with past treat-
ment history $\bar{A}(t-1)$ equal to $\bar{a}(t-1)$. We could measure the degree to which the treatment process is statistically non-exogenous through time $t$ by the time $t$-specific random variable

$$
\begin{equation*}
S W(t)=\prod_{k=0}^{t} \frac{f[A(k) \mid \bar{A}(k-1)]}{f[A(k) \mid \bar{A}(k-1), \bar{L}(k)]} \tag{3}
\end{equation*}
$$

where $\prod_{k=0}^{t} z(k)$ is the product $z(0) \times z(1) \times \ldots \times z(t), f[A(k) \| \bar{A}(k-1)$, $\bar{L}(k)]$ is, by definition, the conditional probability mass function $f[a(k) \mid \bar{a}(k-1), \bar{l}(k)]$ with $(a(k), \bar{a}(k-1), \bar{l}(k))$ evaluated at a subject's data $(A(k), \bar{A}(k-1), \bar{L}(k))$ and $f[A(k) \mid \bar{A}(k-1)]$.

For example, if for a given subject $A(k)$ is zero and there are 55 other subjects with the same $\bar{A}(k-1), \bar{L}(k)$ history of whom 25 have $A(k)$ of zero and 70 subjects with the same $\bar{A}(k-1)$ history of whom 32 have $A(k)$ of zero, then $f[A(k) \mid \bar{A}(k-1), \bar{L}(k)]$ is $20 / 55$ and $f[A(k) \mid \bar{A}(k-1)]$ is $32 / 70$ for the subject. Informally, the denominator in each term in $S W(t)$ is the probability that a subject received his own observed treatment, $A(k)$, at time $k$ given his past antiretroviral treatment and measured prognostic factor history. Informally, the numerator is the probability that a subject received his observed treatment conditional on his past antiretroviral treatment history, but not further adjusting for his past prognostic factor history. Note that the numerator and denominator of $S W(t)$ are equal for all $t$ with probability 1 if and only if the treatment process is statistically exogenous, i.e. $\bar{L}(t) \amalg A(t) \mid \bar{A}(t-1)$. In practice $S W(t)$ will have to be estimated from the data but, for pedagogical purposes, assume for now that it is known.

When $A(t)$ is statistically non-exogenous, we shall consider estimating $\gamma$ by a weighted logistic regression in which a subject is given the weight $S W \equiv S W(K)$. Standard software packages for logistic regression will allow the user to specify the subject-specific weight $S W$. The weighted logistic regression estimator, which we will refer to as an inverse-probability-of-treatment-weighted (IPTW) estimator, is the maximizer of $\prod_{i=0}^{n}\left[L i k_{i}(\gamma)\right]^{s W_{i}}$. This weighted logistic regression would agree with the usual unweighted analysis described above just in the case in which $A(t)$ were statistically exogenous. The IPTW estimator is an extension to longitudinal causal inference models of estimators proposed by Horvitz and Thompson (1952).

If the vector of prognostic factors recorded in $\bar{L}(t)$ constitutes all relevant time-dependent prognostic factors (i.e. confounders) so that there are no unmeasured confounders (as in Figures 28.1(b) or 28.2(b)), then,
whether or not the treatment process is statistically exogenous, the weighted logistic regression estimator of $\gamma_{1}$ will converge to a quantity $\beta_{1}$ that can be interpreted as the causal effect of antiretroviral therapy on the mean of $Y$ (on the log odds ratio scale). In contrast, when $A(t)$ is statistically non exogenous, the usual unweighted logistic regression estimator will still converge to $\gamma_{1}$, but now $\gamma_{1}$ will have no causal interpretation. We now give a formal mathematical meaning to the informal concepts of the causal effect of antiretroviral therapy on the mean of $Y$.

### 4.3 COUNTERFACTUALS AND MARGINAL STRUCTURAL MODELS

To formalize our results, we use counterfactual or potential outcomes. Neyman (1923) introduced counterfactual outcomes to analyse the causal effect of time-independent treatments in randomized experiments. Rubin (1978) championed Neyman's idea and emphasized the usefulness of counterfactuals in the analysis of the causal effects of timeindependent treatments from observational data. Robins (1986, 1987) proposed a formal counterfactual theory of causal inference that extended Neyman's (1923) time-independent treatment theory to longitudinal studies with both direct and indirect effects and sequential timevarying treatments and confounders. In this theory, for any fixed history of antiretroviral therapy $\bar{a}, Y_{\bar{a}}$ is defined to be the random variable representing a subject's outcome had, possibly contrary to fact, the subject been treated with $\bar{a}$ rather than his observed treatment $\bar{A}$. Note that $\bar{a}$ is a possible value of the random variable $\bar{A}$. For each possible history $\bar{a}$ we are assuming a subject's response $Y_{\bar{a}}$ is well defined, although generally unobserved. Indeed we only observe $Y_{\bar{a}}$ for that treatment history $\bar{a}$ equal to a subject's actual treatment history $\bar{A}$, i.e. a subject's observed outcome $Y$ equals $Y_{\bar{A}}$. This identity is the fundamental "consistency" assumption that links the counterfactual data $Y_{\bar{a}}$ to the observed data (Y, $\bar{A})$.

Note that if, at each time $t, A(t)$ can take but one of two values ( 0 for untreated and 1 for treated) and the study duration is $K$ months, then there are $2^{K}$ different $Y_{\bar{a}}$ values associated with each subject as there are $2^{K}$ possible treatment patterns only one of which is actually observed for a given subject. Then, formally, the statement that the effect of treatment history on the mean of $Y$ is a linear logistic function of duration of antiretroviral therapy is the statement that, for each $\bar{a}$,

$$
E\left[Y_{\bar{a}}\right]=g[\bar{a} ; \beta]
$$

where

$$
\begin{equation*}
g[\bar{a} ; \beta]=\frac{\exp \left(\beta_{0}+\beta_{1} \operatorname{Dur}(\bar{a})\right)}{1+\exp \left(\beta_{0}+\beta_{1} \operatorname{Dur}(\bar{a})\right)} \tag{4}
\end{equation*}
$$

$\beta=\left(\beta_{0}, \beta_{1}\right)$, and $\operatorname{Dur}(\bar{a})=\sum_{k=0}^{K} a(k)$ is the subject's duration of treatment under the treatment history $\bar{a}$. We refer to this model as a MSM for the effect of antiretroviral therapy on the mean of $Y$, since it is a model for the marginal distribution of counterfactual variables and, in the econometric and social science literature, causal models (i.e. models for counterfactual variables) are often referred to as structural.

The parameter $\beta$ of our MSM encodes the magnitude of the average causal effects of the treatment on the outcome. By definition, the causal effect of treatment regime $\bar{a}$ on the outcome $Y$ for a given study subject is the difference $Y_{\bar{a}}-Y_{\bar{o}}$ between his outcome $Y_{\bar{a}}$ when treated with regime $\bar{a}$ and his outcome $Y_{\overline{\overline{0}}}$ when never treated. Thus the average causal effect of regime $\bar{a}$ is $E\left[Y_{\bar{a}}-Y_{\overline{\bar{o}}}\right]=E\left[Y_{\bar{a}}\right]-E\left[Y_{\overline{0}}\right]=g(\bar{a} ; \beta)-$ $g(\overline{0} ; \beta)$, which depends on $\beta$. If $\beta_{1}$ is zero, we say that there is no effect of treatment $\bar{a}$ on the outcome since $E\left[Y_{\bar{a}}\right]-E\left[Y_{\overline{0}}\right]$ is the same for all $\bar{a}$. In contrast, the association parameter $\gamma_{1}$ lacks a causal interpretation when treatment is not causally exogenous. Furthermore, the optimal non-dynamic intervention $\bar{a}^{*}$ is the value of $\bar{a}$ for which $E\left[Y_{\bar{a}}\right]=g(\bar{a} ; \beta)$ is the greatest if our goal is to maximize the probability that HIV will not be detected in the serum.

### 4.4 Formal definitions of causal exogeneity and no unmeasured confounders

We are now in a position to offer mathematically precise definitions of causal exogeneity and of no unmeasured confounders. Let $\left\{Y_{\bar{a}}\right\}$ be the set of all counterfactual outcomes $Y_{\bar{u}}$ as $\bar{a}$ varies. Formally, we say that the treatment process $A(t)$ is causally exogenous if,

$$
\begin{equation*}
\left\{Y_{\bar{a}}\right\} C \quad A(t) \mid \bar{A}(t-1) \tag{5}
\end{equation*}
$$

which is mathematically equivalent to the statement that $\left\{Y_{\bar{a}}\right\}$ is independent of $\bar{A}$. Note that even when $A(t)$ is "causally exogenous", if the treatment has an effect on the outcome, then the observed outcome $Y=Y_{\bar{A}}$ will not be independent of $\bar{A}$, since $Y_{\bar{A}}$ is a function of a subject's observed treatment history $\bar{A}$ itself. Given the covariates recorded in $L(t)$, following Robins (1987) we say there are no unmeasured confounders for the effect of $A(t)$ on $Y$ if

$$
\begin{equation*}
\left\{Y_{\bar{u}}\right\} C \quad A(t) \mid \bar{A}(t-1), \bar{L}(t) \tag{6}
\end{equation*}
$$

We shall also refer to the assumption of no unmeasured confounders as
the assumption that treatment $A(t)$ is sequentially randomized given the past. This assumption generalizes Rosenbaum and Rubin's (1983) assumption of ignorable treatment assignment to longitudinal studies with time-varying treatments and confounders. The assumption states that, conditional on treatment history and the history of all recorded covariates up to $t$, treatment at $t$ is independent of the counterfactual random variables $Y_{\bar{a}}$. This will be true if all prognostic factors for (i.e. predictors of) $Y$ that are used by physicians to determine whether treatment is given at $t$ are recorded in $\bar{L}(t)$ and $\bar{A}(t-1)$. That is, as in Figure 28.1(b), the causal graph generating the data has no arrows directly from any unmeasured causal risk factors for $Y$ directly into treatment. For example, since physicians tend to administer prophylaxis to subjects with previous bouts of PCP, and in untreated subjects PCP predicts $Y$, the assumption of no unmeasured confounders would be false if $\bar{L}(t)$ does not contain PCP history. It is the primary goal of the epidemiologists conducting an observational study to collect data on a sufficient number of covariates to ensure that the assumption of no unmeasured confounders will be at least approximately true.

In an observational study, the assumption of no unmeasured confounders cannot be guaranteed to hold even approximately and it is not subject to empirical test. Therefore, it may be useful to investigate the sensitivity to violations of the assumption through a formal sensitivity analysis. Robins et al. (1999, rejoinder) and Robins et al. (1999) provide details.

Robins (1999) proved that when there are no unmeasured confounders, (i) statistical exogeneity implies causal exogeneity, (ii) the weighted logistic regression estimator using the weights $S W$ converges to the parameter $\beta$ of the MSM (4) for $E\left[Y_{\bar{a}}\right]$, and (iii) the probability limit $\gamma$ of the usual unweighted logistic estimator generally differs from the causal parameter $\beta$ of the MSM unless the treatment process is statistically exogenous. Here we provide an informal heuristic argument for (ii). View each person as a member of a pseudo-population consisting of $S W$ copies of themselves. In this new pseudo-population, it can be shown that $\bar{L}(t)$ does not predict treatment at $t$ given past treatment history, and thus we have created a pseudo-population in which treatment is causally exogenous. Furthermore, the causal effect of treatment on $Y$ in the pseudo-population is the same as in the original population. That is, if $E\left[Y_{\bar{a}}\right]=g(\bar{a} ; \beta)$ in the true population, the same will be true of the pseudo-population. Hence, we would like to do ordinary logistic regression in the pseudo-population. But that is what our weighted logistic regression estimator is doing, since the weights create, as required, $S W$ copies of each subject.

We can generalize our MSM (4) slightly and model the marginal distribution of $Y_{\bar{a}}$ within levels of a subset $V$ of the pre-treatment (baseline) covariates $L(0)$. Then, our marginal structural logistic model (4) could
be modified to

$$
E\left[Y_{\bar{a}} \mid V\right]=\frac{\exp \left(\beta_{0}+\beta_{1} \operatorname{Dur}(\bar{a})+\beta_{2}^{\prime} V+\beta_{3}^{\prime} \operatorname{Dur}(\bar{a}) V\right)}{1+\exp \left(\beta_{0}+\beta_{1} \operatorname{Dur}(\bar{a})+\beta_{2}^{\prime} V+\beta_{3}^{\prime} \operatorname{Dur}(\bar{a}) V\right)}
$$

$\beta^{\prime}{ }_{3}$ measures how the magnitude of the effect of $\operatorname{Dur}(\bar{a})$ is modified by the pretreatment covariates $V$. An IPTW estimator of the parameter $\beta$ can be obtained by weighted logistic regression with weights $S W$ except now the logistic model includes $\operatorname{Dur}(\bar{A})$ and $V$ as regressors, and $S W(t)$ is redefined to be

$$
\begin{equation*}
S W(t)=\prod_{k=0}^{t} \frac{f[A(k) \mid \bar{A}(k-1), V]}{f[A(k) \mid \bar{A}(k-1), \bar{L}(k)]} . \tag{7}
\end{equation*}
$$

Note $V$ is already included in the denominator, since $V$ is a subset of the variables in $L(0)$.

Let $d(V)$ be a function (regime) that assigns to each value of the vector $V$ a value of $\bar{a}$ in the set $V$ of possible interventions. If a regime $d$ assigns the same value $\bar{a}$ to each $V$ we refer to the regime $d$ as non-dynamic. Otherwise, we refer to $d$ as a conditional or baseline-dynamic treatment regime, strategy or plan as it individualizes the treatment history a subject receives based on the subject's value of the baseline variables recorded in $V$, a subset of $L(0)$. A wise choice of $d$ should allow us to optimize therapy for individuals and thus should be a better strategy than even the optimal non-dynamic intervention $a^{*}$. Let $E\left[Y_{d}\right]$ be the probability of being without HIV in the serum if all subjects followed plan $d$. For subjects with a given value $v$ of $V$, the conditional expectation $E\left[Y_{d} \mid V=v\right]$ given $V=v$ under regime $d$ equals $E\left[Y_{\bar{a}} \mid V=v\right]$ for the value $\bar{a}=d(v)$ that they receive under the plan. Thus for the population as a whole $E\left[Y_{d}\right]=\Sigma_{v} E\left[Y_{\bar{a}} \mid V=v\right] \operatorname{pr}(V=v)$ is a weighted average of $E\left[Y_{\bar{a}} \mid V=v\right]$ with $a=d(v)$ and weights proportional to the fraction $\operatorname{pr}(V=v)$ of the population with $V=v$. Then $d_{o p}^{\text {baseline }}(v)$ is the treatment plan that maximizes $E\left[Y_{d}\right]$ over all possible baseline-dynamic and nondynamic treatment plans $d$. Now even $d_{o p}^{\text {baseline }}(v)$ only allows one to optimize treatment history $\bar{a}$ based on pretreatment (baseline) variables. However, with time-varying treatments it is usually important to dynamically choose the treatment at each time $t$ based on a subject's entire covariate history up to time $t$. For example, consider drug treatment for a chronic disease. When a drug becomes toxic to a subject, the optimal strategy is to stop the drug (or reduce the dose) at least temporarily. One cannot know when to stop the drug based on baseline covariates. Rather, the optimal treatment strategy must allow treatment decisons to be based on a subject's evolving covariate history. The best methods for estimating the effects of true dynamic regimes are not based on MSMs. Further discussion is provided in section 5.

### 4.5 Marginal structural Cox proportional hazards model

MSMs can easily be extended to failure time outcomes by specifying a marginal structural Cox proportional hazards model such as

$$
\begin{equation*}
\lambda_{T \bar{a}}(t \mid V)=\lambda_{0}(t) \exp \left(\beta_{1} a(t)+\beta_{2}^{\prime} V+\beta_{3}^{\prime} a(t) V\right) \tag{8}
\end{equation*}
$$

where $T_{\bar{a}}$ is the subject's time to death if he had followed anti-retroviral therapy history $\bar{a}, \lambda_{I_{\bar{a}}}(t \mid V)$ is the hazard (force of mortality) of $T_{\bar{a}}$ at $t$ conditional on having pretreatment variables $V, \lambda_{0}(t)$ is an unspecified baseline hazard function, $\exp \left(\beta_{1}+\beta_{3}^{\prime} V\right)$ is the causal rate ratio for the effects of treatment at level $V$ of a vector of baseline regressors including age, calendar year, CD4 count, CD8 count, WBC count, RBC count, platelets, etc. For variety, we have chosen a model which specifies that the hazard of death at time $t$ depends on current treatment status rather than the duration of treatment. Other dose-response models could be used.

Let $T$ be a subject's observed failure (i.e. death) time, so that $T=T_{\bar{A}}$. Arguing as above, Robins (1999) shows that, in the absence of censoring, a consistent estimator of the unknown parameter $\beta=\left(\beta_{1}, \beta_{2}^{\prime}\right.$, $\left.\beta_{3}^{\prime}\right)^{\prime}$ is obtained by fitting the ordinary time-dependent Cox model

$$
\begin{equation*}
\lambda_{T}(t \mid \bar{A}(t), V)=\lambda_{0}(t) \exp \left(\gamma_{1} A(t)+\gamma_{2}^{\prime} V+\gamma_{3}^{\prime} A(t) V\right) \tag{9}
\end{equation*}
$$

except that the contribution of a subject to a calculation performed on a subject $i$ at risk at time $t$ is weighted by $S W_{i}(t)$, as defined in (7) with $T>k$ added to the conditioning event. Note the subject-specific weights change with time. Few standard Cox proportional hazards software programs allow for time-varying weights. To avoid this software problem one can fit a weighted pooled logistic regression treating each personmonth as an observation and allowing for a time-dependent intercept. That is, one can fit, by weighted logistic regression using weights $S W(t)$, the model

$$
\begin{align*}
\text { logit } \operatorname{pr}[D(t) & =1 \mid D(t-1)=0, \bar{A}(t-1), V] \\
& =\gamma_{0}(t)+\gamma_{1} A(t-1)+\gamma_{2}^{\prime} V+\gamma_{3}^{\prime} A(t-1) V \tag{10}
\end{align*}
$$

where $D(t)=0$ if a subject was alive at time $t$ and 1 if the subject died at month $t$, and $\gamma_{0}(t)$ is a time-specific intercept. This method offers the advantage of being easily programmed in any standard statistical package. Under our assumptions we thereby obtain a consistent estimator of the parameter vector $\beta$ of the MSM

$$
\begin{align*}
\text { logit } \operatorname{pr}\left[D_{\bar{a}}(t)\right. & \left.=1 \mid D_{\bar{a}}(t-1)=0, V\right] \\
& =\beta_{0}(t)+\beta_{1} a(t-1)+\beta_{2}^{\prime} V+\beta_{3}^{\prime} a(t-1) V \tag{11}
\end{align*}
$$

When the death rate in any given month $t$ is small, the parameters of (11) and (8) closely approximate one another.

Because of the weights, the standard error estimates outputted by a standard logistic program are invalid and may be either too large or too small. To overcome this difficulty, model (10) should be fit using a generalized estimating equations (GEE) (Liang and Zeger 1986) program which outputs robust variance estimators. The robust variance GEE estimators provide a conservative confidence interval for the $\beta$ (Robins, 1999). That is, the $95 \%$ Wald confidence interval calculated as $\beta \pm 1.96$ $\times$ (robust) standard error is guaranteed to cover the true $\beta$ at least $95 \%$ of the time in large samples.

We now describe how to accommodate censoring in the analysis. We defined a subject as right censored at time $t$ (i.e. $C(t)=1$ ) if by time $t$ he either dropped out of the study or reached administrative end of followup alive.

We say that censoring is ignorable or non-informative if the conditional cause-specific hazard of being censored at $k$ among subjects alive and uncensored up to $k$ does not depend on the failure times $T_{\bar{a}}$ given $\bar{A}(k-1)$, and the time-dependent covariate $\bar{L}(k-1)$ history prior to $k$. Under the assumptions of ignorable censoring and no unmeasured confounding, Robins (1999) shows that we still obtain from fitting (10) consistent estimators of $\beta$ if we weight a subject alive and uncensored at month $t$ by $S W(t) \times S W^{\dagger}(t)$ where (i)

$$
S W^{\dagger}(t)=\prod_{k=0}^{t} \frac{\operatorname{pr}[C(k)=0 \mid \bar{C}(k-1)=0, \bar{A}(k-1), V, T>k]}{\operatorname{pr}[C(k)=0 \mid \bar{C}(k-1)=0, \bar{A}(k-1), \bar{L}(k), T>k]}
$$

is informally the inverse of the ratio of a subject's probability of remaining uncensored up to month $t$ divided by that probability calculated as if there had been no time-dependent determinants of censoring except past treatment history and $V$, and (ii) we modify our definition (7) of $S W(t)$ to add $C(k)=0$ to the conditioning events both in the numerator and the denominator. The denominator of $S W(t) \times S W^{\dagger}(t)$ is informally the probability that a subject would have his own observed treatment and censoring history through month $t$.

Hernan et al. (2000) describe how to estimate the weights $S W(t) \times\left(S W^{\dagger}(t)\right.$ from the data using pooled logistic regression models with treatment (censoring) at each time $k$ as the response variable. Substituting out estimated weights into our IPTW estimators allows us to estimate, under the assumption of no unmeasured confounders, logit $\operatorname{pr}\left[D_{\bar{a}}(t)=1 \mid D_{\bar{a}}(t-1)=0, V\right]=\beta_{0}(t)+\beta_{1} a(t-1)+\beta_{2}^{\prime} V+\beta_{3}^{\prime} a(t-1) V$ and thus the conditional survival curves $S_{\bar{a}}(t \mid V)=\prod_{k=0}^{t} \operatorname{pr}\left[D_{\bar{a}}(k)=0 \mid D_{\bar{a}}(k-1)\right.$ $=0, V]$ that would be observed if all subjects have followed regime $\bar{a}$.

Again let $d(V)$ be a function (regime) that assigns to each value of the
baseline vector $V$ a value of $\bar{a}$ in the set $V$ of possible interventions. Let $S_{d}(t)$ be the survival curve if all subjects followed plan $d$. For subjects with a given value $v$ of $V$, the conditional survival curve $S_{d}(t \mid V=v)$ given $V=v$ under regime $d$ is $\Sigma_{v} S_{\bar{a}}(t \mid V=v) \operatorname{pr}(V=v)$ is a weighted average of $S_{\bar{a}}(t \mid V=v)$ with $a=d(v)$ and weights proportional to the fraction $\operatorname{pr}(V=v)$ of the population with $V=v$. Then $d_{o p}^{\text {baseline }}(v)$ is the treatment plan that minimizes the area under $S_{d}(t)$ over all possible baselinedynamic and non-dynamic treatment plans $d$. Now even $d_{o p}^{\text {baseline }}(v)$ only allows one to optimize treatment history $\bar{a}$ based on pretreatment (baseline) variables. The best methods for estimating the effects of general dynamic regimes are not based on MSMs. Further discussion is provided in section 5 .

We have seen that under the assumption of no unmeasured confounders, IPTW estimation of a marginal structural Cox proportional hazards model can, in contrast with standard methods, be used to estimate the effect of time-varying treatments on survival.

The correctness of the resulting causal inferences is dependent on three key assumptions. First, we must assume that the covariates in $L(t)$ are sufficient to adjust for both confounding and for selection bias due to loss to follow-up. This implies that we have available, in each month, data recorded in $L(t)$ on the history of all time-dependent covariates that (i) are independent predictors of death and (ii) independently predict the probability of changes in treatment and/or of being censored in that month. As in all observational studies, this fundamental assumption cannot be empirically tested. In practice, this would never be precisely or sometimes even approximately true. As described earlier, methods have recently been developed which allow one to evaluate the sensitivity of one's estimates to increasing violation of this assumption.

Second, we must assume that our models for changes in treatment and censoring, given past covariate and treatment history, are correctly specified. Last, we need to assume that our MSM for the effect of antiretroviral therapy on mortality is correctly specified.

Even when estimating the effect of a time-independent treatment using standard statistical models, the same assumptions (no unmeasured confounders, non informative censoring, and no model misspecification) are required to endow the parameters with a causal interpretation. Furthermore, when estimating the effect of a time-varying treatment, our assumptions are less restrictive than those required by standard analyses: an approach based on IPTW estimation of MSMs does not require for validity the absence of confounding by time-dependent covariates.

## 5. Alternatives to MSMs

Before introducing MSMs, Robins and coauthors introduced two
other methods for estimation of the causal effect of a time-varying treatment in the presence of time-varying confounders: the parametric g-computation algorithm formula estimator (Robins 1986), and g-estimation of structural nested models (Robins et al. 1992). When (i) both treatment and the confounders are discrete variables, (ii) they are measured at only a few time points, and (iii) the sample size is large, then estimation can be carried out using fully saturated models (i.e. nonparametrically) and all three methods are precisely equivalent. They differ when, as in observational studies with sparse multivariate data, one must introduce modelling assumptions.

A major advantage of MSMs is that they resemble standard models. For example, the logistic MSM and the Cox proportional hazards MSM described above are the natural way to extend the ordinary logistic and time-dependent Cox models to allow for estimation of causal effects of time-dependent treatments. However a major advantage of the parametric g-computation algorithm formula estimator and g-estimation of structural nested models over MSMs is that these models are much more useful than MSMs for estimating both interactions between treatment and time-dependent covariates and the effect of dynamic treatment regimes (Robins 1999). Due to space limitations, we will focus in this chapter on the parametric g-computation algorithm formula estimator, and we will not consider g-estimation of structural nested models.

Now for any variable $Z$, let $Z$ be the support (i.e. the possible values) of $Z$. Define a treatment regime or plan $d$ to be a collection of $K+1$ functions $d=\left\{d_{0}, \ldots, d_{K}\right\}$ where $d_{m}: \overline{\mathcal{L}}_{m} \rightarrow \mathcal{A}_{m}$ maps histories $\bar{l}_{m} \in \overline{\mathcal{L}}_{m}$ into a treatment $d_{m}\left(\bar{l}_{m}\right) \in \mathcal{A}_{m}$. If $d_{m}\left(\bar{l}_{m}\right)$ is a constant, say $a_{m}$, not depending on $\bar{l}_{m}$ for each $m$, we say regime $d$ is non-dynamic and write $d=\bar{a}, \bar{a} \equiv\left(a_{0}, \ldots, a_{K}\right)$. Otherwise, $d$ is dynamic. We let $\mathcal{D}$ be the set of all regimes $d$. Let $f(o)$ and $F(o)$ represent the density and distribution function of the observed data $O=\left(\bar{A}_{K}, \bar{L}_{K+1}\right)$.

Associated with each regime $d$ is the distribution $F_{d}(o)$ with density $f_{d}(o)$ that represents the distribution of the observed data had, contrary to fact, all subjects in the population been treated with regime $d$. Suppose the assumption of no unmeasured confounders holds for the joint outcomes $\bar{L}_{\bar{a}}=\bar{L}_{\bar{a}, K+1}$ [i.e. Equation 6 holds with $\left\{\bar{L}_{\bar{a}}\right\}$ replacing $\left\{\bar{Y}_{\bar{a}}\right\}$. Then given a regime $d=\left(d_{0}, d_{1}, \ldots, d_{K}\right)$ and the joint density

$$
\begin{equation*}
f(o)=f\left(l_{0}\right) \times f\left(a_{0} \mid l_{0}\right) \times \ldots \times f\left(a_{K} \mid \bar{l}_{K}, \bar{a}_{K-1}\right) \times f\left(l_{K+1} \mid \bar{l}_{K}, \bar{a}_{K}\right), \tag{12}
\end{equation*}
$$

of the observed data (written as a Markov factorization in temporal order), $f_{d}(o)$ is the density $f(o)$ except that, in the factorization (12), $f\left(a_{0} \mid l_{0}\right)$ is replaced by a degenerate distribution at $a_{0}=d_{0}\left(l_{0}\right), f\left(a_{1} \mid l_{1}, a_{0}, l_{0}\right)$ is replaced by a degenerate distribution at $a_{1}=d_{1}\left(l_{0}, l_{1}\right)$, and, in general, $f\left(a_{k} \mid \bar{l}_{k}, \bar{a}_{k-1}\right)$ is replaced by a degenerate distribution at $a_{k}=d_{k}\left(\bar{l}_{k}\right)$. That
is the density $f_{d}(o)$ is given by

$$
\begin{equation*}
f_{d}(o)=f\left(l_{0}\right) \times f_{d}\left(a_{0} \mid l_{0}\right) \times \ldots \times f_{d}\left(a_{K} \mid \bar{l}_{K}, \bar{a}_{K-1}\right) \times f\left(l_{K+1} \mid \bar{l}_{K}, \bar{a}_{K}\right), \tag{13}
\end{equation*}
$$

where $f_{d}\left(a_{k} \mid \bar{l}_{k}, \bar{a}_{k-1}\right)$ is equal to 1 if $a_{k}=d_{k}\left(\bar{l}_{k}\right)$ and $f_{d}\left(a_{k} \mid \bar{l}_{k}, \bar{a}_{k-1}\right)$ is equal to 0 otherwise.

Again suppose the outcome of interest is $L_{K+1}$ which is assumed to be univariate and shall be denoted by $Y$. In the following, let $d\left(\bar{l}_{k}\right) \equiv$ $\left(d_{0}\left(\bar{l}_{0}\right), \ldots, d_{k}\left(\bar{l}_{k}\right)\right)$ and $d_{k}\left(\bar{l}_{k}\right)$ denote values of $\bar{A}_{k}$ and $A_{k}$ respec-
tively. Then the chance of having history $\left(y, \bar{l}_{K}\right)$ under the distribution $F_{d}(o)$ is $f_{d}\left(y, \bar{l}_{K}\right)=f_{d}\left(y \mid \bar{l}_{K}, d\left(\bar{l}_{K}\right)\right) \prod_{j=0}^{K} f\left(l_{j j} \bar{l}_{j-1}, d(\bar{l}\right.$
$\left.{ }_{j-1}\right)$ ) since the
terms
$f_{d}\left(a_{j} \bar{l}_{j}, \bar{a}_{j-1}\right)$ are equal to 1 in expression (13) for $f_{d}(o)$. Thus the chance $f_{d}(y)$ of having $Y=y$ under $f_{d}(\cdot)$ is

$$
\begin{align*}
f_{d}(y) & =\int f_{d}\left(y, \bar{l}_{K}\right) d \mu\left(\bar{l}_{K}\right) \\
& =\int\left\{f\left(y \mid \bar{l}_{K}, d\left(\bar{l}_{K}\right)\right) \times \prod_{j=0}^{K} f\left(l_{j} \bar{l}_{j-1}, d\left(\bar{l}_{j-1}\right)\right)\right\} d \mu\left(l_{j}\right) \tag{14a}
\end{align*}
$$

where $d \mu\left(\bar{l}_{K}\right)$ represents integration (or summation if $\bar{l}_{K}$ is discrete) over all possible $\bar{l}_{K}$ histories. That is if $\bar{l}_{K}$ is discrete we obtain

$$
\begin{align*}
f_{d}(y) & =\sum_{l_{K}} f_{d}\left(y, \bar{l}_{K}\right) \\
& =\sum_{\bar{l}_{K}}\left\{f\left(y \mid \bar{l}_{K}, d\left(\bar{l}_{K}\right)\right) \times \prod_{j=0}^{K} f\left(l_{j} \mid \bar{l}_{j-1}, d\left(\bar{l}_{j-1}\right)\right)\right. \tag{14b}
\end{align*}
$$

with $\bar{l}_{j}$ the initial segment of a given $\bar{l}_{K}$ history. Robins (1986) referred to this formula as the g -computation algorithm formula or functional for the effect of regime $d$ on outcome Y. Similarly, the marginal distribution function of $Y$ under $F_{d}(\cdot)$ is

$$
F_{d}(y)=\int \ldots \int p r\left[Y<y \mid \bar{l}_{K}, d\left(\bar{l}_{K}\right)\right] \times \prod_{j=0}^{K} f\left(l_{j} \bar{l}_{j-1}, d\left(\bar{l}_{j-1}\right)\right) d \mu\left(l_{j}\right)
$$

To estimate $f_{d}(y)$ from the observed data we specify parametric models $f\left(l_{j} \bar{l}_{i-1}, \bar{a}_{i-1}\right)$ and $f\left(y \mid \bar{l}_{K}, \bar{a}_{K}\right)$, fit the models to the observed data to obtain estimates $\hat{f}\left(l_{j} \bar{l}_{j-1}, \bar{a}_{j-1}\right)$ and $\hat{f}\left(y \mid \bar{l}_{K}, \bar{a}_{K}\right)$ and evaluate these estimates at $\bar{a}_{j-1}=d\left(\bar{l}_{j-1}\right)$ and $\bar{a}_{K}=d\left(\bar{l}_{K}\right)$, and substitute them into the above expression for $f_{d}(y)$. Details are given in section 6 below where we consider
estimating the effect of treatment on a survival outcome. Under certain additonal assumptions described in section 6 , one may consider the gformula as giving the overall effect of an intervention regime $d$ on $Y$ by summing up both its indirect effects on $Y$ mediated through the effect of the intervention on the non-intervened on variables $L_{j}$ (which in turn affect the outcome $Y$ ) and its direct effects on $Y$. In contrast, MSMs always model the overall effect of the intervention on $Y$ without any decomposition into direct and indirect effects.

## 6. Analysis of the framingham OFFSPRING STUDY

In this section we describe the use of the parametric $g$-computation algorithm formula estimator of section 5 and IPTW MSM estimator of section 4 to estimate what the cumulative incidence of coronary heart disease mortality in the Framingham Offspring Study would have been

| Content | Time after follow-up | Sample size |
| :--- | :---: | :---: |
| Exam 1 | Year 0 | 5 I24 |
| Exam 2 | Year 8 | 3863 |
| Exam 3 | Year 12 | 3873 |
| Exam 4 | Year 16 | 4019 |

under various hypothetical intervention strategies. The Framingham Offspring public use file contained data on 5124 subjects ( 2483 male and 2641 female subjects). Subjects were examined five times over 20 years. Exam 1 occured in year 0 , exam 2 in year 8 , exam 3 in year 12, exam 4 in year 16 and exam 5 in year 20. The table below shows the number of subjects attending each of the first four exams. Mortality follow-up was essentially complete through exam 5 at year 20. By year 20, 891 subjects were known to have developed coronary heart disease (CHD), including those with pre-existing CHD at exam 1.
For reasons explained below, the follow-up period for our analysis began at exam 2 (which we will refer to as the baseline exam). Each subject contributed person-time from exam 2 to the date of CHD diagnosis, death from any cause, loss to follow-up, or exam 5 , whichever occurred first. Subjects with pre-existing CHD at exam 2 or with missing risk factor values at exam 1 were excluded from all calculations.

In the Framingham public use file, CHD was defined as one of the following events: (i) recognized or unrecognized myocardial infarction (using diagnostic ECG or transaminase/history/autopsy evidence), (ii) angina pectoris (first episode), (iii) coronary insufficiency, or (iv) death from CHD. The following (time-dependent) risk factors were used in various analyses: age at exam (years); sex (male/female); body mass index
$\left(\mathrm{kg} / \mathrm{m}^{2}\right)$; cigarette smoking (current smoker/non-current smoker); alcohol consumption (calculated alcohol index in ounces per week); diabetes mellitus (defined diabetes/no defined diabetes); LDL-cholesterol (mg/dl); HDL-cholesterol ( $\mathrm{mg} / \mathrm{dl}$ ); and systolic blood pressure ( mmHg ).

Missing values of risk factors in later exams were carried forward from the previous exam. Fat intake was not used in our analyses, because the Framingham public use file did not contain that information. Physical activity was also not used, because it was not measured at exams 1 and 4.

### 6.1 INTERVENTIONS

We estimated with the parametric g-computation algorithm formula estimator the cumulative incidence (i.e. risk) of CHD between exam 2 and exam 5 that would have been observed under the following hypothetical intervention strategies.

1. A random half of smokers at baseline (i.e. exam 2) quit smoking forever (no intervention on those who were not current smokers at baseline but started smoking later).
2. All smokers at baseline quit smoking forever (no intervention on those who were not current smokers at baseline but started smoking later).
3. Subjects who were drinkers at baseline were randomly assigned a non-time-varying level of alcohol consumption at baseline drawn at random from a truncated normal distribution with a mean of $20 \mathrm{~g} /$ day for males and $10 \mathrm{~g} /$ day for females, a maximum of $30 \mathrm{~g} /$ day for males and $15 \mathrm{~g} /$ day for females, and a minimum of $10 \mathrm{~g} /$ day for males and $5 \mathrm{~g} /$ day for females (with the sex-specific variance of the normal equal to the sex-specific variance of alcohol consumption at baseline among drinkers in study population). Alcohol consumption was not intervened on thereafter.
4. Body mass index (BMI) was lowered
a) by $10 \%$ at each visit BMI exceeded 22
b) to 22 at each visit BMI exceeded 22
5. The distribution of LDL at baseline was drawn at random from a normal distribution with mean of 90 and a SD of 30 (Chinese distribution both for males and females). LDL was not intervened on thereafter.
6. Interventions 1, 3 and 4a simultaneously.

### 6.2 Data analysis

## PARAMETRIC G-FORMULA

In this section we describe how we estimate the proportion of the Framingham Offspring Study population that would have developed CHD between exams 2 and 5 for each of the interventions mentioned above under the assumption of no unmeasured confounders. Specifically, we used the parametric g-formula (i.e. the g-formula evaluated at predicted values calculated under parametric regression models) to estimate the CHD risk in the Framingham Offspring data under each intervention. We remark that in this exercise we are estimating the cumulative incidence (probability of developing) CHD between exams 2 and 5 under each intervention rather than the expected life years and expected quality-adjusted life-years. Although these latter quantities are more relevant for public health decision-making, we decided to report the results in terms of cumulative incidences because this measure is more familiar to practising epidemiologists. It would be easy to use these same methods to estimate expected life measures. The g-formula to compute the CHD risk (i.e. cumulative incidences between exams 2 and 5) under a particular intervention (i.e. a possibly dynamic regime) $d$ is one minus the probability of "not developing CHD between exams 2 and 5 " which is given by the following g-formula for "survival without CHD."

$$
\begin{align*}
& \sum_{\bar{a}_{4}} \sum_{\bar{a}_{4}^{*}} \sum_{\bar{l}_{4}} \prod_{j=1}^{4} \operatorname{Pr}\left[C H D_{j+1}=0 \mid \bar{l}_{j}, \bar{a}_{j}, \overline{\mathrm{CHD}}_{j}=0, \bar{D}_{j}=0\right]^{I(j>2)} \\
& \quad \times f_{d}\left(a_{j} \mid \bar{l}_{j}, a_{j}^{*}, \bar{a}_{j-1}, \overline{C H D}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right)  \tag{15}\\
& \quad \times f\left(l_{j} \mid \bar{l}_{j-1}, \bar{a}_{j-1}, \overline{C H D}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right) \\
& \quad f\left(a_{j}^{*} \mid \bar{l}_{j}, \bar{a}_{j-1}, \overline{C H D}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right)
\end{align*}
$$

where $j$ denotes exam number for any $\bar{z}_{j}, \bar{z}_{0}$ is defined to be 0 since there is no exam 0 .
$C H D_{j+1}=1$ if CHD was diagnosed between exams $j$ and $j+1,0$ otherwise;
the overbar means history, i.e. $\overline{\mathrm{CHD}}_{j}=\left(\mathrm{CHD}_{1}, \ldots, \mathrm{CHD}_{j}\right)$;
the random variable $L_{j}$ with realized values $l_{j}$ is the set of risk factors not undergoing the intervention;
$a_{j}$ is the intervention value of the risk factors $A_{j}$ undergoing intervention at time $j$;
$\bar{l}_{j}$ is a history of (non-intervened on) risk factors through exam $j$ compatible with a particular $\bar{l}_{4}$ history in the sum;
$\bar{a}_{j}$ is a history of (intervened on) risk factors through exam $j$;
$a_{j}^{*}$ is the value of $A_{j}$ that would be observed at time $j$ if interventions were made through $j-1$ but no intervention was made at $j$ (see below); the sum is overall possible $\bar{l}_{4}, \bar{a}_{4}, \bar{a}_{\stackrel{*}{4}}$ histories;
$\bar{C}_{j}=0$ is the event that a subject remains uncensored through exam $j$;
$\bar{D}_{j}=0$ is the event that a subject has not died from other non-CHD causes through exam $j$;
$f(\cdot \cdot)$ is a conditional density function of the observed data distribution;
$f_{d}(\cdot \cdot)$ is the conditional density function of the exposure $A$ under the proposed intervention. Because we do not intervene until $j=2$, the "intervened on" value $a_{1}$ of $A_{1}$ must equal the "non-intervened on" value $a_{i}^{*}$. To accomplish this, we take for $j=1 f_{d}\left(a_{j} \mid \bar{l}_{j}, a_{j}^{*}, \bar{a}_{j-1}, \overline{\mathrm{CHD}}_{j}=0\right.$, $\left.\bar{C}_{j}=0, \bar{D}_{j}=0\right)=1$ if $a_{j}=a_{j}^{*}$ and 0 otherwise.

Formula (15) includes four generalizations of formula (14). First we allow the intervention treatment $a_{j}$ at time $j$ to depend on the value $a_{j}^{*}$ of $A_{j}$ that would be observed at time $j$ if the planned interventions were made through $j-1$ but no intervention was made at $j$. That is the intervention density $f_{d}\left(a_{j} \bar{l}_{j}, a_{j}^{*}, \bar{a}_{j-1}, \overline{\mathrm{CHD}}_{j}=0, \overline{\mathrm{C}}_{j}=0, \bar{D}_{j}=0\right)$ can be a function of $a_{j}^{*}$. This is meant to reflect the fact that when a subject arrives at visit $j$, his exposure $a_{j}^{*}$ could be noted and if it takes certain values it can, in principle, be intervened on and instantaneously changed to a new exposure labelled $a_{j}$. We assume it is $a_{j}$ that affects outcomes at the next visit $j+1$. For example, the BMI intervention 4 b was modelled in this way. This "instantaneous change" model might be a reasonable approximation to a real intervention in which BMI was reduced to 22 over 3 to 6 months, as the time between visits is four years. To determine at time $j$ the value of $a_{j}^{*}$ for a simulated subject, we need to generate $a_{j}^{*}$ from the non-intervention observed data density $f\left(a_{j}^{*} \mid \bar{l}_{j}, \bar{a}_{j-1}\right.$, $\overline{\mathrm{CHD}}_{j}=0, \overline{\mathrm{C}}_{j}=0, \bar{D}_{j}=0$ ), where it follows from the definition of $a_{j}^{*}$ that the arguments $\left(\bar{l}_{j}, \bar{a}_{j-1}\right)$ in the conditioning event are the values that would be seen had the planned intervention $d$ been made through $j-1$. Second, we now allow probabilistic (i.e. random) interventions so $f_{d}\left(a_{j} \bar{l}_{j}, a_{j}^{*}, \bar{a}_{j-1}, \overline{C H D}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right)$ need not take only the values 0 or 1 . For example, an intervention in which a subject who is a smoker at occasion $j$ (i.e. $a_{j}^{*}=1$ ) stops smoking with probability $1 / 2$ would have $f_{d}\left(a_{j} \bar{l}_{j}, a_{j}^{*}=1, \bar{a}_{j-1}, \overline{\mathrm{CHD}}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right)=1 / 2$ for both $a_{j}=1$ and $a_{j}=0$. Third, we now allow the interventions such as the alcohol intervention 3 and LDL intervention 5 to include active intervention at some times and no further intervention at other times. For example, in intervention $5, f_{d}\left(a_{j} \bar{l}_{j}, a_{j}^{*}, \bar{a}_{j-1}, \overline{\mathrm{CHD}}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right)$ is equal to a normal density with mean 90 and standard deviation 30 at the baseline visit $j=$ 2, but $f_{d}\left(a_{j} \mid \bar{l}_{j}, a_{j}^{*}, \bar{a}_{j-1}, \overline{\mathrm{CHD}}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right)=1$ if $a_{j}=a_{j}^{*}$ at visits $j$ $=3$ and $j=4$ in addition to visit $j=1$. Finally, the end-point is now survival to visit 5 , so the $g$-formula multiplies the probability of not failing at visits 3,4 and 5 .

## Caveats

As discussed above, the g-formula identifies the effect one would have
observed under a particular intervention under the assumption of no unmeasured confounders. But this assumption itself assumes that the counterfactual survival variables $\mathrm{T}_{\mathrm{d}}$ (the time to CHD under intervention d) are well defined. If the intervention is on smoking, it seems only mildly philosophically problematic to assume the existence of a welldefined counterfactual time of CHD had one received a different smoking history than one's actual history. However, for interventions $4 \mathrm{a}, 4 \mathrm{~b}$ or 5 , the counterfactual $\mathrm{T}_{\mathrm{d}}$ may be very poorly defined. For example, the effect on CHD of lowering LDL may possibly depend on the mechanism by which it is lowered. That is, the effect may be different depending on whether LDL were lowered by using drugs that decrease LDL production, drugs that increase its elimination or drugs that impair the absorption of fats from the gut. Similarly the effect on CHD of an intervention to lower BMI may be different depending on the degree of calorie restriction imposed, the foods allowed in the required diet, and the level of increased physical activity imposed. Thus for the interventions 4a, 4b and 5 d that are stated solely in terms of the level of BMI or LDL to be achieved (without further specifying the precise intervention that will be used to realize the reductions), it may be that the counterfactuals $\mathrm{T}_{\mathrm{d}}$ should not be regarded as well-defined. But if the counterfactulas are not well-defined, the assumption of no unmeasured confounders is vacuous (as the assumption is in terms of the counterfactuals). As a consequence, for such interventions, it is unclear what the counterfactual quantity one hopes the g -formula to estimate. However, if one correctly held the belief that LDL in the blood is a dominant cause of CHD and that all reasonable interventions that lowered LDL to a prespecified level would result in equal reductions in CHD risk, then $\mathrm{T}_{\mathrm{d}}$ would be well-defined even for an intervention d that, like intervention 5 , is stated only in terms of the level of LDL to be achieved. Evidence for this belief would be strenghtened if randomized experiments of various lipid lowering therapies that lowered average LDL by an identical amount showed the same improvement in CHD risk, regardless of the mechanism of action of the drug used to lower LDL.

Even when a counterfactual is well-defined, the g-formula only estimates the effect of an intervention under the assumption of no unmeasured confounders. Thus it is important for epidemiologists to try to obtain data on many potential confounders. The assumption of no unmeasured confounders will never exactly be true in observational studies and may not even be approximately true. A particular setting in which substantial unmeasured confounding may be present is when many of the confounders recorded in $L_{j}$ are measured with error. In that case, even if the correctly measured confounders would have served to fully control confounding, the mismeasured confounders may not. This could be the case even when the measurement error is random and nondifferential and the null hypothesis of no treatment effect is true. This
raises an important distinction between studies with time-varying treatments and studies with time-independent treatments. With time-varying treatments, past treatment history itself can be a confounder for the effect of current treatment on the outcome. Thus even when treatment is only subject to random (non-differential) measurement error, confounding bias can exist even under the null hypothesis of no treatment effect. This is in sharp contrast with the results for time-independent treatments in which random (non-differential) subject-specific measurement error does not lead to bias under the causal null hypothesis of no treatment effect. However, there is an additional subtlety in studies with time-varying treatments. If the decision to take a treatment at time $j$ depends on one's past measured treatment history (say as recorded in a pharmacy prescription database) and not on one's true past treatment history, then, under the null hypothesis, there is no bias introduced by the mismeasurement of treatment, even if the measurement error is differential. On the other hand, if the decision to take treatment at a given time depends on past true (but unrecorded) treatment history, then, by only controlling for past measured treatment history in the analysis, bias may exist under the null hypothesis of no treatment effect, even with random (nondifferential) measurement error.

In sections 4 and 5, we assumed the intervened on risk factors $A_{j}$ at time or visit $j$ occurred temporally after the non-intervened on risk factors $L_{j}$. We clearly cannot assume this to be the case in the Framingham Offspring Study for all 6 interventions of interest. Further, because the variables are generally measured but once every four years, each risk factor may potentially causally influence the others over the four years. In this setting, if we let $Z_{j}$ be the set of all measured risk factors at visit $j$ (regardless of whether one wants to consider intervening on any subset), let $D_{j}$ be the indicator variable for survival at visit $j$ with $D_{j}$ prior to $Z_{i}$, then a sufficient and nearly necessary condition for the g-functional (15) to identify the effect on the cumulative incidence of CHD under an intervention on any chosen subset of the measured timevarying risk factors is that (i) no risk factor causally influences any other risk factor over the four years between visits which we formalize as, for each visit $j$, no element of $Z_{j}$ is a cause of any other element of $Z_{j}$ on the causal DAG generating the data, (ii) the set of all measured risk factors $\bar{Z}_{K}$ are jointly causally exogenous for survival in the sense that on the causal DAG generating the data, for all combination of visit times $s$, $t, j$, there is no unmeasured $U_{s}$ that is a common cause of both $D_{j}$ and (an element of) $Z_{t}$ for which there exist directed paths from $U_{s}$ to $D_{j}$ and $U_{s}$ to $Z_{t}$ whose other vertices are all unmeasured variables, and (iii) any unmeasured common causes of variables in $Z_{j}$ are marginally independent of any unmeasured common cause of variables in $Z_{t}$ for $j \neq s$. That is, conditions (i)-(iii) guarantee that there are no unmeasured confounders for the effect of any possible intervention on survival, even without knowledge of time order among the variables $Z_{t}$ measured at
the same visit $t$ and even though there is substantial time (four years) between the visits at which data are recorded. Conditions (i)-(iii) are so restrictive that it seems unlikely that we would ever believe that the g formula will be exactly unbiased for the effect of all the various risk factor interventions in which we might have interest. If conditions (i)-(iii) did hold, then the g -formula gives the overall effect of an intervention regime $d$ on survival by explicitly summing up both its indirect effects on survival mediated through the effect of the intervention on the non-intervened on variables $L_{j}$ measured by the factors $f\left(l_{j} \mid \bar{l}_{j-1}, \bar{a}_{j-1}\right.$, $\overline{\mathrm{CHD}}_{j}=0, \overline{\mathrm{C}}_{j}=0, \bar{D}_{j}=0$ ) (which in turn affect the survival as measured by the dependence of the factors $\prod_{j=1}^{4} \operatorname{Pr}\left[C H D_{j+1}=0 \mid \overline{\bar{l}}_{i}, \bar{a}_{j}, \overline{C H D}_{j}=0, \bar{D}_{j}=0\right]$ on $\bar{l}_{j}$ ) and its direct effects on survival measured by the dependence of $\prod_{j=1}^{4} \operatorname{Pr}\left[C H D_{j+1}=0 \mid \bar{l}_{j}, \bar{a}_{i}, \overline{C H D}_{j}=0, \bar{D}_{j}=0\right]$ on $\bar{a}_{j}$.

## Details of estimation

Because the g -formula is a sum over all possible values of risk factor history $\bar{l}_{4}$ and each $l_{j}$ is a high-dimensional vector of covariates, a direct calculation based on (15) is computationally infeasible. Rather, we approximate the result of the g -formula under a given intervention by Monte Carlo simulation. To see how to conduct the simulation, first note that the g -formula (15) gives the probability of developing CHD between exams 2 and 5 based on a intervention-specific joint distribution of CHD and risk factors. Under the assumption of no unmeasured confounders, this is the joint distribution had all subjects followed the intervention.

Therefore we generate, for each intervention, a simulated population in which the joint distribution of CHD and risk factors is approximately equal to the joint distribution implied by the g -formula. Then the CHD risk in the simulated population (i.e. the expected fraction of subjects in the simulated population who develop CHD between exams 2 and 5) estimates the desired probability. Note that, if we simulate under no intervention (i.e. with $f_{d}\left(a_{j} \bar{l}_{i}, a_{j}^{*}, \bar{a}_{j-1}, \overline{C H D}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right)=1$ if $a_{j}=a_{j}^{*}$ for all $j$ rather than for just $j=1$ ), the expected CHD risk in the simulated population should equal that of the actual study population, because the joint distribution implied by the g -formula for the simulated population is precisely that of the study population.

Let $z_{j}=\left(l_{j}, a_{j}^{*}\right)$. To estimate

$$
\begin{aligned}
& f\left(z_{j} \mid \bar{l}_{j-1}, \bar{a}_{j-1}, \overline{C H D}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right) \\
& \quad=f\left(a_{j}^{*} \bar{l}_{j}, \bar{a}_{j-1}, \overline{C H D}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right) \\
& \quad f\left(l_{j} \mid \bar{l}_{j-1}, \bar{a}_{j-1}, \overline{C H D}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right)
\end{aligned}
$$

where $l_{j}$ includes all non-intervened on risk factors at exam $j$, we chose an arbitrary ordering of risk factors at exam $j$ : such as (i) body mass index, (ii) cigarette smoking, (iii) alcohol consumption, (iv) diabetes mellitus, (v) LDL, (vi) HDL and (vii) systolic blood pressure. The density (15) is invariant to the ordering of the variables in $z_{j}$. This invariance to ordering is one reason why (15) can only have a causal interpretation for all regimes $d$ under that assumption that no risk factor measured at $j$ causes any other risk measure at $j$, as in the above caveats. We then estimate (i) the conditional probability of BMI at $j$ given the past variables through $j-1$, (ii) the conditional probability of smoking given BMI at $j$ and past variables through $j-1$, (iii) the conditional probability of alcohol at $j$ given BMI, smoking at $j$ and past variables through $j-1$, and so on. We estimate each of these conditional densities by maximum likelihood from the observed data. Finally, we estimate $f\left(z_{j} \bar{l}_{j-1}, \bar{a}_{j-1}, \overline{\mathrm{CHD}}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right)$ as the product of these estimated conditional densities.

In detail, the algorithm used to simulate the simulated population exposed to an intervention was as follows:

## Part A: Modelling

1. We fit a pooled (over persons and time) linear logistic regression model to predict the risk of CHD given risk factor history. The outcome was CHD diagnosis between exams $j$ and $j+1$, and the covariates in the model were risk factors at exams $j$ and $j-1$. The model was restricted to those with no diagnosis of CHD at or before exam $j$. The parameters of this model define the estimated conditional probability of CHD risk given the entire past, and thus implicitly assume that risk factors more than 2 time periods previously do not predict CHD risk given risk factors in the past 2 periods.
2. For $j=3$ and $j=4$, we fit pooled regression models to predict each risk factor given past risk factor history, among those with no prior diagnosis of CHD. Risk factors at exam $j$ were the "outcome" in models that include other risk factors at exam $j$ (according to the arbitrary ordering explained above) plus all risk factors at exams $j-1$ and $j-2$ as covariates. We used linear regression for continuous risk factors, and logistic regression for dichotomous risk factors. Continuous risk factor variables with skewed distributions were log transformed. The parameters of these models define the estimated conditional distributions (Bernoulli or Normal or Log normal) of each risk factor.
3. Follow-up started at exam 2 in our analyses because two prior exams are used to predict CHD risk between exams $j$ and $j+1$. Thus, the CHD risk we estimate refers to the 12 -year period between exam 2 (year 8) and exam 5 (year 20).

## Part B: Data generation

4. We simulated a sample of 10000 individuals by sampling with replacement from the study population.
5. The risk factor values at exams 1 and 2 of these 10000 individuals were those actually observed (as interventions beginning at exam 2 could not affect these distributions so we were able to use the empirical distribution of the data).
6. The risk factor values $z_{j}=\left(l_{j}, a_{j}^{*}\right)$ at exams 3 and 4 were generated by sampling a value from the conditional distributions estimated in step 2 above.
7. $a_{j}$ was set equal to $a_{j}^{*}$ for each $j$.
8. The 12 -year probability $1-\prod_{j=1}^{4} \operatorname{Pr}\left[C H D_{j+1}=0 \mid \bar{l}_{i}, \bar{a}_{j}, \overline{C H D}_{j}=0, \quad \bar{D}_{j}=\right.$ 0 ] of CHD for each simulated individual was estimated, based on his/her simulated risk factor values, using the conditional distribution estimated in step 1 above. The population CHD risk is the average of each person's estimated risk.
9. The above procedure simulates the observed study population under no intervention.
10. To simulate a counterfactual population subject to a given intervention $d$ under the assumption of no unmeasured confounders, step 7 had to be modified by drawing $a_{j}$ for $j=2,3,4$ from the intervention density $f_{d}\left(a_{j} \bar{l}_{j}, a_{j}^{*}, \bar{a}_{j-1}, \overline{\mathrm{CHD}}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right)$ rather than setting $a_{j}$ to $a_{j}^{*}$ as under no intervention.

We used nonparametric bootstrap methods (sampling the observed study population with replacement 100 times) to estimate approximate confidence intervals (CI) of the counterfactual CHD risks and risk ratios. The size of each bootstrap sample was that of the original sample. To obtain standard errors and thus Wald-type confidence intervals, we re-applied all of steps $1-8$ to each of the 100 bootstrap samples. All analyses were performed separately by sex.

## MARGINAL STRUCTURAL MODEL

We also analysed the non-dynamic interventions on cigarette smoking using MSMs and compared the results to those obtained using the g-formula. Specifically our goal was (i) to estimate the (potentially time-
varying) hazard of CHD under pre-specified (i.e. non-dynamic) regimes or interventions, and (ii) to use this hazard to compute the 12-year CHD risk.

To estimate the hazard, we fit the discrete-time marginal structural Cox model

$$
\operatorname{logit} \lambda_{T_{\bar{a}}}(t)=\operatorname{logit} \lambda_{0}(t)+\beta a(t)
$$

where $\lambda_{T_{\bar{a}}}(t)=\operatorname{Pr}\left[\operatorname{CHD}_{\bar{a}}(t+1)=1 \mid\right.$ CHD $\left._{\bar{a}}(t)=0\right]$ is the discrete hazard of CHD between $t$ and $t+1$ under intervention $\bar{a}=[a(2), a(3), a(4)]$ at visits 2,3 and $4, \lambda_{0}(t)$ is the baseline hazard, $a(t)$ is smoking status (smoker or non-smoker) at study visit $t$ under regime $\bar{a}$, and, for dichotomous smoking exposure, $\beta$ is the $\log$ relative risk (i.e. odds ratio) for always exposed vs never exposed. Under the assumption of no unmeasured confounders, the parameters of this model can be estimated using inverse-probability-of-treatment weighting. The model used to predict smoking status was the same as the one used for the parametric g-formula. We fit the model by using a weighted pooled logistic model with a time-varying intercept.

To estimate the 12-year CHD risk when everybody quits smoking at baseline (and nobody initiates smoking thereafter), we could have computed one minus the survival probability

$$
\left[1-\lambda_{T_{\bar{a}}}(2)\right]\left[1-\lambda_{T_{\bar{\alpha}}}(3)\right]\left[1-\lambda_{T_{\bar{\alpha}}}(4)\right]
$$

with $\bar{a}=[0,0,0]$. However, this simple way of estimating the risk is potentially very inefficient (large variance) because the inverse-probabil-ity-of-treatment weights can be unstable. Therefore, to improve efficiency we used inverse-probability-of-treatment weighting to estimate the parameters of the model

$$
\operatorname{logit} \lambda_{T_{\bar{a}}}(t)=\operatorname{logit} \lambda_{0}(t)+\alpha^{T} V+\beta a(t)
$$

where $V$ is a vector of baseline (pre-intervention) covariates. The inclusion of $V$ allows us to use $V$-stabilized weights which result in much more efficient estimation.

We then simulated a sample of 10000 individuals by sampling with replacement from the study population and used the parameter estimates of the MSM above to compute the CHD risk between exams 2 and 5 for each simulated individual at each time based on the observed values of $V$ and $\bar{A}(t)$. We then computed the 12 -year CHD risk for each individual and averaged over all individuals to obtain an estimate of the risk under no intervention.

To estimate the 12 -year CHD risk when $50 \%$ of the subjects quit smoking at baseline (and nobody initiates or quits smoking thereafter),
we first assigned a random half of the smokers in our simulated sample to quitting smoking and then proceeded as above to compute each subject's 12 -year CHD risk under his/her smoking history (always smoking or never smoking).

## 7. Results

After exclusions, our analyses included 2230 men (47.8\%) and 2440 $(52.2 \%)$ women with 189 and 68 CHD events, respectively. Thus, the observed 12 -year risk of CHD in the study population was $8.48 \% ~(95 \%$ CI $7.37 \%-9.73 \%$ ) for males and $2.79 \%$ ( $95 \%$ CI $2.19 \%-3.54 \%$ ) for females.

The estimated 12-year risks of CHD (and 95\% CI) based on the parametric g-formula under several public health interventions are shown in Table 28.1. The simulated 12 -year risk of CHD under no intervention was $8.46 \%$ for males ( $95 \%$ CI $5.61 \%-11.30 \%$ ) and $2.82 \%$ for females ( $95 \%$ CI $2.15 \%-3.49 \%$ ). The risk ratios (and $95 \%$ CIs) for each intervention compared with no intervention are also shown in Table 28.1. For example, the estimated risk ratio among men is 0.81 ( $95 \%$ CI $0.73-0.90$ ) for all smokers quitting smoking at baseline vs no intervention. Note in row 6, we obtain the overall effect of combined interventions without explicitly imposing any particular functional form (e.g. multiplicative or additive) for the interaction of the individual interventions 1,3 and 4 a on the overall risk of CHD.

Table 28.2 displays the results when only one repeated measure of each risk factor (the most recent one) was included in the model for CHD risk. Results in Table 28.1 come from a model that includes the two most recent measures of each factor. The results in the two tables are compared and contrasted in the next section. Results of MSM analyses for the smoking intervention are given in Table 28.3 which displays the risk and risk ratio estimates from a MSM for two smoking interventions similar to interventions 1 and 2, respectively.

## 8. Discussion

We have presented two methods-parametric g-formula and MSMs-to estimate the causal effect of hypothetical public health interventions in the Framingham Offspring Study. In the absence of unmeasured confounders and model misspecification, these methods provide causal estimates from observational data that mimic the results of randomized experiments.

The dataset we used has two major limitations: (i) not all relevant confounders are available and (ii) the number of events, especially for women, is small. Because our methods are only valid when the assumption of no unmeasured confounders holds, we would not expect our estimates to be necessarily close to those of a randomized experiment.

Table 28.I G-formula estimates (reference analysis)

|  | Risk |  | Risk |  | BS | BS |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Intervention | (\%) | $95 \% ~ C l$ | ratio | $95 \% ~ C l$ | SE | av. risk |

Male, $\boldsymbol{n}=\mathbf{2} 230$

| 0) | No intervention (simulated results) | 8.46 | 5.61-11.30 | 1.00 |  | 1.45 | 8.72 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| I) | 50 | 7.65 | 5.02-10.27 | 0.90 | 0.86-0.95 | 1.34 | 7.86 |
| 2) | All quit smoking at base | 6.82 | 4.35-9.29 | 0.81 | 0.73-0.90 | 1.26 | 7.00 |
| 3) | Alcohol intake to specified distributio | 8.41 | 5.52-11.30 | 0.99 | 0.97-1.02 | 1.47 | 8.70 |
| 4a) | BMI lowered by $10 \%$ when BMI >22 | 8.09 | 4.57-11.61 | 0.96 | 0.74-1.24 | 1.80 | 8.39 |
| 4b) | BMI lowered to 22 when BMI $>22$ | 7.78 | 3.84-11.72 | 0.92 | 0.65-1.30 | 2.01 | 8.25 |
| 5) | LDL shifts to Chinese distribution | 5.80 | 3.35-8.25 | 0.69 | 0.57-0.82 | 1.25 | 6.17 |
| 6) | Combined interventions I, 3, and 4a | 7.26 | 4.00-10.53 | 0.86 | 0.66-1.12 | 1.67 | 7.54 |

## Female, $\boldsymbol{n}=2440$

| 0) | No intervention (simulated results) | 2.82 | 2.15-3.49 | 1.00 |  | 0.34 | 2.79 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| I) | 50\% quit smoking at baseline | 2.63 | 1.96-3.29 | 0.93 | 0.84-1.04 | 0.34 | 2.62 |
| 2) | All quit smoking at baseline | 2.44 | 1.65-3.22 | 0.86 | 0.68-1.09 | 0.40 | 2.46 |
| 3) | Alcohol intake to specified distribution | 2.83 | 2.15-3.51 | 1.00 | 0.95-1.06 | 0.35 | 2.81 |
| 4a) | BMI lowered by $10 \%$ when BMI $>22$ | 3.11 | 2.20-4.02 | 1.10 | 0.90-1.34 | 0.46 | 3.06 |
| 4b) | BMI lowered to 22 when BMI $>22$ | 3.23 | 1.94-4.51 | 1.14 | 0.82-1.59 | 0.66 | 3.23 |
| 5) | LDL shifts to Chinese distribution | 1.44 | 0.72-2.15 | 0.51 | 0.34-0.77 | 0.37 | 1.51 |
| 6) | Combined interventions I, 3, and 4a | 2.91 | 1.95-3.87 | 1.03 | 0.80-1.33 | 0.49 | 2.91 |

Key: BS SE, bootstrap standard error; BS av. risk, bootstrap average risk.

However, our methods gave coherent results, i.e. when applied to similar interventions, both methods yielded similar estimates. Such coherent results are some evidence against model misspecification but constitute no evidence against confounding by unmeasured factors as even when such unmeasured confounders exist both the parametric g-formula and MSM approaches are estimating the same association parameter. We now discuss the relative advantages and disadvantages of the parametric g-formula and MSMs.

### 8.1 Model misspecification

When using the parametric g-formula, gross model misspecification can be detected by comparing the observed risk and the estimated risk under no intervention. In our example, both risks are similar ( $8.48 \%$ vs $8.46 \%$ in men, $2.79 \%$ vs $2.82 \%$ in women). Note that, whereas dissimilar risks indicate model misspecification, similar risks cannot rule out the existence of model misspecification.

Even in the absence of unmeasured confounding, (i) valid estimation of the effect of the smoking intervention (1) with the parametric g-

Table 28.2 G-formula estimates (risk factor history lagged I visit only)

|  | Risk |  | Risk |  | BS | BS |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| Intervention | (\%) | $95 \% ~ C l$ | ratio | $95 \% \mathrm{Cl}$ | SE | av. risk |

Male, $\boldsymbol{n}=\mathbf{2} 230$

| 0) | No intervention (simulated results) | 8.46 | 5.55-11.36 | 1.00 |  | 1.48 | 8.72 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| I) | 50\% quit smoking at baseline | 7.63 | 4.92-10.34 | 0.90 | 0.86-0.95 | 1.38 | 7.85 |
| 2) | All quit smoking at baseline | 6.79 | 4.20-9.37 | 0.80 | 0.72-0.90 | 1.32 | 6.98 |
| 3) | Alcohol intake to specified distribution | 8.42 | 5.48-11.36 | 1.00 | 0.97-1.02 | 1.50 | 8.69 |
| 4a) | BMI lowered by $10 \%$ when BMI > 22 | 7.74 | 4.49-10.99 | 0.92 | 0.78-1.08 | 1.66 | 8.06 |
| 4b) | BMI lowered to 22 when BMI > 22 | 7.40 | 3.86-10.94 | 0.88 | 0.69-1.11 | 1.80 | 7.75 |
| 5) | LDL shifts to Chinese distribution | 5.80 | 3.24-8.36 | 0.69 | 0.57-0.82 | 1.31 | 6.11 |
| 6) | Combined interventions I, 3, and | 6.94 | 3.87-10.01 | 0.82 | 0.69-0.98 | 1.57 | 7.22 |
| Female, $\boldsymbol{n}=2440$ |  |  |  |  |  |  |  |
| $0)$ | No intervention (simulated results) | 2.80 | 2.15-3.45 | 1.00 |  | 0.33 | 2.74 |
| 1) | 50\% quit smoking at baseline | 2.57 | 1.94-3.21 | 0.92 | 0.83-1.02 | 0.32 | 2.54 |
| 2) | All quit smoking at baseline | 2.36 | 1.64-3.08 | 0.84 | 0.68-1.05 | 0.37 | 2.34 |
| 3) | Alcohol intake to specified distribution | 2.79 | 2.14-3.43 | 1.00 | 0.96-1.03 | 0.33 | 2.72 |
| 4a) | BMI lowered by $10 \%$ when BMI $>22$ | 2.76 | 2.05-3.48 | 0.99 | 0.83-1.18 | 0.36 | 2.69 |
| 4b) | BMI lowered to 22 when BMI >22 | 2.73 | 1.91-3.55 | 0.98 | 0.76-1.26 | 0.42 | 2.65 |
| 5) | LDL shifts to Chinese distribution | 1.36 | 0.82-1.91 | 0.49 | 0.35-0.68 | 0.28 | 1.36 |
| 6) | Combined interventions I, 3, and 4a | 2.53 | 1.82-3.24 | 0.90 | 0.73-1.12 | 0.36 | 2.48 |

Key: BS SE, bootstrap standard error; BS av. risk, bootstrap average risk.

Table 28.3 Estimates from marginal structural model

|  | Risk |  | Risk |  | BS | BS |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| Intervention | (\%) | $95 \% \mathrm{Cl}$ | ratio | $95 \% \mathrm{Cl}$ | SE | av. risk |

Male, $\mathbf{n}=\mathbf{2} 230$

| 0$)$ No intervention (simulated results) | 8.61 | $6.97-10.24$ | 1.00 | 0.83 | 8.55 |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| I) $50 \%$ quit smoking at baseline | 7.93 | $6.61-9.25$ | 0.92 | $0.85-1.00$ | 0.67 | 7.86 |
| 2) All quit smoking at baseline | 6.93 | $5.52-8.34$ | 0.80 | $0.65-1.00$ | 0.72 | 6.89 |

## Female, $\boldsymbol{n}=2440$

| 0$)$ No intervention (simulated results) | 2.57 | $1.96-3.19$ | 1.00 |  | 0.32 | 2.41 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| I) $50 \%$ quit smoking at baseline | 2.37 | $1.78-2.96$ | 0.92 | $0.78-1.09$ | 0.30 | 2.26 |
| 2) All quit smoking at baseline | 2.10 | $1.29-2.91$ | 0.82 | $0.55-1.22$ | 0.42 | 2.06 |

[^113]formula requires that one can correctly model both (a) the conditional probability of developing CHD conditional on risk factor history and (b) the conditional probability of current non-intervened on risk factors given past risk factor history; while (ii) valid estimation with MSMs requires that one (a) correctly model the conditional probability of the intervened on risk factor (smoking) given risk factor history and (b) the structural discrete time proportional hazard model. The parametric g-formula does not require that one correctly models the conditional probability of the intervened-on risk factor (smoking) given risk factor history because for the smoking intervention (1) the density $f_{d}\left(a_{j} \bar{l}_{i}\right.$, $a_{j}^{*}, \bar{a}_{j-1}, \overline{\mathrm{CHD}}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0$ ) does not actually depend on $a_{j}^{*}$; so no model for $f\left(a_{j}^{*} \mid \bar{l}_{i}, a_{j}^{*}, \bar{a}_{j-1}, \overline{C H D}_{j}=0, \bar{C}_{j}=0, \bar{D}_{j}=0\right)$ is required in (15) because the $a_{j}^{*}$ simply get summed out. Thus the models required for valid estimation of the effect of the smoking intervention (1) with the parametric $g$-formula have no overlap with the models required for valid estimation with MSMs. As a result, if estimates of the effect of the intervention agree under the two approaches, one can be somewhat confident that major model misspecification is absent. This represents a great advantage of having two independent methods of estimating the same intervention effect; each method can either reinforce or call into question the results obtained under the alternative method.

A subtle form of model misspecification when using the parametric gformula is due to collinearity among the repeated measures of the risk factors. Collinearity is not a problem for the prediction of CHD risk under no intervention but it may be a problem for predictions of CHD risk under certain interventions. For example, suppose that

1. in the logistic model for CHD risk at time 3, the odds ratio for log BMI at time 2 is 0.2 and the odds ratio for $\log$ BMI at time 1 is 5 , and because of collinearity among the repeated measures of BMI, neither estimate is significantly different from zero, but a chi-squared test on two degrees of freedom of the hypothesis that both coefficients are zero strongly rejects.
2. in the observed data, a subject has a quite elevated BMI of 29 at exams 1 and 2. Nonetheless, the predicted effect of the subjects BMI at exams 1 and 2 on CHD risk at exam 3 is null in the sense that the odds ratio comparing any level of BMI that is the same at the two exams with any other level of BMI also constant at the two exams on CHD risk at exam 3 is $0.2 \times 5=1.0$.

Under 1 and 2, the prediction of CHD risk for the subject under no intervention is quite stable (i.e. the confidence interval for the predicted probability is narrow). Now suppose we intervene to set BMI equal to 22 as soon as it is greater than 22 for times greater or equal than 2 (intervention 4 b ). The overall effect of BMI reduction is predicted to be
harmful since the odds ratio is $5^{\ln 29-\ln 22}=1.56$ compared to a subject with an observed BMI of 29 at both exams. However, in any given bootstrap sample, the odds ratios may be reversed due to sampling variability (because of the high correlation between BMI values over time). Therefore, in some bootstrap samples BMI would appear harmful, whereas in others BMI would appear to have a protective effect. The net result is the large variance we found for the risk (and risk ratio) of intervention $4 b$ and, to a lesser extent, of intervention 4 a (Table 28.1). The variance of the risk estimate for intervention 4 b was $39 \%$ greater than that of the risk estimate under no intervention.

To further assess issues of model misspecification, we conducted two separate sensitivity analyses: (i) we added a quadratic term to the linear term for $\log$ BMI in the CHD model, and (ii) we used cubic splines with 3 knots for log BMI in the CHD model. Both of these strategies slightly increased the variance of the estimates for intervention 4 a or 4 b (data not shown) and the risk ratio estimates did not change significantly.

The high variability arises because the correlation structure of the data is destroyed under the intervention. That is, there is no subject in the sample who has a BMI of 29 at occasion 1 and of 22 at occasion 2. Thus, it is not surprising and wholly appropriate that we are uncertain of what the result would be of an intervention that would create such subjects. Indeed, because there is no subject in the sample who has a BMI of 29 at occasion 1 and of 22 at occasion 2, the uncertainty we see in Table 28.1 for intervention 4 b is really an underestimate; it would have been more appropriate to admit that there was no data evidence with which to estimate the effect of such an intervention. Our apparent ability to estimate the intervention effect at all (albeit with great uncertainty) was wholly based on extrapolation under the, possibly incorrect, modelling assumption that the effect of BMI at the last 2 visits on CHD risk is linear on a logistic scale.

Some investigators, when faced with the problem of highly collinear repeated measures of exposures and the associated high variance of predicted interventions, decide to enter only one of the measures in their model, incorrectly arguing that the collinearity obscures the true exposure effect. To empirically examine some of the consequences of this flawed logic, we included only the most recent measure of each risk factor in the CHD model in Table 28.2 rather than the most recent two measures as in Table 28.1. The variance of the risk estimate for intervention 4 b was now only $24 \%$ greater than that of the risk estimate under no intervention. The difficulty, however, is that this estimate of the intervention effect can be badly biased if, in fact, the second most recent measure of BMI is in truth a risk factor for CHD controlling for the most recent BMI measure.

The lesson to be learned from the above discussion is that interventions that propose big and abrupt changes in the value of a risk factor that otherwise shows little variation over time (under no intervention)
will yield causal estimates with large variances and even these large variances will underestimate the actual uncertainty. This will be true regardless of the methodology used (i.e. MSMs or the parametric g-formula). Intervention involving these risk factors should be formulated in ways that do not imply big and abrupt changes. For example, reducing BMI by $10 \%$ (intervention 4 a ) resulted in a smaller variance than reducing BMI to the value 22 (intervention $4 b$ ).

### 8.2 Efficiency

The parametric g-formula yielded narrower confidence intervals for the risk ratios of interest than the MSM. This was expected based on underlying statistical theory.

### 8.3 Types of interventions

Marginal structural models are not useful to estimate risks under dynamic interventions (i.e. interventions that depend on the evolving values of time varying risk factors), such as interventions 4 a and 4 b . In those cases, the g-formula or structural nested models are needed.

A related subtle limitation of MSMs is that they cannot estimate the risk under no intervention in the presence of effect modification by time dependent non-intervened-on risk factors. This is so because IPTW estimation of a MSM effectively creates a pseudo-population in which the probability of receiving exposure does not depend on one's history of time dependent non-intervened-on risk factors. To clarify this point, we consider the following two scenarios:

1. Exposure is assigned to those who actually received it in the observed data.
2. Exposure is assigned to random subjects (in the same proportion as in the original population).

The proportion of CHD in both scenarios will be equal only in the absence of effect modification. A MSM provides estimates of CHD risk for scenario 2 but not for scenario 1. One consequence of this limitation is that, when using MSMs, we lack a way to detect gross model misspecification by comparing observed and estimated risk under no intervention. In our example, the observed and estimated risks under no intervention were close, so we assumed that effect modification was not an important issue and the relative risks in Table 28.3 are good approximations. Note that the no-intervention MSM results represent scenario 2 while the empirical CHD risk in the cohort represents scenario 1.

Because of these same reasons, we needed to slightly modify the meaning of interventions 1 and 2 when using a MSM. Rather than allowing non smokers to follow their own observed history after baseline (as was done in Tables 28.1 and 28.2), we forced them to remain as non-
smokers. This made little substantive difference as very few subjects initiated smoking after baseline.

### 8.4 Censoring

We defined as censored all subjects that were either lost to follow-up before the last study visit or died from competing causes (i.e. not CHD). Our causal estimates can be interpreted as the CHD risk under each intervention had censoring been abolished (i.e. had nobody been lost to follow-up or died from competing causes), provided there is no unmeasured confounding and censoring is ignorable in the sense that the conditional cause-specific hazard of being censored at $k$ among subjects alive and uncensored up to $k$ does not depend on the time $T$ to CHD given both $\bar{A}(k-1)$ and the time-dependent covariate $\bar{L}(k-1)$ history prior to $k$. However, one might be interested in the CHD risk had censoring by loss to follow-up, but not by competing causes of death, been abolished, especially because one may feel that the time $T$ to CHD were deaths from competing causes abolished is not a well-defined counterfactual quantity. For example, imagine a cohort of 100 people with 10 CHD cases, 20 deaths from other causes, and 70 subjects alive at the end of follow-up. If censoring from other causes was completely independent of failure and censoring by loss to follow-up, our CHD risk estimate would be $10 / 80=12.5 \%$ whereas the risk estimate had other causes of death not been abolished is $10 / 100=10 \%$. Of course, these two risks would only differ substantially if a large proportion of people died from competing causes, which is not the case in our data. We could easily adapt our methods to estimate the effect of interventions when deaths from other causes are included rather than eliminated.

In summary, estimation of non-dynamic interventions should be done both with MSMs and with the parametric g-formula, as agreement between the results is important evidence of lack of major model misspecification. MSMs are only useful to estimate the effect of non-dynamic interventions. The g-formula and structural nested models should be used to estimate the causal effect of dynamic interventions. In future work, we plan to compare $g$-formula and structural nested models based estimates of the effect of dynamic interventions in a manner similar to the comparisons of g-formula- and MSM-based estimates of nondynamic regimes reported here.

## 9. Conclusions

We have tried to show why it is that even with the high quality longitudinal data collected in the Framingham Offspring Study, valid estimation from observational data of the effect of considered interventions requires one to make a large number of unverifiable assumptions that will never hold exactly and may not even hold approximately. In most of the world and for many interventions of interest, relevant high quality
longitudinal data are unavailable, compounding the difficulties. Even if we succeeded in validly estimating the effect of simultaneous interventions on smoking, LDL, BMI and alcohol among the subjects in the Framingham Offspring Study, the question of how to extrapolate such results to other populations would remain unresolved. In this chapter we have stressed the large number of untestable assumptions that must be fulfilled to obtain valid estimates of causal effects and the improbability that all of them hold. We have not done so to discourage attempts to prioritize among potential interventions. Such prioritizing is, and will continue to be, done implicitly or explicitly every day. Our goal was simply to try to help the process by highlighting some of the central issues and potential biases that need to be carefully considered. It is a fact of life that deciding among potential interventions must be done-and must be done under great uncertainty. Methods for making decisions under uncertainty are well established, but the usefulness of such methods ultimately rests on an honest and comprehensive assessment of the uncertainties. It is our hope that this chapter will help in making such assessments.

## References

Greenland S, Pearl J, Robins JM (1999) Causal diagrams for epidemiologic research. Epidemiology, 10:37-48.
Hernán MA, Brumback B, Robins JM (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology, 11:561-570.
Hernán MA, Brumback B, Robins JM (2001) Marginal structural models to estimate the joint causal effect of nonrandomized treatments. Journal of the American Statistical Association-Applications and Case Studies, 96: 440-448.
Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite population. Journal of the American Statistical Association, 47:663-685.
Liang K-Y, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika, 73:13-22.
Manton KG, Stallard E, Singer B (1992) Projecting the future size and health status of the US elderly population. International Journal of Forecasting, 8:433-458.

Neyman J (1923) On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Statistical Science, 5:465-480.

Pearl J (1995) Causal diagrams for experimental research. Biometrika, 82: 669-710.

Robins JM (1986) A new approach to causal inference in mortality studies with sustained exposure periods-Application to control of the healthy worker
survivor effect. Mathematical Modelling, 7:1393-1512, with 1987 Errata in Computers and Mathematics with Applications, 14:917-921.

Robins JM (1987) Addendum to "A new approach to causal inference in mortality studies with sustained exposure periods-Application to control of the healthy worker survivor effect." Computers and Mathematics with Applications, 14:923-945, with 1987 Errata in Computers and Mathematics with Applications, 18:477.
Robins JM (1999) Marginal structural models versus structural nested models as tools for causal inference. Statistical models in epidemiology: the environment and clinical trials. IMA Volume 116. Halloran ME, Berry D, eds. Springer-Verlag, New York.

Robins JM, Blevins D, Ritter G, Wulfsohn M (1992) G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. Epidemiology, 3:319-336. See also the following link for the Errata to the article: Robins JM, Blevins D, Ritter G, Wulfsohn M (1993) Errata to G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. Epidemiology, 4:189. Reproduced with permission of the publisher, Lippincott, Williams and Wilkins.

Robins JM, Greenland S, Hu F-C (1999) Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. Journal of the American Statistical Association, 94:687-700.
Robins JM, Rotnitzky A, Scharfstein D (1999) Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Statistical models in epidemiology: the environment and clinical trials. Halloran ME, Berry D, eds. IMA Volume 116. Springer-Verlag, New York.

Rubin D (1978) Bayesian inference for causal effects: the role of randomization. Annals of Statistics, 6:34-58.

Spirtes P, Glymour C, Scheines R, eds (1993) Causation, prediction and search. Springer-Verlag, New York.

## Chapter 29

# Conclusions and directions for FUTURE RESEARCH 

Alan D. Lopez, Majid Ezzati, Anthony Rodgers, Stephen Vander Hoorn and<br>Christopher J.L. Murray

The analyses of risk factors within a common analytical framework and using comparable methods as outlined in these volumes has ensured greater consistency and comparability in evaluating and using scientific evidence on major risks to health. At the same time, data and knowledge gaps identified in the analyses of these risks illustrate key areas of scientific enquiry necessary to better inform policies and programs that aim to prevent disease by reducing risk factor exposure. The principal findings were discussed in individual risk factor chapters, as well as in those that presented summary results for individual risk factors (chapter 26) and for the joint effects of multiple risks (chapter 27). In this chapter, we use these findings to describe broadly how the analyses might affect pubic health practice as well as research on risk factors.

The analyses of the selected risk factors considered in this work, based on comprehensive reviews of available evidence on exposure and hazards, suggest that a small number of risks, such as childhood and maternal underweight and unsafe sex, accounted for a very large contribution to global loss of healthy life. Further, several risks, such as high blood pressure, tobacco and alcohol, have relative prominence in regions at all stages of development. While reducing all of the above risks to their theoretical minima may not be possible using current interventions, the results illustrate that disease prevention by addressing known distal and proximal risk factors can provide substantial, and under-appreciated, public health gains.

Treatment of established disease will always have a role in public health, especially in the case of diseases such as tuberculosis where treatment contributes to prevention. At the same time, the current devotion of a disproportionately small share of resources to prevention by reducing exposure to major known risk factors, through personal and nonpersonal interventions, should be reconsidered in a more systematic way in the light of this evidence.

Beyond their total hazard, the distributions of risks in a population have major implications for prevention strategies. Risk typically increases along a continuum of exposures. Dichotomous labels such as "hypertensive" and "normotensive" are therefore not a description of the health consequences of risks, but rather an operational convenience. In fact, the "deviant minority" (e.g. hypertensives) who are considered to be at high risk are only part of a risk continuum, rather than a distinct group, leading to one of the most fundamental axioms in disease prevention across risk factors: "a large number of people exposed to a small risk may generate many more cases than a small number exposed to high risk". Rose (1992) pointed out that wherever this axiom applies (see chapter 26 for possible exceptions), a preventive strategy focusing on high-risk individuals will deal only with the margin of the problem and will not have any impact on the large proportion of disease occurring in the large proportion of people who are at moderate risk.

While a high-risk approach may be more appropriate to the individuals and their physicians at any point in time, treating prevention as managing individual, high-risk crises can only have a limited effect at the population level and over long time periods. This is particularly relevant in the context of efforts to improve global health by addressing multiple diseases and risk factors, many of which exhibit continuous associations with disease outcomes. Focusing on high-risk individuals does not alter the underlying causes of illness, relies on having adequate discriminative ability to predict future disease, and requires continued and expensive screening for new high-risk individuals. In contrast, population-based strategies that seek to shift the whole distribution of risk factors have the potential to control population incidence. Such strategies aim to make healthy behaviours and reduced exposures the social norm and thus lower risk in the entire population.

Our exploration of the joint contributions of multiple risk factors suggests that 20 leading risks contributed to considerable loss of healthy life in different regions of the world. In particular, for some of the leading global diseases (e.g. acute lower respiratory infections, diarrhoea, lung cancer, ischaemic heart disease and stroke), substantial proportions were attributable to these selected risk factors. Removing the leading 20 risk factors from among those studied here would not only have resulted in a 9.3 -year ( $17 \%$ ) gain in global healthy life expectancy in 2000 , but also accounted for some of the interregional healthy life expectancy (HALE) differences. The analysis showed that even populations with currently high healthy life expectancy (e.g. developed regions of the Western Pacific and Europe) could further benefit considerably from risk reduction. These results provide a guide for achieving potential gains in (healthy) life expectancy that have been estimated statistically from past trends (Oeppen and Vaupel 2002; Riley 2001) through disease prevention by reducing known risks. The results for multiple risk factors further emphasize that for more effective and affordable implementation of a
prevention paradigm, policies, programmes and scientific research should acknowledge and take advantage of the interactive and correlated role of major risks to health, across and within causality layers. This could be an important step in addressing health inequalities, many of which may arise from concentration of major risks among specific socioeconomic groups.

Health research has at times focused on topics which, while scientifically intriguing, have not always been motivated by broader population health consequences (Anonymous 2001; Gross et al. 1999; Horton 2003; Willet 2002). The collation of evidence on exposure and hazard for different risks in this book, and the existing data gaps have illustrated data and monitoring needs for better quantification of important risks.

Research needs include more detailed and better quality data on exposure to most risks, using exposure variables that capture the full distribution of hazards in the population. Important examples include detailed data on alcohol consumption volume and patterns, dietary and biological markers for micronutrients, and better indicators for physical activity, indoor air pollution and occupational risks, all of which were quantified using indirect measures with limited resolution.

Assumptions and extrapolations were also needed in quantifying risk factor-disease relationships because of gaps in knowledge about the impact of some important global risk factors, particularly in developing countries. Examples include limited quantitative assessment of the hazards of specific sexual behaviours for HIV/AIDS or other sexually transmitted diseases, alcohol drinking patterns (Puddey et al. 1999) or exposure to indoor smoke from solid fuels (Ezzati and Kammen 2002). Equally important are detailed exposure data for risks that have been traditionally studied in developed countries, but have global importance and require more detailed data and hazard quantification in developing regions (e.g. alcohol and obesity).

The limited evidence on the effects of multiple risk factors and risk factor interactions also points to important gaps in research on multirisk and stratified hazards as also discussed in chapters 27 and 28. Including multiple causes in epidemiological research and risk assessment would allow estimating the benefits of reduction in combinations of distal and proximal exposures using multiple interventions (e.g. using education and economic tools to: i) promote physical activity or healthier diet coupled with screening and lowering cholesterol; and ii) address overall childhood nutrition and environment instead of a focus on individual components). In such research, risk factor groups should be selected based on both biological relationships and socioeconomic factors that affect multiple diseases, as discussed in chapter 27. Examples include those risk factors that are affected by the same policies and distal socioeconomic factors (e.g. malnutrition, unsafe water, sanitation and hygiene, indoor smoke from solid fuels and rural development policies) or affect the same group of diseases (e.g. all of the above for child-
hood infectious diseases; smoking, diet and physical activity for cardiovascular diseases). Once risk factors have been selected, the emphasis on reducing confounding should be matched by equally important enquiry into independent and mediated hazard sizes that are stratified based on the levels of other risks.

This is a substantial research agenda, and one for which some progress has been made over the past decades. Yet public health policy needs demand that much greater priority is given to research that more reliably and relevantly identifies the potential for prevention in all countries, including information on exposures and risks among subpopulations, especially among the least well off. Such research will undoubtedly provide a more compelling basis for the massive increase in preventive efforts worldwide that is required if the potential for health gains identified in this book is to be realized.

## References

Anonymous (2001) The human genome, in proportion. The Lancet, 357:489.
Ezzati M, Kammen DM (2002) The health impacts of exposure to indoor air pollution from solid fuels in developing countries: knowledge, gaps, and data needs. Environmental Health Perspectives, 110:1057-1068.
Gross CP, Anserson GF, Powe NR (1999) The relation between funding by the national institutes of health and the burden of disease. New England Journal of Medicine, 340:1881-1887.
Horton R (2003) Medical journals: evidence of bias against the diseases of poverty. The Lancet, 361:712-713.
Oeppen J, Vaupel JW (2002) Broken limits to life expectancy. Science, 296: 1029-1030.
Puddey IB, Rakic V, Dimmitt SB, Beilin LJ (1999) Influence of pattern of drinking on cardiovascular disease and cardiovascular risk factors-a review. Addiction, 94:649-663.

Riley JC (2001) Rising life expectancy: a global history. Cambridge University Press, Cambridge.
Rose G (1992) The strategy of preventive medicine. Oxford University Press, Oxford.
Willet WC (2002) Balancing life-style and genomics research for disease prevention. Science, 296:695-698.


[^0]:    a A: very low child mortality and very low adult mortality; B: low child mortality and low adult mortality; C : low child mortality and high adult mortality; D: high child mortality and high adult mortality; E: high child mortality and very high adult mortality. High-mortality developing subregions: AFR-D, AFR-E, AMRD, EMR-D and SEAR-D. Low-mortality developing subregions: AMR-B, EMR-B, SEAR-B, WPR-B. Developed subregions: AMR-A, EUR-A, EUR-B, EUR-C and WPR-A. This classification has no official status and is for analytical purposes only.

[^1]:    Example is applicable in illustrating gradual reduction in risk after exposure is reduced. The exact functional form of risk reduction may not necessarily be linear or exponential.
    For simplicity of notation, in all these cases we assume that: i) $L=0$. Including a lag is straightforward and can be done by replacing $t$ with ( $t-L$ ) in the corresponding formulas; and ii) there is no threshold for exposure. Including the threshold level is also straightforward using the $\operatorname{TRUE}(x(z) \geq X)$ function.

    In scenarios I, 3 and 4, where the effects of past exposure are absent or decline over time, risk reversibility can take place if exposure is reduced or removed. In scenario I there is
    immediate risk reversibility; in scenario 3 , there is full reversibility after time $K$; in scenario 4 , risk reversibility asymptotically approaches $100 \%$. In scenario 2 there is no risk reversibility and the effects of past exposure remain for an indefinite period.

[^2]:    a Dichotomous estimate.

[^3]:    a Multi-categorical estimate.

[^4]:    No data.
    PEM Protein-energy malnutrition.
    a All crude relative risks reported in Table 2.10 represent risk associated with weight-for-age below the cut-off point compared to risk associated with weight-for-age above the cut-off
    djusted for age, sex, father's education, household crowding and other factors. Crude and adjusted ORs are based on WAZ $<-2$ relative to WAZ $>-1$. Adjusted for age. Adjusted risk ratio is based on WAZ $<-2$ relative to WAZ $\geq 0$.

[^5]:    Table adapted from Rice et al. (unpublished).
    b SD according to North Indian reference.

[^6]:    a AMR-A, EUR-A and WPR-A, comprising the developed countries, are estimated to move further towards overweight and thus the underweight levels are forecasted to decrease stepwise to reach $0 \%$ by 2030. In WPR-A the estimated trend is driven by Japan (comprising $77 \%$ of the population aged 0-4 years in this subregion) and a national survey from 1978-8I already resulted with a low prevalence of underweight (3.7\%). Similarly the 1996 national survey in Australia (contributing 16\% of the population aged $0-4$ years in this subregion) reported $0 \%$ underweight. Appendix $C$ lists by country in alphabetical order publications backing the estimated trend towards overweight in AMR-A, EMR-A and WPR-A.
    b There are very scarce empirical data on the nutritional status of children aged $<5$ years in EUR-B and EUR-C. Based on the little information available to date (see Appendix D) the estimated trends of underweight until 2030 for these two subregions are expected to stagnate and decline, respectively. These overall trend estimations for EUR-B and EUR-C, however, are to be taken with caution.

[^7]:    a Where no authors are listed, documents have been written by multiple authors such as organizations, institutions and governments.
    b Survey data have been reanalysed either by responsible national authorities or by WHO.

[^8]:    Key: $\quad \mathrm{Hb}$, haemoglobin; EP, erythrocyte protopophyrin; BW, birth weight; IDA, iron deficiency anaemia; CNS, central nervous system; DQ, developmental quotient; IM, intramuscular; PEM protein-energy malnutrition; VMI, visual-motor integration; MDI, Mental Development Index.

    Inclusion criteria: IQ/MDI/General Cognitive Index ( GCI ) as outcome measure.
    Source: Table adapted from Grantham-McGregor and Ani (2001), with permission.

[^9]:    a Odds ratio for perinatal mortality associated with a $\mathrm{I} \mathrm{g} / \mathrm{dl}$ improvement in haemoglobin concentration, in the range of $5-12 \mathrm{~g} / \mathrm{dl}$ haemoglobin.
    b Random effects estimate.

[^10]:    a Episode refers to an episode of diarrhoea as per the case definition in the individual study.
    b Follow-up refers to the total number of child-days of follow-up or disease surveillance in each study.

[^11]:    Source: APCSC data (APCSC, personal communication, 2001).

[^12]:    a Name of country correct at time of study.

[^13]:    No data.
    $\downarrow$ Decreased.

[^14]:    Source: Reprinted, by permission of the publisher, from APCSC (2003a) Blood pressure and cardiovascular disease in the Asia-Pacific region. Journal of Hypertension, 21:707-716.

[^15]:    Source: Reprinted, by permission of the publisher, from APCSC (2003a) Blood pressure and cardiovascular disease in the Asia-Pacific region. Journal of Hypertension, 21:707-716.

[^16]:    a The same SBP levels apply to those aged $\geq 80$ years.

[^17]:    ? Specific details not given.

[^18]:    a Russia in original publication.

[^19]:    Key: $\uparrow$, raised; $\downarrow$, decreased.

[^20]:    Source: Reprinted, by permission of the publisher, from Law et al. (1994b) By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? British Medical Journal, 308:367-372.

[^21]:    No data.

[^22]:    Key: $\uparrow$, increase; $\downarrow$, decrease; $\rightarrow$, plateau.
    a Assumes the same absolute change in cholesterol for all age and sex subgroups.

[^23]:    a The same cholesterol levels apply to those aged $\geq 80$ years.

[^24]:    a The same cholesterol levels apply to those aged $\geq 80$ years.

[^25]:    Data unavailable.
    Specific details not given.
    Russia in original publication.

[^26]:    a Based on mean cholesterol levels for individuals aged 60-69 years.

[^27]:    - No data.

[^28]:    Note: These data points relate to population surveys from the countries listed in Table 8.15. The linear regression is shown together with the $95 \%$ confidence intervals of the prediction.

[^29]:    a Overweight only.

[^30]:    Data analysed by l-year age groups in childhood.

[^31]:    Source: Adapted from Shaper et al. (1997).

[^32]:    IHD Ischaemic heart disease.
    a Weighted by person years of follow-up. Total person-years of follow-up is 2148354.

[^33]:    IHD Ischaemic heart disease.
    Studies recording 500 or more cases of IHD.

[^34]:    - No data.

    Notes: The odds ratios relate to the probability of death, with the lowest mortality for any BMI group taken as 1.0 , the odds of the other groups then being calculated as the log of the increased mortality ratio. A $z$-score of $\geq 1.65$ is evidence for a statistically significant increase ( $P<0.05$, onetailed). Note that the minimum mortality occurred at a BMI of $27 \mathrm{~kg} / \mathrm{m}^{2}$ when there was a $10-$ year follow-up, but at $24 \mathrm{~kg} / \mathrm{m}^{2}$ with a 30 -year follow-up. Men and women, smokers and non-smokers, USA and non-USA studies included.
    Source: Data are from Table 4 of Troiano et al. (1996).

[^35]:    Key: Arrow denotes the BMI for minimum risk.

    * Signifies a statistically different risk from that at BMIs of $19-22 \mathrm{~kg} / \mathrm{m}^{2}$.

    Source: Data reformatted by Stevens from Stevens et al. (1998).

[^36]:    a Grams/person per day.

[^37]:    Null Association not statistically significant.
    Inverse Statistically significant protective association of high vs low fruit/vegetable consumption. - No data

    - No data

[^38]:    Key: I, NHS/HPFS; 2, EPIC-Norfolk study; 3, Finnish Mobile Clinic Health Examination study; 4, Massachusetts Health Care Panel study.

[^39]:    Source: USA: Ham (2001a); China: Ham (2001b); Finland: Finbalt data (Luoto et al. 1998).

[^40]:    Note: Zero years represent current smokers. The estimates at 40 years represent cessation of more than 35 years in the subjects. Data are from ACS CPS-II 1998 follow-up.
    Source: American Cancer Society, unpublished data, 2002.

[^41]:    Note: The vertical axis scales are different in the different panels of the figure to increase resolution. The figures in parentheses next to the legend indicate estimates of the subregional prevalence.
    Source: Ezzati and Lopez (2003).

[^42]:    a The exposure variable is smoking impact ratio (SIR), divided into three categories: I) SIR = 0; 2) $0<S I R \leq 0.5 ; 3) 0.5<S I R \leq 1.0$ as described earlier.
    b No risk is estimated for those aged $<30$ years.

[^43]:    a Developing countries include those in AFR, AMR-B, AMR-D, EMR, SEAR and WPR-B subregions.
    b Industrialized countries include those in AMR-A, EUR and WPR-A.
    c Figures in parentheses indicate the proportion of overall disease burden in each category attributable to smoking.

[^44]:    a Independent of intoxication or dependence.

[^45]:    a Prevalences are rounded to full percentages. Thus, a value of zero does not necessarily indicate that there are no people in a certain category, but may indicate a prevalence of less than 0.005 or $0.5 \%$.
    b Drinking categories are defined as follows:

[^46]:    a Pattern values were defined as follows:
    Risk factor: alcohol—second dimension: patterns of drinking.
    Units: range between I and 4; originally derived from optimal scaling and then based on addition of values of each pattern component (see Appendix A for details); subregional averages are populationweighted country averages.

    ## Definitions of categories of risk factor levels

    Level I: based on score of initial pattern components, with values in the lowest quartile reflecting least detrimental patterns of drinking such as least heavy drinking occasions, drinking with meals, no fiesta drinking and least drinking in public places.
    Level 2: based on score of initial pattern components, with values in the second lowest quartile.
    Level 3: based on score of initial pattern components, with values in the second highest quartile.
    Level 4: based on score of initial pattern components, with values in the highest quartile reflecting detrimental patterns such as many heavy drinking occasions, drinking outside meals, high level of fiesta drinking and drinking in public places.

[^47]:    Unstandardized residuals are the raw residuals as indicated by formula I above.

[^48]:    a This corresponds to coefficient $\beta_{3}$ in the model shown in the footnote.
    Level-I model
    Mortality rate $=\beta_{0}+\beta_{1}{ }^{\text {a }}(\mathrm{YEAR})+\beta_{2}{ }^{\text {a }}$
    $\left(\mathrm{GNP}\right.$-PC) $+\beta_{3}{ }^{\text {a }}$ (PC_ALCOHOL) $+\varepsilon$
    Level-2 model
    $\beta_{0}=\gamma_{00}+\mu_{0}$
    $\beta_{1}=\gamma_{10}+\mu_{1}$
    $\beta_{2}=\gamma_{20}$
    $\beta_{3}=\gamma_{30}+\gamma_{31}{ }^{\mathrm{a}}$ (PATB) $+\gamma_{32}{ }^{\mathrm{a}}($ PATC $)+\gamma_{33}{ }^{\mathrm{a}}$ (PATD) $+\mu_{3}$
    where
    mortality $=$ age-standardized IHD mortality
    YEAR = year of observation
    GNP_PC = GNP per capita
    PC_ALCOHOL $=$ per capita adult ( $\geq 15$ years) alcohol consumption
    PATB = dummy variable for pattern 2
    PATC = dummy variable for pattern 3
    PATD = dummy variable for pattern 4.

[^49]:    a Patterns of drinking were entered grand-mean-centred into the equation. Thus, the effect of per capita consumption displayed here is the effect at the grand mean of pattern, which is 2.51 in this sample.

[^50]:    - No data.

    Ranges refer to age-specific AAFs; minimum ( $>0$ ) and maximum estimates are shown.

[^51]:    - No data available, assumed to be negligible.

[^52]:    - No data.

[^53]:    - No data.

[^54]:    - No data.

[^55]:    ${ }^{\text {a }}$ Figures in bold: the median estimates from Table 13.9.

[^56]:    a Sum of the median estimates of the following four causes: AIDS, opioid overdose, suicide via opioids and trauma.
    b Median estimates of all-cause mortality derived from SMR analyses and pooled CMRs.

[^57]:    a Estimate includes the results of an Indonesian survey of ever-married women.
    b Estimate based on DHS of ever-married Nepalese women.
    c Estimate calculated from published medians reported for Polish men in different age groups, only that for men aged $30-44$ years is complete.
    Note: Extrapolated estimates are given in the shaded cells.

[^58]:    a The percentage fecund among the non-users decreases from $40 \%$ to $32 \%$.

[^59]:    a Included in this analysis.
    b Considered to be $100 \%$ due to unsafe WSH.

[^60]:    a Low data coverage.
    Source: Based on data from the Global water supply and sanitation assessment 2000 (WHO/UNICEF/WSSCC 2000), assuming that improved water supplies are most likely to have sanitation coverage.

[^61]:    Source: Map provided by Kiran Dev Pandey, World Bank.

[^62]:    a Slope of the concentration-response (CR) function for air pollution and mortality.
    b Results form regression models in which annual average concentrations measured from 1979-1983 were used as estimates of exposure (Pope et al. 2002).
    ${ }^{c}$ Results form regression models in which the average of annual average concentrations measured from 1979-1983 and 1999-2000 were used as estimates of exposure (Pope et al. 2002).
    ${ }^{d}$ Results from regression models where exposure (i.e. annual average $\mathrm{PM}_{2.5}$ ) is specified on the log scale.

[^63]:    a Dependent variable is the percentage of households using solid fuels.

[^64]:    Source: Adapted from Jere D. Haas' schematic diagram of the causal pathway for indoor cooking smoke and birth weight (Smith et al. 2000).

[^65]:    NA Not applicable.

[^66]:    a These projections only address changes in biomass use, i.e. for India and China, rates of coal use are not predicted to decline in the same manner. Indeed, recent trends in China indicate that coal is being substituted by gas in urban households, but is substituting for biomass in many rural households (Fridley et al. 2001).
    b Children's exposures differ from adult exposures at present in that they are modified by a different ventilation factor, since adults experience the health effects of exposures that took place before improvements in ventilation occurred. In the future, child and adult exposures will converge.
    c We assumed that the Chinese improved-stove programme would reach $90 \%$ penetration for biomass but that rates of coal use would not decrease (Goldemberg et al. 2000). When estimating exposure, the ventilation factor for China was therefore fixed at 0.25 for both adults and children, making the exposures of these two groups the same.

[^67]:    Source: Map based on data provided by M.P. Walsh

[^68]:    NA Not applicable.
    a Children aged 0-4 years.
    b Applied to adults aged 20-79 years.
    BPb : blood-lead concentration.

[^69]:    Source: Climatic Research Unit, Norwich, England.

[^70]:    a All natural disaster outcomes are separately attributed to coastal floods and inland floods/landslides.

[^71]:    a Note that for SEAR-B, there are no suitable studies of the relationship between temperature and cardiovascular disease-specific mortality; they are therefore based on relationships with all-cause mortality.
    b $2=s 550,3=$ s $750,4=$ unmitigated emissions.

[^72]:    $2=s 550,3=s 750,4=$ unmitigated emissions.

[^73]:    a $\quad 2=s 550,3=s 750,4=$ unmitigated emissions.

[^74]:    a $2=s 550,3=s 750,4=$ unmitigated emissions.

[^75]:    a $2=s 550,3=s 750,4=$ unmitigated emissions.

[^76]:    a Results of the simplified version of the HadCM2 model at medium $\left(2.5^{\circ} \mathrm{C}\right)$ sensitivity are shown for comparison.

[^77]:    A total of 597 publications were rejected on the basis of lack of information in abstract, or irrelevance to quantitative assessment (mainly reviews, clinical and laboratory studies)

[^78]:    a When data were not available for all countries, the percentage of the regional working age population ( $\geq 15$ years) represented by data is indicated. Some very small countries, e.g. Grenada, were not included in these calculations.
    Source: ILO (2002a).

[^79]:    Source: Calculated from CAREX (FIOH 1999).

[^80]:    ${ }^{2}$ Derived from major epidemiological studies.
    b Weighted summary relative risk, weighted by the proportion of workers exposed to each contributing carcinogen.

[^81]:    No data.

[^82]:    a Data taken from National Research Council (2001).
    b Data taken from Jin et al. (2000) (China).
    c Summary of data from Ory et al. (1997) (India) and Toroptsova et al. (1995) (Russian Federation).

[^83]:    a Asbestos exposure is the most important cause of mortality from mesothelioma. Cause-of-death statistics coded in ICD-IO allow direct estimation of the total number of mesothelioma deaths. Using this method, recent studies suggest that each year there are about 700 malignant mesothelioma deaths in Australia (Leigh and Driscoll 2003), 700 in Japan (Furuya et al. 2003), 2600 in the United States (Price and Ware 2004), and 4000 in Europe (Peto et al. 1999). A large proportion of these deaths are undoubtedly caused by asbestos exposure, primarily work-related. Combining estimates of asbestos exposure in all 14 subregions with hazards obtained from these studies would result in an estimate of more than 40000 mesothelioma deaths caused by asbestos exposure in the world. Of these preliminary estimates, about 9000 occur in developed countries (AMR-A, EUR and WPR-A), 9000 in SEAR-D, and 16000 in WPR-B. These estimates are subject to uncertainty, especially in developing countries where ICD-IO cause-of-death data and detailed data on history of asbestos exposure are not available. These preliminary estimates are currently undergoing further refinement by authors. Preliminary estimates also indicate that there may have been approximately 9000 deaths from silicosis, 7000 deaths from asbestosis and 14000 deaths from coal workers' pneumoconiosis as a result of exposure to occupational dusts (silica, asbestos and coal dust) in 2000.

[^84]:    a AF was set to zero for $\geq 80$.

[^85]:    Note the disposable syringes rinsed in the tepid water (arrow I) and the multi-dose medication vials (arrow 2).

[^86]:    NA Not applicable.
     stimates for the subregion with the second lowest proportion were extrapolated.
    J. Fitzner, personal communication, 2002.

[^87]:    ${ }^{\text {a }}$ Crude injection frequency estimates. For the purpose of the model, estimates for EUR-B and EUR-C were used after subtraction of the 10 top percentiles.

[^88]:    a AMR-A, EMR-B, EUR-A and WPR-A excluded as reuse of injection equipment is negligible in these subregions.

[^89]:    a AMR-A, EMR-B, EUR-A and WPR-A excluded as reuse of injection equipment is negligible in these subregions.

[^90]:    b Estimates of prevalence not provided for males and females separately. Estimate for all persons displayed in the table. Regression modelling used estimates derived from methods outlined in section 3.
    c Definition of CSA not provided, therefore assumed to be narrow.

[^91]:    Source: World Bank (2001a).

[^92]:    a The results for EMR-D should be interpreted with caution.
    b Summary of the subregions in the table, including the subregional-level estimates for EMR-D.

[^93]:    The results for EMR-D should be interpreted with caution.
    b Summary of the subregions in the table, including the subregional-level estimates for EMR-D.

[^94]:    a The results for EMR-D should be interpreted with caution.
    b Summary of the subregions in the table, including the subregional-level estimates for EMR-D.

[^95]:    ${ }^{\text {a }}$ Summary of the subregions in the table.

[^96]:    a Summary of the subregions in the table.

[^97]:    a Summary of the subregions in the table.

[^98]:    a Summary of the subregions in the table.

[^99]:    a Summary of the subregions in the table.

[^100]:    a Summary of the subregions in the table.

[^101]:    a For these three subregions, the combination of the marginal prevalence of indoor air pollution and our relative risk estimates meant that the algebraic solutions gave prevalence estimates above $100 \%$ and below 0\%. For example, the within US\$ I per day stratum algebraic solution for AMR-D was a cell percentage of $31.1 \%$ for those exposed to pollution and a cell percentage of $-13.7 \%$ for those unexposed. In these instances, the prevalence of exposure to indoor air pollution was set to $100 \%$.
    b Summary of the subregions in the table.

[^102]:    a For these four subregions, the combination of the marginal prevalence of indoor air pollution and our relative risk estimates meant that the algebraic solutions gave prevalence estimates above $100 \%$ and below $0 \%$. In these instances, the prevalence of exposure to indoor air pollution was set to $100 \%$.
    b Summary of the subregions in the table.

[^103]:    a Summary of the subregions in the table.

[^104]:    a Summary of the subregions in the table.

[^105]:    Summary of the subregions in the table.

[^106]:    Summary of the subregions in the table.

[^107]:    ${ }^{\text {a }}$ Summary of the subregions in the table.

[^108]:    a Summary of the subregions in the table.

[^109]:    No data.
    The actual percentage prevalence of the risk factor in the $>$ US $\$ 2$ per day reference category.

[^110]:    * For Turkey and Sri Lanka, data were missing for both electricity and rural/urban status and there were no suitable substitute variables. (See the Methods section for more details.) These two countries were not included in the final asset scores.

[^111]:    Note: The table shows the estimated mortality and disease burden for each risk factor considered individually. These risks act in part through other risks and act jointly with other risks. Consequently, the burden due to groups of risk factors will usually be less than the sum of individual risks (see chapter 27).

[^112]:    Portions of this chapter have been published previously in The Lancet, 2003, 362:271-280, and have been reproduced with permission from Elsevier Science.

[^113]:    Key: BS SE, bootstrap standard error; BS av. risk, bootstrap average risk.

