Cardiovascular survey methods

THIRD EDITION

Russell V. Luepker
Alun Evans
Paul McKeigue
K. Srinath Reddy

World Health Organization
Geneva
2004

WHO Library Cataloguing-in-Publication Data

Cardiovascular survey methods / Russell V. Luepker . . . [et al.]—3rd ed.

1.Cardiovascular diseases—epidemiology 2.Health surveys—methods 3.Epidemiologic methods I.Luepker, Russell V.

ISBN 92 4 154576 3 (NLM classification: WG 100)

TYPESET IN HONG KONG
PRINTED IN MALTA
2002/14796—SNPBest-set/Interprint—5000

Contents

Preface

There is a continuing need for surveys to establish the frequency, distribution and risk associations of cardiovascular diseases (CVD). They define in cross-section the individual and population characteristics related to disease, and they provide base data for follow-up of defined cohorts that allow hypotheses of disease causation and prevention to be tested. Surveillance, including population-based surveys, is also needed to establish time trends of CVD mortality and morbidity, which indicate the impact of medical care, preventive strategies and cultural change in the community.

Used intelligently and in conjunction with clinical and laboratory evidence, surveys provide evidence essential to causal inference. Surveillance provides information on the status of medical care, public health and cultural priorities. The two approaches are complementary and both require sound standard methods and procedures, carefully tested among real populations for reliability, validity, and feasibility.

This new edition of WHO's *Cardiovascular survey methods* goes beyond practical guidelines to detail different useful methods for data collection, editing, analysis and interpretation. Besides being an operational manual for surveys, it provides also the conceptual background and literature base for proposed research approaches and procedures. Above all, this edition brings both method and concept up to date; CVD epidemiology today has greater complexity, technicality and subtlety than it had in 1968 and 1982 when the first two editions of this manual were published. The international epidemiological investigators who compiled the new edition have brought to it new skills and viewpoints, and their rich and diverse experience.

The original edition of *Cardiovascular survey methods* had proved a useful guide over the last 30 years, mainly to clinicians drawn to the study of CVD in the world beyond the clinic. This new edition is a more complete source and critical reference for the many and varied health care professionals and support personnel now involved in cardiovascular research, in evaluation of health care effects and costs, in hospital and population surveillance of trends, and in treatment and prevention trials of new agents, instruments, and procedures. As such, this edition is most welcome.

Henry Blackburn.
Division of Epidemiology, School of Public Health, University of Minnesota, Minneapolis, MN, USA.

Zdenek Fejfar.
Department of Cardiology, Institute for Clinical and Experimental Medicine, Prague, Czech Republic.

Acknowledgements

The authors wish to express their appreciation to the late Professor Geoffrey Rose and Professor Henry Blackburn, authors of the initial WHO monograph on Cardiovascular Survey Methods, published in 1968, for their encouragement in updating their important monograph. They also wish to thank Ron Prineas and Richard Gillum who co-authored the second edition with Geoffrey Rose and Henry Blackburn. Ivan Gyarfas, Ingrid Martin, Shanthi Mendis and Porfirio Nordet of the World Health Organization provided consistent encouragement, editorial advice and support. Professor Frank Key, Professor R. Curtis Ellison and Dr Daniel McGee helped with several of the chapters. Finally, this monograph would never have seen daylight without the administrative support of Jill Berkas, Mary-Jane Watson, Janelle Willard and Shanthi Mendis.

The authors
January 2001

Acronyms and abbreviations

The following acronyms and abbreviations are used in this book:

ACE	angiotensin-converting enzyme
ACC	American College of Cardiology
AMI	acute myocardial infarction
ARIC	Atherosclerosis Risk in Communities
BMI	body mass index
CASS	Coronary Artery Surgery Study
CHD	coronary heart disease
CHF	congestive heart failure
CTR	cardiothoracic ratio
CVD	cardiovascular disease
DALY	disability-adjusted life year
DEXA	dual-energy X-ray absorptiometry
EBCT	electron beam computed tomography
ECHO	Echocardiogram
EDTA	ethylenediaminetetraacetic acid
ESC	European Society of Cardiology
HDL	high-density lipoprotein(s)
ICD	International Classification of Diseases
IDL	intermediate-density lipoprotein(s)
IHD	Ischaemic heart disease
LDL	low-density lipoprotein(s)
METS	metabolic equivalents
MI	myocardial infarction
MHS	Minnesota Heart Survey
MOOP	manual of operations
MONICA	monitoring trends and determinants in cardiovascular disease
MRFIT	multiple risk factor intervention trials
NIST	National Institute for Standards and Technology
OR	odds ratio
PABA	para-aminobenzoic acid
PVD	peripheral vascular disease
RF	rheumatic fever

RHD	rheumatic heart disease
SDR	standardized death rate
SMR	standardized mortality rate
VLDL	very-low-density lipoproteins(s)

Introduction

The worldwide burden of cardiovascular diseases is substantial. In most industrialized countries, cardiovascular diseases are the leading cause of disability and death. Developing countries, with previous low rates are now seeing increased rates as economies develop, infectious diseases are conquered and life expectancy improves. Therefore understanding the etiology origins, distribution and trends of CVD is essential to improving public health in all countries.

Cardiovascular diseases are present in many forms and have differing etiologies. However, among chronic cardiovascular diseases, those related to atherosclerosis, rheumatic fever, and hypertension comprise the greatest burden. These are diseases with insidious development and long latency periods, punctuated by acute clinical presentations such as myocardial infarction, stroke and heart failure. Often the first clinical manifestation of disease is sudden and unexpected death, while chronic vascular insufficiency, arrhythmia or heart failure are common presentations.

Epidemiology is central to understanding cardiovascular disease etiology, patterns and prevention. Mortality data provide an index of the health of a population; morbidity data define the burden of acute illness and chronic disability. Such measures of disease prevalence and incidence are important, but epidemiology provides many other useful insights. Among the more useful and powerful contributions are risk factor measures that predict disease in currently healthy individuals. These include environmental, genetic and individual factors, which provide critical information for disease prevention at both the individual and population levels.

Clinical aspects of cardiovascular diseases are also the subject of intense investigation and new understanding. Advanced technologies have wide application in the treatment of the major cardiovascular diseases, adding dramatically to health care costs. Finally, epidemiology contributes to the proper design of clinical trials to evaluate these new therapies, their costs and outcomes.

The more remarkable aspect in the whole area of CVD is that the patterns of cardiovascular risk factors and CVD rates are dynamic. In some countries, age-adjusted disease rates have fallen steadily as population levels of risk factors decline and advanced treatments are applied. In others, disease rates are rising as affluence and industrialization

increase. Many developing countries are going through an epidemiological transition phase with a decline in infectious diseases, rheumatic heart disease, and haemorrhagic stroke counter-balanced by a rise in coronary heart disease, congestive heart failure and thromboembolic stroke. These trends determine the degree of success or failure in control and prevention. Comparison among populations provides insight into the evolution of epidemics and their control in differing sociocultural circumstances.

There are several objectives to this third edition of *Cardiovascular survey methods.* Since the second edition (1982), the field has advanced with new diagnostic and therapeutic techniques, and new measurement methods (*1*). This volume attempts to provide a broad picture of cardiovascular disease epidemiology, including survey methods, experimental methods and new methods appropriate for use in developed and developing countries.

Like the previous edition, it seeks to be practical and to provide the methods and references necessary for undertaking population studies in a variety of circumstances. It provides a compendium of methods and forms (appendices 1–37) on a CD-Rom, which is included with this edition. It is not a substitute for a basic text in epidemiology and those interested in greater understanding of the field should consult traditional textbooks (*1–9*).

References

1. Rose G et al. *Cardiovascular survey methods*, 2nd ed. Geneva, World Health Organization, 1982 (WHO Monograph Series, No. 56).

2. Lilienfeld AM, Lilienfeld DE. *Foundations of epidemiology*, 2nd ed. New York, Oxford University Press, 1980.

3. MacMahon B, Pugh TF. *Epidemiology: principles and methods.* Boston, MA, Little, Brown & Co., 1970.

4. Mausner J, Kramer S. *Epidemiology: an introductory text*, 2nd ed. Philadelphia, Saunders, 1985.

5. Beaglehole R, Bonita R, Kjellstrom T. *Basic epidemiology.* Geneva, World Health Organization, 1993.

6. Hennekens C, Buring J. *Epidemiology in medicine.* Boston/Toronto, Little, Brown & Co., 1987.

7. Lilienfeld DE, Stolley PD. *Foundations of epidemiology*, 3rd ed. New York, Oxford University Press, 1994.

8. Kelsey JL et al. *Methods in observational epidemiology*, 2nd ed. New York, Oxford University Press, 1996.

9. Labarthe DR. *Epidemiology and prevention of cardiovascular diseases: a global challenge.* Gaithersburg, MD, Aspen, 1998.

1.
History of cardiovascular epidemiology

Introduction

In *Foundations of epidemiology* (*1*), Lilienfeld & Lilienfeld trace epidemiology from Daniel Bernoulli in the 18th century, through LaPlace Poisson and particularly Louis, to the "European students" of the 19th century. The European students include William Guy, John Simon, William Farr and William Budd. A parallel movement in the Americas included Elisha Bartlett, George Shattuck, Francis Delafield and Alonzo Clark. Observations on mortality and demographics had in fact been made earlier by William Petty and John Graunt (*2*). Petty was the father of "political arithmetic", a term that he coined to signify collection of data on population, education, revenue, disease, etc. For his part, Graunt constructed the first "life table" and formulated a principle stating that fashions in the nomenclature of disease can play havoc with mortality rates (*3*). The development of epidemiological methods was influenced from the 17th century by Francis Bacon, who developed inductive logic and by John Stuart Mill, who published *A system of logic* in 1843 (*1*, *4*). Ignez Semmelweis, with his astute observations in 1847 in Vienna on childbed fever and the effect of handwashing and John Snow with his 1849 study on the spread of cholera through the water supply, form the basis of modern hygienic and population interventions. The London Epidemiological Society, founded in 1850, was the first of its kind and focused on the control of waterborne infectious diseases—notably cholera (*5*).

Epidemiological techniques and their application for improving public health emerged in the mid-19th century in response to the rising levels of infectious diseases. Earlier on, however, the development of CVD epidemiology had relied on clinical observations. William Heberden, physician to the lexicographer Samuel Johnson, first coined the term "angina" in 1768 (*6*). According to Proudfit, the origin of the concept of ischaemic heart disease can be traced back to four doctors in the United Kingdom around the turn of the 19th century: Edward Jenner, Caleb Hillier Parry, Samuel Black, and Allan Burns (*7*). Black is mentioned in an article by John Warren entitled "Remarks on angina pectoris" in the January 1812 issue of the New England Journal of Medicine and Surgery reprinted in the sesquicentenary edition (*8*). In his book *Clinical and pathological reports* of 1819 (*9*), Black wrote that "the Physician

who ascertains half a dozen of important facts performs a more valuable, though a less splendid achievement, than he who invents a dazzling theory". Black truly did help develop a dazzling theory—the ischaemic hypothesis of angina pectoris.

A most remarkable feature of Black's work was his adoption of a population approach in asking how individuals who become ill differ from those who do not. Black, however, put it far more elegantly: "Is our knowledge of the remote causes of this disease such as to enable us to classify the liable and the exempt?" He went on to speculate, "I imagine the persons peculiarly liable are those who are of full and plethoric habits who live luxuriously, or at least very plentifully, and do not use a sufficient quantity of exercise. If there be, on the other hand, any persons possessing an exemption from the disease, total or partial, I think we shall be most likely to find them among the poor, the laborious, those who use strong exercise, the foot-soldier and the female sex." Similarly, "We have seen that the disease appears to be connected with a plethoric state of the system and with obesity—that the great majority of the subjects of it have belonged to better ranks of society, who were in the habit of sitting down every day to a plentiful table, in the pleasures of which they may have indulged to a greater extent than was suitable to the tendency of their constitution." (Black also recognized the genetic component of heart disease.)

By the end of the 19th century the battle against many infectious diseases was being won and life expectancy began to increase, resulting in a larger proportion of the population surviving to develop chronic noncommunicable diseases. Myocardial infarction syndrome was first described by an American, James Bryan Herrick, in 1912. Others claim that credit for the first description of coronary thrombosis, in 1910, belongs to two Russians, Obratzow and Straschesko (*10, 11*). In the early decades of the twentieth century the disease was still relatively uncommon; Osler stated that, among 10934 admissions to the Montreal General Hospital between 1900 and 1909, only six cases of angina pectoris were diagnosed (*11*). Things were to change: the reporting by Pardee in 1920 of the evolutionary electrocardiographic changes of myocardial infarction made the diagnosis more certain, and an epidemic of coronary heart disease began to emerge (*12*). This may have been to some extent an example of Graunt's thinking as, according to Bedford, the epidemic was due to a "far greater ability to recognize the disease, thanks to new knowledge and vastly improved means of diagnosis". Angina pectoris had been recognized for a considerable time and atheroma has in fact been reported in an Egyptian mummy (*13*). However, some authorities maintained that the increased incidence was due to a reclassification of disease (*14*).

Credit is due to Sir James Mackenzie for his pioneering work on heart disease. In 1918 he set up the St. Andrew's Institute for Clinical Research in Scotland (*15*). He founded a method for the investigation and prevention of common diseases that would also stimulate research into the earliest manifestations of disease, and believed that disease must be studied in its earliest forms and then observed over time. A pre-

liminary investigation, involving a study of the records of 1000 patients, focused on the different forms of ill health. Thus, Sir James was thinking in terms of the large numbers required for modern epidemiological studies (*16*).

The modern era

The modern era of cardiovascular epidemiology began after the Second World War with the establishment of a number of cohort studies, including the Framingham Heart Study, MA. Clinicians including White—who had spent time with Mackenzie—and Mountain persuaded Van Slyke to initiate the Framingham Heart Study. Commitment to the epidemiological approach to cardiovascular disease research at the new National Heart Institute was far from universal. Many then, as now, believed that the answer to most of the important questions regarding the natural history of atherosclerotic vascular disease would come from basic laboratory research, not from the study of the disease in populations. In any event, the Framingham Study heralded the modern era of the epidemiological study of cardiovascular disease, as did a parallel study at Newton delving into aspects of hygiene, a demonstration programme of which much less was subsequently heard (*17*, *18*). It was at Framingham that the term "risk factor"—a concept borrowed from the world of insurance—was introduced in a paper in 1957 (*19*). The first symposium on cardiovascular epidemiology was given a "leading position" at the Second World Congress of Cardiology in 1954, as it was "one of the most important of all subjects" (*20*).

The Seven Countries Study (*21*) was the first to compare cardiovascular disease incidence and risk factors using a common protocol in different international populations. That study found large differences in dietary fat intake, serum cholesterol and heart disease incidence (*21*, *22*). In his excellent and often humorous account *On the trail of heart attacks in seven countries*, Henry Blackburn records (*23*) that Ancel Keys got the idea during a sabbatical year at Oxford and related travels in 1951 and 1952, which "opened his eyes" to cultural differences in diet, behaviour and disease risk. The Seven Countries Study and other cohort studies developed methodology, questionnaires and validated instruments of measurement for future studies.

Post-mortem studies were also carried out, for example the International Atherosclerosis Project from 1960 to 1965 (*24*), which assessed in many countries the degree of atherosclerosis of the coronary arteries and aorta in more than 31 000 persons who died aged 10–69 years.

These and countless other studies have led to the present level of understanding in cardiovascular epidemiology.

References

1. Lilienfeld AM, Lilienfeld DE. *Foundations of epidemiology*, 2nd ed. New York, Oxford University Press, 1980:23–45.

2. Rosen G. *A history of public health.* Baltimore, MD, Johns Hopkins University Press, 1993:87–88.

3. Morris JN. *Uses of epidemiology*, 2nd ed. Edinburgh, Livingstone, 1964:15–16.

4. MacMahon B, Pugh TF. *Epidemiology: principles and methods.* Boston, MA, Little, Brown & Co., 1970:32–35.

5. Babington C. Presidential address to the Epidemiological Society. *Lancet*, 1850, ii:639–642.

6. Matthews MB. Historical background. In: Julian DG, ed. *Angina pectoris*, 2nd ed. Edinburgh, Churchill Livingstone, 1985:3–12.

7. Proudfit WL. Origin of concept of ischaemic heart disease. *British Heart Journal*, 1983, 50:209–212.

8. Warren J. Remarks on angina pectoris. *New England Journal of Medicine*, 1962, 266:3–7.

9. Black S. *Clinical and pathological reports.* Newry, N. Ireland, Alex Wilkinson, 1819:1–47.

10. Herrick JB. Clinical features of sudden obstruction of the coronary arteries. *Journal of the American Medical Association*, 1912, 59:2015–2020.

11. Osler W. The Lumleian lectures on angina pectoris (I). *Lancet*, 1910, i:697–702.

12. Harvey W. de circulatione sanguinis. KJ. Franklin's translation, Oxford, 1958.

13. Long AR. Cardiovascular renal disease: report of a case 3000 years ago. *Archives of Pathology*, 1931, 12:92–94.

14. Ryle JA, Russell WT. The natural history of coronary disease: a clinical and epidemiological study. *British Heart Journal*, 1949, 11:370–389.

15. Hay J. James Mackenzie and his message. *British Medical Journal*, 1993, 2:1033–1036.

16. Mackenzie J. An address on clinical research. In: *Reports of the St Andrews Institute. Vol I.* London, Henry Frowde and Hodder & Stoughton, 1922:11–29.

17. Dawber TR. *The Framingham Study: the epidemiology of atherosclerotic disease.* Cambridge, MA, Harvard University Press, 1980:14–29.

18. Massachusetts Department of Public Health. Study of heart disease in Massachusetts. *New England Journal of Medicine*, 1948, 239:31–32.

19. Schoenberger JA, Mann GV. Controversies in cardiology. Proposed: low-dose aspirin should be taken daily after age 40 if total serum cholesterol is greater than 160. *Hospital Practice,* 1982, 12:50A–50M.

20. World trends in cardiology. In: Keys A, White PD, eds. *Cardiovascular epidemiology.* New York, Hoeber-Harper, 1956:vii.

21. Keys A. *Seven countries: a multivariate analysis of death and coronary heart disease.* Cambridge, MA, Harvard University Press, 1980.

22. Epstein FH. Cardiovascular disease epidemiology: a journey from the past into the future. *Circulation,* 1996, 93:1755–1764.

23. Blackburn H. *On the trail of heart attacks in seven countries.* Middlesborough, MA, Country Press, 1995.

24. McGill HC. *Geographic pathology of atherosclerosis.* Baltimore, Williams & Wilkins, 1968.

2.
General principles

Concepts

The basic science underlying population studies of cardiovascular diseases is epidemiology. However, many other disciplines are involved, including those of the physical and social sciences. Epidemiology is defined by MacMahon as "the study of the distribution and determinants of disease frequency in populations" (*1*). By measuring population distributions, epidemiology seeks to understand the who, when and where of health and disease. Does the condition occur in men or women, old or young, rural or urban residents, winter or summer? Are there periods when the disease is more common (i.e. epidemic)? Epidemiology also seeks to understand the determinants of disease. It describes factors in the population that are associated with individuals becoming ill or, conversely, being protected from illness.

Two important assumptions underpin epidemiology. First, disease patterns and characteristics in populations are not random events but are describable and quantifiable. Second, much can be learned from focusing on the population; this distinguishes epidemiology from clinical medicine, which focuses on the individual patient.

Epidemiology has numerous and growing uses in the understanding of human health and disease. Traditionally, it seeks to understand the origins and causes of disease. This so-called etiological epidemiology determines environmental and individual characteristics that are associated with the onset or presence of disease. It evaluates characteristics in diseased individuals that help to describe the natural history of their condition, posing the basic question, "How does a person with a disease differ from one without it?"

Epidemiological methods are used to define the health status of populations. Characterization of the population's disease burden facilitates the rational allocation of health care resources. Success or failure in addressing the disease burden in the population can be gauged by evaluating the time trends in disease distribution. Moreover, epidemiological studies of the costs, uses and outcomes of various treatments provide information for health services research. Finally, epidemiological methods are crucial to disease control and prevention. By allowing an understanding of disease distribution and causal factors, they enable the development of methods for the prevention and control of disease.

The origins of epidemiology can be dated to the 19th century, when outbreaks of infectious diseases were common. Control of the major waterborne infectious diseases through public health measures resulted in a striking decline in their incidence. Further advances in public health and treatment of infection have improved health and resulted in increased life expectancy. However, chronic diseases have now become the leading health problems in much of the world as the population lives to an age when these diseases are common. These are diseases with insidious onset, long latent periods, serious clinical manifestations and substantial prolonged morbidity. Many of these diseases, including various cancers, degenerative joint diseases and chronic lung diseases, are common. Cardiovascular diseases, particularly those of atherosclerotic origin, rheumatic heart disease and hypertension, are among the most widespread of the chronic conditions. They are so frequently observed as a cause of disability and death that they can be reasonably called "epidemic" in many countries.

Surveys to determine the distributions, frequencies and determinants of these diseases and their trends are essential to developing prevention and control programmes. Survey methods that characterize populations, disease burdens and trends in these disease conditions form the basis of this book. As a bridging science, epidemiology borrows from basic scientific laboratory methods on pathophysiological processes and from clinical observation of sick individuals to ascertain population disease patterns. However, it also contributes to basic and clinical sciences by making observations in human populations that require further research and elaboration in the laboratory or the clinic.

Epidemiology is particularly linked with concepts of exposure and risk. An understanding of factors that increase or decrease the likelihood of ill-health or disease outcomes leads naturally to interventions for prevention and control.

There are also limits to epidemiology: patterns and outcomes observed in populations do not predict outcomes in individuals. Using statistical methods, epidemiologists can assess the probabilities of a disease occurring or progressing with certain exposures, treatments, or risk characteristics. For example, an epidemiologist can say that patients with an acute myocardial infarction and heart failure have a 25% one-year mortality rate but they cannot state whether or when an individual patient with that combination of conditions will die.

Objectives of cardiovascular disease surveys

Overview

Cardiovascular disease surveys and survey methods are important to the understanding of many aspects of disease and health in human populations. The methods currently used go well beyond describing disease distributions and patterns in the population. Epidemiology also looks at associations to elucidate disease etiology, the

natural history of disease, gene–environment interactions and cost–benefit ratios of medical interventions. Furthermore, the rigorous definitions and diagnostic criteria developed for cardiovascular surveys are also important in experimental epidemiology (e.g. intervention trials).

Frequency and distributions

Cardiovascular disease surveys describe the frequency and distribution of diseases in defined populations. To understand the distribution of a disease it is important to ask the questions who, where and when. Information on both frequency and distribution of disease in the population leads naturally to the observation of disease patterns, which in turn leads to hypotheses on etiology, prevention and treatment. As described later, the study of disease frequency and distribution requires systematic definitions of the disease (cases) and of the population studied.

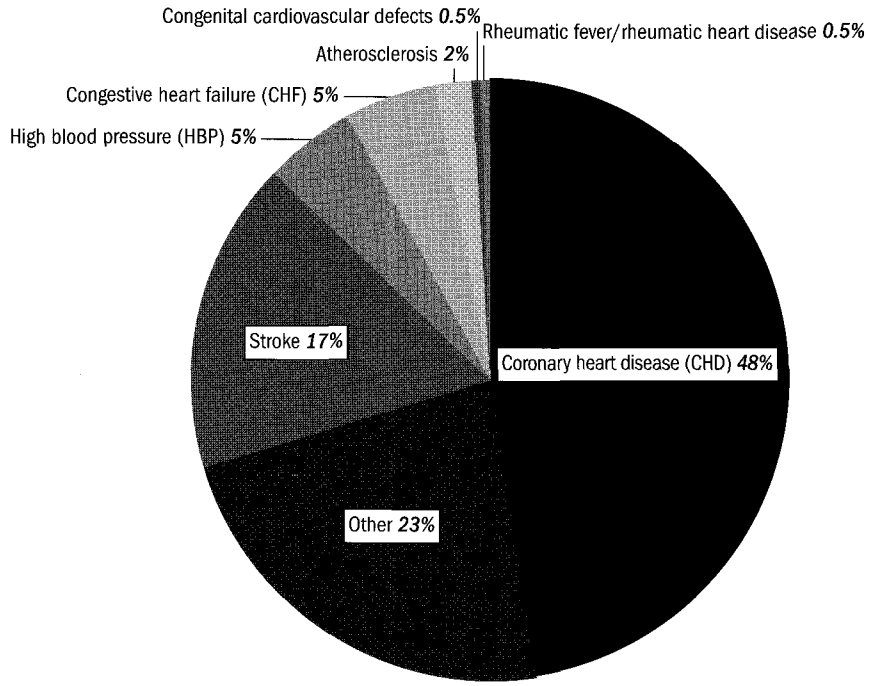
Figure 2.1 shows the distribution of mortality for cardiovascular diseases compared with other causes of death in the population of the United States of America. It represents a use of frequency data derived from death certificates. Figure 2.2 shows the distribution of cerebrovascular accidents in different geographical areas of the USA, with higher prevalence in the south-eastern "stroke belt". Information such as this leads to studies of etiology and allocation of health care resources for prevention and control.

Associations

An understanding of disease distributions in populations can prompt a search for associations or factors that may be related to these patterns. Finding associations is another central task in epidemiological surveys. While associations should not be confused with causation, they can, along with other scientific evidence, provide important information for establishing a causal connection. This is discussed in greater detail elsewhere.

Epidemiological surveys usually include the collection of information on factors already known or thought to be associated with health or disease. Five major areas are usually considered: demographic, physical, behavioural, environmental and genetic. Demographic factors include age, sex, race and socioeconomic status, and are usually ascertained by interview or review of written records. Physical characteristics—height, weight, blood cholesterol, blood pressure, urinary sodium levels and exercise capacity, for example—are usually measured directly. Behavioural characteristics include cigarette smoking, physical activity patterns and eating patterns; and environmental factors can include not only substances in the air and water, but also community factors such as social support. Increasingly, there is interest in information on family members and gene polymorphism as a means of characterizing an individual's genetic make-up.

Fig. 2.1. Percentage breakdown of deaths from cardiovascular diseases, USA, 1998



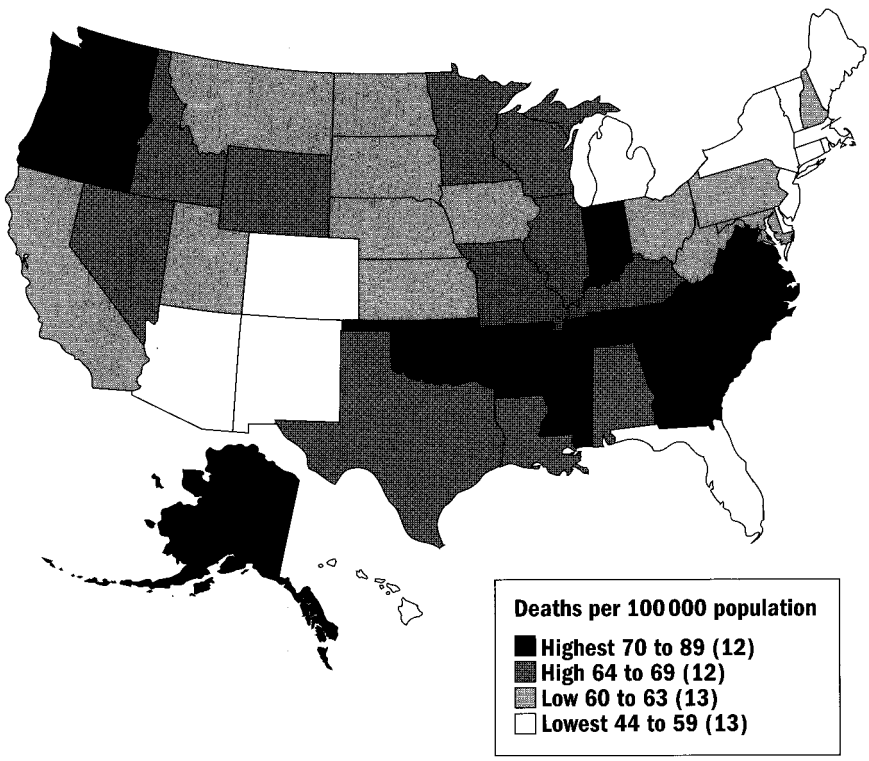
Source: American Heart Association. 2001 Heart and Stroke Statistical Update. Dallas, Texas: American Heart Association, 2000.

An example of the association of CVD mortality with age and sex is shown in Figure 2.3. As with most chronic diseases, frequency increases with age. Similarly, the association of CVD mortality with blood cholesterol levels by age, measured when subjects are still healthy, is shown in Figure 2.4. When associations found in epidemiological studies with factors such as serum cholesterol levels are confirmed by clinical and experimental data, the label "risk factor" is appropriate. A risk factor is a characteristic associated with a disease outcome. It may be positively associated, such as age with increased risk of coronary heart disease (CHD), or negatively associated, such as physical activity with lowered risk of CHD. Risk factors that are part of the direct causal chain for CVD may act in a proximal way, promoting the disease process directly or indirectly through other characteristics.

Natural history of diseases

Epidemiological techniques are also commonly employed to elucidate the natural history of a disease. For example, individuals who may be at high risk but are

Fig. 2.2. Age-adjusted death rates[a] for stroke by state, USA, 1995–1997



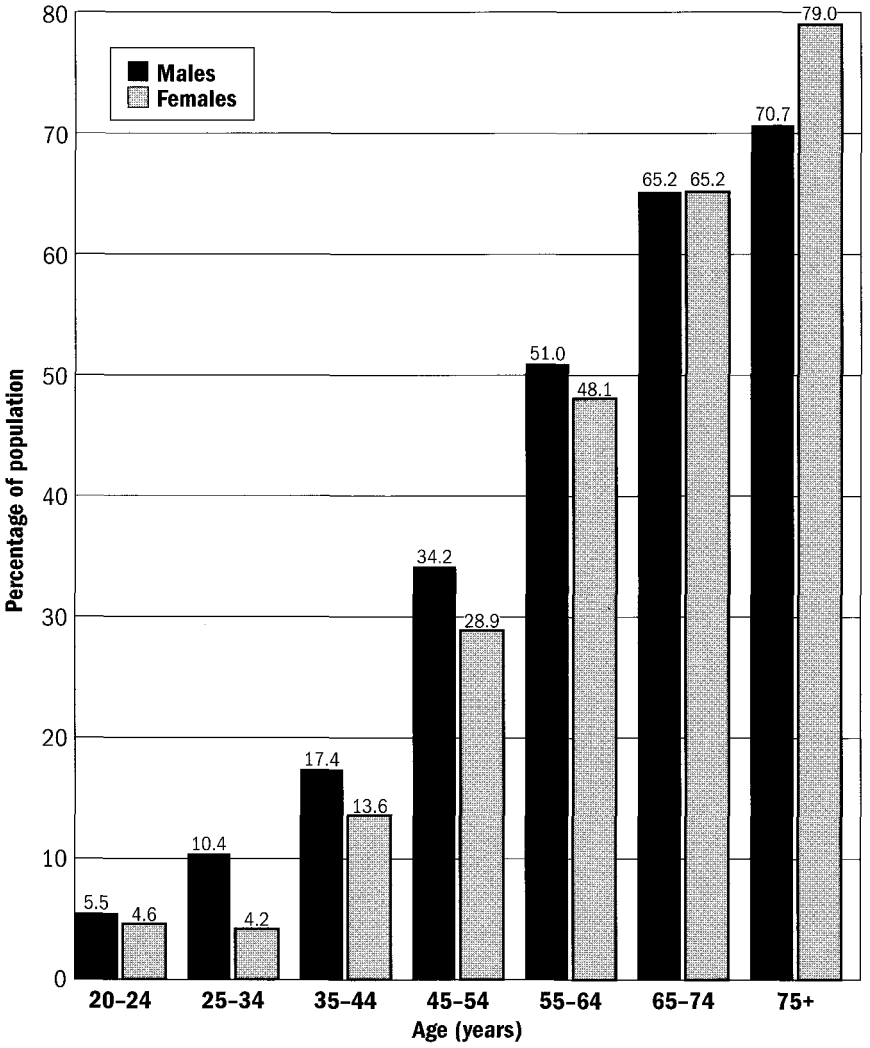
[a] Age-adjusted to the 2000 standard.
Source: National Heart, Lung and Blood Institute (NHLBI), Morbidity and Mortality: 2000 Chart Book on Cardiovascular, Lung, and Blood Diseases.

clinically disease-free are assessed at baseline and then evaluated at regular intervals thereafter for clinical signs and symptoms of the disease and its natural progression. Such an approach is less frequent today because individuals at elevated risk must be treated if acceptable treatments are available, and this can alter the natural history of the disease. Patients who are at high risk or have clinical manifestations and who are receiving treatment are assessed at baseline and then followed at regular intervals to determine the effects of treatment on outcomes. The goal of treatment is to alter the natural history of the disease.

Trends

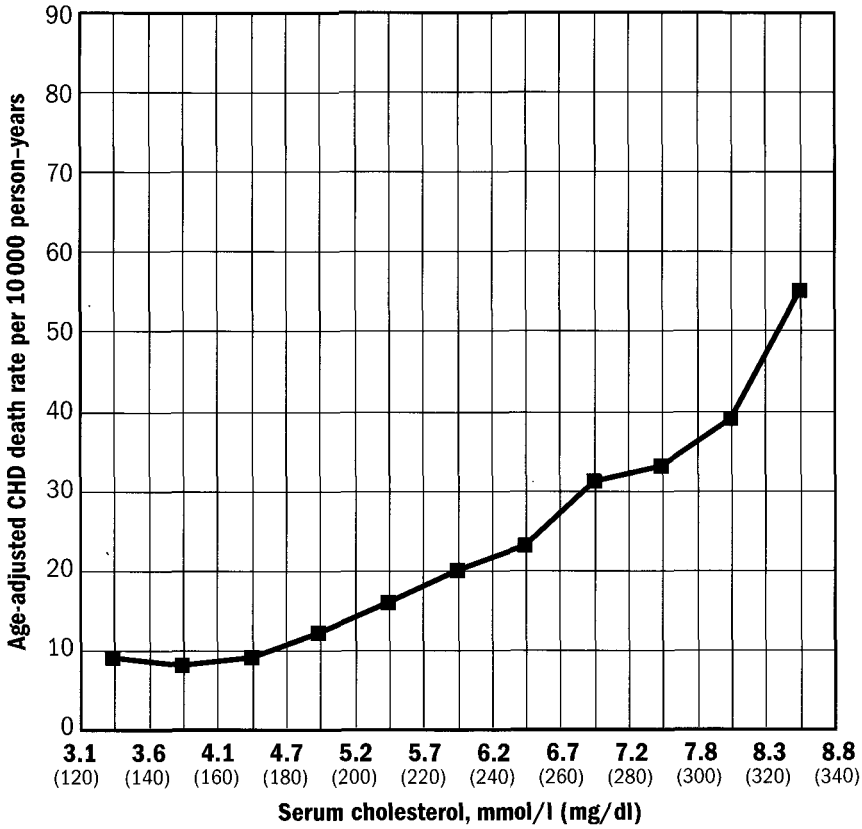
The evaluation of population trends in disease, as well as the evaluation of risk and treatments over time are important CVD survey methods. Knowledge of these trends

Fig. 2.3. Estimated prevalence of cardiovascular diseases[a] in Americans age 20 and older by age and sex, USA, 1988–94



[a] includes: hypertension, coronary heart disease, stroke, congenital heart disease and congestive heart failure.
Source: American Heart Association. 2001 Heart and Stroke Statistical Update. Dallas, Texas: American Heart Association, 2000.

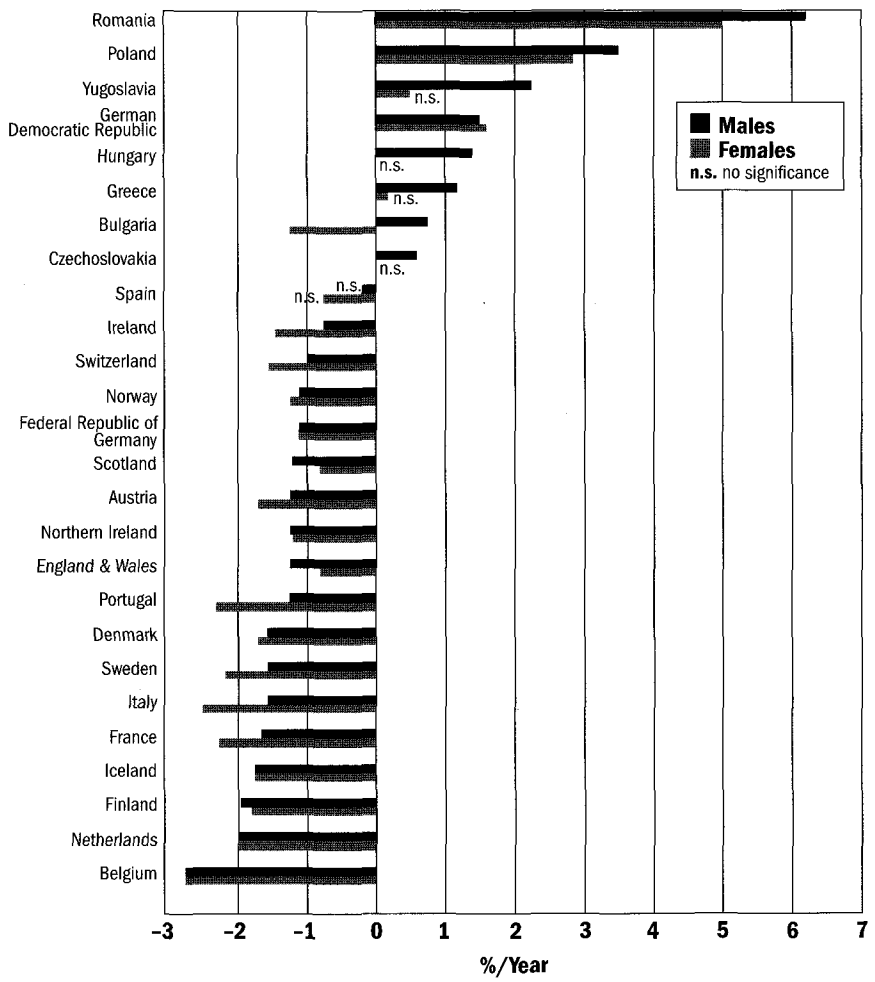
Fig. 2.4. Age-adjusted coronary heart disease (CHD) death rates per 10 000 person–years by level of serum cholesterol for men screened in the Multiple Risk Factor Intervention Trial, 12-year follow-up



Source: Neaton JD, Wentworth D, for the Multiple Risk Factor Intervention Trial Research Group. Serum Cholesterol, Blood Pressure, Cigarette Smoking, and Death From Coronary Heart Disease: Overall Findings and Differences by Age for 316,099 White Men. *Archives of Internal Medicine*, 1992; 152:56–64.

helps to provide a better understanding of the nature of CVD, its projected population burden, and the effectiveness of control measures. As shown in Figure 2.5, CHD mortality trends differ strikingly between countries. In many countries mortality is rising, while in others it is falling or stable (*2*). Similarly, as shown in Figure 2.6, risk factors such as tobacco use in the USA have been falling in the past three decades, suggesting that programmes to reduce or eliminate smoking are succeeding (*3*). Trend analysis also allows evaluation of the adoption of medical treatments. As new treat-

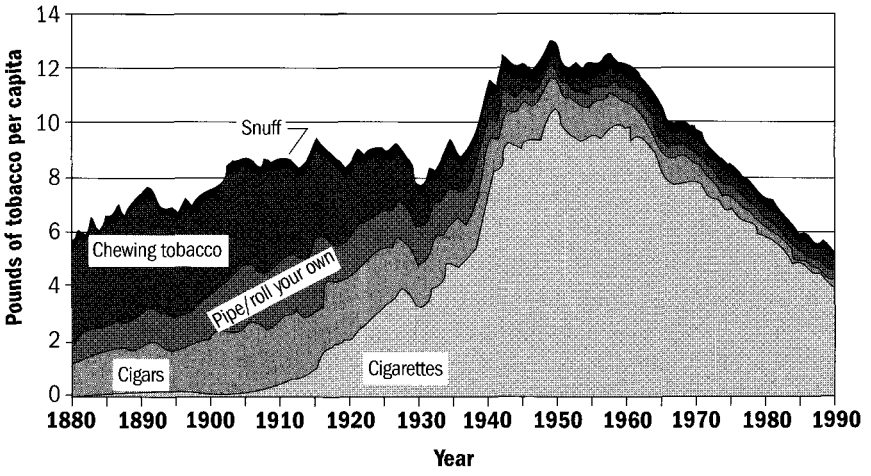
Fig. 2.5. Annual change in coronary heart disease mortality in Europe[a]



[a] Names of the countries are those that were in use at the time of the study. Source: Coronary heart disease mortality in Europe; annual change from 1970 to 1992 in persons 45 to 74 years of age. From Sans S, et al. *European Heart Journal*, 1997; 18:1231–1248.

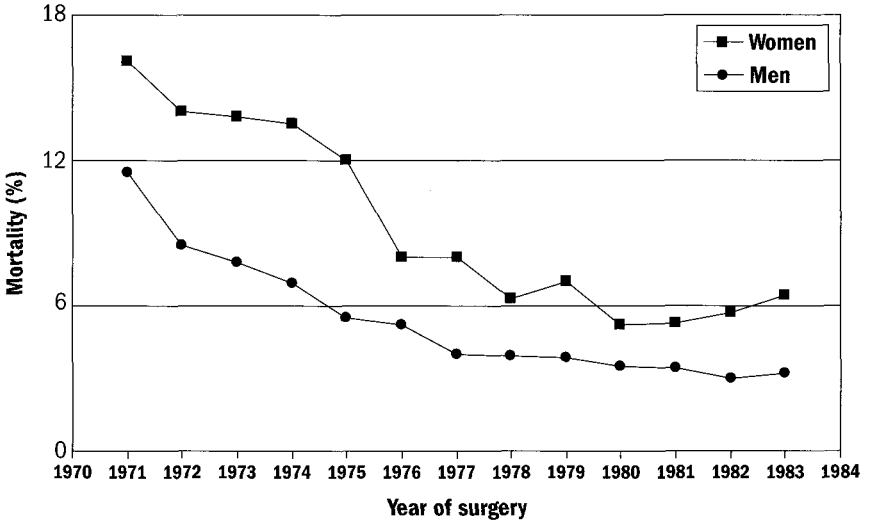
ments are developed and tested, they are used increasingly, while others are rejected. For example (see figures 2.7 and 2.8), coronary bypass surgery and drug use to control high blood pressure has varied over the past 20 years (*4, 5*). Central to the analysis of trends is the reliable and consistent measurement of the characteristic. This is a problem when disease definitions and measurements also change.

Fig. 2.6. Per capita consumption of different forms of tobacco in the USA, 1880–1995



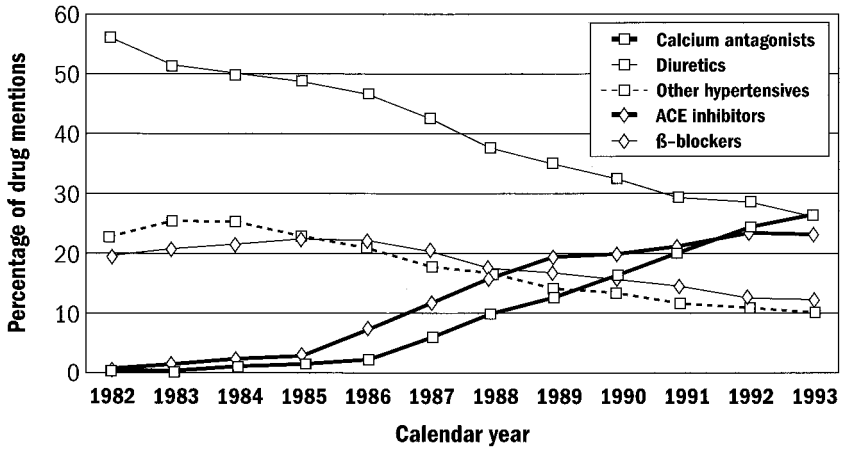
Source: U.S. Department of Agriculture, 1996.

Fig. 2.7. Rates for annual age-adjusted gender-specific operative mortality (3-year running averages) for residents of Minneapolis-St. Paul undergoing coronary artery bypass graft surgery, 1970–1984



Source: Doliszny KM et al. Estimated Contribution of Coronary Artery Bypass Graft Surgery to the Decline in Coronary Heart Disease Mortality: The Minnesota Heart Survey. *Journal of American College of Cardiology*, 1994; 24:95–103.

Fig. 2.8. Percentage of drug mentions by class of antihypertensive agent, 1982–1993



ACE indicates angiotensin-converting enzyme.
Data from IMS America.

Source: Manolio TA et al. Trends in pharmocologic management of hypertension in the United States. *Archives of Internal Medicine*, 1995; 115:829–837.

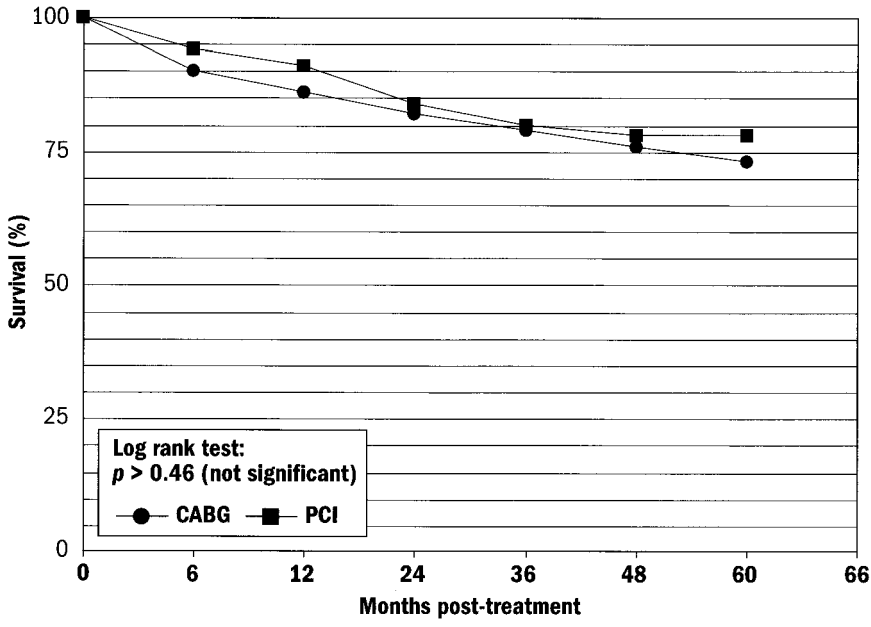
Trials

Trials of cardiovascular disease treatment and prevention measures make effective use of epidemiological methods to evaluate patients, groups or even communities. Consistent methods of measurement and case definition are essential to trial execution and valid comparisons. Figure 2.9 shows an example of a study that randomly allocated patients to medical or surgical therapy once extensive measurements had been made (*6*). Follow-up included measurement of clinical outcomes in a consistent manner.

Health services research

The study of the costs and benefits of medical care and procedures has assumed increasing importance as health care costs have risen dramatically. The combination of frequent CVD and expensive, high-technology treatments threatens the health care budgets of many countries. It is essential that the individual and social costs of CVD are appreciated, as well as the benefits and risks of various treatments. While epidemiology has not traditionally been involved in health services research, the gathering of population data on disease prevalence and trends and on treatment outcomes is important for understanding the costs, risks and benefits included in treatment

Fig. 2.9. Kaplan-Meier survival plot of coronary artery bypass graft surgery (CABG) versus percutaneous coronary intervention (PCI)



The CABG and PCI number of patients (N) and the percentage surviving for each time period are shown at the bottom of the plot.

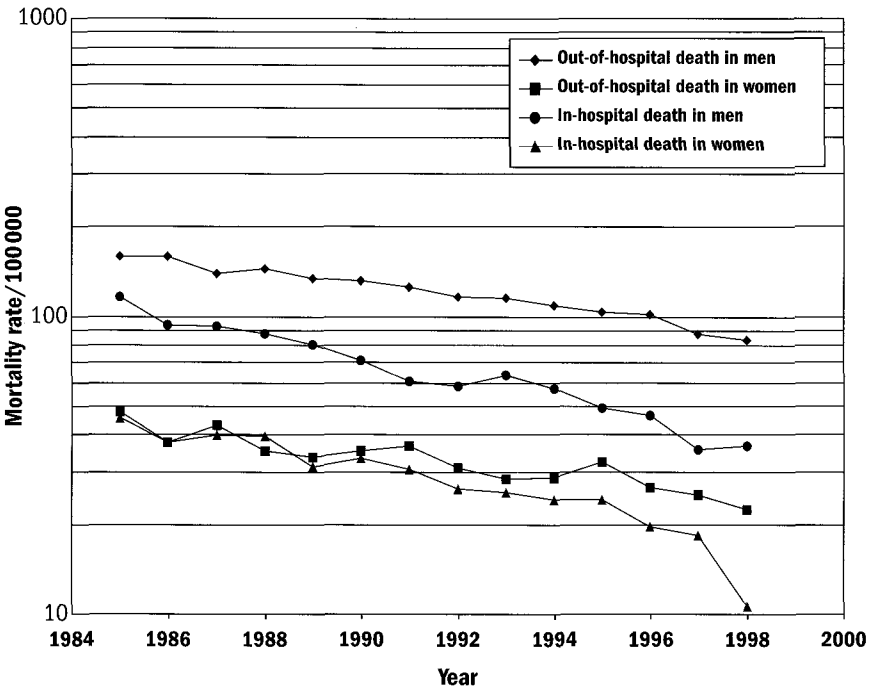
Source: Morrison DA, Sethi G, Sacks J. Percutaneous Coronary Intervention Versus Coronary Artery Bypass Graft Surgery For Patients with Medically Refractory Myocardial Ischemia and Risk Factors for Adverse Outcomes with Bypass: A Multicenter, Randomized Trial. *Journal of American College of Cardiology*, 2001; 38:143–9.

and control of diseases. Data from cardiovascular surveys are increasingly used in planning and evaluating needs and effects of health care strategies. Readers may wish to refer to a more extensive work on this subject (*7*).

Unique aspects of cardiovascular disease epidemiology

One frequent and unique presentation of CVD is sudden out-of-hospital death. Evidence suggests that this is actually more common than in-hospital mortality for CHD (see figure 2.10). Many deaths from CVD are sudden and unattended, frequently

Fig. 2.10. Trends in in- and out-of-hospital death from CHD in men and women aged 30–74 in the Twin Cities



Adjusted to the 1990 USA population.

Source: Minnesota Heart Survey—unpublished.

without warning symptoms, and related to malignant arrhythmias or rupture of major blood vessels. This observation implies the need for primary prevention techniques (*8*, *9*).

Epidemiological research has resulted in a substantial understanding of the major cardiovascular diseases, including rheumatic heart disease, coronary heart disease, stroke and hypertension, their risk characteristics and treatment. Understanding of the physiological factors that increase risk, and of the social, environmental, behavioural, and genetic mechanisms that influence these factors, has allowed the development of prevention strategies. The extensive data available on CVD patterns in populations, risk factors and treatment modalities also permit a detailed understanding of time trends. Such data are available for many countries but are still lacking in others.

Ethics of epidemiological research

Introduction

Much research in cardiovascular disease is observational and the information gathered emphasizes simple standard tests and questionnaires. Consent from both the individual and/or the organization is required before a survey may be carried out—even in observational studies, there is a potential for harm. Simple medical tests may lead to morbidity and, rarely, death. Questionnaires may induce fear, anxiety and emotional distress. Information gathered on individuals may be released inappropriately leading to problems in such areas as employment, life insurance or economic advancement. In all epidemiological research, independent review of proposals, informed written consent and protocol of safeguards for subjects are essential.

The following sections describe the basic concepts and guidelines that are considered appropriate for ethical research in cardiovascular epidemiology. These issues are described in greater detail in the ethical guidelines for epidemiological investigations (CIOMS and WHO) (*10–12*) and in other works (*13–15*).

Basic precepts

All medical research must consider several basic principles, including respect for persons, beneficence, non-malfeasance and justice (*10*).

"Respect for persons" makes several assumptions. Individual subjects in a research project should be capable of understanding both the project and, after full and honest explanation, the risks and benefits involved. They should be free to decide on participation without coercion or other pressure such as financial inducements. Particular care to guard against harm or abuse is required in the case of children, prisoners and individuals who are seriously ill or mentally impaired.

Beneficence implies some benefit for the health of the individual or population. Even those studies that have the potential for large benefit must balance this against possible harm. The overriding consideration must be to do no harm. An ethical study asks important health questions while it maximizes potential benefit and minimizes potential harm.

Justice for participants must prevail in any research project; that is, all subjects must be treated equally. Populations who will benefit from the findings of a study should bear the risks of the study. This is of particular concern when research is carried out in developing countries or among economically deprived populations. Studies that will benefit only subgroups of that society or wealthier countries should be performed only in those groups or countries. Weaker members of society should not bear

disproportionate risks. Finally, research deemed unethical in one country should not be undertaken in another.

Informed consent

Informed consent implies that a competent individual, group or community involved in research understands the risks, benefits and responsibilities of the study. It also implies absence of coercion and inappropriate inducements.

Informed consent is best achieved through a written statement read by or to the participant and signed by the participant and a witness. The study must be presented clearly and honestly, and in terms that subjects comprehend. A basic description of the nature and goals of the project, the risks and benefits to the participant, the responsibilities of the participant, and actions that will be taken in the event of side effects, injury or harm should be detailed. The process should be undertaken in the presence of an informed investigator who can answer questions and provide details when required. Individuals must also be informed of their right to withdraw from the study. Examples of consent forms and formats used in epidemiological studies, including one for genetics studies, can be found in Appendix 1. The complexity of the consent procedure varies by country and institution.

Even higher standards must be applied with certain categories. These include children, those who are dependent upon others (e.g. prisoners), and those who are otherwise impaired or vulnerable. In such cases, informed consent may be given by a parent, guardian or responsible official. That individual should have the dependent subject's interest as the main goal and should be independent of the study.

In certain instances, such as record review or secondary data analysis, informed consent may not be sought. However, it is the responsibility of the investigator to justify to an appropriate review board the reasons for not obtaining informed consent.

In studies of institutions, organizations or communities, informed consent by individuals may not be available or appropriate. In such cases, agreement may be sought from a community or institutional representative, elected officials or other appropriate leaders. However, even when broader approval is obtained by this method, individuals may still exercise their right to withdraw.

Because much medical research is performed by physicians and nurses who have clinical responsibility for and a certain authority over patients, there is the potential for undue influence. A subject who does not want to participate in the study may be uncomfortable in refusing the practitioner who provides their clinical care. This position of power must be neutralized. Informed consent should therefore be obtained by a third party not involved in the subject's usual care. In all cases, subjects should be assured that refusal to participate in the study will not affect future clinical care.

Rewards and inducements

It is increasingly common for inducements to be used to encourage participation in studies. This is particularly true in studies involving many participants, long data collection sessions, numerous tests or multiple visits. As in any study, the risks of the research must be acceptable even if inducements are not available. Inducements, particularly of a monetary nature, pose a particular problem in deprived or poor communities. The use of large amounts of money (by local standards) may encourage individuals to participate in studies that are not in their interest. This practice must be avoided.

Responsibility to diseased or high-risk subjects

Researchers are frequently involved in studies that screen individuals for disease or high risk of disease: incidence or prevalence can be estimated only by identifying cases. Finding cases of disease or high-risk individuals involves a medical duty to assure appropriate treatment. In all surveys the investigators should assure clinical follow-up of individuals found to be ill or at high risk. If facilities are not readily available in the community, they should be provided by the research team. Information on disease or increased risk should be shared clearly, directly, and promptly with the individual and the designated health care provider.

It is the responsibility of investigators to transmit findings of a population study to appropriate officials and health care organizations in the surveyed community. If increased prevalence of a disorder or new findings about that disorder are uncovered, the community that has provided the information must be told promptly so that appropriate changes in public health or medical care can be planned. Similarly, control or placebo subjects in a clinical trial should be informed and offered treatment as soon as the new therapy is determined to be effective. These ethical guidelines preclude the conducting of "natural history" studies in which subjects are observed without treatment.

Confidentiality

Epidemiological studies gather information on risk, other personal characteristics and disease conditions of individuals, and confidentiality of that information is an integral part of the right to privacy of the individuals. It is a primary duty of researchers performing studies to maintain confidentiality.

Some epidemiological studies require no link to the subjects: investigators remain "blinded" to individual data and consent is not needed. Data of this type are devoid of identifying factors (name, age, date of birth, address, telephone number) and are frequently obtained from large insurance or government databanks.

In most observational studies in epidemiology, the investigator is not interested in individual names but rather in analysing grouped information. In this approach identifiers are not needed and it is best to eliminate names from such data sets. In many circumstances, however, follow-up for mortality or morbid events may be necessary, and identifiers are then needed. Nevertheless, care must be taken to ensure that the identifying information is made available only to investigators in the study, who must protect the confidentiality of the information from uses not relevant to the investigation. This is usually achieved by eliminating names or other identifiers from the record and replacing them with a number. That number can then be linked to the identifiers in a separate, protected file available only to approved individuals.

In many countries, there is an increasing call from politicians and the public to limit the use of personal medical information in research projects. Some countries have enacted laws to limit access even for qualified investigators. This is partly the result of increased public concerns about privacy but also of failure by researchers to preserve confidentiality. Clear protocols and vigorously enforced procedures to protect the confidentiality of information are crucial to the survival of survey research.

Ethical review panels

Central to the protection of human subjects in medical research is an independent review panel to oversee the conduct of research, approve study protocols and monitor study conduct on a regular basis. All institutions performing medical research should have such a panel, with the authority to ban studies that do not meet ethical standards and stop those that fail to meet the stated criteria. In most settings, funding agencies and industry will not provide resources for the research unless a review panel has approved the study protocol and design. The panel should be composed of scientists, ethicists and representatives of the public. It should establish standards for ensuring that research applications meet the ethical criteria prevalent in the community concerned. It must have the expertise to determine whether a study that places subjects at risk also offers "adequate" benefit to individuals and the community. The panel should be independent of external pressures, with its core task being the protection of research subjects.

Special circumstances in epidemiological and public health research

Conflict of interest

Conflict of interest is an increasing concern in all scientific research. When researchers may benefit personally or financially from a research outcome or use of a treatment, there is potential for perceived and real conflict. Clinical trials, where the success of a certain drug or device may result in considerable financial gain, are an obvious source of this difficulty. Even observational studies may have outcomes that result in

gain or loss for individuals or institutions. Investigators must eliminate the potential for real or perceived conflict by distancing themselves from any personal gain or other incentives associated with their projects. Some institutions have developed committees and policies to review such potential conflicts of interest (Appendix 2).

Vulnerable or dependent groups

Certain groups in the population are not able to make reasoned or free choices. These include children and severely ill individuals. Others, such as prisoners, employees, medical students, and others may feel obliged to participate in a study because of the authority vested in their wardens or supervisors. Pregnant and nursing mothers always require particular care when involved in research because of danger to their children. Economically and socially deprived communities may not be able to make free choices. All of these groups require special protection, and review panels must be particularly vigilant in asking whether there is free or responsible choice where such groups are concerned.

Control groups

In randomized clinical trials, control subjects are necessary for comparison with the treated groups. They frequently have some disease or increased risk and, during a trial, must receive at least the normal standard of community care for their condition. For example, when effective drugs are available, control subjects cannot be given a placebo to test against a new agent. Controls must be made aware of their condition and treated for acute episodes. When the trial is complete, successful treatments or procedures should be offered to them, usually at no cost as compensation for participating in the trial.

"Primitive" communities

In order to achieve a better understanding of the origins of disease, epidemiologists and medical anthropologists often study "primitive" societies, i.e. communities that have no access to modern technologies and communications and whose lifestyle, diet, and environment are unaffected by the influences common in more industrialized societies. Particular care is needed when working with communities where the understanding of medical research, its benefits and its dangers is limited. The study of such communities requires the investigator to be knowledgeable about and sensitive to their culture.

Interpretation of data: bias and confounding

Overview

In all scientific data, there is always a question of truth. How do we know what we think we know? Statistical analysis can quantify the likelihood of chance affecting outcome but provides no indication of whether the data collected are true. Truth is based on validity, the threats to which must be considered in design and interpretation.

Internal validity involves truth in measurement of risk characteristic, exposure or disease condition. For example, simple measures such as body weight and height are susceptible to errors when poorly collected or misreported: men commonly over-report their height and women under-report their weight (*16*). Exposure to environmental factors can be misreported, particularly with increasing time since exposure, or when a participant may have some reason to over- or under-report an exposure. Diseases can be misclassified when political, social and economic factors influence the data recording. Even pathologists may differ in their interpretation of biopsy specimens. All of these threats to internal validity argue for careful attention to measurement.

External validity allows the results of a study to be generalized to the broader social context. It requires consideration of data in relation to other work on similar populations using other methods, as well as of the representativeness of the population sampled. A representative population mirrors the overall population. Internal validity, which reflects quality of measurement, must precede external validity or generalization.

The various threats to the validity of any study are described in the following paragraphs.

Random error

Random error, above or below the true value, is caused by problems with the sampling or measurement systems. Sampling error can occur when more or fewer "abnormal" individuals are surveyed than actually occur in the population, and is a particular problem when the sample is small. For example, if 50% of a population actually has hypertension and 10 people are randomly sampled, there is a significant likelihood of observing 7 hypertensive and 3 normotensive individuals (70% prevalence) or a variety of combinations other than 50%. Such random effects lead to the misreporting of rates, however accurate the blood pressure measurement.

Random measurement error is a common problem. It includes biological variability as characteristics change over short or longer time cycles. For example, blood

pressure varies with each beat of the heart as well as during a 24-hour cycle. Measurement error also concerns the reliability and reproducibility of measurements: equipment may deteriorate with time, the techniques may vary, or external conditions—including temperature, humidity and barometric pressure—may affect the measurement. Even the most carefully calibrated equipment is susceptible to this type of error. Random error of measurement must be reduced to the extent possible, and should be quantified.

Random variability, too, must be minimized; sampling error and biological variability are reduced by increasing the number of subjects or the number of measurements. For poorly reliable measurements, there may be technical solutions such as using better equipment or repeated measures to stabilize the estimate around the true value. In large studies performed over extended periods with many subjects, random error tends to be eliminated (but not systematic error).

Systematic error or bias

Selection bias

Systematic error or bias is any source of error in design, measurement or analysis that is not random. It is an even more serious problem than random error. Whereas random error (chance) is reduced by increasing sample size or the number of measurements, systematic error may persist and can be obviated only by careful design and conduct of the study. Of the two major categories of systematic error—selection bias and information bias—selection bias is common in population surveys. Here, the methods used to select subjects may result in those who participate differing from those who refuse or are non-participants for other reasons. That difference can affect the outcome, because the participants are not representative of the population. Selection bias is crucial in surveys but is less important in clinical trials where subjects may be selected for certain characteristics or where randomization allows matching of characteristics.

Of a number of different types of selection bias, "referral bias" is common in surveys. It is an error often made by clinicians doing studies of patient populations and selecting patients from those admitted to the hospital or seen in the clinic. The problem is that those who are in the hospital or who are referred are usually far more seriously ill, or have the best access to medical care. Many are excluded because of milder illness, access to services, cost of care or patient choices. The more frequent hospitalization of patients who suffer from more than one disease gives a bias towards an association between diseases in observation of hospitalized patients only (*17*). Their characteristics may differ significantly from the usual disease patterns in the population, which will include many mild and sub-clinical cases. Hospital series may also ignore seriously ill cases who do not have access to medical attention or those who

die before reaching the hospital. Referral bias is a particular problem in tertiary care university hospitals, which receive the most seriously ill cases.

"Detection bias" or "surveillance bias" is another common threat to validity. Patients under regular surveillance are more likely to have their diseases detected because they are tested systematically. For example, if blood pressure is measured more frequently, it is likely that at least one elevated measurement will be found. Individuals who are not so frequently measured may appear to have a lower frequency of elevated blood pressure.

In most population surveys, those who refuse to participate or are excluded are likely to be more ill or to have the risk characteristics of particular interest. For example, individuals with manifest CVD may choose not to participate in a survey because they are already under regular medical care. Their non-participation will lead to an underestimation of disease prevalence. Population surveys of CVD or cancer also under-represent individuals who smoke cigarettes. Many smokers are aware of the deleterious effects of smoking and prefer not to participate, which leads to an underestimation of smoking rates (*18*).

Unlike individuals who avoid participating in surveys because of illness or fear of illness, others are particularly likely to volunteer. These include the "worried well", a group that is health-conscious but needs the constant reassurance of negative tests (e.g. chest X-ray, blood pressure). Individuals who exercise regularly may seek to participate in a study of the effects of exercise because they are proud of their lifestyles. Volunteers for a study of exercise may produce a population with vigorous, but atypical, exercising patterns. In general, volunteer populations do not adequately represent the general population.

Studies of groups may be prone to bias where groups are comprised of individuals who are affluent or highly educated, or have other common attributes but who have different health behaviours and characteristics in relation to CVD.

Similarly, the "healthy worker effect" is frequently observed in businesses, and is characterized by better health and lower relative mortality from all causes. This occurs because healthy individuals are more likely to gain and retain employment. These effects were first described in 1885 by William Ogle, who observed that "some occupations may repel, while others attract, the unfit at the age of starting work" and noted the "considerable standard of muscular strength and vigour to be maintained by those who pursued various occupations". Thus, the observed frequency of disease may be low in such a cohort, despite exposure to a putative risk factor, since persons who are less fit have been excluded from the workplace.

Another category of selection bias is "survival bias", which involves the selection of prevalent cases rather than incident cases (cases reporting with acute coronary disease

for the first time). This is also called the prevalence-incidence bias or Neyman's bias. In CHD, for example, a large proportion of patients die during the first days of an initial acute coronary event. The survivors represent people with less severe disease and/or other factors that favourably influence the outcome. In addition they are likely to have modified their lifestyle and to have lowered their risk factors. A patient's risk factors could also be affected positively or negatively by treatment with cardioactive drugs. Thus, the prevalent cases may be unrepresentative of the natural disease in the population, and a study restricted to prevalent cases may misidentify factors that lead to better survival. Risk factor surveys of prevalent cases will certainly lead to underestimation of the link between disease and lifestyle. Incident cases should be free from this source of bias.

Other forms of selection bias must be considered in any non-random selection of a population. To control selection bias, objective measures of exposure and disease are established and consistently applied, and characteristics for inclusion and exclusion of cases are carefully defined before the study. Maximal participation rates in well-defined populations are sought to minimize bias due to differential participation. Refusal and non-participation should be kept to a minimum. For most survey work, self-selected volunteers are not acceptable. The investigator strives for a clear definition of the population and high participation rates to the survey, while recognizing the limitations of generalizability if the population is restricted.

Information bias

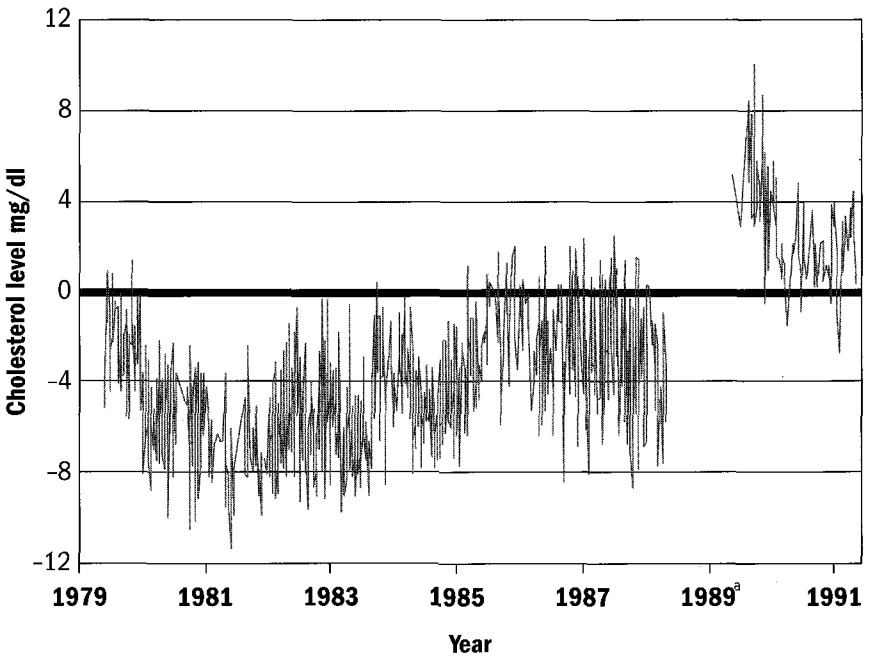
Even surveys with high participation rates are subject to bias in data collection. Most surveys use questionnaires to define exposures, treatments and other characteristics. There may be recall bias—people have difficulty remembering distant events or may suffer from distortion or loss of memory of events related to their illness. For example, individuals with a recent myocardial infarction are more likely to cite stressful events associated with the onset. The common belief that stress causes attacks encourages a memory search by the patient for associated episodes. Control patients do not have similar concerns and may under-report identical stressful events.

In addition to bias introduced by the participant, the interviewer may also add distortion. Interviewers who are aware of an individual's disease status may inquire more vigorously about associated factors than they do with non-diseased subjects. The age, sex and race of the interviewer affects participants' responses (*19*). For example, women may be uncomfortable giving information on menstrual history to male interviewers. In some settings, participants do not want to admit actions or behaviours that they regard as socially or personally unacceptable. For example, the use of objective measures such as expired carbon monoxide significantly increases the "prevalence" of smokers in groups for whom this is a socially unacceptable habit (*20*).

There are many methods of reducing bias in data collection. Aids to memory, such as counting backward in dates, can help the recall of distant events. Interviewers blinded to subject status can be taught to ask questions in a consistent manner. Closed-ended questions that have numerical or categorical answers minimize probing and differential responses. Finally, a study that blinds not only the interviewers but also the participants to interview goals may result in information of higher quality; neither those being interviewed nor the interviewers know the "correct" answer.

In addition to random error, physical measurements may also introduce further biases. Poorly calibrated measurement devices, unstandardized chemical reagents and inadequately trained technicians give rise to systematic errors. These factors may change over time and be particularly difficult to detect in the absence of continual standardization and quality control. Figure 2.11, a standard sample was run several times daily over a period of 12 years. The figures showed daily random variability for

Fig. 2.11. Daily laboratory drift from Abell-Kendall reference method from 1979–1991



[a] **No data available for 1989.**

Source: Minnesota Heart Survey—unpublished
Comparison of automated machine values to reference method (Abell-Kendall)

the machine throughout the period. However, a change in reagents to a new lot resulted in a dramatic increase in the average cholesterol measurement. These differences were large enough to be biologically significant. Evaluation with a reference technique demonstrated that the initial values were systematically low while the later values were systematically high. These examples of erroneous measurements indicate the absolute need for calibration and quality control.

Confounding

Confounding occurs when exposed and non-exposed groups are not comparable because of differences in background disease risk. This is due to a mixing of effects between the exposure, the disease and a third factor associated with the exposure, independently affecting the risk of disease. The extraneous third factor is the "confounder" or "confounding variable". Three conditions are necessary for a factor to be a confounder. First, it must be associated with the disease, even in the absence of the exposure under study. Second, it must be associated, positively or negatively, with the exposure (the proposed risk factor in the study). Third, the potential confounder must not be an intermediate step in the causal pathway between exposure and disease. In summary, the confounding variable is independently associated with both the exposure and the disease and is not a mediator.

For example, in a study evaluating the relationship between oral contraceptives and CHD in women, smoking is an independent risk factor for the disease; moreover there is an increased frequency of smoking in women consuming oral contraceptives (association with the exposure). Smoking is thus a confounder. The true effect of oral contraceptives can be assessed only after controlling for the differential frequency of smoking. Otherwise, the increased frequency of smoking-related disease in users of oral contraceptives will result in an overestimate of the relationship between contraceptives and the disease. Similarly, truck driving may appear to be associated with increased coronary risk unless an adjustment is made for the higher smoking rates among truck drivers. Age, sex and social class are frequent confounders related to both exposure and disease.

Confounding must be distinguished from effect modification. An effect modifier is a factor that modifies the effect of a putative causal factor under study. Age, for example, is a modifier for the effect of systolic blood pressure on stroke, in that older people have a lower relative risk than younger people for the same level of systolic blood pressure. Similarly, sex is a modifier of the effect of serum cholesterol on the risk of CHD, since at younger ages women have a lower relative risk than men for the same level of serum cholesterol at younger ages.

Confounding should be controlled by appropriate design and analysis. If a factor does not vary between those who are exposed and those who are not exposed, or between those who are diseased and those who are not, confounding does not occur.

Specific methods to control confounding include restriction of the study population, matching, randomization of exposure, stratification and multivariate analysis (discussed below).

Overview

Much of health research aims to ascertain cause-and-effect relationships. Epidemiology helps in identifying causal relationships from the associations noted among the many variables of interest.

The cardinal requirement for developing effective preventive strategies is knowledge of causation. The term "risk factor" is used to describe characteristics that are agents in the chain of causation and also show statistical association. Before any associated factors are declared to be the cause of a disease, it is necessary to apply the criteria for causal inference.

Two questions that must be asked in evaluating the causal role of any factor are:

1. Is there a valid statistical association?
2. Can this association be judged as cause and effect?

Statistical associations

Statistical association refers to statistical dependence between two variables, which in the context of disease causation is the degree to which the proportion of disease in persons with a specific exposure is either higher or lower than the proportion of disease among those without that exposure. The first step in evaluating causation is to establish such an association statistically. The next is to determine whether this association is due to bias or confounding.

Since the samples (preferably random samples) being studied are samples of the total "universe", the role of chance in observing a particular association needs to be evaluated. A large sample reduces the role of chance and enhances precision of estimates. Quantification of the extent to which chance may account for the results observed in an individual study is obtained through hypothesis testing and the derived probability statement (*P* value).

The *P* value is a composite measure that reflects both the sample size and the magnitude of the difference between groups. Even a small difference may be statistically significant if the sample size is sufficiently large; conversely, a large effect may not be significant if the sample size is insufficient. A related and informative measure is the confidence interval, or the range within which the true magnitude of effect lies with a certain degree of assurance. A statistically significant result does not mean that chance cannot account for the findings, only that such an explanation is unlikely.

A result that is not statistically significant does not mean that chance is responsible for the results, only that it cannot be excluded as a likely explanation. While tests of statistical significance merely estimate the role of chance as an explanation for the observed result, the confidence interval estimates the likely range of size of the effect.

Systematic error or bias (described above) is any source of error in design, measurement or analysis that results in deviation from the truth. It reflects the difference between what the study is actually estimating and what it is intended to estimate. Whereas random error (chance) is reduced by increasing sample size, systematic error may persist and can be obviated only by careful design, measurement and conduct of the study.

Criteria for causal inference

Once a statistical association has been identified, it is essential to ascertain whether it can be judged a cause. This is more complex than determining statistical significance, and the guidelines for causality have been framed by Hill and modified by Evans (*21*, *22*).

The following questions need to be answered to infer causality:

1. Is the association strong?
2. Is the association consistent across studies?
3. Does the "cause" precede the "effect"? That is, is the temporal relationship correct?
4. Is there a dose–response relationship? For example, is there a relationship between the number of cigarettes smoked and CHD rates?
5. Is the association specific? That is, is a single causative factor directly related to a single disease?
6. Is there experimental evidence in humans?
7. Does the association make biological sense? Is it concordant with laboratory, animal and clinical studies?
8. Does the association make epidemiological sense? Is it reasonable to infer these associations?
9. Is the association analogous to a previously proven causal association? Are there other known causal chains that are similar?

Most of these criteria must be met before it is possible to infer that X causes Y. Elevated serum cholesterol in atherosclerotic coronary artery disease, for example, fulfils all of these criteria. Long-term prospective studies have revealed a strong association (in relative risk), consistency of association, a dose–response relationship, and a temporal relationship, the cause preceding the outcome (*23*). Therapy with cholesterol-lowering drugs and diet has been shown, in controlled clinical trials, to result in a reduction in the incidence of coronary events (*24*, *25*). Angiographic regression of

coronary lesions has also been demonstrated (*26, 27*). Exclusion of bias, chance and confounding as explanations is well demonstrated, and biological proof has been established from animal and human studies demonstrating cholesterol deposition in atheromatous plaques.

Design of research studies to determine causation

Many types of epidemiological studies aid in the demonstration of causation. They include case–control studies, cohort studies and randomized controlled trials, with a hierarchy of methodological strength in the three designs. In contrast, descriptive epidemiological studies are methodologically weaker because bias and confounding may be unavoidable or unassessable. Merely stating that 65% of Indian patients below the age of 40 years with myocardial infarction were smokers does not establish causation; data on prevalence of smoking in sex- and age-matched controls, and matching for confounders such as blood pressure, serum cholesterol, body weight and diabetes are also essential. However, descriptive studies often help to generate hypotheses. Based on the hypothesis, a more powerful study design can be employed to estimate the strength as well as the validity of this association. Case–control studies are susceptible to several sources of bias. Cohort studies are methodologically stronger but are expensive and also susceptible to bias. Randomized controlled trials are difficult to perform and expensive, but yield the most unbiased estimates of risk and the effects of risk modification.

Use and misuse of epidemiological data

Misuse of survey data

Data from cross-sectional descriptive and other surveys are frequently used to draw causal inferences about associations. This is particularly true in the popular press and is invalid for several reasons. A temporal relationship is usually impossible to establish in cross-sectional data. "Survival bias" is a major problem since the survey can identify only the survivors and not high-risk cases who may have succumbed to the disease. For example, if acute coronary events cause higher mortality among smokers than among non-smokers, a survey would identify more non-smokers among patients with a history of myocardial infarction, leading to the erroneous conclusion that smoking is protective. Thus, cross-sectional surveys provide information only about prevalent cases and not about incident cases. The onset of the disease or its clinical recognition may have led to modification of the risk factor status, for example, angina pectoris leading to sedentary habits, or a lipid profile altered by diet changes or lipid-lowering medications. There are thus many reasons to avoid inferring causation from cross-sectional surveys.

Regression dilution bias

In most surveys, single measurements of variables are obtained. The phenomenon of regression towards the mean, however, reduces the variance in repeat estimates of any variable in stable subjects. Regression to the mean implies that any measurement above or below the true mean will tend towards that mean (i.e. regress) on subsequent measurement. For example, if individuals with high blood pressure have their pressure remeasured, it will be more likely to be lower. Multiple averaged measures are more likely to represent the true value. The strength of association in studies where repeat sampling or repeat measures are carried out is increased. Where only single measures are made, the strength of association is diluted. This has been demonstrated for variables such as serum cholesterol and blood pressure and vascular events. It is therefore advisable to carry out repeat measures in surveys.

Generalizability

Can a causal relationship that has been determined in a single age, sex or racial group be generalized to other age, sex or population groups? Are the relative risks at various risk factor levels the same in every population? Are risk factors likely to differ with the presence and extent of other risk factors in the population? For example, is the coronary risk of smoking different in an alcohol-consuming population from that in a community that traditionally abstains?

While most of the CVD risk factors are universally relevant, as demonstrated by the consistency of association in different populations, the sex-specific or race-specific nature of some risk factors suggests that they should be evaluated in as many groups as feasible. Similarly, risk factor interactions and risk gradients may differ across populations. It would be ideal, therefore, to estimate these in each population group. In developing countries, this issue becomes particularly relevant. Public health action is urgently needed to contain the emerging cardiovascular epidemic, yet there have been few risk factor studies in these populations. An appropriate strategy might be:

- to conduct cross-sectional surveys in different geographical or ethnic groups to estimate average risk factor levels and to study the ecological associations with disease rates, so that prevention priorities can be identified;
- to conduct case–control studies to evaluate risk gradients of "known risk factors" and to generate hypotheses about putative risk factors of relevance in these populations; and
- to plan and conduct cohort studies in the samples assembled in the cross-sectional surveys. However, such a plan may be impossible to implement or may take many years, in which case it may be necessary to generalize from other data and then implement the indicated public health measure.

References

1. MacMahon B, Pugh TF. *Epidemiology: principles and methods.* Boston, MA, Little, Brown & Co., 1970.

2. McKee M, Zatoñski W, Sklodowska-Curie M. How the cardiovascular burden of illness is changing in eastern Europe. *Evidence-based Cardiovascular Medicine,* 1998, 2:39–41.

3. Burns DM et al. Cigarette smoking behavior in the United States. In: Burns DM, Garfinkel L, Samet JM, eds. *Changes in cigarette related disease risks and their implication for prevention and control (Monograph 8).* Bethesda, MD, National Institutes of Health, 1997 (NIH Publication No. 97–4213).

4. Doliszny KM et al. Estimated contribution of coronary artery bypass graft surgery to the decline in coronary heart disease mortality: the Minnesota Heart Survey. *Journal of the American College of Cardiology,* 1994, 24:95–103.

5. Manolio TA et al. Trends in pharmacologic management of hypertension in the United States. *Archives of Internal Medicine,* 1995, 155:829–837.

6. Morrison DA et al. Percutaneous coronary intervention versus coronary artery bypass graft surgery for patients with medically refractory myocardial ischemia and risk factors for adverse outcomes with bypass: a multicenter, randomized trial. *Journal of the American College of Cardiology,* 2001, 38:143–149.

7. Shi L. *Health services research methods.* New York, Delmar, 1997.

8. McGovern PG et al. Recent trends in acute coronary heart disease: mortality, morbidity, medical care, and risk factors. *New England Journal of Medicine,* 1996, 334:884–890.

9. Kuller LH. Magnitude of the problem. In: LaRosa JH, Horan MJ, Passamani ER, eds. *Proceedings of the National Heart, Lung, and Blood Institute Symposium on Rapid Identification and Treatment of Acute Myocardial Infarction.* Rockville, MD, NIH Publication No. 91–3035, 1991:3–24.

10. *International guidelines for ethical review of epidemiological studies.* Geneva, Council for International Organizations of Medical Sciences, 1991.

11. Ethical guidelines for epidemiological investigations. In: Vaughan JP, Morrow RH, eds. *Manual of epidemiology for district health management.* Geneva, World Health Organization, 1989:169–174.

12. International ethical guidelines for biomedical research involving human subjects. Geneva, Council for International Organizations of Medical Sciences (CIOMS) in collaboration with the World Health Organization, 1993.

13. Coughlin SS, Beauchamp TL. *Ethics and epidemiology.* New York, Oxford University Press, 1996.

14. Last JM. Association News. Guidelines on ethics for epidemiologists. *International Journal of Epidemiology*, 1990, 19:226–229.

15. Fayerweather W, Beauchamp TL, Higginson W. *Ethics and epidemiology.* Oxford, Pergamon, 1991.

16. Pirie P et al. Distortion in self-reported height and weight data. *Journal of The American Dietetic Association*, 1981, 78:601–605.

17. Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biometrics*, 1946, 2:47–53.

18. Luepker RV et al. Population risk of cardiovascular disease: the Minnesota Heart Survey. *Journal of Chronic Diseases*, 1985, 38:671–682.

19. Kraus JF et al. Relationship of demographic characteristics of interviewers to blood pressure measurements. *Journal of Community Health*, 1982, 8:3–12.

20. Ockene JK et al. Comparison of smoking behavior change for special intervention and usual care study groups. *Preventive Medicine*, 1991, 20:564–573.

21. Hill AB. Statistical evidence and inference. In: *Principles of medical statistics*, 9th ed. New York, Oxford University Press, 1971:309–323.

22. Evans AS. Causation and disease: the Henle-Koch postulates revisited. *Yale Journal of Biological Medicine*, 1976, 49:175–195.

23. Neaton JD et al. Serum cholesterol level and mortality findings for men screened in the multiple risk factor intervention trial. *Archives of Internal Medicine*, 1992, 152:1490–1500.

24. Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet*, 1994, 344:1383–1389.

25. Shepherd J et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *New England Journal of Medicine*, 1995, 333:1301–1307.

26. The Post Coronary Artery Bypass Graft Trial Investigators. The effect of aggressive lowering of low-density lipoprotein cholesterol levels and low-dose anticoagulation on obstructive changes in saphenous-vein coronary-artery bypass grafts. *New England Journal of Medicine*, 1997, 336:153–162.

27. Haskell WL et al. Effects of intensive multiple risk factor reduction on coronary atherosclerosis and clinical cardiac events in men and women with coronary artery disease: the Stanford Coronary Risk Intervention Project (SCRIP). *Circulation*, 1994, 89:975–990.

3.
Study designs

Overview

Surveys of cardiovascular disease use several study designs broadly categorized as either observational or experimental. In the former, disease patterns and their associations are evaluated using systematic observations. In the latter, one or more study factors are deliberately manipulated to test the effects of an intervention or treatment on randomly selected individuals or populations. The study design used depends on the aim of the survey and the current state of scientific knowledge about the disease or characteristic. Observational studies generally precede intervention studies as an understanding of the disease characteristics, risk factors and natural history is needed to develop treatments.

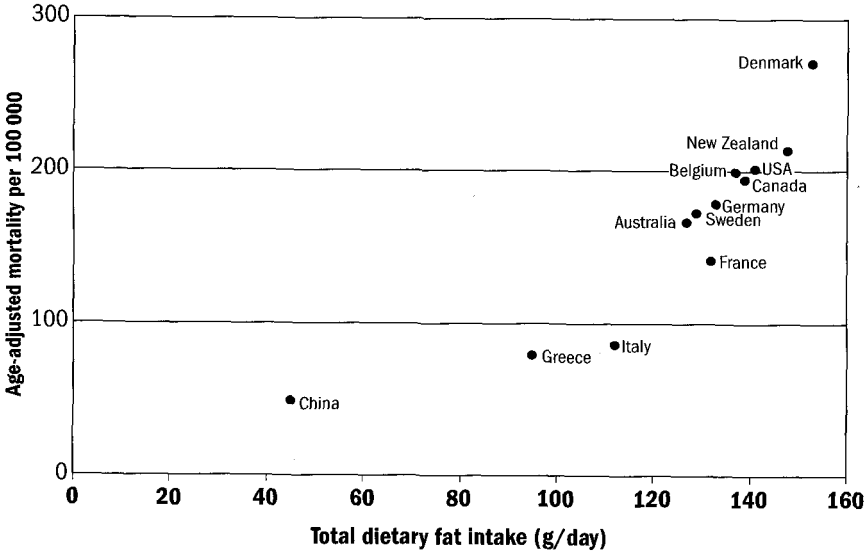
Ecological studies

Ecological studies are common, basic methods for describing associations between diseases and other factors of interest: disease rates in a population are correlated with levels of exposure or risk. The unit of analysis is the whole population rather than individuals.

One type of ecological study computes the correlations between disease rates in populations and the level of a certain characteristic in those populations. In Figure 3.1, CHD mortality rates are correlated with fat intake in different countries. The mortality rates are derived from national data submitted to WHO. Fat consumption is based on agricultural data submitted to the Food and Agricultural Organization of the United Nations (FAO). The figure shows a positive correlation between mortality rates and fat consumption, indicating that there is a true association between fat intake and CHD. However, it is not possible to infer causality from these data. There are many possible explanations for this association and ecological correlations such as this can be used only in generating hypotheses for further investigations. They may also provide confirmatory or non-confirmatory data for other hypotheses.

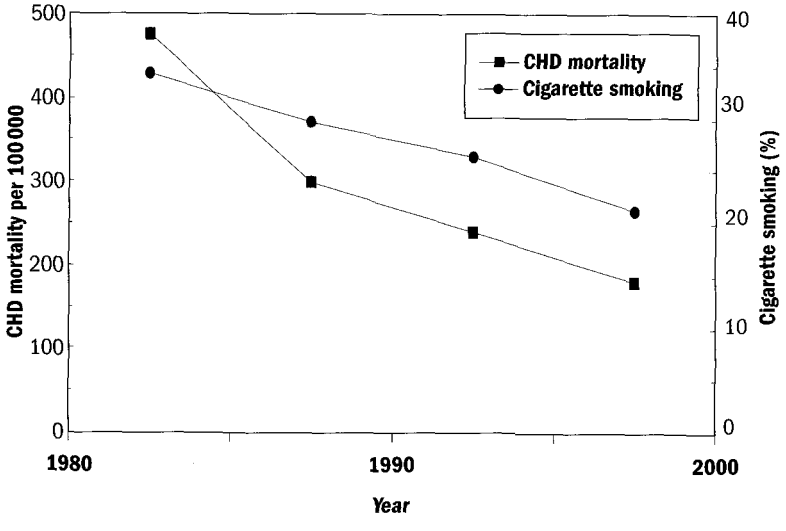
Another type of ecological study is a study of trends. Here changes in disease rates are related to changes in other characteristics in a population. In Figure 3.2, for example, CHD mortality and average population cigarette smoking rates are

Fig. 3.1. Coronary heart disease and dietary fat



Source: CHD from official statistics, FAO fat consumption

Fig. 3.2. Smoking rates and CHD mortality in men



Source: Adapted from Minnesota Heart Survey Data for Minneapolis-St. Paul Area. Unpublished material.

compared. Again, these observations are thought-provoking but alone they do not establish cause and effect.

Similar associations derived from ecological studies, which may contain some truth, include those between television licences (as an indicator of sedentary lifestyle) and CHD, or between intake of dietary fat and CHD in Norway during the Second World War. However, without other confirmatory data or a demonstrated biological mechanism, there is no reason to assume a causal association.

Ecological studies may well generate hypotheses about disease causes and trends. Because the data being compared are frequently gathered by different organizations, primarily for other purposes, they are inexpensive and they provide quick but partial answers to exploratory questions. They tell something about the phenomena of interest, define the burden and support the potential of population strategies.

Ecological studies also have limitations. Information about the population as a whole, frequently expressed as averages or means, does not apply to individuals: the methodology does not allow for this linkage. Thus, individuals who develop the disease may not be those who were exposed to the risk factor of interest. The possibility of lack of association between the exposure and those affected is termed the "ecological fallacy" (*1*). Ecological data frequently involves large numbers of subjects resulting in statistical significance without biological significance. Second, ecological data provide little or no information on associated or confounding factors. For example, the ecological association between cigarette sales and the fall in CHD could be confounded by age, with mortality occurring in the older group but cigarette consumption falling only among younger individuals. Data on age distribution of CHD mortality and [the age] of cigarette smokers are obviously necessary to analyse this relationship, but may not be routinely or readily available. Finally, the use of secondary data sources in ecological studies may demonstrate basic weaknesses in these data. Even essential public health data, such as mortality from death certificates, may be poorly collected in certain countries or areas.

Ecological studies are frequently misquoted in the popular press. They are useful but need careful interpretation.

Case reports and series

Reports by clinicians of observations in individual patients (case reports) and groups of patients (case series) are commonly used as a basis for characterizing diseases. These sometimes anecdotal and frequently unique observations serve to generate hypotheses that lead to more substantive studies.

Individual case reports may lead to new thinking and to the collection of case series, in which more consistent and systematic information is collected. Cardiovascular

survey methodology can be useful at an early stage in defining systematic and rigorous data collection in groups of patients to facilitate internal comparison as well as comparisons with other case series. The astute clinician will make use of these systematic methods to improve the information gathered by case series.

Case series and case reports have many strengths. They may function as an early warning system to alert the medical community to a new disease or its manifestations. There is no other system at present for many new diseases. Such reports are inexpensive and easy to prepare from the information that is readily available to most clinicians.

Case reports and case series, however, are only an initial step in understanding disease patterns. They rarely reflect disease in the general population and can be strongly affected by selection or referral bias. For example, individuals with rare and new diseases may be sent to a particular hospital and thus become common there but remain infrequent in the broader population.

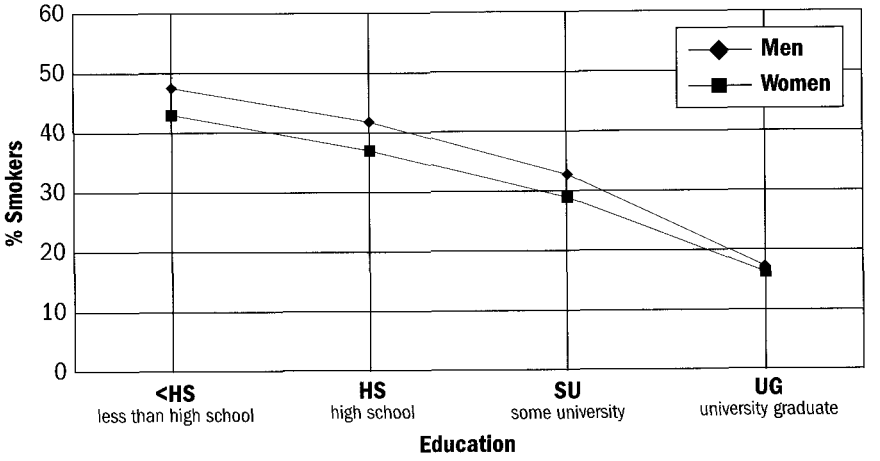
Cross-sectional surveys

A cross-sectional study seeks to define disease status, exposures and other characteristics in a population at a given point in time. Associations between disease states and other characteristics can be evaluated using this method. The prevalence of disease in a population can be ascertained along with the prevalence of associated factors. Although a cross-sectional study technically covers a discrete point in time, practical considerations mean that surveys last over a defined interval, but rarely more than a year. Longer survey periods are subject to changing trends in disease and associated characteristics (secular trends). On the other hand, shorter periods may introduce seasonal or other biases.

There are numerous examples of cross-sectional surveys. Figure 3.3, for example, shows a cross-sectional study of a large urban area demonstrating an association between cigarette smoking and educational level (*2*). A population sample was surveyed and individuals were asked about their smoking habits and their level of education: in many industrialized countries cigarette smoking is increasingly associated with the less educated and less affluent groups.

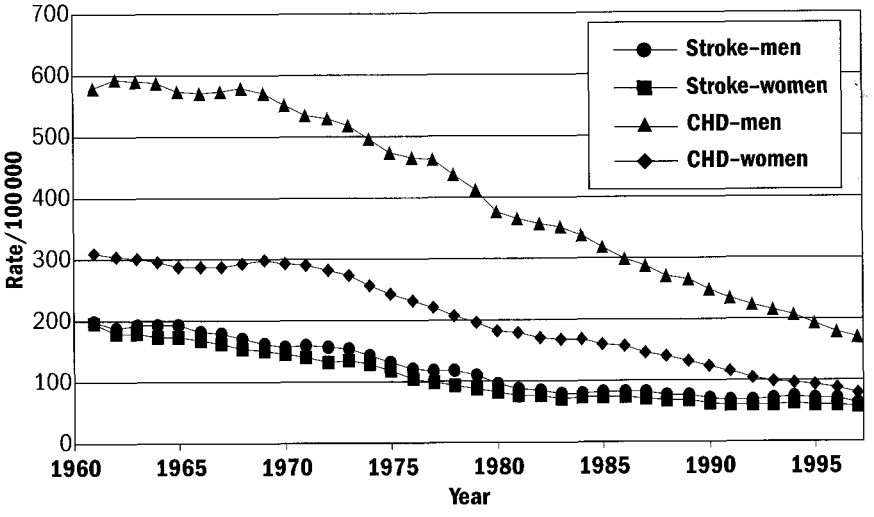
Periodic cross-sectional surveys of a population employing identical measurement methods are used to evaluate disease trends. Mortality statistics, for example, represent a form of multiple annual cross-sectional survey of causes of death in a population (Figure 3.4). Other examples of regular cross-sectional surveys are found in data from the WHO MONICA Project (Monitoring trends and determinants in CVD) where data on CHD incidence and treatment surveillance were maintained over a decade (*3*). Figure 3.5 shows data from several sites in the MONICA study for active

Fig. 3.3. Smoking prevalence and education level



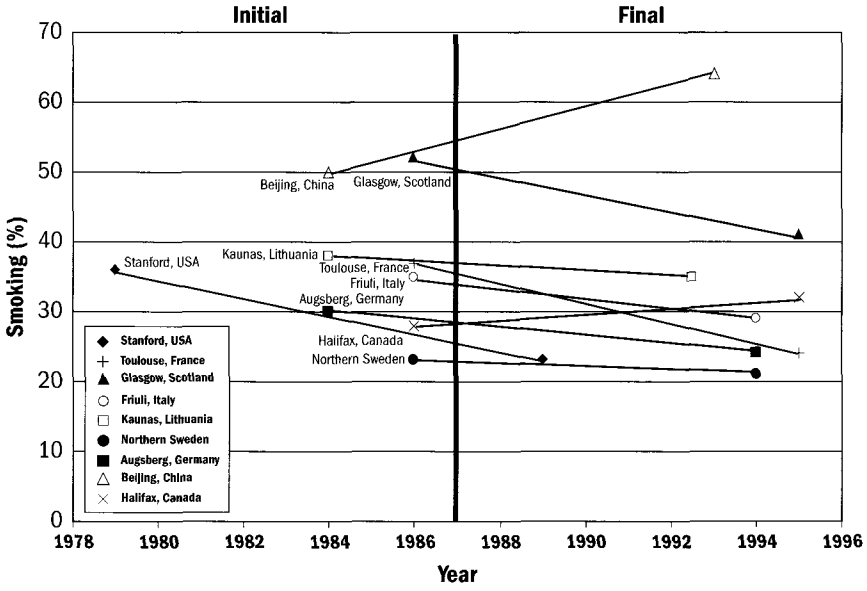
Source: Iribarren C et al. Twelve-year trends in cardiovascular disease risk factors in the Minnesota Heart Survey: Are socioeconomic differences widening? *Archives of Internal Medicine*, 1997; 157:873–881.

Fig. 3.4. Age-adjusted stroke and CHD mortality 1960–1998: all Minneapolis–St. Paul residents



Source: Minnesota Heart Survey, unpublished material.

Fig. 3.5. Coronary incidence rates per 100 000 population (age 35–64)



Source: Molarius A et al. Trends in cigarette smoking in 36 populations from the early 1980s to the mid-1990s: Findings from the WHO MONICA project. *American Journal of Public Health*, 2001; 91:206–212.

smoking in men. At each site, two independent surveys were conducted approximately 10 years apart to describe cigarette smoking trends (*4*).

Although cross-sectional studies usually focus on specific and defined populations, the unit of measurement is the individual, with information on various health and disease characteristics being collected from each person surveyed. This approach allows comparisons to be made between individuals and groups.

There are many uses for cross-sectional surveys. They are essential to describing the disease status and health of populations, and this information in turn is critical to the development and planning of health services and medical care facilities. The study design also allows determination of population characteristics and the association of environmental and individual risk factors with disease status. Cross-sectional studies have allowed the development of several hypotheses on exposure–disease relationships. Repeated cross-sectional studies are the basic method for determining disease rates, exposure and risk factor trends in a population, and are therefore, essential for assessing the effectiveness of prevention and control strategies.

Cross-sectional surveys also have limitations. For instance, they are of little use in studying rare diseases where a large and costly population survey may reveal only a few prevalent cases. They are also unsuitable for studying diseases of short duration, except in outbreaks in small, well-defined population groups. In the case of CVD, rarity and short duration are not usually issues—most diseases are both common and chronic.

While cross-sectional studies can provide important information on associations and suggest causal connections, the approach also has limitations for chronic diseases. The simultaneous measurement of disease status, exposures and risk factors may distort the picture of the disease. This is particularly true of CVD since the risk factors and exposure may precede the condition by many years and treatment advice after diagnosis may alter associated factors in the presence of disease. For example, individuals with a previous myocardial infarction may smoke less because they have been encouraged to quit after their event, or the heaviest smokers may have died as a result of their first infarction. A typical cross-sectional survey may therefore suggest that smoking and myocardial infarction are unrelated or negatively related because survivors appear to have lower smoking rates. This example illustrates the importance of a detailed smoking history and of determining when a patient ceased to smoke.

Case–control studies

Case–control studies are commonly used to study factors associated with disease. Because they are particularly suited to uncommon or rare diseases and must assess exposure retrospectively, they are less frequently used in CVD than for cancers or other rarer diseases. Nonetheless, the case–control methodology can be useful in the study of CVD.

In a case–control study, cases (those with disease) are compared with controls (those without disease) to determine whether the exposure of interest is more, or less, common in the cases. The initial step is the selection of cases and controls for comparison. Cases are individuals who, according to a strict definition, have a specified illness or condition; they are chosen to be as alike as possible in their disease status. Ideally, cases should include all new or incident events in a defined population over a specified period of time. A comprehensive disease registry, on a population basis, is the "gold standard" in assembling cases. The WHO MONICA Project has been designed in this way (*3*). In most circumstances, however, this ideal does not prevail. In practice, cases are more commonly collected from one institution or clinic and do not represent any specific population other than patients who attend that clinic. Frequently the cases are patients whose conditions have already been diagnosed and who have undergone extensive treatment before initial contact.

Selection of appropriate control patients presents a similar challenge. Controls are individuals without the disease but with the same background characteristics as the

patients with disease. Controls should come from the same community as the cases. In practice, controls are generally recruited from patients referred to the same hospital, usually with other diseases, and they are matched with cases on basic demographic characteristics such as age and sex. The result of a case–control study is expressed as an odds ratio (OR) which describes the association of the diseased (cases) having a certain exposure or presumed risk factor with the presence of that characteristic in controls.

Table 3.1 shows an example of a case–control study in CVD. An association was postulated between coffee intake and increased rates of myocardial infarction. Patients with incident non-fatal myocardial infarction were drawn from 78 hospitals. Controls were drawn from patients at the same hospitals admitted for trauma or non-respiratory infections but with no history of myocardial infarction. Those with other evidence of CHD were excluded. As shown in Table 3.1, individuals with high intake of coffee, with or without caffeine, had an adjusted OR suggesting an association with myocardial infarction. Other factors such as smoking, obesity, alcohol use and physical activity were controlled in the adjusted analysis.

The case–control study design has strengths and weaknesses. It is quick, easy and relatively inexpensive, and it can be repeated in a number of centres using identical case and control definitions. It is particularly useful when the disease of interest is uncommon and when it is not easy to recruit cases. However, the approach also has limitations. Since many case–control studies are performed in referral hospitals, it is difficult to know whether the cases or the controls are representative of any population other than those referred to the hospital. Because these studies are usually retrospective or based on prevalent cases of the disease, it may be difficult to ascertain whether the associated exposure is important in the causal chain or the result of the disease. These selection factors for patients and controls may result in the odds ratio incorrectly estimating the true relative risk. For example, myocardial infarction and stroke survivors provide an incomplete sample when they are evaluated by the case–control methodology because high case-fatality rate for these diseases eliminates

Table 3.1. Coffee intake and myocardial infarction

Coffee Type	Cups/day	Cases	Controls	Crude OR	Adjusted OR
None	0	147	205	1.0	1.0
Regular	1–2	235	275		1.4
	3–4	351	219		2.0
	5–9	437	199		2.7
	10+	288	70		4.9
Decaffeinated	1–2	48	55		1.0
	3–4	41	26		1.8
	5+	36	15		3.1

Source: Rosenberg L et al. Coffee drinking and nonfatal myocardial infarction in men under 53 years of age. *American Journal of Epidemiology, 1988; 128:570–578.*

an important portion of the population. Finally, diseased individuals are more likely to over- or under-evaluate their previous exposures and experiences on interview when they know that they are ill: controls do not have similar recall bias.

Cohort studies

The cohort survey is the most commonly recognized and widely used study design in CVD. It is uniquely suited to the understanding of common conditions with long development periods and delayed clinical manifestations. Cohort studies offer important evidence about the relationship to subsequent disease manifestations of risk factors found when the patient is healthy.

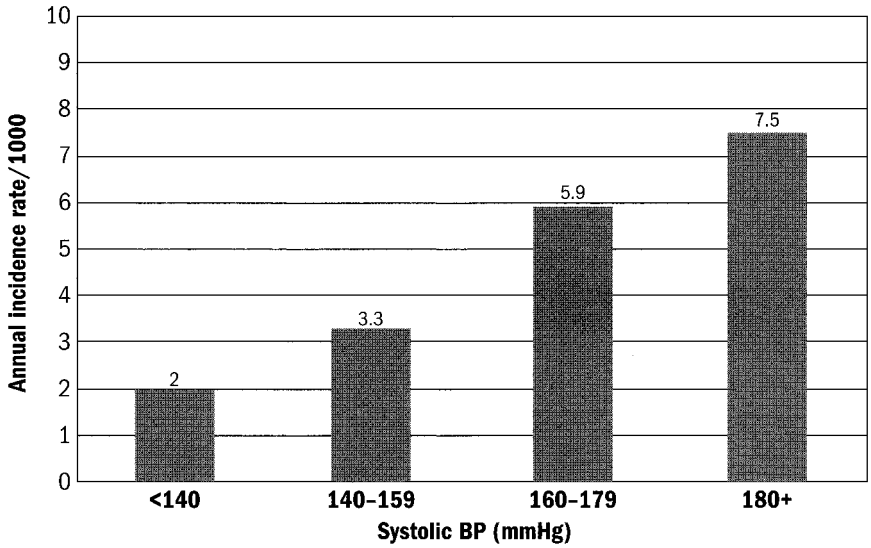
The cohort study starts much as a cross-sectional study. A population is selected and extensive data are collected from a sample or from the entire group, including disease status so as to allow separate analysis of individuals who are already affected. Cohort members are then followed for evidence of disease, the appearance of or change in risk characteristics, and exposure to medical treatment. Follow-up of the cohort after the baseline measurement is an essential feature of the strategy, in contrast to the cross-sectional design.

There are many well-known historical examples of CVD cohort studies, including the Seven Countries Study, Minnesota Business and Professional Men's Study, Framingham Study, Men Born in 1913 Study, Whitehall Study, Israeli Heart Study, Tecumseh Study, Tromsø Study, and the Honolulu Heart Study (*5–13*).

For example, the Framingham Study, which began in 1949, evaluated a group of 5209 adults, aged 30–59, in Framingham, Massachusetts (*7*). This cohort of individuals was examined every second year after the initial evaluation; surveys included a re-evaluation of characteristics thought to be associated with CVD as well as evaluation for disease end-points. Questions about medical history were supplemented by physical examinations and other testing. In addition, local hospitals were kept under constant surveillance to ascertain whether cohort members were admitted. Rigorous criteria were established to define disease end-points, and new events were recorded. For those who died, information on cause of death was collected. This information was then related to the participants' characteristics found in the surveys before the onset of disease. In this way, it was possible to demonstrate the effect of such characteristics as systolic blood pressure on subsequent sudden death (Figure 3.6). Data from multiple measurements also showed the effect of change over time in physical characteristics, such as weight, on disease end-points (Figure 3.7).

At the international level, the Seven Countries Study enrolled cohorts of healthy middle-aged men around the world (*5*). Baseline measurements were collected in the late 1950s. The men were then evaluated at 5-year intervals for risk factor levels, disease endpoints and other health characteristics. The study provided information

Fig. 3.6. Systolic blood pressure at initial exam and sudden death



Age-adjusted incidence of sudden death according to systolic blood pressure. The number of sudden deaths associated with each group's incidence rate are 43, 32, 22 and 12.

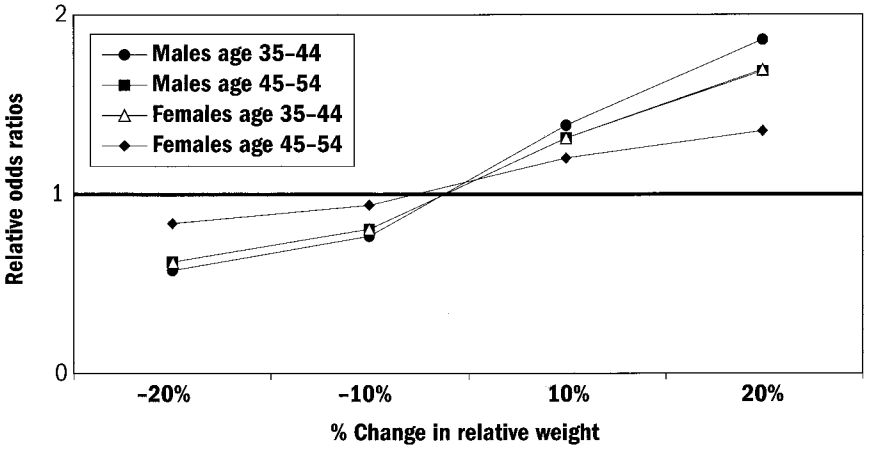
Source: Kannel WB et al. Precursors of sudden coronary death: Factors related to the incidence of sudden death. *Circulation*, 1975; 51:606–613.

on risk factors and disease within countries but also allowed comparisons across countries based on identical information. The study demonstrated the universal importance for CVD of risk characteristics such as lipids, smoking and blood pressure, and provided an understanding of the widely differing health habits and disease prevalence in different populations.

Cohort designs may include a variety of population groups. They may include groups that share high exposure to specific factors, e.g. cigarette smokers or industrial workers; others that share the same occupation, behaviours, or level of education, e.g. physicians, college graduates, military pilots and athletes; and others that share a common geographical origin or residence, e.g. Honolulu, Hawaii, or Bogalusa, Louisiana (*13, 14*).

Cohort studies are usually prospective, with individuals surveyed at baseline and then followed through time, but may be retrospective, looking back in time. In a retrospective study, a cohort was previously defined, usually for another purpose, and its health characteristics were evaluated. The cohort is then surveyed or measured for

Fig. 3.7. Coronary heart disease risk and weight change



Relative odds ratios corresponding to changes in relative weight in Framingham Cohort Exams 2-9.

Source: Ashley FW, Kannel WB. Relation of weight change to changes in athrogenic traits: The Framingham Study. *Journal Chronic Diseases*, 1974; 27:103–114.

subsequent disease. Young men evaluated for induction into military service (*15*), or individuals who underwent evaluation on their entrance into university commonly lend themselves to this type of retrospective cohort study (*16*), since records are usually kept of these initial evaluations and are available for review. A historical study is one that includes elements of both retrospective and prospective studies. The cohort is defined at a time before the start of study and then evaluated at regular intervals thereafter to determine disease experience (*17*).

There are several advantages to cohort studies. They allow a direct determination of risk characteristics and the relative risk of developing disease. Because people are classified according to strict evidence of disease, the relationship between measured characteristics and subsequent events is viewed as strong causal evidence (temporality). Cohort studies provide information about the lag time between risk characteristic appearance and disease manifestation. Depending on the selection of the population, it is usually much easier to generalize the findings of a cohort study than those of a case–control study.

One technique that is frequently used as an integral part of cohort studies is the nested case–control study. Here the cohort study provides both the new cases and the controls. Specialized tests may be performed on this smaller number of subjects

to test a hypothesis at reduced cost. For example, genetic testing may be performed on stored DNA, comparing cases with controls in a cohort subgroup.

Among the limitations of cohort studies of chronic diseases are many years of effort and sustained support that are necessary. The studies are costly and difficult to implement and maintain. Loss to follow-up, where cohort members become unavailable or are unwilling to be re-examined, is a particular problem. Multiple measurements may cause subjects to change their behaviours or seek additional medical care. Finally, cohort studies are most effective for common diseases, and less effective for rare diseases because of the very large populations that must be assembled and measured to generate a few cases.

Trials

There is often a need to evaluate the effects of an intervention on individuals or populations, to ascertain its real benefits and risks, to confirm causal relations, or to assess interventions with potential clinical benefit. Scientists who perform observational studies such as those described above may also perform controlled trials using many of the same research methods.

Controlled trials consist of the administration of an intervention to a study group and comparison of defined outcomes with those in a similar group that has not received the intervention. The two groups should be comparable, in all respects, except for the intervention under evaluation. This is usually achieved through the method of randomization that should result in an equal allocation of confounders between the two groups. Such an experimental design provides the most unbiased evaluation of the independent effects of any intervention. The presence of a control group is essential to distinguish between the true effects of the intervention and changes that may occur as part of the natural history or other concurrent influences. This may be an untreated group, placebo controls (in a drug study) or randomization to the usual type of care (to compare with a new approach).

Community trials

Community intervention approaches are being increasingly used in CVD prevention research. There are examples of such studies in many countries; they include the North Karelia Project, the Stanford Three Town and Five Cities Studies, the Minnesota Heart Health Project, the Pawtucket Heart Health Project, and the German Cardiovascular Project (*18–24*). In each of these, intervention communities were selected and compared with matched reference areas.

Community projects offer distinct advantages but present unique methodological problems. Usually they are long-term projects to provide adequate time for the evaluation of CVD outcomes. Demonstrating the effectiveness of primary prevention

interventions aimed at modifying the risk of CVD and its consequences requires many years. Furthermore, the intervention is applied to the whole community which is thus more likely to represent the complete spectrum of the disease and its determinants in that population. The results of a community intervention have great generalizability in their applications at the population level. Community-based studies also provide for "effectiveness" evaluation under operational "real world" conditions in contrast to most clinical trials which take place in artificial experimental settings and which provide for "efficacy evaluation". Impact on total mortality is also easier to assess in community trials, which may have sufficient "power" to evaluate coronary mortality and total mortality.

However, community projects also suffer from several drawbacks. First, the sampling unit is the whole population, not the measured individual, so that the number of units for comparison is limited (*25*). Secondly, the community units may not be fully comparable with respect to known and unknown confounders that may influence the outcome. In a clinical trial, the process of random allocation of individuals provides for maximal comparability of known and unknown confounders between the intervention and control groups. Such a balance is unusual in a community project because it is usually impossible to deal with more than a few units. Even if adjustments are made for known confounders, differences in unknown confounders may influence the outcome. Thus, an unbiased assessment of the independent effect of the intervention cannot be assumed.

Another methodological problem in community projects is "contamination". When an intervention starts in a community, its effects are likely to diffuse into the control community through increased public awareness (*18*). For example, health messages in North Karelia also diffused to the control community in Kuopio (*18*). This contamination tends to reduce the net impact of an intervention. The strict separation of the two groups in terms of the intervention of interest, which is a strength of an experimental study, may not be fully achievable in community studies.

Finally, because of their scope and complexity, community trials are very costly and difficult to implement.

More about the design and analysis of community trials can be found elsewhere (*25, 26*).

Clinical trials

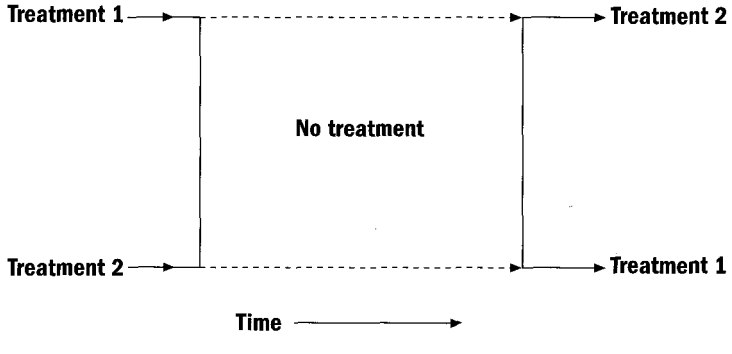
Most clinical trials are randomized, that is, the participants are randomized to an intervention group receiving the treatment and a control group receiving "standard" treatment or an inert placebo. An "uncontrolled" trial, in which the effects of intervention are assessed, as a "before" and "after" comparison in a single group is a much weaker design.

Clinical trials may be of "parallel group" or "cross-over" design (Figure 3.8). In the former, a population is randomized to different regimens (new therapy versus standard therapy, for example) throughout the trial period; in the latter, each randomized group is exposed sequentially to both regimens. The trials may be "open" (unblinded, unmasked) or "blinded" (masked), depending on whether the subjects and physicians are aware of the nature of the intervention being administered to each subject. If only the subjects are unaware, the trial is said to be single blind; if both subjects and research staff are unaware, the trial is spoken of as double blind. Where possible, those involved in treatment assignment and in the assessment of outcomes should be unaware of the nature of the intervention.

Randomization of treatment allocation is mandatory for unbiased comparisons. Systematic schemes (quasi-randomization), in which every alternate subject is assigned to the intervention, are suspect because future assignment is predictable. Such methods are vulnerable to manipulation by staff and to other biases. Physicians may manipulate assignment because they want, more or less consciously, to make their preferred therapy look better, or because they genuinely believe that one form of therapy is better for their patient. In any case, the desired balance between the two groups is lost. Other forms of assignment, using birth dates, odd or even date of admission, or telephone numbers, are also flawed as the staff can ascertain a subject's treatment assignment. Selective enrolment may be manipulated and a lack of blinding may lead to a biased assessment of outcome events.

Compliance with the treatment protocol is essential in clinical trials. Measures used to promote compliance with therapy should be applied equally in both treatment

Fig. 3.8. Crossover experiment



Two treatments are compared by a design that allows the two randomized groups (1 & 2) to be exposed to each treatment with a no-treatment washout phase before crossing over.

and placebo groups. Blinding of the subjects and of research staff protects against selective application of compliance-enhancing measures.

A good trial report must provide full details of the follow-up of patients in each group so that all subjects are fully accounted for. While withdrawals before randomization (due to lack of consent, for example) are permissible, all patients who have been randomized must be included in the analysis (intention-to-treat principle). Post-randomization withdrawals are to be avoided as far as possible. If unacceptable side-effects lead to drug withdrawal, this fact must be reported, and the patients must continue to be counted in the group of original allocation: in effect, they constitute treatment failures and merit inclusion in that category. Ideally, there should be no loss to follow-up. Any such losses may be analysed by applying the group averages to them. The test treatment is still most likely proven to be better, despite the worst-case of assuming that all of its follow-up losses have unfavourable results.

Non-compliance or unintended cross-overs are usually analysed as though they are still a part of the group to which they were originally allocated ("intention to treat" analysis). This may appear illogical but is necessary to maintain the balance created by the randomization. If they are excluded, the groups will no longer be comparable in terms of baseline characteristics and the trial may be flawed. Therefore, in order to preserve the structure of randomization and comparability, these protocol violations are regarded as originally assigned. Such an analysis is also referred to as an "effectiveness trial" in which the net benefit from the therapy is assessed, in operational conditions that include drop-outs and cross-overs. A trial may report both types of analysis. An "intention to treat" analysis must, however, be reported in all trials. In some situations, such an analysis is the only logical one despite the impression that it is not.

Consider a trial of coronary artery bypass surgery versus medical treatment for unstable angina pectoris in which 20% of the patients assigned to medical treatment cross over to surgery by the end of the trial period. Any "intention to treat" analysis that includes these patients as part of the medical therapy group seems open to question. However, if the research aims to determine whether the policy of initially treating the patients with unstable angina with medical therapy is comparable to the policy of initially treating them by surgery, the situation becomes clearer. Initial medical therapy does not preclude surgery as a later option. Cross-over to surgery one year later does not mean that surgery should have been the initial option. If both groups fare equally well at the end of 5 years, then the policy of initial medical treatment, with the option of later surgical referral when necessary, is not flawed. Here the "intention to treat" analysis does justice to the question of choosing the initial line of management. A comprehensive description of clinical trial methods can be found elsewhere (*27*, *28*). A checklist for evaluation of clinical trials is shown in Table 3.2.

Table 3.2. A checklist for evaluating a clinical trial

1. Was a primary hypothesis clearly stated?
2. Were the groups randomized?
3. Was comparability of the groups ensured at entry and/or analysis?
4. Were effective blinding procedures followed?
5. Was there any contamination between the intervention groups?
6. Was the follow-up complete, with equal compliance and diagnostic surveillance in treatment and control groups?
7. Were the outcome criteria clearly defined and their reproducibility assessed?
8. Were appropriate statistical tests of the hypothesis applied?
9. Was the sample size adequate to test for a clinically relevant difference?
10. Are the results of the study generalizable?
11. Was the study ethically sound?

Clinical trials are complex and costly, requiring sustained, substantial effort on the part of the investigators.

Genetic epidemiology

There is a growing realization that the study of genetic variation will contribute to the understanding of common disorders such as CHD and cancer. Interest in this area has been kindled by an important study from the Swedish Twin Registry in which 21 004 monozygote twins were investigated (*29*) for the risk of death from CHD. Relative hazard estimates were obtained in a multivariate survival analysis. The study assessed the relative hazard of a twin's death from CHD when the other twin had died from this cause before the age of 55 years, compared with the hazard when the twin had not died. The relative hazard in monozygotic twins was roughly double that of dizygotic twins and, although the effect decreased with age, an excess persisted well into old age.

It is now well recognized that susceptibility to most CVD is polygenic, that is, it involves many genes, and that it is also influenced by interactions with environmental factors. Newer methods permit the search for any significant contribution of single genes to diseases of complex etiology and to variation in quantitative traits such as plasma cholesterol. For complex diseases, as for rare single gene defects, statistical analysis of family data can provide clues to the existence of a basic metabolic defect. However, it is considerably more difficult to demonstrate that a familial pattern of disease is consistent with Mendelian segregation ratios for a disease with a complex etiology than for a simple autosomal dominant or recessive disorder. Although numerous genetic loci may determine susceptibility to CVD, most may have only a minor effect. There may be one or a few that independently, or in combination, have a major effect. In addition, complex diseases such as CHD may exhibit etiological heterogeneity, with one locus having a major impact in some families, and another locus or an environmental variable having a bearing in others. Because of these intricacies, statistical analyses of family patterns of such diseases may sometimes yield ambiguous results. The statistical methods used include association studies, linkage

analysis and segregation analysis. A straightforward account of these techniques can be found in *Molecular genetics for the clinician* (*30*); *Genetic variation and human disease* (*31*) provides a more advanced text.

The study of DNA structure has been greatly facilitated by the use of restriction endonucleases—naturally occurring enzymes that will cleave DNA only at sites of a specific nucleotide sequence. Nucleotide insertions, deletions or substitutions, which either create or destroy a recognition site, will alter the size distribution of the DNA fragments and these will be detectable after hybridization to a specific cloned or synthesized hybridization probe (*32*). DNA variations can be arbitrarily divided into two classes. One class comprises DNA variations that have a direct cause–effect relationship with the disease phenotype; the other represents those variations that have no such cause–effect relationship with disease, either because they occur outside coding regions or because they are mutations which have no effect. The majority of these "restriction fragment length polymorphisms" fall into the latter class. Although they result from events independent of those directly responsible for the disease phenotype, they can be used as a marker of a segment of the chromosome on which a gene functions in an abnormal fashion. The frequency of association between a gene marker and a "disease" gene will depend on variations in the frequency of the marker in different populations. In addition, it is important to remember that genetic markers are based on a concept of statistical association rather than on physical proximity of the relevant loci.

Another issue often not addressed in studies of association among genetic markers and disease is that of biological interaction between the genetic marker and other genetic and environmental factors. In the presence of gene–environment interaction, a truly causal association between a marker and disease may be diluted by the proportion of the population that is or is not exposed to environmental factors that interact with the susceptible genotype (*32*).

While these types of study can provide evidence of association at a population level, the contribution of a major locus to disease susceptibility is more convincingly demonstrated when alleles at a marker locus co-segregate with disease in families. The lack of proven association at a population level does not preclude the marker concerned associating with a disease locus in a particular family. Linkage analysis thus involves calculation of the likelihood of obtaining an observed transmission pattern of a disease and a marker in a set of families (*33*). When two loci are situated closely together on the same chromosome they tend to be transmitted together from parent to offspring. If recombination events occur at this site during meiosis—when DNA replicates during the production of germ cells—the resulting chromosome will contain a paternally derived allele at one locus and a maternally derived allele at the other. The probability of recombination, the recombination fraction, is a function of the distance between the two loci, taking a maximal value of 0.5 for loci on different chromosomes or 0 for loci that are so closely linked that recombination (a

combination of genes in the offspring, not seen in the parents) is never observed. In linkage analysis, the likelihood of the observed transmission pattern in a set of families is calculated for various values of recombination fraction Ø. In order to distinguish between coincidence and significance, the probability of a set of observations representing true linkage is expressed as the "logarithm of the odds" or the lod score (Z). Methods for estimating the power of a family structure to produce significant linkage tests have been worked out for various conditions and are beyond the scope of this text (*34*).

Segregation analysis involves calculating the likelihood that a particular familial pattern of disease would be observed under alternative hypotheses concerning genetic and environmental contributions to susceptibility. This is done by comparing the observed proportion of affected sibs and offspring with the proportion expected according to a particular genetic hypothesis.

For a "major locus" model, there are essentially two key parameters: the proportion of individuals of each major locus genotype who are affected (that is to say the penetrance for each genotype) and the frequency of the disease allele. The values of these parameters are estimated using a likelihood approach: the likelihood of observing pedigree data under a sporadic, or alternatively a major locus model, is determined and compared (*35*).

Other factors that can lead to misinterpretation of the results of segregation analysis are the presence of undetected age or sex effects, population and etiological heterogeneity, and shared environmental factors that contribute to disease susceptibility. These problems are compounded in the study of CHD because the disease end-point, however defined, is usually the result of an interacting sequence of events, only some of which are directly related to a specific gene product. Thus the analysis of quantitative, rather than dichotomized, traits is considerably more powerful. Frequently, the levels of quantitative risk factors for atherosclerosis are adjusted for age and sex effects before analysis; alternatively, these factors may be included as covariates. In addition, because the distribution of some traits (such as the triglyceride level) is skewed, log-transformation may be required before analysis. In the "measured genotype" approach the average effect of each allele is computed as the difference in the trait value between individuals who carry the allele and the population mean. The contribution of the locus to quantitative variation of the trait is the ratio of the variance among the mean trait values for the separate genotypes to the total phenotypic variance (*36*).

These methods were elegantly demonstrated in the Nancy Study, France, (*37, 38*) by Cambien et al., who investigated the origin of familial resemblance of plasma angiotensin converting enzyme levels. In this study, plasma-angiotensin converting enzyme (ACE) levels were correlated with age in offspring but not in parents; height, weight and blood pressure were uncorrelated in both parents and children. After

adjusting for age, the father–offspring, mother–offspring and sibling correlations were 0.166, 0.323 and 0.303 respectively. The question remains as to why, if the trait is deleterious, it is so common. The probable reason is that it conferred an advantage at some time during human evolution. There are other examples of this: sickle-cell disease is very common in some parts of the world where it protects against malaria (*39*).

Not surprisingly, analysis of genetic data from different populations sometimes yields discrepancies in results. Disease susceptible loci detectable in one population may not be found in another because of different environmental risk or other genetic or racial factors, or because of differing methods of ascertainment. While these potential shortcomings must always be borne in mind, more advanced statistical approaches are continually being developed that can at least partially take account of such factors. For recent reviews of this area see references (*40, 41*).

Sample size and power

An adequate sample size capable of detecting relevant differences is an essential prerequisite to any study. For all observational and trial studies the description of the methods should include pertinent details of the calculation of sample size.

Sample sizes are estimated by using appropriate formulae for detecting statistical differences between proportions (for dichotomous events such as death, stroke or myocardial infarction), or differences between means (for continuous variables such as exercise time, blood pressure and serum cholesterol). A study sample size is based upon:

- An estimate of the event rate or the mean level of the continuous variable of interest in the population, usually taken from previous work in the literature.
- Δ (delta)—biologically relevant differences that the study seeks to detect, if present.
- α (alpha)—probability of the observed difference arising by chance (probability of a type I error or false positive).
- β (beta)—probability of missing a true difference (probability of a type II error or false negative). 1-β is the ability of the trial to detect a difference, if it truly exists, and is known as the "power" of the study.

The conventional significance level is 0.05 but may vary from 0.1 to 0.01. The sample size will also vary according to whether the test is one- or two-sided. If the alternative hypothesis is at all possible (the new treatment could be better or worse than the control therapy, or one group might have a higher or lower value than the other), a two-sided or two-tailed alpha is used. A two-tailed alpha is preferred because it recognizes that alternative outcomes are possible. A one-tailed alpha requires a smaller sample size. The beta (β) is conventionally set between 0.05 and 0.20, i.e. power of 0.80 to 0.95.

An appropriate formula for the calculation is as follows:

$$SE = \sqrt{pq/n}$$

where SE = standard error
p = proportion affected
$q = (1 - p)$
n = number in sample.

This simple formula is slightly inaccurate when rates are low.

In comparisons between two studies, the significance of the differences between estimates is given with sufficient accuracy by:

$$\frac{p_1 - p_2}{SE\ difference}$$

where p_1 = proportion affected in first sample
p_2 = proportion affected in second sample.

This function has a normal distribution.

The SE difference is given by the expression:

$$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

In order to calculate the required minimum values for n_1 and n_2 (the sample sizes), a second type of error must be considered—namely, that of failing to demonstrate a difference that really exists. If an investigator is prepared to accept an 85% chance of recognizing a real difference between two populations with various specified true rates, some of the required sample sizes may be obtained from Table 3.3.

Table 3.3. Required sample sizes for giving an 85% chance of recognizing a specified difference in rates (1-β power) between two populations, significant at the 5% level, (two-tailed test)

Estimated true rates (%)		Required sample sizes
p_1	p_2	($n_2 = n_1$)
5	2	670
10	5	490
15	5	160
20	5	80
20	10	230

Studies of continuous variables

The error in estimating the mean value of a continuous variable depends not only on the sample size but also on the amount of variation of individual values between subjects:

$$\text{Standard error} = \sqrt{SD^2/n}$$

where SD = standard deviation of individual values for each subject
n = number of subjects.

If this variability is known approximately, either from previous studies or from a pilot study, then it may allow the advance estimate of either the confidence range that will be obtained from a specified sample size or the sample size that will yield a specified confidence range.

The variance of individual values between subjects has three components: true between-subject variation; biological variation within subjects; and measurement error. The effects of the last two may be reduced by replicate measurements, and measurement error may also be reduced by attention to technique. Such steps will reduce the necessary sample size (or improve confidence limits).

In comparisons between two studies, the significance of a difference observed between the two estimates of mean values is given with sufficient accuracy by:

$$\frac{x_1 - x_2}{SE\ difference}$$

where x_1 and x_2 are the two mean values ($x_1 - x_2 = \Delta$), and SE difference = $\sqrt{SE_1^2 + SE_2^2}$

This function has normal distribution.

In planning a comparative study the investigator may wish to get some idea of the necessary sample sizes. First there is a need to specify: the magnitude of difference that it is desired to detect; the significance level at which any difference can be demonstrated; the acceptable risk for failing to demonstrate a real difference; and estimates of the variability of individual values.

Table 3.4 illustrates the sample sizes needed to meet various specifications. It may also be used to indicate the magnitude of difference that an investigator may reasonably hope to detect between two samples of specified size. The table should be used only as a rough guide, since the means and standard deviations on which it is based will not actually be known until after the study is completed. In addition, it

Table 3.4. Required sample sizes for giving an 85% chance of recognizing a specified difference in mean values between two populations, significant at the 5% level (two-tailed test)

Standard deviation of individual values	Required sample size in each sample for demonstrating the following differences in mean values:					
	1	2	5	10	25	50
1	20	<10	<10	<10	<10	<10
2	70	20	<10	<10	<10	<10
5	450	110	20	<10	<10	<10
10	1800	450	70	20	<10	<10
25	11000	2800	450	110	20	<10

assumes equal variability in the two populations and ignores the risk of biased measurement errors.

Other considerations

In the case of a quantitative variable, any random measurement error naturally affects the accuracy of results and may necessitate an increase in the size of the study. For qualitative variables, on the other hand, random error does not affect the standard error of a prevalence estimate, and it is important only in so far as it tends to be associated with systematic error. In either situation it is systematic error that is really dangerous; the effects of systematic error are not reduced by increasing the size of the sample.

The choice of sample size is of course affected by many practical issues besides the statistical estimates. This does not mean that an investigator should begin a study without first making the best possible calculations of whether or not it is likely that reasonable answers to the study questions will be obtained, but there are other considerations. Too large a study is laborious, and the overstretching of resources may impair the quality of results. Many groups have found that the detailed examination of 500–1000 subjects is ample work for one study. If large numbers are necessary, it may be better to carry out two smaller studies than a single large one.

More details on sample size calculations are found in statistical textbooks (*42–46*).

Statistical analysis

Table 3.5 provides a brief guide to commonly applied statistical tests.

A common problem is multiple significance testing. If a large number of variables are tested for differences between two groups, some variables may show statistically significant differences by chance alone. Subgroup differences may emerge even when

Table 3.5. Commonly applied statistical tests

Design	Data	Statistical test
Two independent groups	Proportions (nominal) Rank ordered (ordinal) Measured quantities (interval ratio) Survival data	Chi-square; Fisher's exact Mann-Whitney U Unpaired t-test Mantel-Haenszel
Two related groups (before/after; matched pairs; cross-over)	Proportions Rank ordered Measured quantities	McNemar's chi-sq; Binomial test Sign test; Wilcoxon signed rank Paired t-test
More than two independent groups	Proportions Rank ordered Measured quantities Survival data	Chi-square Kruskal-Wallis Analysis of variance Log rank
More than two related groups	Proportions Rank ordered Measured quantities	Cochran Q Friedman Analysis of variance
Multivariate	Proportions Measured quantities	Log linear models Multiple regression Discriminant function
Cohort design Case–control	Rates exposed and unexposed Unpaired or paired	Relative risk Odds ratio

the overall study shows no difference. Such subgroup analyses should always be considered as hypothesis-generating and not as hypothesis-testing. Each positive subgroup difference needs to be tested afresh with a study designed for that purpose. Differences often vanish when tested.

The best way of avoiding this problem is to state clearly the primary and secondary hypotheses before the study is done (a priori) and to design the study on the basis of these prior hypotheses. Secondary hypotheses may also be stated and analysed but these should be specified in advance and not be the results of "post hoc data dredging".

Health research does not always consist of hypothesis testing (statistical significance). It also includes estimating the magnitude of the biological effect (clinical significance). This is done by estimating the 90% or 95% confidence interval around the observed treatment effect, and it has two important implications. First, in a "positive study" it is possible to obtain a statistically significant result ($P < 0.05$) by increasing the sample size to such an extent that even minor differences can be detected as significantly different. It is said, that "with small studies you can prove nothing, with large studies you can prove almost anything". The 95% confidence interval may show the range of differences to be irrelevant despite the statistical significance. Second, in a "negative study" with inadequate power, an estimate of the likely benefit versus risk

may still be useful. For example, if it is found that the 95% confidence interval ranges from a possible 3% increase in adverse clinical outcome to a possible 27% reduction in that outcome, thanks to an inexpensive and easily administered mode of therapy, this will encourage efforts to test the therapy in a larger trial. A good study must therefore report the confidence intervals for all its estimated end-points. Statistical textbooks provide detailed information (*42–46*).

Risk assessment

Estimation of risk of a disease outcome is a common result of epidemiological studies. There are several statistical indices for quantifying risk at the individual level as well as at the population level.

Relative risk

	Disease (present)	**Disease (absent)**	
Exposure (present)	*a*	*b*	(*a* + *b*)
Exposure (absent)	*c*	*d*	(*c* + *d*)
	(*a* + *c*)	(*b* + *d*)	

$$\text{Relative risk} = \frac{a}{a+b} \div \frac{c}{c+d}$$

Although this estimate views exposure as dichotomous (absent or present), different levels of relative risk can be calculated for different cut-off levels of a continuous variable such as blood pressure or serum cholesterol.

Odds ratio

The odds ratio for a set of case–control data is the ratio of the odds in favour of exposure among cases (*a*/*c*) to the odds in favour of exposure among non-cases (*b*/*d*). This reduces to *ad*/*bc*. The disease odds ratio for a cohort or cross-sectional study is the ratio of the odds in favour of disease among the exposed (*a*/*b*) to the odds in favour of the disease among the non-exposed (*c*/*d*). This also reduces to *ad*/*bc*.

Attributable risk

Attributable risk is the rate of an outcome in exposed individuals that can be attributed to the exposure. It is derived by subtracting the rate of the outcome (usually incidence or mortality) among the unexposed from the rate among exposed individuals.

Population attributable risk

Population attributable risk (PAR) is the incidence of disease in a population attributable to exposure to the risk factor. It is often expressed as a percentage:

$$\text{PAR} = \frac{P_e(I_e - I_u)}{P_t \times I_t} \times 100$$

where:

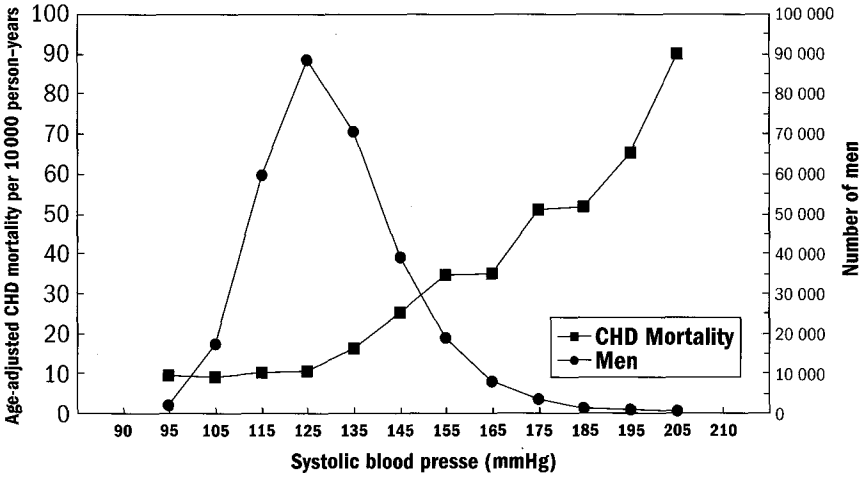
P_e is the number of persons exposed
P_t is the number of persons in the population
I_e is the incidence rate among exposed
I_u is the incidence rate among unexposed
I_t is the incidence rate for the total population

PAR is a more meaningful figure than relative risk in determining the need for population-based preventive programmes. Even if the relative risk of a particular risk factor is not high, but the risk factor is widely prevalent in the population, the total number of cases attributable to the risk factor may be high. Conversely, a rare risk factor with a high relative risk will not contribute much to population levels of disease. Thus, relative risk is useful for estimating individual risk due to a risk factor exposure, while population attributable risk provides a useful profile of the cumulative risk across the entire population. For example, while severe diastolic hypertension is associated with a higher relative risk of vascular events than is mild diastolic hypertension, there are far more cases of mild than severe diastolic hypertension in the population. Thus, many more vascular events at the population level are attributable to mild hypertension than to severe hypertension. Similarly, data from the Multiple Risk Factor Intervention Trial (MRFIT) cohort reveal how more coronary cases arise in the borderline systolic blood pressure group than in those with severe elevation (Figure 3.9). Preventive strategies seeking control of disease at a national level must pay great attention to population attributable risk (*47*).

Absolute risk

Relative measures such as odds ratio or relative risk may distort the actual danger of proven risk associations. Attributable risk is one method to clarify these associations. Another method is to look at absolute risk, expressed as rate/100 000 or similar values. Actual or absolute risk enhances the relative measures making clinical relevance more clear. For example, it has been observed that sexual intercourse has a relative risk for sudden death of 2–3 compared with sedentary activity (*48*). While this may be initially of concern, the absolute risk of sudden death during intercourse is so small that the activity can still be recommended. In this case, absolute risk measures clarify relative risk.

Fig. 3.9. Systolic blood pressure and CHD mortality



Source: Neaton JD, Wentworth D, for the Multiple Risk Factor Intervention Trial Research Group. Serum cholesterol, blood pressure, cigarette smoking, and death from coronary heart disease: Overall findings and differences by age for 316099 White Men. *Archives of Internal Medicine*, 1992;152:56–64.

References

1. Piantadosi S, Byar DP, Green SB. The ecological fallacy. *American Journal of Epidemiology*, 1988, 127:893–904.

2. Luepker R et al. Socioeconomic status and coronary heart disease risk factor trends. *Circulation*, 1993, 88:2172–2179.

3. WHO MONICA Project Principal Investigators. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. *Journal of Clinical Epidemiology*, 1988, 41:105–114.

4. Molarius A et al. Trends in cigarette smoking in 36 populations from the early 1980s to the mid-1990s: findings from the WHO MONICA Project. *American Journal of Public Health*, 2001, 91:206–212.

5. Keys A. *Seven Countries—a multivariate analysis of death and coronary heart disease*. Cambridge, MA, Harvard University Press, 1980.

6. Keys A et al. Coronary heart disease among Minnesota business and professional men followed for fifteen years. *Circulation*, 1963, 28:381–395.

7. Dawber TR. *The Framingham Study: the epidemiology of atherosclerotic disease.* Cambridge, MA, Harvard University Press, 1980:14–29.

8. Welin L et al. Analysis of risk factors for stroke in a cohort of men born in 1913. *New England Journal of Medicine*, 1987, 317:521–526.

9. Smith GD et al. Plasma cholesterol concentration and mortality. *Journal of the American Medical Association*, 1992, 267:70–76.

10. Goldbourt U, Yaari S, Medalie JH. Factors predictive of long-term coronary heart disease mortality among 10 059 male Israeli civil servants and municipal employees: a 23-year mortality follow-up in the Israeli Ischemic Heart Disease Study. *Cardiology*, 1993, 82:100–121.

11. Ostrander LD, Lamphiear DE. Coronary risk factors in a community: findings in Tecumseh, Michigan. *Circulation*, 1976, 53:152–156.

12. Thelle DS et al. The Tromsø Heart Study: methods and main results of the cross-sectional study. *Acta Medica Scandinavica*, 1976, 200:107–118.

13. McGee DL et al. Ten-year incidence of coronary heart disease in the Honolulu heart program: relationship to nutrient intake. *American Journal of Epidemiology*, 1984, 119:667–676.

14. Webber LS et al. The interaction of cigarette smoking, oral contraceptive use, and cardiovascular risk factor variables in children: the Bogalusa heart study. *American Journal of Public Health*, 1982, 72:266–274.

15. Kocher DB et al. Hyper-low density lipoproteinemia in United States Air Force recruits. *Circulation*, 1973, 48:1304–1310.

16. Hearn MD, Murray DM, Luepker RV. Hostility, coronary heart disease, and total mortality: a 33-year follow-up study of university students. *Journal of Behavioral Medicine*, 1989, 16:105–120.

17. Ruberman W et al. Ventricular premature beats and mortality after myocardial infarction. *New England Journal of Medicine*, 1977, 297:750–757.

18. Puska P et al. *The North Karelia Project: 20 year results and experience.* Helsinki, University Printing House, 1995.

19. Farquhar JW et al. Community education for cardiovascular health. *Lancet*, 1977, 1:1192–1195.

20. Fraquhar JW et al. Effects of community-wide education on cardiovascular disease risk factors: the Stanford Five-City project. *Journal of the American Medical Association*, 1990, 264:359–365.

21. Luepker RV et al. Community education for cardiovascular disease prevention: risk factor changes in the Minnesota Heart Health Program. *American Journal of Public Health*, 1994, 84:1383–1393.

22. Luepker RV et al. Community education for cardiovascular disease prevention: morbidity and mortality results from the Minnesota Heart Health Program. *American Journal of Epidemiology*, 1996, 144:351–362.

23. Carleton RA et al. The Pawtucket Heart Health Program: community changes in cardiovascular risk factors and projected disease risk. *American Journal of Public Health*, 1995, 85:777–785.

24. Greiser E et al. Risk factor modification by primary prevention—final results of the German cardiovascular prevention study in Bremen city. *Circulation*, 1992, 86: I401.

25. Murray DM. *Design and analysis of group-randomized trials.* New York, Oxford University Press, 1998.

26. Bracht N. *Health promotion at the community level 2 new advances.* Thousand Oaks, CA. Sage Publications, 1999.

27. Spilker B. *Guide to clinical trials.* New York, Raven Press, 1991.

28. Friedman LM, Furberg CD, DeMets DL. *Fundamentals of clinical trials, 3rd ed.* New York, Springer-Verlag, 1998.

29. Marenberg ME et al. Genetic susceptibility to death from coronary heart disease in a study of twins. *New England Journal of Medicine*, 1994, 330:1041–1046.

30. Brock DJH. *Molecular genetics for the clinician.* Cambridge, Cambridge University Press, 1993.

31. Weiss KM. *Genetic variation and human disease: principles and evolutionary approaches.* Cambridge, Cambridge University Press, 1995 (Paperback ed. with corrections).

32. Sing CF, Moll PM. Genetics of variability of CHD risk. *International Journal of Epidemiology*, 1989, 18:S183–S194.

33. Morton NE. Linkage and association. In: *Genetic epidemiology of Coronary Heart Disease: Past, present and future*. New York, Alan Liss, 1984:245–265.

34. Risch N. Linkage strategies for genetically complex traits. The power of affected relative pairs. *American Journal of Human Genetics*, 1990, 46:229–241.

35. MacClure JW. Statistical approaches to identifying major locus effects on disease susceptibility. In: Luis AJ, Sparkes SR, eds. *Genetic factors in atherosclerosis: approaches and model systems.* Monograph of Human Genetics, 1989, 12:50–78.

36. Bonney GE, Lathrop GM, Lalouel JM. Combined linkage and segregation analysis using regressive models. *American Journal of Human Genetics*, 1988, 43:29–37.

37. Cambien F et al. Familial resemblance of plasma angiotensin-converting enzyme level: the Nancy Study. *American Journal of Human Genetics*, 1988, 43:774–780.

38. Villard E et al. Identification of new polymorphism of the angiotensin I-converting enzyme (ACE) gene, and study of their relationship to plasma ACE levels by two-QTL segregation-linkage analysis. *American Journal of Human Genetics*, 1996, 58:1268–1278.

39. Fleming AF. *Sickle-cell disease: a handbook for the general clinician.* Edinburgh, Churchill Livingstone, 1982:21–26.

40. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*, 1996, 273:1516–1517.

41. Cambien F et al. Coronary heart disease and genetics, an epidemiologist's view. *Molecular Medicine Today*, 1997, May:197–203.

42. Lemeshow S et al. *Adequacy of sample size in health studies.* New York, Wiley, 1990.

43. Paul SL, Lemeshow S. *Sampling of populations—methods and applications.* New York, Wiley, 1991.

44. Machin D, Campbell MJ. *Statistical tables for the design of clinical trials.* Oxford, Blackwell Scientific Publications, 1987.

45. Allan B. Approaches to sample size estimation in the design of clinical trials—a review. *Statistics in Medicine*, 1984, 3:199–214.

46. Kahn HA, Sempos CT. *Statistical methods in epidemiology.* New York, Oxford University Press, 1989.

47. Rose G. Sick individuals and sick populations. *International Journal of Epidemiology*, 1985, 14:32–38.

48. Muller JE et al. Triggering myocardial infarction by sexual activity. Low absolute risk and prevention by regular physical exertion. Determinants of myocardial infarction. *Journal of the American Medical Association*, 1996, 275:1405–1409.

4.
Principles of study design and measurement

Overview

In addition to the basic principles that underlie scientific inquiry, there are others that are specific to population studies. These principles, ranging from the explicit statement of a hypothesis to appropriate data analysis, are briefly reviewed. Later chapters (Chapters 5 and 9) describe the practical implementation of these principles in cardiovascular surveys.

Hypotheses

Any scientific study begins with the statement of the hypothesis, that is, the question to be investigated. Such statements are generated by a thorough understanding of published and unpublished research, observations by the investigators, and discussions with colleagues working in the field. They are also driven by a need for answers to health problems. Many studies are guided by a single hypothesis while others seek to answer several questions. Whether the study rests on one or several hypotheses, the questions need to be clear, focused, relevant, ethical and answerable by the proposed study.

Study design

The study design chosen is the best or most practical approach to the research issue and frequently requires taking into account and balancing different considerations, including available resources, populations, measurement instruments and feasibility. Under certain circumstances, a large, complex, costly ideal study can be accomplished. Generally, however certain compromises must be made but must not jeopardize the scientific integrity of the research.

Objectives and specific aims

The selection of the hypotheses should translate readily into specific objectives. These in turn need to be restated into clear activities and targets both for monitoring

progress and, importantly, for determining whether the plan is realistic given actual resources.

Population definition

CVD surveys include a definition of the study population, including size and other characteristics. These are essential in the planning of the study and ascertainment of it feasibility. Populations may be defined by geography, work site, school, membership, demographic, physical, disease or other characteristics such as age and sex. The population is chosen based on the hypotheses to be tested, the feasibility of the study, study objectives and population size. A close examination of the population selected may reveal significant and unexpected barriers. For example, the participants may be suspicious of the investigators or their institution or the measurement may call for undressing the participants, which may be culturally unacceptable. Equally, it may be difficult to define the population because of inaccurate census data, a high degree of mobility, or other factors. A clear definition of the population sample, its availability and its size is a crucial step.

Sample size (see Chapter 3)

While a defined population may be available and a survey feasible, the study may lack adequate numbers of participants to allow the questions to be answered. Failure to consider sample size adequately is a major flaw of many studies, frequently discovered only after the data are collected. For example, the number of cases collected in a local hospital might be lower than expected. Statistical consultation is essential and helps investigators to determine the appropriate sample size before data collection. While sample size calculation is partly a statistical exercise, calculations also depend on the estimated characteristics of the group, which are often imputed from experience or from the literature. Greater numbers of participants allow detection of smaller differences between groups and more stable estimates of mean values or rates.

Measurements

Investigators need to select or develop appropriate tests and measurements to enable the questions to be answered. They should first choose measurement instruments previously used in similar studies. This allows for greater comparability with other work in the field, and offers the benefits of previous testing and experience. In certain instances, the questions proposed will require new measurements, which may involve design and evaluation of new instruments, and pilot testing of the suitability of the tests in the intended population. Development of a new measurement demands consideration of the validity of the test and its accuracy in measuring the variable of interest. It also includes reproducibility, i.e. the ability to measure accurately a similar value more than once in the same individual.

Quality control

A formal quality control protocol is essential for any data collection system. All levels of data collection, from sample selection to interviews to measurement of collected samples, require formal standardization with both internal and external quality control. Internal control refers to methods performed within the study by the staff, such as double entry of data or re-interview of a sample of participants to evaluate interview quality. External controls require resources such as a reference laboratory to evaluate duplicate samples. Such quality procedures assure that the instruments used are performing to their original specifications throughout the study, and allow comparisons of values over time within the study. Formal quality control procedures need to be documented and performed at specified intervals.

Data handling

Current survey data are usually entered on-line and tabulated on computers. The availability of powerful portable and desk-top machines at reasonable cost makes this approach universally feasible. Computerization of information allows ready access and editing as well as rapid data analysis. However, powerful statistical software packages now available are not substitutes for good statistical consultation and thoughtful data analysis.

5.
Data sources

Overview

Approaches to collecting information in a CVD survey range from routine collection of vital statistics to active performance testing of participating subjects. Each of these is considered for quality, availability, ease of collection, cost and relevance to answering the questions. Data sources available for cardiovascular surveys are described below, along with the strengths and weaknesses of each method.

Vital statistics

Mortality statistics based on classification of cause of death are basic to the study of the health of any population. Many countries collect such data routinely, and have done so for decades, while others are only now developing comprehensive systems. In areas where mortality data are not routinely collected new systems for collecting this information are crucial.

Almost as important as the fact of death is the cause of death. This information varies in quality by region and country. Autopsy or postmortem examination is the accepted "gold standard" by which cause of death is most accurately certified. However, even autopsy may not always be helpful in cases such as sudden out-of-hospital death, a common form of demise in CVD. Autopsy is practised with decreasing frequency, except in cases where criminal behaviour or unnatural causes are suspected. Hence, most certification of cause of death depends on assessment by a local physician, medical examiner, coroner or other health professional. This individual may or may not have been involved in the patient's care. The absence of accurate or informed information in the completion of a death certificate leads to misclassification. Social, political and religious considerations may make some causes of death more or less acceptable, irrespective of the real cause.

Appendix 3 shows a standardized WHO death certification form and the form used in the USA. Each requires demographic information on the deceased, on relatives, on underlying and associated causes of death, and on the certifying individual. Such information is coded using both a systematic method and the WHO International Classification of Diseases (ICD-10) System (*1*). The 10th Revision, the International Statistical Classification of Diseases and Related Health Problems, introduced in

1993, provides a comprehensive classification system for cause of death. It is the global standard and forms the basis of national reporting to WHO and of statistical comparisons.

While the absolute number of mortal events can be determined from death certificates, along with cause and age at death, it is impossible to determine disease rates in a population without knowing the population base or denominator. National and regional census data may be accurate in many areas and inadequate in others. Even those census data known to be regularly and rigorously collected have weaknesses, including incomplete data on mobile, economically deprived or minority populations. Estimating the population base in years between census surveys presents a problem: most countries do not conduct annual census surveys. Population size between census points must be estimated based on information available on births, deaths, immigration and emigration (*2*).

Routine patient data

Morbidity data on patients seen in health care facilities are routinely collected in many countries, and range from demographic characteristics (age, race, sex), number of visits and hospitalized days, to diagnoses, treatments and outcomes. Such information may be less comprehensive and less rigorously collected than the death certificate information. It can vary greatly between countries, regions and individual health care facilities. Information on hospitalization is usually more completely collected and classified than that on outpatient services.

Diagnostic classifications based on the WHO ICD codes (*1*) are collected in many hospitals and clinics. For hospitalized patients, the discharge diagnosis is a basic means for determination of disease patterns. Similar data may be collected on outpatient visits. A diagnosis is assigned based on the clinician's diagnosis of the underlying disease that resulted in the patient's contact with the health care system. These data are commonly available to the researcher.

Data from hospitals and clinics are subject to various sources of bias and misuse. In some systems, the data are inconsistently recorded and the diagnoses not rigorously assigned. In other systems, where reimbursement depends on the diagnosis, information is likely to be complete, but diagnoses may be influenced by insurance reimbursement levels for different diagnoses (*3*). Information may be complete in one institution in a city but partly missing in another, and certain regions may have comprehensive collection schemes while others have nothing. Of course, individuals who are ill but do not consult the health care system will not be found by this method.

Morbidity patterns are an important element in determining the health of a population. These data are particularly important for chronic CVD which creates popu-

lation disability and consumes a large portion of health care resources. However, before these data are used in scientific studies of CVD, their quality, completeness and availability must be established.

Insurance data

National, local and private health insurance systems frequently collect information on the disease patterns of individuals insured in their systems.

Information from insurance plans may be useful as it is usually complete. However, if the insurance scheme is not national or population-based, the data may not be generalizable and may be biased. The data from insurance organizations are usually much less comprehensive than those in hospital or clinic records.

Interview data

Information gathered by interviewing patients, or subjects selected from a population, is a primary collection method in CVD surveys. Self-completed questionnaires, direct interviewer-administered questions, and telephone interviews are three of the more common methods used to collect information. Questionnaire design depends on the method of administration, and is a critical task for the investigator developing a survey.

Numerous sources of bias are associated with interviews and the collection of good quality information from interviews depends on factors detailed above (Chapter 2, p 38) and in later sections. Good information can be obtained with careful design of instruments, training of interviewers and control of the interview setting.

Observation of subjects

Most cardiovascular surveys have concentrated on direct interviews or testing. Increasingly, however, subjects are observed in their natural setting because health habits may be erroneously reported by participants on interviews or questionnaires. Individuals tend to over-estimate behaviours that they view as positive and minimize those that they feel the interviewer will not approve of. Direct observation is one way to avoid these sources of bias. For example, physical activity among youths in schools or adults at work can be observed directly and quantified (*4*). Both eating and cigarette smoking behaviours may be unobtrusively observed in such settings as a cafeteria or restaurant.

In some cases, bias may occur if the subjects know that they are being observed; this is the so-called Hawthorne Effect. In addition, they may feel offended. While informed consent is one way of ensuring acceptance, it also alerts the individual to the intended study, which in turn may lead to a change in behaviour. Finally, direct

observation is costly in time and personnel, it requires considerable work to organize, and is not suitable for all settings.

Physical measurements

Basic physical measurements such as height, weight and waist circumference are commonly collected using a strict protocol in CVD surveys. Other more complex tests include electrocardiograms, exercise tests for fitness, underwater weighing to estimate body fat, and measurements on blood and urine. The collection of these samples requires careful adherence to a strict protocol to ensure quality and consistency. In addition, precise handling, storage, and analysis of samples and measurements require careful staff training and consistent methodology. Quality control protocols are necessary: information regarded as adequate for clinical purposes is frequently inadequate for scientific studies because of variable collection and analysis methods.

Unobtrusive measures

The ability to monitor risk factors and trends in CVD inexpensively and without undue effort is an attractive concept. Certain indicators are available, for example food consumption estimates collected by the FAO, based on production, import and export. These allow for the international comparison of broad trends.

Perhaps the best population health indicator is a country's gross national product. It has been known for a very long time that economic prosperity is good for a nation's health. In terms of the total community perspective of burden of disease, it is now possible, with well developed health information systems using medical record linkage, to produce a clear indication of what is happening to the burden of some diseases in the population. This approach requires some form of registry for the purpose of periodic validation (see Chapter 8).

Information may also be gathered from customs and excise tax data about the sale of tobacco and cigarettes but, without details of the consumers' age/sex breakdown, the data may be less helpful. Information on hypertension and on cholesterol may be derived from relevant drug sales and prescribing data. In developed countries there are usually several large retailers who command a large proportion of the market and it is possible to gain information from computer price scanning directly from them, for example, on trends in low fat milk use. Anthropometric data are usually available from school health services and from conscripts to the armed forces. It has been suggested that the monitoring of weight reduction products might be useful.

In conclusion, it is stimulating to try to devise innovative methods of easy monitoring of CVD, but while there are many things that could be monitored, there is

no cheap easy substitute for direct collection of high quality epidemiological data by survey.

References

1. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems,* Tenth Revision. Geneva, World Health Organization, 1992.

2. Bureau of the Census. *Current population reports. Population estimates, Series P-25.* Washington, DC, United States Government Printing Office, (issued periodically).

3. Assaf AR et al. Possible influence of the prospective payment system on the assignment of discharge diagnoses for coronary heart disease. *New England Journal of Medicine,* 1993, 329:931–935.

4. McKenzie TL et al. Children's activity levels and lesson context during third-grade physical education. *Research Quarterly for Exercise and Sport,* 1995, 66:184–193.

6.
Conducting the research

Overview

Key steps in the development and implementation of a CVD survey include hypothesis formulation, data collection, presentation and interpretation. The following sections describe specific steps.

Hypotheses

The essential first step is to formulate a hypothesis that encapsulates the questions for which answers are sought. The questions should be few, brief, clear and answerable by the study design proposed.

For example, in a regional study of stroke and high blood pressure, the hypothesis might be that:

- Region A will have a higher incidence of stroke than Region B.
- Associated with this higher rate of stroke will be:
  - — higher mean population levels of blood pressure;
  - — lower rates of detection, treatment and control of high blood pressure.

This example of a survey postulates differences in disease patterns and proposes an explanation based in turn on past work showing that hypertension is causally related to stroke.

Alternatively, in a clinical trial of a new cholesterol-lowering drug, it might be hypothesized that:

- Individuals taking new drug x will achieve a lower blood cholesterol level than subjects given a placebo.
- A lower rate of acute myocardial infarction (AMI) will be found in the group taking active drug x compared with the placebo group.
- The lower rate of AMI will be related to lower levels of cholesterol.

Or again, in an ecological study that compares population sodium consumption to stroke rates, the hypothesis might be that:

- Differences in sodium consumption between countries, determined by sales of salt for human consumption, explain differences in the hypertension-related diseases of stroke, congestive heart failure and renal disease.

This is an ecological study that relates national salt use in different countries to hypertensive or related disease rates. The hypothesis attempts to support the debated relationship between salt intake and blood pressure.

In all these examples, the hypotheses are positive, suggesting that the relationships exist. Some argue that hypotheses should be phrased in negative terms—the so called "null hypothesis"—for example, salt intake is unrelated to hypertension complications. This is a matter of practice, and either approach is acceptable.

Specific aims and objectives

Specific aims and objectives are meant to translate the hypothesis into the actual tasks that need to be accomplished to deliver answers to the scientific questions. They should be brief, succinct and describe anticipated schedule.

Specific aims
For example, to answer the question about differences in stroke and hypertension between two regions, specific aims might include:

a) Collection of accurate and identical data on stroke mortality in each region,
b) Measurement of systolic and diastolic blood pressure;
c) Determination of prevalence of hypertension and its detection. The detection, treatment and control of hypertension in the population.

To obtain these data, the objectives or practical tasks will include:

Objectives for aim (a)
- Data on stroke mortality in the two regions on adults aged 25–84 years will be collected for the period of 1990–1995 from death certificates.
- Death certificates will be obtained and analysed to ascertain the prevalence of stroke diagnoses ICD-9: 420–427;
- Stroke mortality on a sample of death certificates will be validated using available medical records, health professionals' records, and family interviews; a specified diagnostic algorithm and an expert panel will be used to classify all cases of stroke mortality.

Objectives for aims (b) and (c)
First a population sample of adults aged 25–74 years ($n = 3000$) from each of the regions will be selected by a random process using census lists.

Then for this population:

— blood pressure will be measured using a standardized method;
— demographic information, history of hypertension and current medication use will be ascertained.

These aims and objectives are expressed in short definitive statements, and give an overview of the tasks in the study.

Selecting and defining a population

Selection of a population for CVD survey depends on the study design. For some studies, some or all of the data may already be available. For others, the population may need to be defined and sampled, and the data collected. In either case, a clear definition of the population is essential. Is it well characterized? Are there missing subgroups? If it is a sample, how does it relate to the larger population and is it representative?

Available data

Many cardiovascular studies use existing data described in Chapter 5. These may include mortality data from death certificates, morbidity data from the health care system, census information and consumption data. Selecting data from available sources involves several considerations. Are the data relevant to the hypothesis? Is the quality of the data suitable for research? (It cannot be assumed that data from other sources, even official sources, is accurate and adequate for research needs.) Will the source make the data available for the study?

Questions concerning available data are given in Table 6.1.

It is important to know how the data were collected. It is critical to define the population to determine whether it is representative of the particular group of interest to the survey and the extent to which the data can be generalized. There is a need to

Table 6.1. Characteristics of the available data

1.	How was the population base established? Is it representative?
2.	How was sampling done?
3.	What was collected (i.e. what methods and instruments were used)?
4.	What were the participation rates?
5.	Are there missing data?
6.	What quality control methods were used to assure good data and are these results available for review?
7.	What are the weaknesses of this dataset?

know what was collected and how. What were the participation rates in the survey? Generally, in a population survey more than 80% participation is considered good and less than 60% poor. How was quality control of the study assured? Were there regular quality control measurements, and are those data available? This information on the methodology may already be in the published literature, but if it is not, investigators should be willing to share it with anyone seeking to use data from the study. Unavailability of methodological information suggests that there may be serious problems with the study and makes the data unreliable.

In many instances, there will be no available data on the questions being asked. In this case, the population will have to be selected and defined and data will have to be collected. This allows more flexibility in defining the population and the information needed than when using available data sources. However, much more work is required.

Population surveys may include clinical populations where individuals served by an outpatient clinic or a hospital form the population base. In many countries where health care systems are centrally organized, a single inpatient and outpatient system may serve the entire population with private practice being non-existent or representing a minor component. In this setting, patients attending that facility may indeed represent all those who have a specific illness in that particular geographical area. Where health care is less centralized, factors such as physician practices, insurance plans and referral patterns may influence the "selection" of the patient population by clinics and institutions. That population may therefore not be representative of the community but instead be a reflection of those who attend or are referred to each specific physician practice. This pattern can severely limit generalization, unless it can be demonstrated that the observed population is representative of the larger community. This can be achieved by a systematic comparison of data from the observed population with data from the larger community or with a sample known to be representative of the population. For example, if the characteristics of the observed population match census data on age, sex, race and income, this would provide assurance of the representativeness of the observed population. It is critical to ascertain the source of a clinical population, given the potential for selection or referral bias.

Many cardiovascular surveys are performed on large populations with a particular affiliation, including military recruits, work sites (Western Electric Study, London Civil Servants, United States Railroad Workers), or insurance enrolment (Prudential Study, Lutheran Brotherhood Study). Data on such populations may be readily available or the populations may be willing to participate in a new survey. Again, knowing the origin of the population is critical in order to gauge its representation. While some groups such as military recruits may be representative of young men of that age group, others, such as industry-based populations, may be selected by the employer for

reasons of background, skills, education, physical health or other characteristics. This may give rise to the recognized "healthy worker effect". Employed people are less likely to have chronic diseases and disabilities than the unemployed and are more likely to have a lower morbidity and mortality. Yet other groups, such as Seventh Day Adventists who have foresworn cigarettes, meats and alcohol, are of interest because of specific health habits that allow comparisons with other populations.

The most representative studies are based on a geographical area. Town or city dwellers usually make up a broad cross-section of society, and a large enough community may be representative of all social elements. Smaller towns and/or parts of cities may be less representative. The larger the population or sample, the greater the possibility of generalization. Geographic boundaries with well defined borders are used. Thus, the Framingham Study represented a medium-sized town in central Massachusetts (*1*). Tecumseh, Michigan is also a middle-sized community (*2*), while the Minnesota Heart Survey (MHS) represents the large Minneapolis–St. Paul Metropolitan Area of approximately 2.5 million people (*3*). In each of these studies, CVD morbidity, mortality and risk factors are characterized using a sample of all adult residents in the community.

In cases of a large work site, or a big community, sampling may be necessary to conserve resources while characterizing the population adequately. Kish (*4*) describes population sampling techniques. There are two basic principles. The first is to choose an adequate sample size to represent the population. The definition of "adequate" depends on the variables being measured and/or differences being sought: the bigger the sample, the less is the variability of the outcome estimate and the greater the reliance that can be placed on the mean value. The second principle is to ensure representativeness of all elements of a population. For example, studies that rely on volunteers in a population lead to biased estimates of the population characteristics since volunteers usually do not have the same health characteristics as the general population.

Besides the origin of the overall population, other characteristics are considered depending on the nature and distribution of the disease. These include age, sex and race. Additional specifications can be built into the population sampling, at the cost, however, of reducing generalizability.

The following are examples from established studies of population selection.

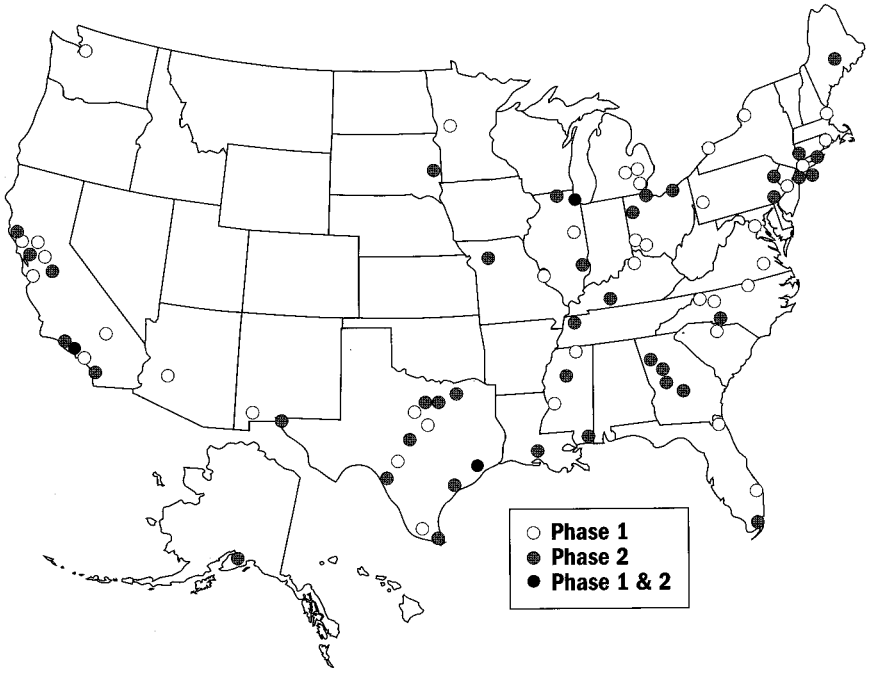
1. The Honolulu Heart Program recruited all men of Japanese descent born between 1910 and 1919 who were conscripted during the Second World War. The geographical area was the entire island of Oahu. The assumption was that military registration during the war was quite complete (*5*).
2. The Group Health Cooperative of Puget Sound is a large, prepaid medical practice with 500 000 enrollees living in a defined geographical area. The members

of the Cooperative have their health insurance plan purchased for them by their employers. While the Cooperative's membership is not population-based, its large size and diverse enrolment allow comparisons on a population basis (*6*).

3. The National Health and Nutrition Examination Survey is carried out at intervals by the United States National Center for Health Statistics. Its aim is to evaluate the health and disease status of a representative sample of residents in the United States (*7*), and it uses a cluster sampling technique. This technique consists of randomly selecting geographical areas for evaluation as shown for one city in Figure 6.1. It includes all ages, and samples of minority ethnic groups are added to ensure adequate characterization of these populations. This study is costly and takes several years to complete, but it provides an overall picture of the health of the United States population.

Selecting and defining a population sample are critical in CVD studies. Whether studies are observational or experimental, the origins of the population and the selection factors must be clear.

Fig. 6.1. Primary sampling units



Source: CDC/NCHS, Third National Health and Nutrition Examination Survey.

Choosing data collection instruments

The selection of the measurement instruments for a study represents a critical and time-consuming step. Instruments, in this context, encompass a wide spectrum of methods, from determining the age of the participant to the most sophisticated clinical testing. All must be given serious consideration, regardless of the degree of complexity of the measurement. Does the instrument provide information essential to the study hypothesis? Measurement tools that collect core data must be considered before other items of "ancillary interest" are included. Do they measure accurately (i.e. is the measurement valid)? Is the measurement reproducible? A proposed new instrument must undergo rigorous testing before being used in the field.

Every investigator planning a new study should first evaluate available instruments. This has several advantages. First, it can eliminate the many difficulties of developing, testing and validating a new instrument: the developmental work has been done by others. Second, and of equal importance, it allows comparision of new data with published data. This provides convergent validation when data are similar, or helps in seeking an explanation if they differ. The use of previously developed instruments and tests implies faithful duplication of the conditions of administration and evaluation. Even trivial differences in survey wording or sample handling may produce divergent results.

There are a number of approaches for finding and evaluating data collection instruments and tests. The first is a thorough review of the literature, which frequently will provide the information needed to duplicate the measurement tool. Many commonly used and tested instruments are described in the appendices of this book.

When the published literature fails to provide adequate detail about the measurement tool, it may be useful to contact the authors. Most health researchers are pleased to provide their instrument to other investigators at no cost; others may charge a modest fee. It may also be possible to obtain insights into data collection instruments that are not reported in the published literature from others working in the specific area.

For many new and evolving areas of enquiry, there may be no published literature. However, there may be studies in progress and attendance at scientific meetings and review of published abstracts from these meetings may help to identify investigators with similar interests. Again, direct contact with those investigators will usually yield information about their approaches to solving problems of data collection.

After a review of the literature, contacts with investigators and collection of various methods, the best instrument to address the research needs must be selected. Of the available instruments or tests, which provide the best information to test the hypothesis? Next, which instrument will allow comparison of data from this study with rel-

evant data from other studies in the field? Finally, past instruments need to be evaluated in the context of the study and population. Some instruments, although of the highest quality, may not be appropriate because of cost, differences in cultural characteristics, or differences in the context of administration.

Because CVD is a worldwide problem, the study of its characteristics is international in scope, involving many cultures and languages. Some tests, such as the measurement of blood cholesterol, blood pressure and other physical indicators, can be universally applied. Other instruments, such as questionnaires, require consideration of both cultural contexts and language. It is well known that questionnaires developed in one language may lose their meaning when translated into another. The usual method for evaluating this issue involves two steps. First, the questionnaire is translated into the new language by a skilled bilingual individual. Then a different and equally skilled bilingual individual translates it back into the original language. This method verifies comparability of meaning in the setting of the new language. However, although it ensures accurate translation it may not compensate for cultural biases that can be associated with an interview. This may require a more focused approach, with validation of the instrument in the new culture/language against some reference method or "gold standard".

In many instances, the question being asked may be new, and there is no appropriate measurement instrument or test to collect the information. It may also be that available tests are not applicable to the population or culture to be studied, in which case a new instrument must be developed. Table 6.2 shows several essential tasks in this process.

The scientific questions of the study determine the specific health characteristic to be measured, which in turn dictate the need for a new instrument. This instrument can range from a new set of questions to physical measurements to a new form for collecting information from patients' records. In each case the steps are the same. First, the test is developed based on the investigator's experience of similar tests, with technical consultation as necessary.

A draft questionnaire is usually then piloted on volunteers. Staff reviews will indicate the time required, and the suitability and acceptability of the instrument. An

Table 6.2. Factors to consider when developing a new instrument/tool

1. Specification of information sought.
2. Development of testing questions and/or procedure.
3. Pilot testing.
4. Determining reproducibility or reliability.
5. Validating the test against a "gold standard".

abstraction of medical data from available records will determine whether the usual record provides clear and accurate data. If it is a laboratory test, blood, urine or other samples can be collected from volunteers to determine whether the handling of samples and measurements are possible.

The instrument is now ready to test for reliability, repeatability and validity. Different interviewers may interview the same individual over a short period of time. For record abstraction, different abstractors may collect the data from the same record to test that the procedure is being done in an identical manner. For laboratory tests, split, blinded samples are used to determine the reproducibility of the sample handling and of the chemical tests. In each instance, an adequate sample is drawn to check that measurements made on the same individual or sample over a short period of time produce the same result.

Of particular interest is the stability of measurements over time. A particular test is repeated after some predefined interval. A blood or urine test, for example, should give the same result before or after storage.

Validation of historical information such as clinical records or death certificates is performed by obtaining more detailed information from diverse sources. For example, validation of hospital discharge diagnoses may involve a review of surgical reports and of pathological reports, and interviews with physicians or other informants.

Laboratory tests are frequently employed to confirm the presence or absence of a disease or risk factor. While no one test can provide certainty about disease status, it is clearly desirable to choose diagnostic instruments that offer the highest probability of accurate classification with the lowest individual inconvenience and economic costs. Whether the test is performed as part of a study or the test's results collected from available clinical records, similar questions must be asked of the test's validity and reliability.

The validity or accuracy of a new test can be best established by "independent" comparison with a reference "gold standard", namely a test that clearly identifies the disease or characteristic under examination. Such a "gold standard" is usually more costly or complex than the survey measurement under consideration and is therefore generally unsuitable for large surveys. Validation may directly evaluate the disease process (e.g. a biopsy) or physiology (e.g. intra-arterial blood pressure measurement). Where no "gold standard" exists, a composite standard of several measurements, or an expert consensus based on the results of several types of tests, is used. However, care should be taken to ensure that the observers interpreting the test are blinded to the results of the "gold standard". The test being evaluated should also not be a part of the composite standard, as this will produce a bias towards agreement between results of the two tests.

The results of the test should also not influence the extent to which the standard test is likely to be applied. "Verification bias" or "work-up bias" occurs when a test is restricted to patients in whom the disease status is already known. In such cases, the magnitude of the bias is directly proportional to the association between the disease and the result of the test under study. For example, a study validating echocardiography as a decision-making instrument for cardiac surgery in valvular heart disease might report excellent diagnostic agreement between the findings at echocardiography and at surgery. However, this would not consider the fact that surgery has been performed only in those selected by echocardiography and that surgery might have been denied to some deserving patients (false-negatives not identified). It is imperative that the standard be uniformly applied to all persons in the study. For example, Goldschlager et al. (*8*) studied the accuracy of treadmill tests in the diagnosis of coronary artery disease using coronary angiography as the "gold standard". They included 80 "healthy" male volunteers, who were presumed normal on the basis of clinical evaluation and treadmill testing, in their category of true negative patients even though no angiography had been performed. The assumption that all unverified patients are disease-free leads to false-negatives being classified as true-negatives.

When a "gold standard" is available, it is best to categorize test results into true-positives (disease present by both tests), false-positives (disease present only by the test but not by the "gold standard"), true-negatives (disease absent by both) and false-negatives (disease absent by the test but present by the "gold standard") is best done by constructing a 2×2 table (Figure 6.2).

It is also essential to evaluate the reproducibility of test interpretation by different observers. Wide disagreement between observers in interpreting the same set of test results renders the test useless for clinical or epidemiological application. Two or more observers should independently evaluate the test results without having access to the clinical data. If the test results are dichotomously classified, a chance-corrected index of agreement (Kappa) should be calculated. If the data are continuous (as in a scoring system), intra-class correlation coefficients are the preferred method of evaluation.

Pilot studies

A pilot field study to test the entire data collection protocol and schedule is necessary before active data collection begins. This differs from a pilot study of new measurement instruments, and it is best done on a population similar to the intended population of the actual survey, but involving only a small number of subjects. The objective of the pilot field study is to check that the data collection system is integrated and flows smoothly, thus optimizing the use of time by both the staff and participants. Pilot studies serve to reduce problems when the actual survey begins, and allow evaluation of the feasibility and acceptability to both staff and subjects of the planned survey. To perform the pilot the various data collection instruments and tests are first ordered sequentially. A commonly used order of tests is shown in Table 6.3.

Fig. 6.2. Validity

	+	−	
+	**True positive** ***a***	**False positive** ***b***	***a + b***
−	***c***	***d***	***c + d***
	a + c	***b + d***	***n*=total number** ***(a + b + c + d)***

$$\frac{\text{True positive}}{\text{True positive + False negative}} = \frac{a}{a + c}$$

Sensitivity
positivity in disease, that is, the ability of the test to correctly diagnose disease when present

$$\frac{\text{True negative}}{\text{True negative + False positive}} = \frac{d}{d + b}$$

Specificity
negative for the disease, the ability of the test to correctly identify the absence of disease

$$\frac{\text{True positive}}{\text{True positive + False positive}} = \frac{a}{a + b}$$

Positive predictive value
probability of a patient with a positive test having the disease

$$\frac{\text{True negative}}{\text{True negative + False positive}} = \frac{d}{d + c}$$

Negative predictive value
probability of a patient with a negative test not having the disease

Table 6.3. Order of data collection

1. Reception and identification of participant.
2. Explanation of study and procedures and signing consent forms.
3. Collection of data:
   - Interviewer-administered questionnaires
   - Self-administered questionnaires
   - Physical measurements
   - Collection of blood and other samples.
4. Exit interview for questions and participant satisfaction.

The order of the tests following identification and informed consent depends on the number and types of tests. For example, tests that require disrobing, such as an electrocardiogram or skinfold measurements, should be ordered such that undressing is convenient. The nature of the information being collected should also dictate the order of the tests.

In addition to the right sequence of tests, it is important to determine the length of each examination and to follow a schedule, recognizing that different measurements take longer with some individuals than others. Estimates can be determined during the pilot study and be used to plan flow patterns to make the best use of space and the staff's and subjects' time.

Finally, a pilot study will reveal the suitability of the environment planned for various tests. For example, quiet areas may be needed for interviews or self-administered questionnaires, as well as for the measurement of blood pressure. Privacy is important if clothing has to be removed and in areas where blood is drawn or urine collected. The pilot study provides information on all these factors and may dictate a restructuring or even relocation of the survey site.

Similar, though less complex, pilot studies are also applicable to medical record review surveys. In this case they may provide information about record availability, quality and quantity of data, and effort needed to extract the data.

Time spent in pilot testing brings significant rewards by allowing detection and correction of problems before the study goes into full operation.

Manual of operations

A well-designed study documents data collection methods in a manual of operations (MOOP) that clearly specifies the methods used for data collection. The MOOP is essential in training and includes the elements shown in Table 6.4. An example of survey manuals for a study of acute myocardial infarction is shown in Appendix 4. Manuals for MONICA are found on the Web site (http://www.ktl.fi/publications/monica/manual/).

Hypotheses, aims, objectives and background. The questions being asked and the reasons for asking them should be briefly discussed. Understanding the rationale of a study will make staff more interested and enhance their performance.

Table 6.4. The manual of operations

Hypotheses, aims, objectives and background for the study.
Data collection forms.
Detailed explanation of the data collection forms, study instruments and tests (including a description of the intent and method of application for each question or test).
Description of the qualifications and training of the staff needed to collect the information.
Physican work space and environmental conditions required for the survey site.
The sequence and schedule for collection of data.
Quality control procedures..
Questions commonly asked by subjects and staff.

Study Forms. All of the forms used in the study, from the initial collection of identifying information to incident reports, should be included here. A general listing is shown in Table 6.5.

Explanation of the use of forms. For each form, a question-by-question explanation of the appropriate use of the form is given in the MOOP. The meaning and definitions of every answer should be described in detail (Appendix 4). Data-handling procedures for field editing of forms—for accuracy and completeness—are also described. This gives staff a clear idea of the nature and extent of information required and it equips staff with the knowledge to extract important information and to answer participants' questions appropriately.

Staff requirements. Staff with different skills and training, ranging from receptionists to technical personnel, will be needed to collect information. A clear description of each job, including training requirements and responsibilities, is given. Supervisory responsibilities and lines of authority are also detailed. In this regard the MOOP contributes to effective recruitment, hiring, supervision and substitution of staff. The pilot work will have suggested the time and skills needed to administer each research instrument.

Space. The MOOP should contain information on the space needed for the survey and its configuration. For fixed survey sites only the initial configuration and participant flow need be detailed. Multiple survey sites require a description of suitable space rearrangements and subject flow patterns for each site (Table 6.6).

Table 6.5 Examples of form

Subject identification information.
Consent forms: 1. consent for testing 2. consent for results being sent to clinic or physician 3. special consents (e.g. HIV or genetic testing).
Interview forms.
Physical measurement forms.
Forms related to collection and handling of specimens.
Incident reports (e.g. injury of a subject; unexpected occurrences).
Quality control protocol forms.

Table 6.6. Space requirements

Reception area and waiting room.
Work stations for data collection.
A room for handling body fluid samples.
A locked and secure supply room.
Staff and meeting room.

There should be a reception area where subjects are welcomed, non-confidential information is collected, and the consent form is described and signed. It can also serve as a place for people to wait between examinations. It should be large enough to seat the maximum number of subjects expected at any one time and should have reading materials available.

Room size is a function of the equipment needed and of the number of technicians and participants in the room at any one time. A room for blood pressure measurement or phlebotomy also requires visual and sound isolation to provide a calm setting. In addition, temperature control and privacy are essential if subjects need to undress for examination or testing.

Sample processing. Processing of blood and urine samples should take place out of sight of survey subjects. Tables and electrical outlets are required for laboratory equipment such as centrifuges and refrigerators. Appropriate containers are needed for disposal of needles and contaminated supplies. A secure supply area with a locked cabinet should be available.

It is also important to keep a lounge area separate from subject flow where staff can relax or perform work that does not involve participants.

Supplies. The MOOP should contain a list of all equipment and supplies needed for the survey clinic operations. Each station of the survey should have a listing of all the items necessary to operate that station, from pencils to furniture. To ensure adequate supplies for each working day, a staff member should be assigned to maintain stocks.

Flow pattern. The participant flow pattern should be detailed, including appropriate room assignments and the time needed at each station.

Quality control. Quality control is described in detail later on (see page 148). However, the MOOP should include information on the specific procedures for quality control in the field. Details of the protocol should be included in this section of the MOOP.

Subject questions. As part of their training, staff should be made aware of commonly asked questions and suggested responses. These should be written down and included in the training material. This saves considerable time and reduces participant and staff anxiety. Experienced staff can frequently foresee the questions before the survey actually begins. The pilot study will also help to determine questions, which should then be noted. As the survey progresses, new questions may arise and these should also be recorded, together with the correct responses.

The MOOP provides the blueprint for implementation of the study and for training. It is also a record of how the work is done, and is useful in large field studies where many different staff will have investigative and technical roles.

Sample participation and collection

Obtaining complete information is essential. If the sample is to be representative of a larger population, extensive efforts must be made to ensure participation. Completeness of information applies not only to population surveys but also to record abstraction and use of secondary data sources. In each instance, participation and complete data are needed, including characterization of non-participants and of missing data.

Record abstraction

In the abstraction of data collected by others for secondary data analyses, it should not be assumed that the data are complete. The extent and quality of data should be carefully evaluated, and missing records or data should be characterized. Several examples of problematic situations are described below.

It is a mistake to assume that death certification is complete. Even where there is a legal requirement for registration of death, data may be missing. In countries with a history of comprehensive mortality surveillance, it is still possible to find that people known to be dead, with a local death certificate available, are not registered as such at the state or national level. This may be the result of inefficiencies or fraud.

Hospitals and clinics too, which have a legal obligation to maintain accurate medical records, may not always collect comprehensive information. Occasionally, medical records also are unobtainable. More importantly, mergers of hospitals or lack of recent attendance by a patient at a clinic or patient death may result in a record being discarded. Legal obligations to maintain medical records also vary by region; in some areas there are none.

Finally, even when the medical record is available, there is no guarantee that it contains the necessary information. Recording systems vary between and within countries and institutions. They rely on local and individual practice. For example, many medical records may contain no information on blood pressure, height or weight. Even basic and essential measurements are commonly missing.

When using secondary sources such as health insurance data, it is necessary to determine the source and quality of the data. Good sources will describe the methods used and the completeness of the data. If this information is unavailable, it is useful to contact the organization that provided the data to ascertain their quality and completeness.

Sampling and recruitment

Recruitment and participation of subjects are substantial and critical tasks in population surveys and clinical trials. The selection of a representative sample for a survey depends on current lists of names, addresses and telephone numbers, and on other demographic information (the sampling frame). Such information can be inaccurate or unavailable, in which case a special census may be needed to characterize the population to be sampled. A census involves a survey of households in a geographical area to establish accurately who is living there. Such work is tedious, time-consuming and costly, but may be necessary to ensure a complete and representative sample.

The ability to select and define the population of interest does not ensure that people will be willing to participate. Again, representativeness requires high participation rates; a 100% response rate is the aim of any study. While such a response is virtually unobtainable, it should remain the goal because any non-participation introduces bias. Experience shows that non-participants differ in health characteristics from participants (*9*). For interview-based surveys it is not uncommon to see participation rates above 90%. For more complex surveys, which involve direct measurements, participation rates above 80% are obtainable. When rates are lower (below 60%), it is likely that the sample is biased. One technique for ascertaining the extent of this bias involves the intensive recruitment and assessment of a sample of original non-respondents.

Clinical trials are particularly vulnerable to participation problems. Individuals who initially elect to join a study may leave later, causing a significant loss of information. Many investigators employ early "run-in" periods to pick out individuals who are likely to withdraw from the study. These are asked to complete the requirements of the study for some weeks or months, which may reduce generalizability of the experiment but minimizes subject loss. Many trials or cohorts also develop newsletters and other inducements to maintain participation.

A good cardiovascular survey requires high rates of participation, good data collection and thorough follow-up. Accurate record-keeping of recruitment and participation rates is essential to monitor progress of the study and to introduce remedial measures. Without regular monitoring of recruitment progress, it may be impossible to maintain acceptable participation rates.

Staff selection, training and supervision

The selection, training and supervision of the technical staff who collect data are an important responsibility of the investigator. The quality of the information collected depends heavily on the staff and their skills. The type of procedures planned, the time required for different data collection procedures, and the expected number of subjects or patients to be evaluated each day will determine the number of staff needed and their skills.

A survey supervisor should have previous experience of health surveys at a supervisory level and knowledge of the technical procedures to be used. The supervisor should have good interpersonal skills, preferably come from a health professional background and have a clear appreciation of the importance of collecting scientific information in a consistent and rigorous manner. The identification and recruitment of a supervisor first allows that individual to take part in the hiring of other technical staff, thus helping to build a team that will work well together.

For each of the technical staff positions, job descriptions should include the formal education required, previous experience, responsibilities and work expectation. A well-defined job description is useful when evaluating candidates and later, when assessing performance.

For certain technical posts formal education is necessary. For example, echocardiographic or laboratory technicians or phlebotomists have to undergo specific training programmes. For medical record abstraction, nurses have the training necessary for evaluating and interpreting medical charts. However, for many other tasks commonly associated with CVD surveys, such as recruitment, interviewing, blood pressure measurement, electrocardiographic recording, a training programme specifically designed by the investigators may produce better technical staff than formal medical training. For example, most health professionals learn routine blood pressure measurement in medical or nursing school. However, this may be inadequate for research purposes where procedures are to be done in identical and precise ways.

A rigorous programme of training in the skills necessary to implement the protocol follows the selection of staff. This may take a week or more. The MOOP serves as an invaluable training tool providing staff with an understanding of the aims and details of the study. This general training should be provided for all staff, from recruitment personnel to receptionists and laboratory technicians, since all will have questions asked of them by the subjects and they should be able to explain the study accurately.

Following the general training, specific training for data collection protocols should be undertaken with relevant staff. Established training programmes exist for many of the instruments commonly used in CVD surveys. For example, training in blood pressure measurement according to scientific data collection standards can take several days (*10*). It starts with hearing tests (because many individuals have hearing losses in the frequency range necessary for blood pressure measurement), and includes didactic presentations and practice in using various blood pressure devices. Students are then tested to confirm that they have acquired the skills.

Similarly, there are specific training programmes for a number of commonly used protocols for data collection by interview. Among the best known is that for collection of 24-hour food recall data (*11*), which takes several days.

For most of the data collection instruments used in cardiovascular surveys, retraining is important. Staff may subtly and unconsciously alter the collection techniques that they learned during the initial training programme or adopt "short-cuts" in the handling of samples or the taking of certain measurements. This may result in a "drift" in measurement quality over time that may be interpreted as a real change. Quality control protocols help in detecting this problem but regular retraining is an effective preventive tool. For many procedures this may be necessary every 3 or 6 months.

Cross-training of certain staff to carry out more than one procedure is a useful investment. When staff miss work because of illness or holiday, other individuals can replace them.

It is important to establish lines of authority for supervision. Individual job descriptions and organization charts should contain this information so that staff know whom to report to, and who reports to them. In small studies or surveys such a system may be informal, but it should exist nonetheless.

Data handling

Information collected in surveys—be it from documents, subject interviews or various measurement devices—is brought together and computerized for evaluation and analysis. The data collection process begins with the development of forms and ends with storage of information. In between, every step requires attention to data quality and suitability for analysis. A formal written protocol should be developed for each step and included in the MOOP to document appropriate data handling techniques.

The process begins with the development of printed forms. Whether these are for record abstraction, secondary data analysis or primary data collection in a field survey, due attention must be paid to the format and wording. With modern software and desktop publishing, it is possible to develop forms of high quality that are easy to complete and convenient for subsequent editing and data entry. All forms must include identifying information on the subject and the data collection site, plus space for the signature or other identifier of the individual responsible for completing and editing the form. A sample form for a number of measurements including blood pressure is provided in Appendix 5.

The forms should be printed on high-quality paper to allow long-term storage. Some investigators are now entering the "paperless world" and using computer-generated forms and direct entry of information. This technique is desirable but is in its early stages and requires sophisticated portable computer equipment and programming expertise.

A central element in data collection is the assignment to each subject of a unique identifying number (see Chapter 9). This may be done when initial recruitment occurs or during the first interaction with the receptionist. The identifier should then be applied to every piece of information on that participant. Preprinted adhesive labels are useful as they can be applied to forms, sample containers or other information unique to that subject. Bar-coding or other machine-readable identifiers can be included on these labels.

A protocol should also be developed to ensure that all portions of all forms are completed, with a paper-based or computerized logging system to record every form or item of information (Appendix 6). Such a system is very useful to determine survey progress or identify missing information.

The initial editing of most forms should take place in the field. It is preferable that someone other than the primary data collector inspects the forms for completeness of and appropriateness of the information. This should be done in a timely fashion, preferably while the participant is still in the survey area and can be readily contacted in the event of incomplete information. Editing in the field reduces later problems.

Some data may be added later. For example, laboratory information is usually completed days to weeks after the subject is seen. It is entered on individual forms with subject identifiers. In some cases direct electronic communication between laboratory analysis equipment and the study computer is possible. The original data collection survey forms should not be sent to the laboratory as primary data may be lost.

Computer handling of large surveys is the most common and convenient approach. Forms designed for computer entry with easily visible marks and logical placement of questions are a distinct advantage. In some instances, computer-scanned forms may be used. These various methods ensure accurate entry of data. The most rigorous method is the "double entry technique", in which the form is entered once on the computer and then re-entered later by a different operator. A matching program compares the two entries and notes inconsistencies, which are then checked against the original data. If data entry personnel are accurate and well trained, double entry may not be necessary for all forms, although a random sample is usually taken for double entry to ensure accuracy.

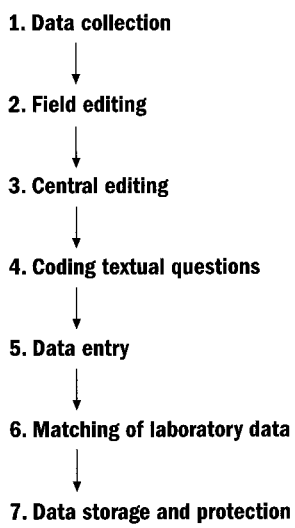
The computer is also used to search for missing data or entries that are outside of a programmed range. In many available database programs, it is possible to set limits on all numerical values that ensure "reasonable" answers. For example, it would be important to verify information on an adult who weighed less than 40 kg or more than 200 kgs ("constraint violation"), or on an individual classified as male who responded to questions on menstrual cycles.

Usually all information on data collection forms is entered into the computer. However, regulations on data protection and confidentiality may require that only identifying numbers are entered with the data. A protected file linking the numbers to names and other identifiers is kept separately and with limited access. In addition, qualitative information—such as responses written in text form—may not be suitable for computer entry until it undergoes a coding or classification procedure; this requires an additional step in the editing process. A trained individual categorizes these responses using a written protocol. The collection and coding of textual information is costly and time-consuming and is generally kept to a minimum in large field studies.

Information for long-term use is usually stored in one of two forms. Traditionally, the original paper forms are stored in a protected area where they will be safe from fire, water and other damage. They should be stored in an organized system that permits easy retrieval of information: filing by identifying numbers is a common approach. Computerized data are stored in an electronic form. It also requires back-up and safe storage. It is usual to transfer data from an active file to a second copy in electronic or optical format, which is then stored in a secure area away from the original data. In most settings, computerized data are backed-up at least weekly and sometimes daily.

It is clearly important to establish protocols and procedures for the collection, editing and computerization of data (see Figure 6.3) and to protect the study from inaccurate data and data loss.

Fig. 6.3. Data processing steps



Quality control

The maintenance of high-quality data is a challenge for all scientific investigations and requires particular effort for CVD surveys. The selection of study instruments must be accompanied by assurance of their validity and reliability. An initial assurance of quality, however, does not imply that quality will be maintained throughout the study. Changes in staff, site, equipment or reagents can lead to a deterioration of quality over time, with consequent loss of validity or deviation from true values, increased variability, and problems of comparability over time and within population surveys. To prevent this deterioration in data quality, a formal and continuous quality control system is essential. As with other survey procedures, it is essential to document all methods of quality control in the MOOP. It is then the responsibility of the investigator to ensure that those procedures are followed. The quality control methods commonly used in surveys are described below, but specific protocols may need to be developed for new and unique collection methods.

Field survey centres commonly encounter deterioration in data collection techniques and deviation from the protocol detailed in the MOOP. A visit to the data collection site is a common method for maintaining quality. There are two common approaches to a "site visit" to a field centre. The first involves sending a trained individual unannounced, to participate in the survey as a normal subject, field staff should be unaware of the person's identity. The individual then files a detailed written report on the survey procedures and adherence to the protocols of the MOOP. This approach can be particularly useful as it gives all the information from a subject's perspective.

A second approach is to send a team, ideally unannounced, to observe the survey centre. However, the mere presence of the team serves as notification of a quality control visit. Team members follow subjects to each measurement station and observe the methods used. While this system would be expected to result in maximum adherence to the MOOP, staff may be unable to follow protocol standards even under this kind of close observation, suggesting that their usual performance is even poorer. At the end of the site visit, the team files a report describing procedures and adherence to the study MOOP, and providing recommendations for correction in lapses. It is the responsibility of the principal investigator and survey supervisor to act on this report.

Specific measurements require different approaches with regard to quality assurance.

1. Record abstraction. When primary data are being used, it is important to ensure that they are correctly abstracted and recorded. This is usually done by taking a sample of records (5–10%) and re-abstracting them using a different staff member. Comparison of the recorded information for concordance of collected data reveals both errors in recording and different understandings of the objec-

tives of the questions. Such a re-abstraction procedure is used to evaluate staff on a regular basis during the study.

2. Interview. For data collection by interview, quality control is particularly difficult. One commonly used approach is editing of the interviewers' work. Interviewers whose data consistently require editing and correction are probably not adhering to the MOOP and may require retraining.

   A re-interview procedure is also commonly used. A sample (5%) of subjects who have been interviewed are re-interviewed, usually by a supervisor or a very experienced interviewer, shortly after the initial data collection using the same questionnaire. The second interview is then compared with the first interview for concordance.

   Another method involves the supervisor observing or listening during the interview. A voice recording for later review achieves the same purpose. Although this may create some anxiety in the staff and subject, it will also reveal deviations from the protocol.

3. Duplicate samples. For testing of blood or other collected specimens, a split sample technique is commonly used. For example, an extra sample is taken from the subject at the time of phlebotomy. It is labelled with a new case number, which, for quality control purposes, blinds the technician to the identity of the subject involved. It is then analysed with the other routine samples in the laboratory. In later evaluation, the blinding is broken and the laboratory value compared with that of the subject's correctly labelled sample. Differences are analysed in terms of the known variability of the laboratory technique. As a general rule, 5–10% of all samples should be included in this quality control method.

   This approach may also be applied to other measurements, such as ECGs or pulmonary function tests. Here a second measurement is performed shortly after the initial test. Both the quality of data collection and the interpretation can be checked.

4. Calibration. Many cardiovascular surveys use measurement equipment such as weighing scales, blood pressure devices and skin calipers in the field setting, for each of which there is usually a standard calibration method. For example, scales may lose their calibration when they are moved or as environmental conditions change. A standard weight is kept with the scale and the device calibrated daily. A calibration log is kept with the device and records the calibration procedure and the name of the technician performing it. Quality control during site visits should include evaluation of these logs.

   Calibration and quality control methods for other devices used in the field are usually provided by the manufacturer. Calibration should be performed at regular intervals and recorded on a specific document.

5. Sample handling. Because many surveys are carried out in the field, the processing of biological samples from distant sites is critical. Handling of samples can affect values measured in the laboratory and must be consistent both at the site and during transport. Observation of handling techniques during a site visit is one way of standardizing handling. However, this is not enough to ensure that

samples are shipped in the appropriate manner. It is also useful to monitor the samples for inappropriate handling during transport. For example, many samples must be transported over long distances at cold or freezing temperatures, and indicators are available to determine whether warming has occurred during shipment.

6. Measurements by technicians. Quality control of measurement techniques during site visits is one way to ensure good measurements. However, it is recognized that individuals who are familiar with the protocol will seek to adhere to it when they know they are observed. Another method that is therefore being used more frequently is surveillance of the distribution of measurements recorded by technicians. In large surveys, each technician should see enough assigned individuals to produce a normal distribution curve of any measurement value. For example, technicians who have performed a substantial number of blood pressure measurements should record distributions similar to those recorded by other technicians doing the same task. Analysis of those distributions may suggest that certain technicians are either biased in their measurements, that is, their mean value is significantly different from the grand mean of all measurements, or that there is increased variability, suggested by a large standard deviation (Appendix 7). In either event, the measurement technique is inconsistent and an improvement in quality is needed. Alternatively, they may exhibit so-called digit preference (the tendency to make zero readings), which is a marker of biased blood pressure recording. This approach requires statistical evaluation and the collection of adequate data.
7. Training programmes. The availability of quality control data provides assurance that measurements are consistently performed as intended in the study design. When there is variation from the expected quality, steps such as the introduction of regular retraining programmes must be taken. If measurement devices fail to meet adequate quality control standards, they should be removed from the survey immediately and sent for repair. Additional instruments should be available so the survey can continue.

When an individual technician consistently fails to meet quality control standards despite retraining, the data from that technician should be excluded from the study; and the technician should be given notice or assigned another task.

In summary, there are three essential elements of quality control. The first is ongoing surveillance to assure that high quality data are being collected. The second is clear documentation of the quality control procedures and of available data so they may be reviewed at regular intervals by the senior staff. The third is direct and timely action by senior staff to correct lapses in quality.

Running the study

Qualified staff together with a definitive MOOP should ensure a smoothly functioning study. However, the experienced investigator will recognize that even a

well-planned survey may suffer unexpected problems. While no one can foresee all mishaps, close monitoring will help ensure that small problems do not grow.

Formal systems should be established for detecting and preventing problems. The first includes regular review of study progress by monitoring recruitment rates and data collection. Recording systems at each step in the process can provide information on progress toward pre-set study goals and deadlines. The development of this system is important for management of the project and is also very helpful to the investigator.

The second system consists of regularly scheduled meetings with the supervisor of the study. This will help to detect problems early and assure the supervisor of the investigators' continued support and interest. There should also be regular meetings at which senior investigators update the entire staff on progress with the study, and reinforce the importance of the work being done. These meetings, which need to be held only once or twice a year, can be important for the morale of the study staff and require only small amounts of senior investigators' time.

References

1. Dawber TR. *The Framingham Study: the epidemiology of atherosclerotic disease.* Cambridge, Harvard University Press, 1980:14–29.

2. Ostrander LD, Lamphiear DE. Coronary risk factors in a community: findings in Tecumseh, Michigan. *Circulation,* 1976, 53:152–156.

3. Gillum RF et al. Cardiovascular mortality trends in Minnesota, 1960–1978: the Minnesota Heart Survey. *Journal of Chronic Diseases,* 1984, 37:301–309.

4. Kish L. *Survey sampling.* New York, Wiley, 1965.

5. Kagan A. *The Honolulu heart program: an epidemiological study of coronary heart disease and stroke.* Newark, NJ, Harwood Academic Publishers, 1996.

6. Siscovick DS et al. Diuretic therapy for hypertension and the risk of primary cardiac arrest. *New England Journal of Medicine,* 1994, 330:1852–1857.

7. Ezzati TM et al. Sample design: third National Health and Nutrition Examination Survey. *Vital Health Statistics 2,* 1992, 113:1–35.

8. Goldschlager N, Slezer A, Cohn K. Treadmill stress tests as indicators of presence and severity of coronary artery disease. *Annals of Internal Medicine,* 1976, 85:277–286.

9. Wilhelmsen L et al. A comparison between participants and non-participants in a primary preventive trial. *Journal of Chronic Diseases,* 1976, 29:331–339.

10. Prineas RJ. *Blood pressure sounds: their measurement and meaning.* Minneapolis, Minnesota, Gamma Medical Products Corporation, 1977.

11. Feskanich D et al. Computerized collection and analysis of dietary intake information. *Computer Methods and Programs in Biomedicine*, 1989, 30:47–57.

7.
Mortality surveillance

Introduction

Mortality is the yardstick on which most causal relationships in human health are measured, whether it is of exposure and effect or of intervention and benefit. Age- and sex-standardized total mortality rates, and cause-specific mortality rates, provide measures for the evaluation of health status in different countries or at different times in the same country.

Cause-specific mortality rates are fundamental in estimating CVD burden and are of special relevance in evaluating epidemiological patterns of CVD. For example, deaths attributable to infectious and nutritional causes of death are currently declining in most developing countries while CVD and cancer-related deaths are increasing, as both absolute and proportional contributors to total mortality. In many developed countries, CVD rates are currently showing a decline but remain the leading contributor to total mortality.

Even in countries where total mortality data are accurate, valid cause-specific mortality rates relevant to CVD can be difficult to obtain. Comparison of mortality rates among populations and the study of secular trends require comprehensive data collected with standardized methods. Adequacy of data acquisition as well as uniformity of criteria for ascribing the primary cause of death are essential for making valid comparisons.

Currently, the extent of death registration and medical certification as well as the practice and the process of assigning the cause of death vary greatly between countries and even between regions within countries. Under-reporting of deaths and inadequate certification are major problems in a number of developing countries where improved mortality data and census are urgently needed for planning medical and public health initiatives.

Background

Mortality rates are the most basic indicators of the health of populations. Both the fact of death and its cause are available in many countries. In addition, information on age, residence, sex, date of birth and occupation is frequently included (*1*). These

data are most often recorded in a standardized death certificate using the format provided by WHO (Appendix 3). Information is usually available by country or region and more detailed data may be available from special registries that survey specific populations. Mortal events are linked with population census data to provide crude rates. Data on age and sex distribution provide group information to allow adjusted rates. Good-quality local or national mortality data depend on the comprehensive and accurate collection of death certificate information and accurate census data.

While information on mortality is essential, systems for collecting these data are far from perfect. In many developed countries, death is recorded at rates approaching 100%, with laws requiring death certification before burial. However, in some countries, even the fact of death is not recorded. Reliable data are only available for 35% of the world's population. Resources and trained personnel may also be limited. Cultural and religious practices may inhibit the collection of this information. Data are particularly difficult to obtain in poor and rural areas. In many areas, these factors are associated with poor census data, making mortality rates difficult to interpret. For example, the World Bank estimates up to 30% under-reporting of mortality in China (*2*).

The fact of death is usually indisputable but the cause of death presents larger problems for all countries. Accurate information requires an experienced diagnostician with training in death certification and knowledge of the patient's circumstances or clinical course. An autopsy or post-mortem examination, clinical data and information from relatives can be crucial in determining the cause of death. Unfortunately, this combination of characteristics is rarely met consistently in any country. Death certification is frequently performed by clinicians or medical examiners or coroners who have no knowledge of the patient's clinical history. Autopsies are less commonly performed than they were previously except when death occurs in suspicious circumstances.

The need for a system of coding the causes of death was recognized over a hundred years ago by William Farr (*3*). The first international classification of diseases[1] was published in 1899 and since then the system has been revised each decade. The World Health Organization (WHO) assumed the responsibility in 1948. The 10th revision of ICD[2] (ICD-10, see page 103) has been introduced and should be uniformly applied by all countries (*4*).

ICD-10 is based primarily on disease categories. It assumes the availability of diagnostic facilities that may not be uniformly present. Changing diagnostic technology and the use of different editions of the ICD may contribute substantially to variations in cause-of-death assignment and amplify or obscure the true differences in mortality patterns between countries and especially over time.

[1] The Bentillon Classification or International List of Causes of Death.
[2] The International Statistical Classification of Diseases and Related Health Problems, 10th Revision, 1993.

Table 7.1. Estimated deaths due to ischaemic heart disease

	Population (×1 000 000)				IHD (×100 000)			
	Men		Women		Men		Women	
	1990	2020	1990	2020	1990	2020	1990	2020
[a]Established Market Economies	390	434	407	455	8.3	12.1	8.4	11.1
Formerly Socialist Economies	165	170	181	191	4.7	7.1	5.6	7.0
China	439	608	410	589	6.2	14.1	5.6	12.0
India	585	727	548	721	3.9	8.1	3.8	6.8
Other Asia and Islands	343	497	340	505	2.3	5.8	2.3	5.5
Sub-Saharan Africa	252	555	258	565	0.9	2.2	1.2	2.6
Latin America	222	331	223	336	1.8	4.4	1.7	4.1
Middle East Crescent	256	496	247	487	3.2	8.7	2.9	7.2
Total World	2654	3819	2614	3848	31.3	62.6	31.3	56.3

Adapted from: Ôunpuu S, Negassa A, Yusuf S, for the INTER-HEART Investigators. INTER-HEART: A global study of risk factors for acute myocardidal infarction. *American Heart Journal*, 2001; 141:711–21.

[a] Regional groupings as cited in Murrary CJL, Lopaz AD. *The Global burden of disease*, Cambridge, MA.

Despite these limitations, estimates of deaths attributable to CHD, stroke and rheumatic heart disease (RHD) have been extensively studied to provide comparative global assessments as well as to document secular trends within countries. The decline in coronary mortality rates in the USA and the stroke mortality rates in Japan have been impressively recorded in recent years. The rise in CVD-related deaths, from 12.1% of all deaths in 1957 to 35.8% of total mortality in 1990, in the urban population of China is convincing evidence of the accelerating CVD epidemic in that country. RHD continues to be an important contributor to premature mortality in developing countries (Table 7.1) though declining trends are suggestive of an epidemiological transition from infectious to other chronic CVDs.

Declining or differing CVD mortality rates could be due to a decline in the incidence of disease, a reduction in case-fatality rates, or both. Incidence data on morbid events and their fatality rates can help the contributors identify a decline in mortality. If case-fatality rates decline without a change in incidence, prevalence rates of disease in the community should rise. Declining cause-specific mortality rates accompanied by a fall in, or unchanged disease prevalence rates, indicate a true decline in incidence as a contributing factor to the mortality change.

In the category of CVD, mortality data present unique opportunities and challenges. Because CVD is a common cause of death in most countries, these data are particularly important in understanding the health of a country.

Mortality issues specific to CVD

Sudden death. One of the more difficult areas in cardiovascular mortality surveillance is the phenomenon of sudden out-of-hospital deaths, usually precipitated by a malig-

nant ventricular arrythmia. Reports suggest it accounts for anywhere from 20% to 70% of all coronary heart disease deaths and a smaller fraction of stroke deaths. Sudden death frequently has no premonitory symptoms and may be unwitnessed. It is commonly defined as death occurring within one hour of onset of symptoms or in the absence of no symptoms (*5*).

Because sudden death involves cardiac arrest, it is usually attributed to CVD. However, autopsy studies of sudden death victims suggest that there may be other underlying causes that are missed (*6, 7*). While cardiac arrest may be the immediate cause of death, the underlying cause may be cancer or other serious health conditions. When these are ignored, an overestimate of CVD mortality occurs. Even autopsy, usually the final arbiter of the cause of death, may be unhelpful when the event is rapid in onset. Most adults have some coronary disease but the electrical nature of the mortal event rarely leaves a trace for the pathologist.

The challenge of appropriately classifying sudden out-of-hospital death as due to cardiovascular or other diseases is a particularly difficult one. Better understanding of these events is an important challenge for CVD surveillance. One approach to assembling data from a variety of sources after death to improve classification is shown in Appendix 8.

Elderly. Much of the burden of CVD, specifically CHD and stroke, is found in the elderly, who frequently have many and serious co-morbid conditions. There can be many contributors to the final mortal event, in addition to underlying CVD. As populations age, the appropriate classification of cause of mortality on death certificates to cardiovascular or other conditions becomes more difficult. A coding for multiple causes of death, now available in a number of countries, may help in improving accuracy but is unlikely to completely solve the problem.

Just as co-morbidities in the elderly may lead to the overestimation of CVD mortality, many cases of CVD death in the elderly are classified as nonspecific causes (ICD 726) leading to underestimation. In the elderly, death may be considered as due to old age as opposed to a specific organ system failure. Because CVD is so common in the elderly, the classification of cases as death due to old age or nonspecific causes may underestimate the disease burden.

Rheumatic heart disease (RHD). In developing countries, RHD is an important and frequently under-reported cause of premature death. Though community-based age-standardized mortality rates attributable to RHD are not available, the majority of RHD-related deaths occur before the fifth decade of life in developing countries where the disease occurs early and progresses rapidly. This contrasts with developed countries, where the disease is less frequent and has a more benign natural history.

Misclassification of the cause of sudden death, RHD, co-morbidities, and ageing populations is a serious obstacle to understanding cardiovascular rates and trends. The absence of trained personnel or adequate resources to improve the accuracy of diagnosis is an important contributing factor. If death certification is to improve, better training of personnel who record the information is essential. Until then, many of those performing population studies of CVD may need to do their own validation studies.

Methods

Evaluation of available data

While complete ascertainment of deaths is essential for valid health statistics, cause of death and associated demographic data are also crucial in the use of this information. There are many threats to the validity of cause of death data. While some may be unavoidable problems of local and national data collection systems, the researcher can use several methods to help ensure the quality of information being collected. Such information may also aid in attempts to improve the system to deal with these weaknesses.

Evaluation of those who are responsible for completion of death certificates is key to determining the quality of the data. In different parts of the world, certificates may be completed by a physician, medical examiner, coroner, nurse or other health professional. The training and skill of this certifying individual provide insight into the quality of the available information. The clinician who attended the deceased in life may be most aware of the patient's medical history and have considerable information about the underlying cause of death. Such information is frequently unavailable to emergency ward physicians or others who first see the patient in the terminal state. The availability of earlier medical records on the patient at the time of death certification is very valuable in ascertaining the underlying cause of death.

The availability of complete data on each death certificate is an important indicator of the care taken in filling out the form and of the quality of data. Death certificates with missing information are likely to have been poorly considered. Those that include only one cause of death may be suspect, particularly for CVD in older adults where complex causation is common. The use of a single "cause" such as sudden death or "ventricular fibrillation" may indicate a lack of comprehensive information on the patient. Those who classify the death certificates are left to make assumptions about the actual underlying cause.

Mortality classification systems frequently include the overuse of certain nonspecific codes including: (ICD-10: R00, R07, R54, R55, R57; ICD-9: 780.2, 785, 797, 798).

Other circumstances of death may be termed "old age", which fails to define a disease process. Widespread use of these nonspecific codes is an indicator of poor knowledge about the patient and poor death certification.

Specific approaches

Post-mortem examination

Post-mortem examination is the best method to validate the cause of death. Autopsy is less frequently performed than in the past but continues to be important and should be encouraged.

For accurate classification of CVD, a careful examination of the heart and blood vessels is essential and, in the case of stroke, examination of the brain and its blood vessels. Death due to arrhythmia may not be apparent at autopsy, particularly when the fatal event is sudden.

When post-mortem examination data are available, they can be used to improve the quality of death certification. The natural delay in obtaining these data should not prevent them from being combined with other information to ascertain and classify the underlying cause of death.

Hospital and clinic records

The assignment of cause of death may be verified in part by scrutiny of all relevant medical records. Hospital and clinic records are likely to provide information on important underlying conditions as well as on antecedent events. Guidelines for deriving cause of death from such records should be developed through expert consensus with a panel of physicians. These criteria should then be assessed for validity (through comparison with an autopsy "gold standard"), and reproducibility (through measures of inter- and intra-observer agreement).

Interview data

Interview data obtained by questioning the physicians, relatives or associates of the deceased individual may be used to validate the cause of death assigned by death certification or lay reporting. These data, however, need to meet the test of agreement with other data sources such as hospital records. As with hospital records, a structured approach should be adopted, with specified rules on quality and a hierarchy of information (Appendix 8).

Lay reporting

Lay reporting has been employed in some developing countries to obtain estimates of cause-specific mortality. In India, for example, the Survey of Causes

of Death (Rural), commissioned by the Registrar General, collects information on causes of death through a structured questionnaire on a sample of 0.5% of all rural deaths. This is based in about 1300 primary Health Centres throughout the country.

The relative merits of lay reporting compared with medically trained interviewers in ascertaining the causes of death need to be evaluated. Medically trained interviewers are costly and may bias the data by attempting to interpret responses during the interview. Validated lay reporting systems may therefore be cost-effective in countries where surveillance systems remain to be established (Appendix 9).

Verbal autopsy

Ascertainment of the cause of death from information obtained through retrospective interview of relatives or associates of the deceased is used in settings where vital registration systems are poor and the proportion of persons dying under medical care is low. The verbal autopsy can provide cause-specific mortality data in population samples (*8*). It could be employed in community surveys or in sentinel surveillance systems in areas where the usual data are not available.

This approach has been validated and applied to ascertain causes of childhood deaths. Instruments for assigning cause of death, especially CVD-related events, in adults are under evaluation. If validated instruments are cross-culturally adaptable, they would facilitate acquisition and monitoring of CVD-related mortality in developing countries (Appendix 9).

The verbal autopsy is usually administered by trained lay persons. Algorithms derived from the validation studies provide probability estimates of CVD-related deaths, based on responses to questions about events preceding death. Studies of specific applications of this technique for CVD mortality ascertainment are under way.

Data obtained from such interviews could be applied in countries without better systems:

- to estimate the relative public health importance of CVD-related deaths in different populations;
- to identify intervention priorities;
- to monitor trends in mortality over time, especially to evaluate the response to specific interventions;
- to evaluate risk factors associated with particular causes of death.

Expert panels

When data on mortality are collected from diverse sources, a standard method should be used for determining the cause of death. This permits the consistency of approach necessary when data are combined. However, divergent sources often produce conflicting information. The use of a panel of experts, usually clinicians or medically trained epidemiologists, may be the best approach. The panel should have a predetermined set of rules and a hierarchy of information (e.g. autopsy supersedes the clinical record) to determine cause of death. Where disagreements arise, the panel may need to arrive at a conclusion by vote (and therefore should have an uneven number of members). Alternatively, an expert panel may develop an algorithm with specified rules for application by a technician. A final "cause of death" is based on those rules. The Atherosclerosis risk in communities study (ARIC) provides an example of the type of information gathered in an expert approach, and a method using an expert panel and a computer-based algorithm are given in appendices 10 and 11.

Sentinel populations

Where national data for cause-specific mortality are unavailable because of inadequate death certification, sentinel surveillance systems (defined as a sample that reflects the larger population) can provide representative information. Here, a defined target population is selected for close observation. China, for example, does not have a nationwide vital registration system. A random sample of the Chinese population is monitored through a Disease Surveillance Point system in which teams (including a physician) review hospital records or interview the family to ascertain the likely cause of each death (*2*). However, even the Disease Surveillance Point system has not been able to capture all deaths in the surveillance areas, and estimates of underreporting vary from 11% to 30%.

ICD coding

The International Classification of Diseases published by WHO is the basic system used worldwide for the categorization of cause of death (*4*). A similarly derived system is used to classify hospital-based diagnoses. Both are updated approximately every decade to reflect changes in medical knowledge. In addition to disease categories, the ICD system also provides a specific system for the training and use of nosologists who evaluate and classify death certificate data for most governments (Appendix 12).

Sudden or out-of-hospital death

Deaths from CVD that occur outside of the hospital or clinic are both common and difficult to classify. Methods have been developed to categorize these out-of-hospital events more accurately. They are considered as sudden (less than 1 hour), rapid

(between 1 and 24 hours), or delayed (more than 24 hours), based on the onset of symptoms. Death sites are non-acute medical care facilities, outside hospitals (e.g. home, work site), or dead on arrival, or death in the emergency room of the hospital. Sometimes, if attempted resuscitation is prolonged, it may be unclear when or where the patient actually died.

In addition to the death certificate, several sources of information are sought to validate the timing and the circumstances surrounding the final sudden event. The sources of information include family or coworkers who may have observed the acute event or are familiar with the patient's medical history, as well as medical records and interviews with the responsible clinician. A standardized interview conducted by telephone or in person by an interviewer is shown in Appendix 8. Based on these data and on information from the death certificate, a physician for the survey then classifies the cause of death as:

1. Definite fatal myocardial infarction (MI)
2. Sudden cardiac death
3. Definite fatal out-of-hospital CHD
4. Hospital fatal CHD
5. Non-CHD event
6. Not classifiable

A system used for this classification is shown in Appendix 8.

Evaluation of mortality data

Standardized death rates

Evaluation of mortality rates depends on effective registration of deaths and causes. Effective comparisons between rates and over time (trends) depends on a population census providing population size estimates and sex and age distributions. Where high-quality census information is unavailable for countries, trends and rate estimations are impossible. Even in countries with effective census systems, it is necessary to develop methods for population estimates in years between census surveys.

The mortality pattern of a population depends on its age and sex composition, thus making the crude death rate a poor comparative measure. Therefore, there is a need for an index of total mortality that can be used to compare mortality rates among populations with differing age profiles or sex distributions. The Standardized Death Rate fulfils this need.

An age-standardized mortality rate eliminates the effects of age differences and the dual effects of differences in the age and sex composition of the populations compared. Thus, standardization is a process by which mortality rates are adjusted to permit com-

parison of data sets among populations. This is achieved by estimating the overall death rate that any population would have if it had the age structure of a "standard population". Standardization is achieved by two methods, direct and indirect.

Direct standardization

Direct standardization is applied when age-specific death rates are available. The age-standardized death rate for the whole population is obtained by calculating the weighted average of the age-specific death rates of composite age strata. The proportions of persons in the various age groups in a "standard population" are used as the weights (*9*).

Indirect standardization

Indirect standardization is used when age-specific death rates are unavailable but the age structure and the crude death rate of the population are known. The method involves obtaining an "adjustment factor" for the effect of the difference between age structures of the study population and the standard (*9*). The crude death rate of the standard population is the numerator of the adjustment factor. The denominator is called an "index death rate" and is computed by applying the age-specific death rates of the standard population to the age structure of the study population. The crude death rate of the study population is then multiplied by the "adjustment factor" to give the standard death rate.

The standardized mortality ratio (SMR) is the ratio of the observed number of deaths in the study population to the number of deaths that would be expected if the population had the same rates as the standard population. This permits the standardization death rate (SDR) to be derived by adjusting the crude death rate for the differences in age distribution between the study and standard populations. The product of SMR and the crude death rate of the standard population gives the SDR of the study population.

Comparison of different standardized rates is valid only if they are based on the same standard population. The main limitation in the use of standardized indices is that there is no single standard population that can be used in all standardization procedures. The Segi World Population and the European Standard Population are two commonly used "Standard Populations" (*10*).

Time lost as a consequence of premature death

While age-standardized death rates provide the means for comparing mortality rates of different countries or time periods, the overall burden of years of life lost because of premature mortality is not adequately conveyed. Four different methods have been used to measure time lost due to mortality: potential years of life lost, period expected

years of life lost, cohort expected years of life lost, and standard expected years of life lost. The last method has been used, in conjunction with morbidity-related estimates, to derive the unit of disability adjusted life year (DALY). The DALYs lost because of different diseases have become descriptive units for estimates of the burden of disease in recent World Bank and WHO publications (*9*).

References

1. First annual report of the Registrar-General on births, deaths, and marriages in England in 1837–1838. *Journal of the Statistical Society of London,* 1839, 2:269.

2. Murray CJL, Lopez AD. *Global comparative assessments in the health sector: disease burden, expenditures and intervention packages.* Geneva, World Health Organization, 1994.

3. Eyler JM. *Victorian social medicine. The ideas and methods of William Farr.* Baltimore, Johns Hopkins University Press, 1979.

4. World Health Organization. *International statistical classification of diseases and related health problems,* 10th Revision. Geneva, World Health Organization, 1992.

5. Goldstein S. The necessity of a uniform definition of sudden coronary death: witnessed death within 1 hour of the onset of acute symptoms. *American Heart Journal,* 1982, 103:156–159.

6. Barron HV, Lesh MD. Autonomic nervous system and sudden cardiac death. *Journal of the American College of Cardiology,* 1996, 27:1053–1060.

7. Wellens HJJ. Key references on sudden death. *Circulation,* 1994, 90:2547–2553.

8. Chandramohan D et al. Verbal autopsies for adult deaths: issues in their development and validation. *International Journal of Epidemiology,* 1994, 23:213–222.

9. Hennekens C, Buring J. *Epidemiology in medicine.* Boston, Little, Brown & Co, 1987.

10. Doll R, Muir C, Waterhouse J. *Cancer incidence in five continents.* Berlin, Springer, 1970.

8.
Morbidity surveillance

Introduction

The collection of morbidity data is complex. It requires systems covering both hospital cases and cases that do not gain admission to hospital, which usually entails establishing some form of register. As long ago as 1974, Brooke, in his review of Registrars for WHO, observed (*1*) that, "the concept of registrars remains the same even though the name itself is not used". What Brooke implied was that registration can take several forms. Thus, morbidity surveillance as part of a cohort study is also a form of registration as much as traditional population-based surveillance.

Hospitalized patients are covered by diagnostic classifications through hospital activity analyses, patient administration systems, and discharge summaries. Information may also be available from hospital outpatient departments as recommended by Florence Nightingale more than a century ago (*2*). Even in developing countries, hospital morbidity data may be of high quality in certain centres. However, it is always important to consider the inherent biases that affect hospitalizations, such as the availability of beds, private versus public sector, and the varying quality of diagnoses in different hospitals. The frequency of a diagnosis will depend on the quality of the diagnostic facilities; for myocardial infarction, accuracy will depend on adequate electrocardiogram (ECG) tracings and cardiac enzyme estimations. A disease that is rare, for example CHD in women in southern Europe, tends to be underdiagnosed (*3*). Obviously, with a spectrum of the acute coronary attack ranging from mild angina through massive infarction, the final diagnosis will depend on the availability of evidence and will be influenced by various forms of bias, including clinical bias. In addition, there are the therapeutic interventions that modify or even abort more serious myocardial events. While this is to be welcomed, it creates difficulty in trying to monitor morbidity without validation. Cerebrovascular disease exhibits a similar spectrum, ranging from transitory cerebral deficit, produced even by a migraine, to a catastrophic stroke. Firm diagnosis may be possible only where advanced diagnostic equipment is available for stroke. Obviously, the introduction of such equipment may introduce spurious trends in the available data and account for differences between areas that have the equipment and those that do not.

Outside hospital, the quality and availability of data become less adequate. Information from registries is essential where stroke and myocardial infarction patients,

particularly the elderly, are cared for at home. The only way to gain knowledge of these cases is through family practice. Some countries have well developed family practice information systems, with sentinel practices that monitor the burden of cardiovascular morbidity in the population, provided that they are representative of the background population. In developing countries, however, the provision of primary care may be variable. As already indicated much valuable information can be obtained from prevalence surveys, but these have limited value in establishing trends reliably unless they are repeated, employing very strict quality control.

Registries

Overview

To achieve a broad yet accurate perspective of CVD incidence, attack, morbidity and mortality rates, some form of registry is essential. Valuable information on CVD can be gained from registries set up for other purposes, for example the Swedish Twin Registry (*4*), which has shed light on the genetic aspects of CHD. Registries are set up in defined areas for defined populations. It is important to detect resident cases treated outside the area as well as to reject cases from outside the area that are treated within the area. A strict definition of CVD cases under study is also essential. The definition may be taken from ICD-10 (*5*) or may require specific and unique criteria. The sex and age range to be kept under surveillance should also be strictly defined. Too young an age group may produce insufficient cases and therefore be liable to fluctuations due to small number effects; in too old an age group, the cases will have mixed pathology, making firm diagnoses difficult. The size of the age–sex group of the population to be kept under surveillance is crucial to detect change reliably over time; this is dependent on an adequate number of events.

The term "registration" implies something more than "notification"; it indicated that a permanent record will be established, cases followed up, and basic statistical tabulations on both frequency and survival prepared (*6*). In addition, the patients in a registry should frequently be the subject of special studies. What is clear is that the registry will involve case finding, follow-up and statistical analysis of the data obtained. Registration implies the identification of each person who presents as a case of the disease; and the establishment of a permanent record for all such people, and includes the recording of data necessary for identification purposes and the collection of subsequent data on the course of the disease and relapses (*1*).

Uses of registries

There is no doubt that a seriously biased picture can be obtained if CHD is viewed from the perspective of the coronary care unit. As McNeilly & Pemberton eloquently observed, "Those who are admitted to hospital represent, in fact, the survivors of a storm which has already taken its main toll" (*7*). In the WHO MONICA Project

(*3*) the 28-day case-fatality was just under 50% in males and 54% in females (35–64 years), with around 80% of these deaths occurring within the first 24 hours. Many sufferers do not survive long enough to be admitted to hospital.

The gathering of surveillance data serves to validate statutory mortality data. This is desirable because, as Stamler has pointed out, pitfalls in international comparisons of ischaemic heart disease (IHD) mortality rates are due to differences in death certification and coding methodology (*8*). Through coronary event registration, using consistent criteria, international differences can be examined. This has been one of the main objectives of the WHO MONICA Project (*9*). Registries have been used to compare areas with different levels of acute services for myocardial infarction and to compare case-fatality rates and survivorship (*10*). They are also useful in monitoring competing forms of CVD, such as CHD versus heart failure, and in assessing the effects of intervention programmes of diverse types. Registries are an excellent source of cases for case–control studies (*11*), and are useful in follow-up studies of cohorts.

Registries can be recommended for monitoring the introduction of new therapies and for following the population impact of community intervention programmes. It must not be forgotten, however, that—without adequate survey data—whatever is observed in incidence cannot be ascribed causally to changes in levels of therapy or to lifestyle changes in the population. A comparison of the Coronary Artery Surgery Study (CASS) Registry and the Balloon Valvuloplasty Registry in the USA illustrated the difficulties in using registry data to evaluate medical procedures, the primary problem being the lack of a valid comparison group. The authors correctly concluded that, "It is certain that a registry is not a replacement for a well-designed clinical trial" (*12*). It is worth emphasizing that the framework of a registry set up, for example, for myocardial infarction, can be used to monitor other conditions, such as heart failure.

Examples of registries

The systematic surveillance of chronic disease was first attempted in 1918. Continuous recording of cancer cases began in Mecklenberg in 1937 (*13*). Whereas, cancer is a heterogeneous group of diseases, CVD is less so and a stronger case could be made for its registration than that of cancer. Numerous community registrars have been made of CVD, particularly CHD.

The WHO myocardial infarction register

After the Second World War, research into the causes of CVD was concentrated on national, regional, ethnic and individual differences in rates. Death rates from CHD had increased in most industrial countries during the first two decades after the War but stroke rates were usually falling. The Edinburgh Study was a 1-year pilot

registration study for the WHO Myocardial Infarction Community Register (*14*), an international collaborative study coordinated by the WHO Regional Office for Europe. In 1971, 19 communities made an intensive study of every person below the age of 65 years believed to have suffered an acute myocardial infarction (AMI). A total of 3.6 million people were included in the study. Completeness of case ascertainment was ensured by a system of notification from doctors and health workers, through the scrutiny of documents from hospital records and by making appeals, both on a personal basis and via the mass media. Every suspected case was registered, and information was recorded on the medical history, the timing of events, the course of the patient's illness over the following 12 months, and any complications that occurred (*15*). These efforts demonstrated the feasibility of morbidity surveillance across many centres.

The WHO MONICA Project

During the 1970s it became apparent that CHD mortality was declining in several countries particularly Australia and the USA, whereas trends were stable or rising in many others. The reasons for these different trends were not known. Changes in mortality might have been due to a decline in disease incidence or to better survival of those affected, but consistent population data on morbidity and risk factors were scarce and not collected in a uniform manner. In most cases, the trends in mortality were unsupported by validation of death certificates, trends in non-fatal CVD, or data on cardiovascular risk factor levels and sociocultural behaviours such as diet and exercise. Such information is fundamental to the development and monitoring of strategies for CHD prevention and control. Following a conference convened by the National Heart, Lung and Blood Institute in Bethesda, USA, in 1978 (*16*), WHO's CVD Unit brought together a working group of interested participants and multinational monitoring of trends and determinants in CVD, and the MONICA Project was born. Data collection in all participating centres was under way by January 1985 (*9*). Ten years later there were 32 MONICA centres with a total population of 10.6 million persons aged 25–64 years in defined geographical areas under surveillance (*3*).

Populations were large enough to display 10-year trends, typically providing 100–300 annual deaths due to CHD in men aged below 65 years; the larger the number of deaths the smaller the trend that is detectable. The selected areas were entities with continuing availability of demographic and mortality data and had medical personnel with the cardiological and epidemiological expertise to identify and register coronary events. Demographic and statutory mortality data were reported on a regular basis "with the numbers of coronary deaths indicating what needed validation" (*3*). Pilot studies were conducted in each population to find the best methods of identifying coronary events. Hospital cases could be identified through either the active screening of admissions ("hot pursuit"), or identifying and extracting records after discharge ("cold pursuit"), or a combination of both. Efforts to identify non-fatal cases that were

managed outside hospital involved contact with the family practitioners, surveillance of the general practitioner requests for cardiac enzymes, and hearsay.

Qualifying events had to satisfy MONICA criteria for non-fatal (or fatal) "definite myocardial infarction" or "possible coronary death", or to be fatal events, usually sudden, in which CHD was the potential cause. Core data included sex, birth date, date of onset, management, survival at 28 days (allowing classification of sudden death, i.e. within 1 hour), diagnostic data, previous CHD, MONICA diagnostic category, and clinical diagnoses, coded to the International Classification of Diseases (ICD-9) (*17*). Details of the death, death certificate diagnoses, and findings at autopsy were recorded. Coronary events were studied from onset until midnight between the 27th and 28th day, when they were designated fatal or non-fatal. The same period distinguished first from recurrent events. Non-fatal events not suspected or investigated within this time were ineligible. Non-fatal events were classified as "definite", "possible" "ischaemic cardiac arrest" or "no myocardial infarction". The major categories came from the European registers in the 1970s but were redefined quantitatively, incorporating criteria from the USA, in particular, Minnesota coding of ECGs. For a non-fatal event to be definite there had to be:

- either a progression of Minnesota codes on serial ECGs, that is,
  - — progression from no Q wave to a definite Q wave;
  - — a lesser Q wave progression combined with a progressive ST-segment depression, newly developing ST-segment elevation, or progressive T-wave inversion; or
  - — persistent ST-segment elevation with progressive T-wave inversion in sequential daily ECGs; or
- cardiac enzyme levels twice the limit of normal, either with typical symptoms and an ECG that was not normal, or with an ECG progression labelled "probable" and lesser symptoms.

Further details from the MONICA case definitions and event registration are found in Appendix 13.

Atherosclerosis risk in communities (ARIC)

The Atherosclerosis Risk in Communities study has been conducted in four American communities to measure the natural history of CVD, medical care and associated risk factors by race, sex, place and time (*18, 19*).

ARIC has a cohort study and also a community surveillance component. The community surveillance component stemmed from the 1978, National Heart, Lung, and Blood Institute workshop on CHD mortality (*20*). Hospitalized non-fatal myocardial infarction and CHD deaths are ascertained in men and women aged 35–74 years. Community surveillance for hospitalized myocardial infarction cases involves a review

of hospital records for residents with a diagnosis of myocardial infarction or related illness. All ICD-9 410 and 411 (*17*) discharge diagnoses are included, and other diagnoses sampled. Hospital records identified through this process are abstracted for information relating to history, symptoms, signs, times of onset and admission, enzymes, ECG and treatment. This information is used in a diagnostic algorithm that classifies each event as confirmed myocardial infarction, possible or absent. Selected events are reviewed for validation purposes. The surveillance of CHD deaths is accomplished by abstracting all age- and residence-eligible death certificates with various manifestations of IHD coded as the underlying cause of death. An additional subset of death certificates is sampled from a group with related ICD codes. Sources of validation for out-of-hospital deaths include interviews with physicians and the next of kin, the coroner or medical examiner's reports, and hospital records. Deaths occurring in the hospital are classified by abstracting information from the medical records. CHD deaths identified underwent review by a committee of investigators. The data collection forms and a diagnostic algorithm are applied providing a preliminary classification and identify events with insufficient information or equivocal diagnostic information that requires further interpretation (Appendix 14).

The Minnesota Heart Survey (MHS)

The Minnesota Heart Survey performs community surveillance of trends in CHD mortality, morbidity and risk factors for coronary heart disease and stroke in the population aged 25–74 years starting in 1970 and still ongoing (*21*, *22*). It seeks to establish trends and their associations by monitoring mortality, morbidity and population risk factor levels in this urban centre of 2.5 million inhabitants. All death certificates from CVD and other associated diagnoses are collected. Comprehensive surveillance of all area acute care hospitals is carried out, monitoring for CHD and stroke diagnoses. These records are retrospectively abstracted (cold pursuit) by trained nurses who collect information needed for consistent validation of events. The resident population of adults, age 25–84, is randomly sampled and surveyed for demographic, risk factor and clinical data at 5-year intervals. These population data are combined to provide a clearer picture of the magnitude of, and potential reasons for, trends. Age-adjusted CHD has steadily declined since 1970, in parallel with falling rates of smoking, cholesterol levels, and hypertension, and with improved medical care (*23*). See Appendix 15 for myocardial infarction case definition.

American College of Cardiology/European Society of Cardiology definition

A combined task force of the American College of Cardiology (ACC) and the European Society of Cardiology (ESC) met in 1999. They developed a new case definition of AMI based on the USA and MONICA experience but with new biomarker data (23a).

Stroke surveys

The WHO community stroke registers were established in the 1970's (*24*). MONICA has continued this method with the monitoring of mortality and incidence rates of acute stroke in 21 populations from 11 countries (*25, 26*).

Two early stroke registers in developing countries deserve special mention. The first was set up in Ibadan in Nigeria and covered the period from 1973 to 1975 (*27*). It was part of the WHO stroke registers programme (*28*) and registered 318 patients over the 2-year period. Stroke was more common in Nigerian men than women. Two major risk factors were hypertension and diabetes mellitus, and this was age-dependent. In the city of Salvador in Brazil, data on stroke were collected over the period 1979–1980 (*29*) from all health facilities treating stroke; mortality was also kept under surveillance. Data on 1320 male and female cases of all ages were assembled. Around the age of 45 years incidence started to rise abruptly and early mortality was high.

Stroke surveillance has advanced in recent years with the advent of computerized tomography and magnetic resonance imaging techniques, which have improved diagnostic sensitivity and specificity. The MONICA and MHS stroke surveillance forms are two such examples of registers (Appendix *16, 17*).

Heart failure

Registers for heart failure have increasing relevance as acute CHD mortality declines in developed countries and heart failure becomes more common in an ageing population (*30, 31*) with increasing health care needs. This represents a trade-off of mortality for morbidity. In addition, progress in heart failure is improved by medical and surgical intervention and because survival after myocardial infarction is now more likely (*32*). One of the problems of a heart failure register is that heart failure is a syndrome with a broad spectrum of clinical signs and symptoms, for which many different scoring systems have been developed (*33*). A simple "gold standard" system that is non-invasive and can be employed in the field has yet to become available. Echo Doppler methods may advance this field but studies employ different systems so that comparisons are often invalid (*33*) and the positive predictive value for each factor tends to be poor. Even in the CASS Register, of the 4034 CHD patients with a normal Minnesota coding of the ECG only 0.6% had an ejection fraction ≤ 0.35 (*34*). It is estimated that 2–3 million people in the USA have congestive heart failure (CHF), half of whom are over 65 years of age (*35*). The Framingham study showed a prevalence of 1% CHF at ages 25–54 and 4.5% at age 65–74 years. Above 75 years, about 10% of the population is affected (*36*). CHF carries a poor prognosis, with steeper downward slopes to the survival curves than for some cancers and only a 35–40% 5-year survival rate.

Few cohort studies have employed surveillance for CHF: one from Finland (*33*) two from the Netherlands, and one from the USA (Rochester, MN). Another from Sweden studied an area inhabited by 1.64 million people in a retrospective survey based on hospital records (*37*). A total of 2711 patients, aged 16–65 years, were classified as having CHF over a 6-year period. The diagnosis was based on the presence of heart failure by the patient's physician and at least one of six signs had to be satisfied:

— pulmonary rales
— pulmonary congestion visible on X-ray film
— peripheral oedema of probable cardiac origin
— 3 kg of weight loss in 5 days in response to diuretic therapy
— cardiogenic shock
— signs of congestion on autopsy.

These and other studies have used a combination of symptoms, signs and laboratory values to classify CHF. In addition to Framingham criteria others including those used at the Mayo Clinic, are also available and are reviewed elsewhere (*33*) (Appendix 18).

Peripheral vascular disease (PVD)

In the UK, it is estimated that approximately 2% of adults in late middle age have intermittent claudication (*38*). The WHO questionnaire on intermittent claudication is insensitive for epidemiological studies of PVD. Therefore, it has been suggested that epidemiological studies of PVD of the lower limbs should also include three non-invasive tests: measurement of the ankle-brachial systolic pressure ratio, treadmill exercise test, a reactive hyperaemia test, and assessment of toe-pulse reappearance time (*38*). A subsequent cross-sectional study of 1592 subjects in Edinburgh showed that 4.5% of 55–74-year-olds, had intermittent claudication, whereas 8% had major asymptomatic disease causing significant impairment to blood flow (*39*).

Other registries

Registries have been set up for angina pectoris (*40*), cardiac surgery (*41*), angioplasty (*42*), cardiogenic shock (*43*), congenital heart defects (*44*), interventions (*45*), carotid surgery (*46*) and other cardiovascular conditions.

Registries in developing countries

CVD is an increasing problem in developing countries as infectious disease control enables life expectancy to increase (*28*). Stroke registers have been established in a number of developing countries. Although stroke is relatively more prominent than CHD in many countries, as diets change and cigarette consumption rises steeply, an epidemic of CHD is expected. Full-scale registration of CHD is costly, and particu-

larly so where the disease is uncommon. In developing countries hospital morbidity data can be a useful but limited indicator. It might be useful for researchers to set up registers for short periods of time or carry out national audits of beds by diagnoses, for example, for a designated week each year. Another approach is to establish a system of lay-reporting for out-of-hospital events (see Chapter 7).

In the early part of the 20th century, rheumatic fever (RF) and rheumatic heart disease (RHD) were considered rare in tropical countries. Several large series failed to find any significant amount of RF or RHD but it appears that the problem has always been there. In 1932, Hodge unequivocally stated that RF was not rare in India, and by 1938, streptococcal antibody in sera was highly prevalent (*47*). Indeed, between the 1930s and the late 1970s, hospital data suggested that approximately one-third of total cardiac admissions of adults and children to a teaching hospital in India were due to RF and that nearly half of the patients were less than 20 years of age. RF and RHD remain of major health importance in the developing world, particularly in India (*47*). The registration of cases to establish prevalence and incidence and to monitor the progress towards prophylactic and surgical control of RHD is of undoubted relevance.

Establishing and running a register

In setting up a register, the investigators should first define it and ask which diseases are to be monitored, and which diagnostic criteria mapped, how the background population is to be defined, what the qualification period is for residency, what the age range is and so on. If a disease is relatively rare, a very large population will have to be followed to establish true prevalence with precision. Thus, the population should be capable of generating sufficient events in the age/sex stratum under study so that sufficient power is imparted to detect changes over time. If this is not the case, the epidemiological utility of the register will be jeopardized. Another area where caution is urged is in choosing what to register. For example, in some western countries there has been an exchange of fatal myocardial infarction diagnoses for heart failure (*31*). If only one of these conditions is monitored a biased picture will result. Again, for arguments sake, the possibility that thrombolytic therapy is aborting myocardial infarction. By only registering definite myocardial infarction an important therapeutic effect might be missed. Another point in setting up a register is whether it will be possible to acquire the information. Similarly, an important consideration is whether census and socio-demographic data are available for the area. And what is to be done about patients who have their attacks outside the study area and how will a change in autopsy rates affect the data?

In many countries, stroke and myocardial infarction patients, particularly the elderly, may be cared for in nursing homes; information from general practice is therefore essential, and should be one of the main preoccupations of anyone involved in registering these diseases. Moreover, a considerable proportion (up to one-third) of all infarcts may be silent (*48*, *49*), ECG signs may regress, leaving no lasting sign of infarction after 2 years (*50*). This is important, because this kind of infarction carries

as poor a prognosis as the fully evident syndrome. Incidence of silent infarction increases over the age of 40 and does not start to fall until after age 65. These cases can be found only by regular screening of all individuals over the age of 40. A similar situation may apply to stroke: brain scans carried out on an elderly population commonly reveal lesions yet the significance of these is unclear (*51*).

In summary, there are difficulties in establishing a register and due care is essential. It is a wise precaution to run pilot studies first. For the establishment of a register, the disease in question should have important public health consequences. This is certainly true of CVD, which is a universal problem. Registries are more likely to be successful if set up in conjunction with local agencies, which should ensure the availability of adequate sufficient long-term funds (see below).

Location and setting of the register

It is usual for a register to be housed in some academic institution. Certainly, it is a good idea for it to be located close to the major source of events such as a large hospital, which will increase contact with clinicians and improve the esteem in which the register is held.

Funding

Funding will depend on local circumstances, but health service providers should have an interest in such data, giving assurance of sustained funding.

Management

It is usual to have an advisory committee on which the interests of physicians, other providers of data, and the users of the data, are represented. Such a committee might meet once or twice a year. In addition, it may be necessary to have specialist medical and statistical committees that meet more regularly to evaluate data.

Staffing

The principal investigator should have epidemiological training, a medical background, and good general knowledge of CVD, plus skill in quality control. Moreover, the investigator must attract funding and ensure that high quality reports and papers are produced. Another key function is the hiring and training of staff. Nurses make good register workers because of their medical knowledge. However successful registers have been established using only trained clerical staff.

Other staff should include a secretary and a research officer/supervisor. The number of staff will depend on the workload and size of the area/population to be covered. For some tasks, a medical background may actually be a disadvantage. For example,

staff with a clinical working knowledge of ECGs may find the chore of careful Minnesota coding irritating because they believe they can instantly recognize whether or not an ECG shows infarction.

Good relationships with other departments

It is essential to establish and maintain good relationships with other departments such as medical records departments. Staff will be dependent on many sources for the assembly of good-quality data. A register will be seen as a research-orientated task and may meet resistance from departments that have a strong service orientation and dislike being monitored. To function satisfactorily, the information should be collected in as unobtrusive a fashion as possible and the clinician's advice sought.

Statistical analysis

Ideally a register would have its own dedicated computer equipment with rapid data-entry to assist in quality control checks. Patient confidentiality should be a paramount concern: although CVD may not be as emotive a diagnosis as cancer, patient confidentiality must be sacrosanct. Follow-up for survival analyses should be undertaken in relation to various forms of treatment requiring record linkage.

Data collection

Completeness of data is critical, and searches must be made periodically for cases that may have been missed through erroneous diagnosis or out-of-hospital management. There should be regular contact with outpatient clinics as well as with hospitals. Strict criteria must be established so that any episode of disease is recorded as an event in the register. The best strategy is to have strict criteria pre-specified to determine a case as described in Appendices 14 and 15. Very often there are borderline cases and the investigator will be consulted. This happens in all registries and is not of importance when the rate is low. Competing causes for hospitalization also need to be taken into account and rules should be made to cover such instances as a patient having both a stroke and a myocardial infarction. In making these difficult decisions it is wise to keep a log or "precedent" book.

Interviews with relatives of deceased individuals may be necessary when there are no other sources of the necessary information, and it may be possible for a member of the registry staff to interview the relatives in their homes. Nurses or social workers have been successfully used in these circumstances. The bereaved family will often welcome the visit as it gives them the opportunity to ask questions. Collecting information on cases requires a data-collection monitoring system. Data-based management systems and statistical packages such as Epi-info (*52*) have greatly facilitated registration (*www.cdc.gov/epo/epi/software.htm*).

"Hot" vs "Cold" pursuit

"Hot" and "Cold" pursuit are terms that refer to prospective or retrospective acquisition of data. "Cold" pursuit is satisfactory provided that the information is readily available in the hospital or clinic diagnostic records. "Hot" pursuit is the collection of data from the patient at presentation at hospital or clinic, and is more difficult because it is costly and staff-intensive. In "cold" pursuit, hospital morbidity data cannot be relied upon until they have been properly validated and shown to be accurate. Pilot studies are essential to assess the diagnostic categories that should be screened. If the yield of certain diagnostic categories is tiny, they can be omitted, although periodic checks should be carried out to ensure that there have been no drifts in diagnostic "fashion" among the clinicians or that no coding errors have crept in. It may in certain circumstances be possible, once good relationships have been built up, for example with coronary care unit staff, for additional items to be routinely collected by the unit in a "hot" pursuit. Good relationships may be improved by demonstrating the usefulness of registration to clinicians. This is best accomplished by facilitating their studies and through the publication of high quality papers.

Quality control

Avoidance of bias, particularly a changing level of bias, is critical. Bias can create spurious trends in incidence and attack rates and case-fatality. Conversely, it may conceal true trends. It cannot be stressed too highly that avoidance of bias requires quality control procedures to be carried out by staff who are trained, certified and recertified. This can be through standard test case histories, laboratory valves and ECGs, or through recoding a sample of medical charts randomly selected from the registry's data base.

Some of the problems of changes in diagnostic tools, for example the changing sensitivity of biomarker estimation in AMI, have already been discussed. One observer's coding of records may change over time or there may be discrepancies between observers. If the registry is assembling longitudinal data, it is very important to have continuity of trained staff, so staff must be well treated. Too great a reliance on hospital inpatient data may be a problem if there is a change in the way the data are captured. Changes in the rates of autopsy may affect the number of cases because of the better diagnostic information. To provide early warning, it is wise to compare the data as they accumulate throughout the year with the previous year's data. For example, if halfway through the year only 50% of the events observed in the same period last year have been collected, something is probably wrong. Regular training, reabstraction of a sample of records (5–10%) and test cases will also help to maintain the quality of data collection.

A "Happy" register

Morale is important and success with good quality publications will foster this. Registration can be monotonous and other studies going on in the background provide stimulation. Register staff should also be encouraged to attend training and scientific meetings and should be rewarded by social functions. It is wise to maintain links with other registers, and international links are particularly gratifying.

Knowing when to stop

Lastly, a good principal investigator should recognize that, once objectives have been achieved and the study questions are fully answered, the registry should be closed down. As Tunstall-Pedoe (*53*) has observed, "The collection of information for its own sake is of doubtful value unless it is acted upon. Community Registers should not become the medical equivalent of village war memorials".

References

1. Brooke EM. *The current and future use of registers in health information systems.* Geneva, World Health Organization, 1974.

2. Nightingale F. Hospital statistics. In: *Notes on Hospitals,* 3rd ed. London, Longman, Green & Co., 1863:159–180.

3. Tunstall-Pedoe H et al. Contributions of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA Project populations. *Lancet,* 1999, 353:1547–1557.

4. Marenberg ME et al. Genetic susceptibility to death from coronary heart disease in a study of twins. *New England Journal of Medicine,* 1994, 330:1041–1046.

5. *International Statistical Classification of Diseases and Related Health Problems,* Tenth Revision. Geneva, World Health Organization, 1992.

6. *Epidemiological methods in the study of chronic diseases.* Geneva, World Health Organization, 1967 (WHO Technical Report Series, No. 365).

7. McNeilly RH, Pemberton J. Duration of last attack in 998 fatal cases of coronary artery disease and its relation to possible cardiac resuscitation. *British Medical Journal,* 1968, 3:139–142.

8. Stamler J. Opportunities and pitfalls in international comparisons related to patterns, trends and determinants of CHD mortality. *International Journal of Epidemiology,* 1989, 18:S3–S18.

9. *World Health Organization MONICA Project: Manual of Operations.* Geneva, World Health Organization, 1990.

10. Mathewson ZM et al. Coronary care and community mortality from myocardial infarction. *Lancet*, 1985, 1:441–444.

11. Parra HJ et al. A case-control study of lipoprotein particles in two populations at contrasting risk for coronary heart disease. The ECTIM Study. *Arteriosclerosis Thrombosis*, 1992, 12:701–707.

12. Davis K. Use of data registries to evaluate medical procedures. *International Journal of Technology Assessment in Health Care*, 1990, 6:203–210.

13. Wagner G. History of cancer registration. In: Jensen OM et al., eds. *Cancer registration: principles and methods*. Lyon, International Agency for Research on Cancer, 1991 (IARC Scientific Publications, No. 95).

14. Armstrong A et al. Natural history of acute coronary heart attacks: a community study. *British Heart Journal*, 1972, 34:67–80.

15. *Myocardial infarction community registers*. Copenhagen, World Health Organization Regional Office for Europe, 1976 (Public Health in Europe, No. 5).

16. Havlik RJ, Feinleib M. *Proceedings of the Conference on the Decline in Coronary Heart Disease Mortality. Bethesda, October 24–25, 1978*. Bethesda, MD, US Department of Health, Education and Welfare, Public Health Service, 1979 (NIH Publication No. 79–1610).

17. *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death Ninth Revision*. Geneva, World Health Organization, 1977.

18. ARIC Manual 1: General description and study management. *NHLBI*, 1987.

19. ARIC Manual 2: Cohort component procedures. *NHLBI*, 1988.

20. White AD et al. The community surveillance of coronary heart disease in the Atherosclerosis Risk In Communities (ARIC) Study: methods and initial two years experience. *Journal of Clinical Epidemiology*, 1996, 49:223–233.

21. McGovern PG et al. Trends in survival of hospitalized myocardial infarction patients between 1970 and 1985: the Minnesota Heart Survey. *Circulation*, 1992, 85:172–179.

22. McGovern PG et al. Trends in mortality, morbidity, and risk factor levels for stroke from 1960 through 1990: the Minnesota Heart Survey. *Journal of the American Medical Association*, 1992, 268:753–759.

23. McGovern PG et al. Recent trends in acute coronary heart disease mortality, morbidity, medical care and risk factors. *New England Journal of Medicine*, 1996, 334:884–890.

23a. Myocardial infraction redefined – a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infraction. *European Heart Journal, 2000, Sept; 21(18): 1502-13* and also in *Journal of the American College of Cardiology. 2000 Sep; 36(3): 959-69.*

24. *Hypertension and stroke control in the community.* Proceedings of a WHO meeting, Tokyo 11–13 March 1974. Hatano S, Shigematsu I, Strasser T, eds. WHO, Geneva, 1977.

25. Thorvaldsen P et al. Stroke incidence, case fatality, and mortality in the WHO MONICA Project. *Stroke*, 1995, 26:361–367.

26. Asplund K et al. Multinational comparisons of stroke epidemiology: evaluation of case ascertainment in the WHO MONICA stroke study. *Stroke*, 1995, 26:355–360.

27. Osuntokun BO et al. Incidence of stroke in an African city: results from the stroke registry at Ibadan, Nigeria, 1973–1975. *Stroke*, 1979, 10:205–207.

28. Dodu SRA. Emergence of cardiovascular diseases in developing countries. *Cardiology*, 1988, 75:56–64.

29. Lessa I, Bastos CAG. Epidemiology of cerebrovascular accidents in the city of Salvador, Bahia, Brazil. *Pan American Health Organization Bulletin*, 1983, 17:292–303.

30. Erikson H. Heart failure: a growing public health problem. *Journal of Internal Medicine*, 1995, 237:135–141.

31. Bonneux L et al. Estimating clinical morbidity due to ischemic disease and congestive heart failure: the future rise of heart failure. *American Journal of Public Health*, 1994, 84:20–28.

32. Clarke AL, Coates AS. New evidence for improved survival in chronic heart failure. *Clinical Cardiology*, 1994, 17:555–558.

33. Cowie MR et al. The epidemiology of heart failure. *European Heart Journal*, 1997, 18:208–225.

34. Rihal CS et al. The utility of clinical, electrocardiographic, and roentgenographic variables in the prediction of left ventricular function. *American Journal of Cardiology*, 1995, 75:220–223.

35. National Heart, Lung and Blood Institute Data Fact Sheet on Congestive Heart Failure. Bethesda, MD, United States Department of Health and Human Services, Public Health Service, 2000.

36. Kannel WB. Need and prospects for the prevention of cardiac failure. *European Journal of Clinical Pharmacology*, 1996, 49:S3–S9.

37. Anderson B, Waagstein F. Spectrum and outcome of congestive heart failure in a hospitalized population. *American Heart Journal*, 1993, 126:632–640.

38. Fowkes FGR. The measurement of arterial atherosclerotic peripheral arterial disease in epidemiological surveys. *International Journal of Epidemiology*, 1988, 17:248–254.

39. Fowkes FRG et al. Edinburgh Artery Study: prevalence of asymptomatic and symptomatic peripheral arterial disease in the general population. *International Journal of Epidemiology*, 1991, 20:384–392.

40. Wood DA. Clinical epidemiology. In: Cleland J, ed. *Asymptomatic coronary artery disease and angina*. London, Science Press, 1996:1–12.

41. Emond M et al. Long-term survival of medically treated patients in the Coronary Artery Surgery Study (CASS) Registry. *Circulation*, 1994, 90:2645–2657.

42. Holmes DR et al. Coronary perforation after excimer laser coronary angioplasty. The Excimer Laser Coronary Angioplasty Registry experience. *Journal of the American College of Cardiology*, 1994, 23:330–335.

43. Hochman JS et al. Current spectrum of cardiogenic shock and effective early revascularization on mortality: results of an international registry. *Circulation*, 1995, 91:873–881.

44. Olshan AF, Schnitzer PG, Baird PA. Paternal age and the risk of congenital heart defects. *Teratology*, 1994, 50:80–84.

45. Baim DS et al. Evaluating new devices: acute (in hospital) results from the New Approaches to Coronary Intervention Registry. *Circulation*, 1994, 89:471–481.

46. Washburn WK et al. Late stroke after carotid endarterectomy: the role of recurrent stenosis. *Journal of Vascular Surgery*, 1992, 15:1032–1036.

47. Vijaykumar M et al. Incidence of rheumatic fever and prevalence of rheumatic heart disease in India. *International Journal of Cardiology*, 1994, 43:221–228.

48. Sigurdsson E et al. Unrecognized myocardial infarction: epidemiology, clinical characteristics, and the prognostic role of angina pectoris. The Reykjavik Study. *Annals of Internal Medicine*, 1995, 122:96–102.

49. Barsky AJ et al. Silent myocardial ischemia: is the person or the event silent? *Journal of the American Medical Association*, 1990, 264:1132–1135.

50. Kannel WB, Abbott RD. Incidence and prognosis of unrecognized myocardial infarction: an update on the Framingham Study. *New England Journal of Medicine*, 1984, 311:1144–1147.

51. Manolio TA et al. Magnetic resonance abnormalities and cardiovascular disease in older adults: the Cardiovascular Health Study. *Stroke*, 1994, 25:318–327.

52. Dean AG et al. Epi Info: a general-purpose microcomputer program for public health information systems. *American Journal of Preventive Medicine*, 1991, 7:178–182.

53. Tunstall-Pedoe H. Uses of coronary heart attack registers. *British Heart Journal*, 1978, 40:510–515.

9.
Population surveys

Introduction

This chapter describes field methods applicable to both cross-sectional and cohort surveys, and to other studies of population samples. Cross-sectional surveys often lay the basis for a cohort study, using the data collected at baseline; for this reason it is desirable to plan for a possible follow-up study even if there is no guarantee that resources for the follow-up will be available. Methods for measuring disease outcomes are described first, then methods for measuring risk factors or other relevant determinants.

The methods appropriate for a survey intended to measure health or risk factors in individuals are not necessarily appropriate for a survey intended to measure the average disease rate or exposure level of groups. For studies of associations between risk factors and disease at the individual level, it is necessary to use measurement techniques that accurately characterize individuals. In ecological studies, where the objective is to study associations between disease rates and the average of some characteristics in groups, it may be possible to use methods that measure the average exposure of groups but not the exposure of individuals. For instance, to assess the average dietary intake of groups, it is possible to use household food inventories that record all food consumed by a household in a defined period.

It is important to use methods that are affordable, with modest budgets for equipment (including anthropometric, electrocardiographic, and M-mode echocardiographic). Methods that are likely to become cheaper in the next few years, such as Doppler ultrasound measurements of flow velocity, are discussed briefly on page 163.

Planning a survey should distinguish between *exposure variables* and *outcome measures.* The most important outcome measures in cardiovascular field surveys are estimates of the prevalence of disease or symptoms. For measuring the prevalence of CVD, all current techniques have serious limitations. Symptom questionnaires have poor specificity for myocardial ischaemia and cannot be relied on for cross-cultural comparisons of prevalence. Comparisons of prevalence estimates based on history of diagnosed heart disease may be biased by differences in access to medical care and diagnostic facilities. ECG criteria for prevalence of ischaemic heart disease are more

likely to yield unbiased comparisons of prevalence, but have low sensitivity for the detection of myocardial ischaemia in the non-acute setting.

The range of exposure variables (risk factors) measured in cardiovascular surveys is extremely wide: many social, behavioural and biological measurements have been found to predict CVD in one or more surveys.

Questionnaires and medical records

The general principles of questionnaire design have been dealt with earlier. Where possible, each questionnaire item should be structured so that the response is assigned to one of several mutually exclusive and exhaustive categories. The questions can then be arranged so that items that are not applicable (for instance, the age of starting smoking where the respondent is a non-smoker) are skipped. Examples of forms used in contemporary studies are found in Appendices 19 and 20.

Angina pectoris and other cardiovascular symptoms

Angina pectoris is the commonest symptom of CHD. The standard questionnaire developed by Rose for measurement of chest pain on exertion has been widely used in cardiovascular field surveys. This questionnaire was originally validated in men against clinical diagnoses (*1*). This instrument should not be regarded as an "angina questionnaire", but as a questionnaire that records symptoms (chest discomfort on exertion, relieved by rest) in a standardized manner. Recent studies using exercise thallium scanning as reference standard have shown that the chest pain questionnaire has relatively low specificity for exercise-induced myocardial ischaemia (*2, 3*). Especially in women, the questionnaire fails to distinguish coronary from non-coronary symptoms (*3*). The Rose chest pain questionnaire is included in Appendix 21, but we recommend that it be interpreted critically, if at all, as a measure of the prevalence of ischaemic cardiac symptoms. In fact, no study should be designed to use a chest pain questionnaire as a main outcome indicator. Similar points apply to questionnaires used for other symptoms of cardiovascular disease, such as claudication and breathlessness on exertion. These questionnaires measure symptoms rather than the presence of disease, and their validity as measures of disease in the population under study should be assessed critically against whatever reference standard measurements may be available.

Medical history and prevalence of diagnosed disease

A question such as, "Have you been told by a doctor that you had heart disease?" can be followed with questions about the diagnostic and therapeutic procedures performed—such as exercise testing, angiography and revascularization—in populations where access to secondary medical care and invasive cardiological techniques is good. The basis for the diagnosis, clinical symptoms, and the results of diagnostic proce-

dures such as angiography should be recorded. An obvious limitation of relying on prevalence of diagnosed disease is that the measure may be biased towards positive answers in those with better access to medical care.

It is difficult to use prevalence surveys to measure morbidity from stroke: many patients with stroke are unable to participate, or fail to appear in a sample of private households because they are in nursing homes. For this purpose, prospective stroke registers are more likely to yield valid data (Chapter 8). Such registers have been successfully set up in developing countries (*4*, *5*).

Measurements of health status and quality of life

Measurements of health status and quality of life are relevant to economic analyses of health policy and disease burden in the population. For instance, disability scores are used to derive estimates of the burden of disease in terms of disability-adjusted life years (DALYs) lost (*6*), which can then be used to compare the cost–effectiveness of various health interventions. Quality of life measurements can be used similarly, although measurement of quality of life involves more assumptions than simple measurement of functional capacity and disability. For the same reasons, measurements of health-care utilization are useful in assessing the contribution to health service costs associated with a given condition.

Various standard questionnaires are available for measurement of quality of life or functional capacity. Many of these are designed for use with groups that have serious functional impairments, and include questions about disabilities such as the inability to dress or undress unaided. Such individuals are unlikely to appear for examination at a local clinic in a cardiovascular field survey. The choice of instrument will depend upon the objectives of the study and the type of population being surveyed. For instance, if an objective is to derive estimates of the burden of a disease in terms of DALYs lost, the instrument used should be capable of assigning participants to one of the six classes of disability used by this approach (*7*).

The Medical Outcomes Study Short Form-36 (SF-36) (*8*) is an instrument developed by researchers in Boston, USA, to measure the outcome of medical care. As the instrument is sensitive to less severe levels of morbidity than most other quality of life-scoring instruments, it is more suitable for general population surveys. The instrument consists of 36 items that measure eight sub-scales: physical functioning, role limitations because of physical health problems, social functioning, pain, general mental health, role limitations because of emotional problems, vitality, and perception of general health. The instrument has been shown to be sensitive enough to detect the effects of antihypertensive therapy on quality of life in patients with hypertension (*9*).

Demographic data

Essential items to be recorded in any survey are full name, sex, marital status, date of birth, identification numbers, date of interview and identity of the interviewer. To maintain confidentiality, the name and address of the participant can be recorded on a front sheet that can be removed from the questionnaire when it is returned to the survey office, after the serial number of the participant has been entered on the next sheet of the questionnaire.

Identification numbers

Each participant in a survey should be assigned a serial number at the time the sample is assembled. Where the survey is based on households, serial numbers can be a combination of a household number and a number denoting the participant within the household. Serial numbers should have a built-in check mechanism so that most errors in transcription or data entry will be detected by the computer. A simple method of doing this is to add a check digit at the end, calculated by multiplying each digit in the main part of serial number by some constant, dividing by 10, and taking the remainder.

An example of an algorithm for calculating check digits to be added to 4-digit serial numbers is shown below. In each 4-digit serial number, the first digit is multiplied by 3, the second by 1, the third by 7, and the fourth by 3. The sum of these products is divided by 10, and the remainder is added as a check digit to give a 5-digit serial number.

	Original 4-digit serial number				Check digit (remainder after dividing 72 by 10 and rounding)
Full serial number	1	4	8	3	7
Multiplier for each digit	3	1	7	3	
Product	$1 \times 3 = 3$	$4 \times 1 = 4$	$8 \times 7 = 56$	$3 \times 3 = 9$	72

To maximize the chance that an error in transcription of the serial number will be detected, the numbers 2 and 5 should not be used as multipliers because they are factors of 10.

Programs written for data entry should include a subroutine to check the validity of serial numbers, so that entry of a serial number in which the check digit does not evaluate correctly will produce an immediate error message. Common transcription errors such as digit transposition can thus be detected instantly.

Recording identity numbers, such as social security numbers, which are used by the health service or by government population registers, makes it possible to link data collected in the survey with other records such as hospital admissions and death certificates. Written consent for the recording of linkage, tracing through population registers, or tracing through the health care system should be sought when the consent form is explained to participants.

Date of birth is required both for identification purposes and to calculate calendar age. Place of residence and place of birth are relevant in studying the effects of migration between regions or countries that have different rates of disease. For those participants who are migrants, age at migration should be recorded.

Ethnic origin

Ethnic groups are defined by shared identity based on parentage, religious and cultural characteristics. Some, but not all, boundaries between ethnic groups define shared gene pools. As rates of cardiovascular disease or levels of risk factors usually differ between ethnic groups, recording of ethnic origin is relevant. Assignment of ethnic origin may require recording of self-described ethnic origin; native language; country of birth of participant, parents and grandparents; name; religion; and physical appearance as assessed by an interviewer. Self-described ethnic origin can usually be pre-coded to one of several categories, but it may be necessary to pilot the question first to establish that the categories chosen correspond to people's preferred description of their ethnic identity. For example, piloting of census questions in the United Kingdom before the 1991 census established that "black Caribbean" was preferred to "West Indian" as a self-description.

Occupation and socioeconomic status

Social and occupational classification is important because rates of ill-health display marked social gradients in most societies. Understanding the causes of these gradients is a key area of research into the epidemiology and control of cardiovascular disease (*10*). Occupational status has traditionally been used to classify socioeconomic status, but with rapidly changing job descriptions and a more flexible labour market, occupational status is becoming less useful as a single measure. Where a national classification of occupational status exists, the questions should be designed to elicit sufficient information to classify individuals according to this system. For instance, the system traditionally used by the Registrar-General for England and Wales assigns occupations to one of six classes: professional (I), managerial (II), skilled non-manual (III-N), skilled manual (III-M), semi-skilled manual (IV), and unskilled manual (V) (11). If no official classification exists, it should at least be possible to classify occupations as manual ("blue-collar") or nonmanual ("white-collar").

As a general rule, the following questions are required to classify occupational status:

- What is the title of your usual job?
- What do you actually do in this job?
- Does your job require any special qualifications or experience?
- Are you self-employed or working for an employer?
- Do you have to supervise other people?

Coding occupational status on the basis of these questions is a separate exercise that can be carried out at a later stage. If no manual of coding occupations to occupational status groups exists, one may have to be prepared.

The participant's current economic position (working, unemployed, or not participating in the labour market because of retirement, long-term sickness, full-time study or child care) is a separate item.

In the USA and many other countries, income and education are better indicators of socioeconomic status than occupation (*12*). Income can be measured as individual or household earnings. While loss of data from non-participation can be a problem, a card with income categories on it has been found to be successful in obtaining data in >90% of cases. Highest level of formal education is an excellent indicator of socioeconomic status and is easily and reliably collected. It is strongly associated with occupational status and income. In some surveys, it is a better predictor of CVD risk than either (*13*).

Questions on consumer spending and housing costs (rent or mortgage repayments) may be more reliable guides to disposable income than questions about reported income. Thus questions about ownership of consumer durables, such as cars or televisions yield information about the resources at the disposal of the household. In countries where most households have access to these items, questions about spending on luxuries such as holidays may be more effective discriminators. In low-income communities, access to more basic amenities such as piped water and electricity may be more useful. Questions about housing tenure are useful in countries such as the United Kingdom, where socioeconomic status is strongly associated with owner-occupation.

It is important to collect information about past socioeconomic status where possible, as illness often leads to downward social mobility. Simple measures of past socioeconomic status may be based on the amenities available to the participant's household at a specified age in childhood (for instance, did the household have bath/shower, running water only, no running water).

Family history

Questionnaire items on family history in cardiovascular field surveys typically ask whether a first-degree relative—parent, sibling or offspring—was ever diagnosed with

heart disease, diabetes or hypertension. Positive family history is usually defined as the occurrence of disease in one or more first-degree relatives. This simple method of categorizing family history fails to take into account the number of first-degree relatives or their age and sex distribution. For instance, ischaemic heart disease at age 50 years in an only sister is more informative than ischaemic heart disease at age 70 years in one of four brothers. The probability of a positive family history in a sibling will depend on the number of sibs: if the risk to a sib is k, the probability of a negative family history is $(1 - k)^n$, where n is the number of sibs. Thus, if family history data are to be collected, the questionnaire should ask about the number of first-degree relatives of each type, their current ages or ages at death, and the age at onset of disease. Family history can then be scored to take into account the number of relatives of each sex and their age at onset can be employed when the data are analysed. Methods for scoring family history quantitatively are described in detail elsewhere (*14, 15*).

The general principle of these methods is to compare the observed with the expected number of first-degree relatives affected. The respondent is asked for the following information on each first-degree relative: age, sex, age at diagnosis (if affected) and age at death (if deceased). This information can be used to compute the total number of person-years at risk in each age–sex category. The expected number of affected relatives can then be calculated for each respondent using incidence rates derived either from the survey dataset itself or from some population survey or registry. The family history score F for each participant is then computed from the observed number O and the expected number E of affected relatives as $F = \frac{O - E}{\sqrt{E}}$.

Data on cause of death or history of illness in parents may be unreliable, especially in developing countries or migrant communities where access to medical care in the previous generation has been poor. In some populations, data on family history obtained from women appear to be more accurate than data obtained from men (*16*).

Size at birth

Recent studies report that measurements of size at birth are associated with blood pressure, diabetes, plasma lipids and mortality from cardiovascular disease in adult life (*17*). Associations are identified with weight, length, head circumference and abdominal circumference at birth. Unfortunately there are few countries where records of size at birth are available on people aged over 30 years. Opportunities to study these relationships in older adults in these groups are few.

Where possible, data on gestational age, maternal age and medical history; and socioeconomic status of parents should be collected in addition to size at birth. It may be possible to identify cohorts for study by searching for hospitals or maternity units

where old records have been retained or by searching for earlier studies of newborn infants where survey records and identification data may still exist.

Participant's own medical history

The minimum information to be collected should include diagnoses of hypertension, hyperlipidaemia, diabetes and details of all medications currently used. A checklist of relatively common illnesses can be devised for the population and age group under study. It is usually helpful to ask participants to bring with them all the medications they are taking, so that brand names or generic names can be accurately recorded.

The drugs taken by participants should be coded according to the pharmacological classes to which they belong; for instance, patients taking antihypertensives and angina medication may be coded as taking a beta-blocker, a calcium antagonist and a thiazide diuretic. The condition for which each drug was prescribed should be recorded if known; for instance, beta-blockers may be prescribed for angina, for hypertension, for migraine headaches or for other medical indications.

The information collected on women participants should include parity, as this is strongly related in many societies to obesity in later life. Current or past use of exogenous sex hormones as contraceptives or for post-menopausal oestrogen replacement should also be recorded.

Rheumatic heart disease is the most important childhood illness to ask about in relation to adult cardiovascular disease. However, the validity of a reported past history of rheumatic fever is likely to be poor, especially in people from backgrounds where access to medical care in early life was limited.

Smoking

Questions about smoking should include age of starting smoking, age of stopping in those who have given it up, the type of smoking (manufactured cigarettes, hand-rolled cigarettes, pipe or cigar), and the quantity used. The brand of cigarette may also be useful, especially where this can be related to data on tar and nicotine yields. Asking participants whether they inhale and how much of the cigarette end they leave, may also be relevant to the risk of ischaemic heart disease. A typical smoking questionnaire is given in Appendix 22. Standard questionnaires should be adapted to local circumstances where necessary; for instance where smoking of hand-rolled cigarettes is common, quantity of tobacco consumed is more useful to assess than simply number of cigarettes smoked. Passive smoking exposure, a growing issue, can be assessed by questions found in Appendix 23.

The simplest and cheapest biomarker of smoking is the amount of carbon monoxide in breath, measured using a portable monitor. End-tidal carbon monoxide

concentration is measured after the participant has held a breath for 15 seconds. Thiocyanate levels, another indicator of burning tobacco exposure, can be measured in serum, urine or saliva. Thiocyanate levels can be raised by dietary sources of cyanide, but this is unlikely to result in misclassification of smoking habit unless the population's diet contains substantial amounts of food with unusually high cyanide content, such as cabbage or cassava. Cotinine levels can be measured in plasma or urine: unlike carbon monoxide levels, this is sensitive enough to measure passive smoking as well as active smoking, but measurement requires an expensive radioimmunoassay or liquid chromatography. Exposure to other sources of tobacco such as chewing tobacco or snuff can also raise cotinine levels. Carbon monoxide appears to be as accurate as urinary cotinine measurements or serum thiocyanate measurements in distinguishing adult smokers from non-smokers and it has the advantages over other methods of lower cost and immediately available results (*18*).

Measurement of smoking status by biomarkers is more important in intervention studies, where participants may be motivated to report that they smoke less than is the case, rather than more in field surveys where questions are generally asked in a neutral manner. In measurement of exposure to tobacco smoke, questionnaire data on smoking may fail to capture differences in smoking behaviour such as depth and frequency of inhalation. Measurement of biomarkers may thus be more accurate than questionnaire data in quantifying tissue exposure to substances in tobacco smoke.

Alcohol

The relationships between alcohol drinking and cardiovascular disease are complex. Regular light drinking may be associated with lower risk of coronary disease, whereas heavy drinking, even occasionally, may be associated with increased risk of stroke and sudden cardiac death. There is uncertainty about the effects of the type of alcoholic beverage consumed: beer, wine and spirits may have different effects on cardiovascular disease risk. In cardiovascular surveys, it is therefore desirable to determine the type of alcoholic beverages consumed, the quantity of each beverage consumed, and the frequency of drinking.

Alcohol questionnaires must generally be devised to be specific for the population under study, since the type of alcoholic beverages consumed and the pattern of drinking vary between communities. In populations where alcohol drinking is socially unacceptable or illegal, one or two initial questions can be used to identify those who are total abstainers before details of alcohol consumption are sought. For instance, where beers of varying alcohol content are available, there must be separate enquiries about each type of beer. Essential data items to be collected include the frequency of consuming each form of alcoholic beverage and the quantity usually consumed. A simple questionnaire of this kind is the Quantity–Frequency Index developed by Straus and Bacon (*19*). An example of this type of index is shown in Appendix 24.

Respondents are asked about their drinking of each main type of alcoholic beverage; for instance, beers, wines and spirits. For each beverage, the usual frequency of drinking is scored on a scale ranging from "three or more times a day" to "never"; the usual quantity drunk is also scored on an appropriate scale using units familiar to the respondent (for instance, cans of beer, bottles of wine).

A limitation of the Quantity–Frequency questionnaire is that it asks only about the usual quantity drunk on any single occasion. Alcohol consumption and the frequency of heavy drinking will thus be underestimated if respondents who drink heavily on one or two days of the week report their consumption on other days as their "usual" consumption. One approach has been to modify the Quantity–Frequency questionnaire to obtain more detailed information about the frequency of heavy drinking and the quantity drunk on such occasions. An alternative approach is to ask in detail about alcohol consumption on each day during a "typical" week (*19*, *20*), using a questionnaire matrix with one column for each day of the week and one row for each type of alcoholic beverage.

The last time that alcohol was drunk should be recorded when blood pressure is measured, as alcohol has a blood pressure-raising effect that wears off after a few days of abstinence. Participants should be asked whether they have reduced their consumption or given up drinking and, if so, the reason for this change in behaviour. Lifetime non-drinkers can thus be distinguished from former drinkers who have given up drinking because of illness.

Short questionnaires on alcohol-related problems of social adjustment or early signs of alcohol dependence, such as the CAGE[1] questionnaire given in Appendix 25, may also be useful in defining the prevalence of alcohol-related morbidity in the population (*21*).

As with other questionnaires developed and validated in Europe or North America, however, the questionnaire cannot be assumed to be equally valid in other populations. For instance, questions about whether others have criticized the respondent's drinking have different meanings in a society where alcohol drinking is socially unacceptable. Estimates of national average alcohol consumption in surveys generally yield values that represent only about half the total supplies of alcohol moving into consumption (*22*). This may reflect both underestimation of alcohol consumption by responders and high rates of non-response among heavier drinkers. Biomarkers such as blood alcohol level, plasma gamma-glutamyl transferase and mean corpuscular volume do not provide useful measures of light to moderate drinking but may help to detect heavy drinking. Supplementing the CAGE questionnaire with these measures, however, has not been shown to improve the ability to discriminate alcohol-dependent patients from controls (*23*, *24*).

[1] CAGE: questionnaire devised by Dr John Ewing

Psychosocial factors

Social support measures and life events

Social support measures are among the most consistent predictors of health outcomes. Several forms are included in Appendix 25 but these instruments are culturally specific and may need to be individually designed for a particular population.

In communities where social support comes mainly from the extended family, items about social support from individuals outside the family may be less relevant than in developed countries where nuclear families are usual. Questionnaires to measure *life events* are widely used in psychiatric research, but have not generally found wide use in surveys of CVD. Their principal disadvantage is that the scoring of life events used by some researchers depends upon subjective ratings of the meaning of an event to the participant.

Job demands and control over work

Swedish researchers have developed a scheme for assessing occupational characteristics according to the level of job demands and the individual's control over work (*24*). These questions can be easily applied to employed workers, but may be less relevant to those who are self-employed. The level of control over work is closely related to occupational status, and questions on this can be combined with questions about the level of skill and responsibility required for participants to perform their usual occupations.

Type A behaviour and minor psychological symptoms

In some studies in the USA, a pattern of behaviour characterized by hurrying, impatience and preoccupation with deadlines has been found to predict heart disease (*25, 26*). Type A behaviour is measured by a standardized interview. However, this assessment is difficult to apply outside the population in which it was originally developed, and more recent replication studies have failed to confirm that Type A behaviour predicts mortality from CVD. Standard questionnaire instruments such as the General Health Questionnaire (*27*) can be used to screen for minor psychological symptoms in population surveys. More recently hostility has been proposed as the important element in Type A behavior (*28*).

Acculturation

Measurement of acculturation is especially relevant to studies of migrant groups. Ad hoc scales can be easily constructed by compiling a series of items such as language spoken at home, frequency of consuming foods typical of the community of origin or the host community, and frequency of socialization with other migrants from the same

country or with natives of the host country (*29*). A pilot study will establish which of these items discriminate between migrants and natives of the host community.

Health beliefs and behaviours

Asking about health beliefs and behaviours can help to identify issues relevant to health promotion strategies. For instance, participants can be asked to score which factors on a list they think cause heart disease, and whether they believe that heart disease is usually preventable (Appendix 26).

Questions about attitudes to body size are useful in populations with high prevalence of obesity: participants can be asked to rate themselves on a scale ranging from underweight to very overweight.

Physical activity and energy expenditure

Measuring physical activity in cardiovascular surveys usually involves trying to estimate two underlying variables: total energy expended in physical activity and vigorous activity that raises cardiorespiratory fitness. The design of the questionnaire will depend on which of these objectives is important and on the levels of activity that it is desirable to identify; a questionnaire suitable for young adults is unlikely to be suitable for an older population.

Measurement of energy expenditure makes it possible to assess the extent to which dietary intakes are under-reported, and to examine the relationship of obesity to energy expenditure and physical activity. The main determinants of total energy expenditure are likely to be the distance walked or cycled per day, and levels of occupational activity. Activity that raises cardiorespiratory fitness is likely to be vigorous: in sedentary individuals half an hour of vigorous activity three times a week is sufficient to raise maximal oxygen uptake by 10–15% (*30*). Thus, questions about activity at work and distance walked per day are likely to show relationships with obesity, whereas questions about vigorous activity during leisure-time are likely to show relationships with cardiorespiratory endurance.

Intensity of activity

Exercise intensity can be expressed in terms of energy expenditure, oxygen consumption or metabolic equivalents (METs). One MET corresponds to the oxygen consumption of the body at rest: about 3.5 ml $O_2 kg^{-1} min^{-1}$. Using published data on the energy costs of various activities (*31*), most everyday activities can be grouped by intensity (Table 9.1).

This classification in METs is based on the categories defined by Haskell & Pollock for the United States Surgeon General's report on Physical Activity and Health (*32*).

Table 9.1 Classification of physical activity intensity, based on physical activity lasting up to 60 minutes

	Endurance-type activity							Strength-type exercise	
		Relative intensity		Absolute intensity (METs) in healthy adults (age in years)				Relative intensity*	
Intensity	VO_2 max (%) heart rate reserve (%)	Maximal heart rate (%)	RPE†	Young (20–39)	Middle-aged (40–64)	Old (65–79)	Very old (80+)	RPE	Maximal voluntary contraction (%)
Very light	<25	<30	<9	<3.0	<2.5	<2.0	≤1.25	<10	<30
Light	25–44	30–49	9–10	3.0–4.7	2.5–4.4	2.0–3.5	1.26–2.2	10–11	30–49
Moderate	45–59	50–69	11–12	4.8–7.1	4.5–5.9	3.6–4.7	2.3–2.95	12–13	50–69
Hard	60–84	70–89	13–16	7.2–10.1	6.0–8.4	4.8–6.7	3.0–4.25	14–16	70–84
Very hard	≥85	≥90	>16	≥10.2	≥8.5	≥6.8	≥4.25	17–19	>85
Maximal‡	100	100	20	12.0	10.0	8.0	5.0	20	100

* Based on 8–12 repetitions for persons under age 50 years and 10–15 repetitions for persons aged 50 years and older.
† Borg rating of Relative Perceived Exertion 6–20 scale (Borg 1982).
‡ Maximal values are mean values achieved during maximal exercise by healthy adults. Absolute intensity (METs) values are approximate mean values for men. Mean values for women are approximately 1–2 METs lower than those for men.
U.S. Department of Health and Human Services. Physical Activity and Health: A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, 1996.

Activity diaries

Diaries can be used to record all physical activity over a period of a few days. A summary index of physical activity can be derived from the record by summing the duration of each activity and multiplying by the estimated rate of energy expenditure for that activity. Record-keeping imposes a considerable burden on participants, and may temporarily influence physical activity. Because the duration of record-keeping is usually limited to a few days, diaries are unlikely to yield reliable rankings of individuals by levels of physical activity within a population, although they may be useful for measuring the average physical activity of groups. Activity diaries are most useful when combined with direct monitoring of physical activity or energy expenditure by devices such as motion sensors.

Recall questionnaires

Recall questionnaires are less likely to influence behaviour than activity diaries. Activities can be separated into four main contexts: work, transport (walking or cycling), everyday housework and leisure-time activity including sport. Questions about activity at work should distinguish different kinds of activity: standing, walking around, lifting loads or using equipment such as shovels. Specific questions about distances walked or cycled each weekday and during each weekend can be used to estimate energy expenditure during transport: walking for 1 km on the level at 3–4 km/hour typically expends about 30 kcal more than driving the same distance. Questions about leisure-time activity should cover the frequency and duration of participation in sport, and other leisure-time pursuits such as dancing or gardening.

A single question—asking whether, at least once a week, the participant exercises sufficiently to work up a sweat—has been found to predict measurements of cardiorespiratory fitness as well as a more lengthy questionnaire (*33*). In hot climates, where even slight exertion causes sweating, this measurement may be less discriminatory.

Questions about physical *inactivity* may be more informative than questions about physical activity. Thus, questions about time spent watching television, afternoon napping or using the computer can show stronger associations with obesity than questions about physical activity (*34*, *35*). Simple questions about inactivity can include number of hours spent watching television or video each week, number of hours spent reading and proportion of time spent sitting while at work. Examples of questionnaires used to measure physical activity are shown in Appendices 26 and 27.

Direct recording of physical activity and energy expenditure

Various monitoring devices have been developed to measure physical activity by direct recording.

Motion sensors. Electronic motion sensors generally perform better than mechanical sensors and have been validated against physical activity records completed over one year (*36*). Pedometers count steps and thus measure distance walked; accelerometers track motion and are more expensive and less easily used to derive estimates of energy expenditure.

Heart-rate monitors. Heart-rate monitors have been used to measure energy expenditure by calibrating the relationship of heart rate to energy expenditure for each participant in an exercise test, using indirect calorimetry (*37*). It may be necessary to continue heart-rate recordings for at least 5 days to obtain valid measurements of average energy expenditure. At present this technique remains a research tool, and further developments are necessary before it can be used to measure energy expenditure in population surveys. Heart-rate monitors must be fitted carefully if they are to measure heart-rate over a period of several days, and require more cooperation from the participant than actometers. Ambulatory heart-rate monitors capable of recording for 5 days at a time cost about US$ 500 each.

Double-labelled water. The reference standard for measurement of energy expenditure in free-living individuals is the double-labelled water method, in which the production of carbon dioxide is estimated from the rates at which stable isotopes of hydrogen and oxygen (^{2}H and ^{18}O) are replaced in the body water space. Unfortunately at a cost of at least US$ 500 per participant for the double-labelled water and the isotopic analyses, the method is prohibitively expensive for all but the smallest surveys. However, the method can be used to validate less expensive indirect methods for estimating energy expenditure.

Dietary intake measurements

Methods of dietary assessment have been reviewed in detail elsewhere (38–41); all current methods have limitations. Until recently, evaluation of dietary assessment instruments was generally based on test–retest reliability or agreement with weighed records as a reference standard. Now, it has become clear that all methods that depend on participants' own reporting are liable to bias. Studies using biomarkers to measure energy expenditure and nitrogen excretion have shown that under-reporting of dietary intake is common (*42*), and that this under-reporting is not randomly distributed but tends to occur systematically in certain groups such as obese individuals (*42, 43*). In random population samples, energy intake by weighed records is typically only about 80% of recorded energy expenditure. Intake of fat and sugar intake is under-reported more than that of micronutrients such as vitamin C: in other words, under-reporters report what they think they should be eating. Under-reporting appears to have a less serious effect on measurements of fat intake expressed as percentage of total energy than on absolute measurements of fat intakes.

One way to identify incomplete records is to reject all those in which the energy intake estimated from the dietary records is less than 1.2 times the predicted basal

metabolic rate, based on the assumption that average energy requirement is 1.6 times the basal metabolic rate predicted from body weight and height. Published equations for predicting basal metabolic rate from age, sex and body weight are described in Table 9.2 (*44*).

Exclusion of those who under-report, however, may introduce further bias: weight-conscious individuals are the most likely to underestimate their intake. Questionnaires that ask about dieting or "restrained eating" (*45*) may be useful in identifying such individuals, and more generally in studying the extent to which people rely on voluntary actions to control their energy intake.

An alternative means of validating diet records and identifying under-reporting is to measure 24-hour urinary nitrogen excretion. This method relies on the assumption of nitrogen balance in the participant. To enable the completeness of 24-hour urine collection to be checked, *p*-amino-benzoic acid (PABA) should be given as a marker (see page 171).

In calculating nutrient intake from food records, food tables are used to assign the appropriate nutrient values. However, the nutrient composition of foods in the tables may be missing or inaccurate; therefore only a few examples of those which are comprehensive, up-to-date and accurate are available in industrialized countries (*46*). Where no code exists, it may be necessary to choose the nearest equivalent. When records are coded in duplicate by two observers, inter-coder variation may be sufficient to produce quite large differences in variables such as the fatty acid composition of the diet (*40*). If no food tables are available for the population under study, tables produced for some other population must be used, supplemented if necessary with data for specific foods derived by chemical analyses or calculation from recipes. Some nutrients such as sodium are very difficult to assess using food tables.

One way to overcome the limitations of dietary intake measurements is to use biomarkers—for instance, measuring levels of a vitamin in plasma or red cells, rather

Table 9.2. Predicting basal metabolic rate from age, sex and body weight

	Predicted basal metabolic rate (kJ/min) as a function of body weight (*W*) in kg	
Age group (years)	**Men**	**Women**
18–30	$0.063W + 2.896$	$0.062W + 2.036$
30–60	$0.048W + 3.653$	$0.034W + 3.538$
Over 60	$0.049W + 2.459$	$0.038W + 2.755$

Source: Schofield WN. Predicating basal metabolic rate, new standards and review of previous work. *Human Nutrition—Clinical* Nutrition, 1985; 39:5–41.

than measuring intake. This has the advantage of being free from recall bias, and there may also be advantages to using a direct measure of body stores rather than relying on dietary intake data which depend on bioavailability. Urinary excretion measurement is established as the standard technique for sodium and potassium. Unfortunately no biomarker for total dietary fat intake exists, although analysis of the fatty acid composition of plasma lipids, erythrocyte membranes or subcutaneous adipose tissue provides a useful marker of the dietary intake of fatty acids that are not synthesized in the human body (*47*) (which include linoleic acid and *trans*-fatty acids produced in the food industry). Measurement of *n*-3 fatty acids such as eicosapentaenoic acid in plasma lipids is also a useful guide to dietary *n*-3 fatty acid intake, mainly from marine sources. Measurements on adipose tissue biopsies reflect the fatty acid composition of the diet over several months or years; thus they are less likely than plasma measurements to be affected by recent changes in diet (e.g. in people diagnosed with heart disease). However, all laboratory measurements of fatty acids are complex and costly and may exhibit significant variability.

24-hour recall

Single 24-hour recalls of dietary intake are widely used in epidemiological surveys as they provide a good measure of recent intake. Studies using 7-day diet records have shown, however, that a single 24-hour recording covers far too short a period to measure habitual intake of individuals with diverse food intake in industrialized countries. On the other hand, 24-hour recordings on a large group of participants are an efficient way to measure the average dietary intake of a group. A standardized approach to a dietary recall is included in Appendix 28. To rank participants within a population by habitual intake, at least 10 records should be collected on separate days (*48*, *49*).

Diet history

The diet history is a detailed interview from which the participant's habitual diet is assessed. In an open-ended interview, the interviewer attempts to construct a typical week's food intake. Each meal is examined and each period between meals to construct a 7-day pattern of expenditure. Quantities can be assessed in terms of familiar units such as cups of milk, or with the aid of models or photographs showing a choice of portion sizes. On balance, the diet history tends to give higher estimates of intake than diet records, but agreement with measured energy expenditure is no better than with diet records (*43*). The diet history requires a highly-trained interviewer, and is especially liable to bias because of the reliance on interviewing technique and participants' recall. The technique is difficult to standardize, and the effects of participants' perception of their food intake on their reporting are not well understood.

Food frequency questionnaires

In recent years, food frequency questionnaires have been used widely for large-scale population surveys (*50*). The questionnaires are described as "semi-quantitative", meaning that precoded responses on usual portion size and frequency of consumption are converted into estimates of average daily consumption. For studies of occupational groups that are skilled in filling out forms, it may be possible to use self-administered questionnaires and automated scoring. This makes it practicable to study large populations cheaply. Food frequency questionnaires have obvious attractions for field surveys: they do not require trained field staff or lengthy interviews, standardization is not a problem, and the method is not onerous to participants.

The accuracy of food frequency questionnaires depends on the dietary variable under study. In the development of a food frequency questionnaire, it is possible to define food items that can be omitted or grouped together with similar foods, without loss of accuracy in the estimation of intake of a specific food or nutrient. If, for instance, the objective is to measure the fatty acid composition of the diet, detailed questions about the types of cooking oil and the brands of margarine used may be appropriate. If the objective is to measure total fat intake, these distinctions are less relevant. Food frequency questionnaires are thus most likely to be accurate in measuring nutrients that they were specifically developed to measure, and least likely to be accurate in measuring nutrients that were not evaluated at the time the questionnaire was developed (*51*). In studies of health professional groups, the correlations of estimates of total fat intake and saturated fat intake derived from food frequency questionnaires with estimates of the same variables derived from repeated diet records are in the range 0.5–0.7 after adjustment for total energy intake (*52*). The correlations between repeated food frequency questionnaire measurements 1 year apart are similar. For nutrients measured by biomarkers, such as vitamin E in plasma or *trans*-fatty acids in adipose tissue biopsies, the partial correlations of intake estimated from food frequency questionnaires with biomarker measurements are in the range 0.2–0.4. Compared with diet records, estimates of food intake derived from food frequency questionnaires are biased towards the dietary patterns that are considered desirable; thus consumption of fruits and vegetables is over-reported, while meat, snacks and dairy products are under-reported.

In general, food frequency questionnaires are not sufficiently accurate to detect associations between disease and variables such as vitamin E intake within a homogeneous population in which most people eat a similar diet. They are, however, accurate for measuring gross differences in the type of food consumed—for instance in distinguishing vegetarians from non-vegetarians and identifying people who are at the extremes for dietary intake of specific foods. They are also dependent on accurate food composition tables. Where food frequency questionnaires are used in a study, their ability to assess intake of the specific nutrients concerned should, if possible, be validated against diet recalls, records or biomarkers in a sub-sample (*53*, *54*). An example of a questionnaire is shown in Appendix 29.

Diet records

In contrast to the diet history, standardization is easier for diet records than for other methods, as the objective is a complete record of food intake on specified days. Diet record methods may be classified into three types: precise weighing, weighed intake and food diary. The precise weighing method, in which all ingredients of all recipes are weighed, as well as the portions consumed, is the most accurate but also the most laborious for both participants and research staff. Each record generally requires two or three home visits by a trained field worker, and close cooperation from all members of the household who are involved in preparing food. Precise weighing is the only method able to distinguish variation in dietary intake resulting from variation in the recipes used to prepare standard dishes. Weighed intake methods, in which portions are weighed but not the ingredients of individual recipes, are less laborious, and do not require the close cooperation of household members other than the participant. Use of portable scales with a built-in tape recorder simplifies the capture of data: participants place the food on the scales, then press a button and speak into the microphone to record a description while the weight of food is recorded automatically. This system can be used even with participants who are not literate.

For large surveys, the best compromise between accuracy and practicability may be a 7-day or 9-day food diary, in which food intake is recorded prospectively but portions are not weighed. Weights or portion sizes can be assessed by household measures, or by using photographs. If household measures are used, it is important to calibrate the measuring devices—the capacity of a mug or a spoon, for instance varies markedly. Records with estimated weights or estimated portion sizes do not show a systematic bias compared with weighed records.

From repeat measurements using diet records, it is estimated that classification of 80% of those participating into the correct tertile of the distribution at 95% confidence requires 7 days of records for dietary energy and nine days for dietary fat (*41*).

A serious source of error in studies using diet records is that the act of keeping a diet record, especially if it entails weighing portions, alters food habits. Over a 4-day period of recording weighed intakes, average weight losses of 0.3 kg to 0.6 kg have been reported (*55*, *56*), equivalent to average negative energy balance of 2000–4000 kcal. Changes in body fat stores of this magnitude are too small to detect reliably in individuals, but can be measured as group averages. An alternative to keeping diet records is for participants to prepare duplicate portions for analysis of composition in the laboratory. However, this is even more likely to affect participants' intake even than keeping diet records, as many participants are understandably reluctant to see food wasted (*57*).

Household food inventory

In the household food inventory, all food entering the home in a defined period—usually 1 week—is recorded by the person responsible for buying most of the food. This is then divided by the number of adults and children in the household to obtain the average intake of the household. The results are reported as group means, and can be stratified by the number of adults and children in the household so that comparisons of average intake are not confounded by household composition.

The household food inventory is relatively easy to record and standardize, and may be useful in surveys where the objective of dietary assessment is to calculate means for groups rather than individuals. Even for this purpose, however, it has several limitations. It does not include food eaten outside the home, which now constitutes a major portion of intake in some countries. Although separate questionnaires about food eaten outside the home during the recording period can be included, the method then has many of the disadvantages of using diet records. Another problem is that estimates of average consumption of any food items purchased in bulk or infrequently will be inaccurate unless a very large sample of families is used. The method assumes that all food purchased is actually eaten, and it fails to allow for wastage. With these limitations in mind, the household food inventory can be used to estimate the average diet of groups where almost all food is consumed within the home and bulk buying is uncommon. It is also possible to measure the consumption of specific foods: for instance, household consumption of cooking oils can be measured, if necessary, by measuring stocks at the beginning and end of the recording period as well as purchases during the recording period. Use of low-fat dairy products can be ascertained (*58*).

Food analysis

The most accurate method of determining nutrient content of foods is by chemical analysis. Duplicate portions of single items or whole meals are saved and analysed for the nutrients of interest. Accurate sampling and appropriate preparation and storage are important as some components may decrease before analysis. The method is very intrusive and costly but provides a "gold standard" for determining the amount and nutrients in the meals of interest. It is most easily used in cafeterias or other large feeding settings where composition of food items is constant and equal portions are served. The method cannot determine bioavailability, however, only what is in the food eaten.

Physical examination

Measurements based on physical examination alone are generally difficult to standardize accurately. Where quick and cheap measurements using special equipment are available, we recommend that they should supplant physical examination. For

example, electrocardiography is more accurate than clinical examination for detection and diagnosis of arrhythmias. Where reference methods for a measurement exist but are expensive, they can be used in a sub-sample to validate estimates based on less costly techniques—anthropometric measurements of body fat distribution can be validated against computed tomography.

Anthropometry

Anthropometry in cardiovascular surveys has three main uses: to standardize for body size, to estimate body composition as defined by percentage of body fat, and to measure the distribution of body fat. Estimation of body composition is discussed later (see pages 155–157). To standardize for body size, body surface area is sometimes more appropriate than height or weight. Body surface area (A) in square metres can be estimated from height (H) in metres and weight (W) in kg according to the method of Dubois as: $A = 0.2026 W^{0.426} H^{0.726}$. Others use body mass index (BMI) to standardize body size. It is calculated as BMI; weight in kg divided by square of height in metres.

Other formulae based on weight and calf circumference have also been developed (*59*).

The use of anthropometry to measure body composition is discussed later. Body fat distribution is measured by girths and by skinfolds. Waist girth correlates well with the intra-abdominal (visceral) fat mass measured by computed tomography (CT) (*60*). The sagittal abdominal diameter at the level of the fifth lumbar vertebra (L5) is an alternative measure, which appears to have slightly better correlation with CT-derived measurements of visceral fat than the waist girth.

These measurements are generally standardized by expressing them as ratios. Waist:hip girth ratio standardizes for body size, and also takes into account gluteal fat deposits. The correlation between waist:hip girth ratio and the proportion of total body fat sited intra-abdominally is typically about 0.6 (*60*). A disadvantage of the waist:hip ratio is that hip girth is not entirely satisfactory either as a measure of body size or as a measure of lower body fat stores. Waist:thigh ratio is less dependent on bone structure, and its correlation with the ratio of intra-abdominal to extra-abdominal fat appears to be as high as that for waist:hip ratio (*60*). An alternative measure is the waist:height ratio, which appears to predict clinical outcomes better than waist:hip ratio. A suggested minimum for cardiovascular surveys is to measure height and weight, plus waist, hip and thigh girths.

Ultrasound measurement of the distance from the front of the aorta to the back of the abdominal wall muscles in the sagittal direction yields high correlation with visceral fat mass in lean individuals (*61, 62*). Unfortunately it is difficult to detect clear echoes from these structures in obese individuals, so that the technique

appears to be useful as a measure of visceral fat only in surveys of relatively lean individuals.

Where participants are being examined at a field station, they can be asked to remove all clothes, or all clothes except undergarments. A separate changing area should be provided to speed the process. Paper underpants can be supplied to standardize the thickness of undergarments included in the measurement of hip girth. Where participants are examined in their homes, it is usually necessary for the measurements to be performed without removal of clothes apart from heavy outer garments.

Height and weight

Standing height can be measured with a stadiometer, or with simpler equipment such as a measuring scale fixed to a vertical wall, using a right-angled block to determine the level of the top of the head. Height should be measured with the head in the Frankfort plane—in this position an imaginary line passing through the external ear canal and across the top of the lower bone of the eye socket immediately under the eye is parallel to the floor (i.e. horizontal). The participant is instructed to "stand as tall as you can" while keeping the head tilted in this plane.

Measurement of sitting height as well as standing height may help to correct for differences in leg length.

Girth measurements

The following recommendations for technique are adapted from a WHO report (*63*). The measuring tape measure should be 1 cm wide and made of fibreglass or some other material that does not stretch. To standardize the tension in the tape, a small spring balance can be attached to one end and the observer instructed to pull the balance so that a specified reading (about 600 g) is shown on the balance (Appendix 30). This level can be marked on the scale.

Waist girth should be measured at a level halfway between the iliac crest and the costal margin in the mid-axillary line; the subject should be in the standing position.

Hip girth should also be measured with the subject in the standing position, with both feet together at the level of the greater trochanters. If the greater trochanters are not palpable, this level will usually be the level of the largest horizontal girth around the buttocks.

Thigh girth can be most easily measured with the subject's right foot resting on a chair or other support positioned so that both the hip and the knee are flexed at right angles. The subject is asked to twist to the left, and the tape is slid up to the crease

at the top of the thigh where it meets the perineum. An alternative is to measure thigh girth when the subject is in the standing position, but it is then difficult to position the tape at the top of the thigh.

"Sagittal abdominal diameter" is measured with the subject lying supine, using an anthropometer with one prong slid underneath the participant. The level of L5 corresponds to the level of the iliac crest in the mid-axillary line.

Skinfold thicknesses

Skinfolds should be measured at multiple sites, as measurements are quick to obtain and random error is less when data from multiple measurements are combined. Standard instruments such as the Holtain-Harpenden calipers generate a pressure of $10\,g/mm^2$ between the jaws which is independent of the separation between the jaws (Appendix 30). The choice of sites for skinfold measurements depends upon the objectives of the survey: if, for instance, the objective is to measure body fat distribution in women, it is important to include measurement sites on the lower limb. As a minimum, the following measuring sites are recommended:

- *Triceps.* Halfway between the tip of the acromion and the tip of the olecranon. The fold should be raised exactly over the middle of the triceps muscle, where the fat layer is thickest. The degree of rotation at the elbow should be specified.
- *Subscapular.* 2 cm below the tip of the scapula.
- *Supra-iliac.* 1 cm above the iliac crest in the mid-axillary line.
- *Anterior thigh.* on the front of the thigh, halfway between the mid-inguinal point and the upper border of the patella.
- *Other possible sites.* Biceps, suprapatellar, para-umbilical, mid-calf.

The orientation of each skinfold (horizontal, vertical, oblique) should be specified in the protocol. In general, the orientation should be chosen to align with the underlying muscle and fascia; thus, the subscapular skinfold is easiest to raise if it is angled obliquely downwards and laterally. The sites of each skinfold measurement should first be identified and marked on the skin. Holding the calipers in the right hand, a skinfold is lifted between the left thumb and forefinger and the calipers are applied without releasing the fold so that there is no tension on the skin underneath the jaws. Calipers should be applied to the thinnest part of the skinfold—not the base or the crest. The thickness reading tends to fall in the first few seconds after the calipers are released. The interval between releasing the calipers and taking a reading should be standardized; an interval of 5 seconds has been found to correlate best with radiological estimates of subcutaneous fat thickness. Measurements can be recorded to the nearest 0.1 mm. Marked variation between observers is likely unless they are frequently trained and retrained against one experienced observer who acts as the reference standard. With attention to technique and training of observers, most values will be reproducible to within 10%.

Clinical examination of the heart

Physical examination of the heart by palpation and auscultation has been widely used in epidemiological surveys to screen for rheumatic or congenital heart disease. As an epidemiological tool for examination of population samples, physical examination of the heart has several drawbacks:

— it requires experienced physicians
— even with skilled observers, the reliability of the findings and the agreement between observers is generally poor
— unnecessary anxiety may be caused to healthy participants when the detection of a murmur necessitates referral.

In a recent study using transthoracic and transoesophageal ultrasound as the reference standard, physical examination by an experienced cardiologist was estimated to have a sensitivity of 70% and a specificity of 98% for the detection of valvular heart disease (*64*). Although this level of accuracy is excellent for clinical practice, where a high proportion of referred patients have disease, it is inadequate for research surveys. In a population with 2% prevalence of valvular heart disease, more than half the participants found to have signs of disease at clinical examination would be disease-free. It is recommended that auscultation should be included only in surveys in which the prevalence of rheumatic or congenital heart disease is an outcome of interest. Where auscultation is used as part of a field survey it should be supplemented with ultrasound examination of the heart so that participants do not have to make their own arrangements for ultrasound examination.

A simple protocol for clinical examination of the heart will include palpation of the carotid pulse and precordium, and auscultation at three sites: the apex in expiration in the left lateral position for murmurs associated with mitral valve dysfunction; during expiration in the fourth intercostal space at the left sternal edge with the subject leaning forward for an aortic early diastolic murmur; in the second intercostal space at the right sternal edge for an aortic ejection systolic murmur, after isometric handgrip exercise to raise the heart rate to approximately 100/minute.

Criteria for distinguishing functional murmurs from signs of valve disease should be defined in advance: intensity, site where the murmur is best heard, radiation to other sites, and effects of manoeuvres such as exercise or change in posture.

Scoring sheets should record the presence or absence of murmurs at each of these three sites, as well as their timing and pitch (Appendix 31).

Examination of retinal blood vessels

Direct fundoscopy with an ophthalmoscope has been replaced in epidemiological studies by fundus photography, which does not require mydriasis and allows assess-

ments to be made under standardized conditions away from the field site. Hypertensive changes are classified according to the criteria of Keith et al. (*65*). Observer agreement for scoring of hypertensive changes in the retinal blood vessels is poor. Epidemiological surveys have demonstrated a modest association between blood pressure and hypertensive retinal changes in the population (*66*), but the association is probably not strong enough for this to be a useful measure of hypertensive end-organ damage. Inclusion of fundus photography in cardiovascular field surveys is not recommended unless eye disease or diabetic complications are key conditions of interest. Where fundus photographs are obtained, corrected visual acuity should also be measured.

Blood pressure measurement

The measurement of blood pressure by sphygmomanometry is central to CVD surveys but is not always straightforward. Resting blood pressure fluctuates from beat to beat and depends on the environment, acute participant factors and the effect of the examiner's presence. Occasionally participants have a vasovagal faint when the cuff is first applied. Observers tend to record certain terminal digits more frequently than others, and generally have a bias towards recording levels that are just below the threshold requiring intervention. Observers may use other cues, such as the oscillation of the mercury column, rather than Korotkoff sounds alone.

There has been controversy over the merits of using ordinary mercury sphygmomanometers, random-zero sphygmomanometers and automated instruments. The decision about what to use depends on several issues. Physiologists require instruments that yield measures as close as possible to the reference standard of intra-arterial recording. This is not necessarily desirable in a field survey; if the purpose of the survey is to identify what levels of blood pressure should be treated, it is preferable to use a method that gives readings as close as possible to those obtained with the method recommended for use in clinical practice (usually mercury sphygmomanometry), even if this method is biased in relation to measurements made by intra-arterial recording. On the other hand, if a key objective of the survey is to make comparisons with other surveys that have used a random-zero sphygmomanometer or an automated device, it is desirable to use the same method, even if the measurements are biased in relation to those obtained in clinical practice with an ordinary mercury sphygmomanometer.

The greatest advantage of using a mercury sphygmomanometer is that readings are directly comparable with those made in clinical practice. It is inexpensive, portable and widely available, and can be used on all respondents for whom Korotkoff sounds are audible.

The main disadvantage of the mercury sphygmomanometer is the potential for observer bias and observer error. It is often difficult to eradicate incorrect technique

that observers have learned in their clinical practice. Observers must be trained in auscultation, in correct release of the valve so that the column falls at 2 mm/s, in avoiding bias, and in avoiding reliance on fluctuation in the mercury column rather than sound to designate systolic blood pressure. Training requires a high trainer-to-trainee ratio, and it is difficult to ensure that good technique observed in a training session is actually practised in the field. Retraining at fairly frequent intervals is usually necessary. Data quality control checks can identify terminal digit preference. Evaluation of mean values of individual technicians can identify others (see Chapter 6). The mercury sphygmomanometer should be cleaned every six months if it is in regular use.

Another disadvantage of the mercury sphygmomanometer is that it requires a quiet room, and must be used on a level surface at the correct height. During a survey in which measurements are made in participants' homes, these conditions are often unachievable.

A third disadvantage is that leakage of mercury poses a health hazard. Mercury must be removed from sphygmomanometers before they are shipped by air. At present a ban on the use of mercury in sphygmomanometers is under discussion in the European Union.

Automated oscillometric devices are easy to use, require less training of observers, have less potential for observer bias, and give reliable results even when used in environments not conducive to good technician practice. All available instruments, however, yield results that are systematically different from those obtained with a mercury sphygmomanometer. Another concern is that models are continually updated, and there is no guarantee that a particular model will still be available after a few years. This is a serious disadvantage if, for instance, the objective of the survey programme is to monitor long-term changes in the blood pressure of the population. There can be significant differences between different automated devices or even within a specific type.

If subjects are to be examined in their homes, and fieldwork is being undertaken by a scattered team of field workers who cannot be closely supervised, use of an automatic sphygmomanometer is more likely to give reliable and consistent results than reliance on auscultation with mercury sphygmomanometers. It is recommended that any instrument used for blood pressure measurement should be calibrated regularly beforehand against a standard mercury sphygmomanometer.

The choice of instrument should be based on questions listed in Table 9.3.

Table 9.3. Queries influencing the selection of blood pressure devices

Has the instrument been validated against an appropriate reference standard (intra-arterial pressure recordings, mercury sphygmomanometry)?
Is there a systematic difference between the measurements obtained with the instrument and those obtained by other techniques, and has this difference been quantified?
Are the readings obtained repeatable?
Does the instrument depend heavily on the observer's training, reliability and lack of bias?
Does performance drift?
Is the instrument portable and does carrying it affect its calibration?
Is the device likely to be available throughout the period covered by the survey programme?
Are any hazards associated with its use?
Can the instrument be used on all respondents?
What requirements are there for maintenance and calibration of the instrument: how often and by whom?
What is the cost of using the equipment, including maintenance, calibration and training of observers?

The following general recommendations for blood pressure measurement protocols apply to both mercury sphygmomanometers and automated equipment.

- Before blood pressure measurement, a cuff of the correct size should be fitted and the participant should rest quietly for 5 minutes, sitting on a chair with the arm resting on a table (*67*).
- Many other factors that affect blood pressure measurement should be considered, including time of day, temperature of the room, alcohol and caffeine consumption, and tobacco use (*68*).
- The cuff should be fitted so that the centre of the bladder is in line with the brachial artery, and the cuff should be large enough to encompass two-thirds of the circumference of the arm. Several cuffs, from paediatric to thigh sizes, should be available for the survey to ensure proper placement. Measurement of arm circumference facilitates proper cuff use. Participants should not cross or uncross their legs during the resting period or the measurement, as this alters venous return and cardiac output (*69*, *70*) (Appendices 5 and 32).

Measurement using a mercury sphygmomanometer

In most surveys in which blood pressure is to be measured by a few highly-trained observers under controlled conditions in a quiet room, mercury sphygmomanometers are used and the systolic and diastolic pressures are determined by auscultation. The approximate pressure at which the radial pulse is obliterated (pulse obliteration pressure) should first be measured by inflating the cuff rapidly while palpating the radial pulse. This makes it unnecessary to inflate the cuff to a very high pressure in every individual. The required peak inflation pressure should be at least 20 mmHg above the pulse obliteration pressure. After inflating to peak pressure, the mercury column should be allowed to fall at 2 mm/s.

The systolic reading is taken as the mercury level at which the first sound (Korotkoff I) in a succession of sounds is heard. The diastolic pressure (Korotkoff V) is taken as the first level at which there is an absence of an expected sound. These definitions can also be used to measure blood pressure in a participant with marked pulsus paradoxus (fall of blood pressure during inspiration).

Measurement using a random-zero mercury sphygmomanometer

The random-zero mercury sphygmomanometer has been widely used in epidemiological surveys. Before a reading is taken, a wheel on the instrument is spun two or three times to set the level of mercury in the reservoir to a value that is unknown to the observer. According to the model used, the level in the reservoir can vary between 0–20 and 0–40 or 0–60 mmHg. After the observer has recorded the systolic and diastolic blood pressure readings, the cuff is released to determine the mercury level in the reservoir. Subtraction of the level in the reservoir from the systolic and diastolic readings yields the true systolic and diastolic readings. This procedure helps to reduce digit preference and makes it harder for the observer's expectation of the value to bias the result; however, it does not fully eliminate the possibility that the readings will be biased by the observer's expectations. The Hawksley instrument has been shown to give values for systolic and diastolic pressures 2–3 mmHg (267–400 Pa) lower than those obtained with an ordinary mercury sphygmomanometer (*70–72*). Larger discrepancies have been attributed to two defects in technique: failure to allow the reservoir to fill completely after inflating the cuff before closing the tap, and failure to allow the level to settle fully before taking the zero reading. The instrument has recently been modified by the manufacturer, but published evaluations of this modified instrument are not yet available. When using the random-zero instrument, the observer should wait 5 seconds after inflating the cuff before closing the tap, and should tap the machine to ensure that the mercury has settled before taking a zero reading. Regular servicing of the instrument and cleaning of the mercury according to the manufacturer's recommendations are essential.

Whether a random-zero or a standard mercury sphygmomanometer is used, training of observers is essential. Observers should practise controlling the valve so as to maintain the rate of fall of the mercury column at approximately 2 mm/s; this can be done with the cuff wrapped around a bottle or other rigid object. Video tapes of falling mercury columns accompanied by a sound track can be used to give a realistic simulation of blood pressure measurement (*73*).

Another method of standardizing observers is to use audio tapes of Korotkoff sounds; observers record the time of systolic and diastolic pressures with a stopwatch (*68*). However, this does not fully simulate the combination of visual and auditory inputs on which measurement of blood pressure with a mercury sphygmomanometer relies.

Measurement using an automatic sphygmomanometer

Many types of automatic sphygmomanometer are available. They differ in the extent of systematic differences between their recordings and those of the mercury sphygmomanometer. Calibration studies have already been performed on large samples of some instruments. In comparison with the mercury sphygmomanometer, oscillometric instruments typically underestimate or overestimate systolic and diastolic pressures by up to 5 mmHg (677 Pa) (*74*). Use of automated devices is easy and requires only about 1 hour of training. The readout is in digital form, and on some models includes pulse rate; no calculation is required and a standard interval between measurements is ensured. Most devices will run either on batteries or on mains electricity. Annual calibration by the supplier is usually required. The oscillometric measurements are not dependent on the detection of sounds, so background noise is not a problem. Some devices cannot reliably measure blood pressure in those with a tremor or atrial fibrillation.

It is important to emphasize that automated blood pressure readings require the same training and scrutiny in environment, participant preparation and cuff placement as manual machines.

Ambulatory blood pressure monitoring

Ambulatory blood pressure measurements have been widely used in research and in clinical practice to identify those at risk of hypertensive target-organ damage and therefore eligible for antihypertensive therapy. Ambulatory blood pressure measurements are more accurate than single resting blood pressure measurements but reproducibility is not markedly higher than for single measurements of resting blood pressure (*75*). Thus, to improve the reliability of blood pressure measurements, repeat measurements of resting blood pressure twice on different days are likely to be cheaper and more convenient than ambulatory measurements. The reproducibility of within-day fluctuation in blood pressure is poor, making it somewhat difficult to show correlations between blood pressure variability and target-organ damage. It currently has some clinical indications (e.g. white coat hypertension), but is still in the research and development phase for population studies. It should be stressed that ambulatory blood pressure measurement is a research tool only.

Ankle blood pressure recording

Measurements of ankle blood pressure can be used to detect peripheral arterial disease in field surveys. Blood pressure cuffs around the ankle and arm are used, with pulse sensors placed on digits or other sites below the cuff to measure resting blood pressure in the brachial artery and the ankle simultaneously. Measurements of ankle pressure 30–90 seconds after 1 minute of walking on a treadmill identify additional cases of peripheral arterial disease. A resting ankle–brachial blood pressure ratio of less than

0.9, or a fall in ankle blood pressure of 30 mmHg (400 KPa) or more in one or both legs is taken as evidence of peripheral arterial disease (*76*) (Appendix 33).

Measurement of body fat mass

Ideally, measurements of body fat mass include both total body fat mass, from which percentage body fat can be calculated, and the proportion of body fat contained in visceral (omental and mesenteric) depots. Methods for measuring body composition and body fat distribution have been reviewed in detail elsewhere (*77–79*).

Body mass index

Body mass index is widely used as a simple measure of obesity. In general population samples, the correlation between BMI and percentage body fat measured by more direct techniques is generally between 0.6 and 0.8. BMI is not necessarily valid when comparing obesity between populations with different average body frame size and lean tissue mass (*80*); a population with high average body mass index is not necessarily more obese than a population with lower average body mass index. In studies of comparisons between populations it is recommended that BMI is used with other indicators.

Skinfold measurements

Skinfold measurements can be used as measures of subcutaneous fat, or in a regression formula to total body fat. Many equations are available for prediction of percentage body fat from skinfold measurements, but it is uncertain whether these are valid outside the circumstances in which they were derived. Skinfold measurements are quick and easy to obtain, but difficult to standardize for differences between observers and drift of measurement technique within observers. In some participants it is difficult to raise a skinfold because the skin and superficial fascia are tightly tethered to deeper tissues. Skinfold thicknesses can measure only the subcutaneous fat depot, which is not always a guide to total body fat stores (Appendix 30).

Subcutaneous ultrasound

Subcutaneous ultrasound (A-scan mode) has not been shown to be more accurate than skinfold thicknesses for prediction of percentage body fat. Measurements of the thickness of the subcutaneous fat layer are easy to obtain on the limbs, but more difficult to obtain on the trunk where there is no deep fascia to give a clear echo. Ultrasound measurements take more time than skinfold measurements.

Total body water

One approach to estimating percentage body fat is to measure total body water. On the assumption that water comprises 73.2% of fat-free mass, fat mass can be calculated if body weight and total body water are known. Total body water can be measured by deuterium oxide (heavy water) dilution. These measurements are easy to perform, but the costs of measuring deuterium by mass spectrometry or infrared absorption are high.

Body electrical impedance

Electrical impedance measurements can be used to estimate total body water, or to predict percentage body fat directly (*81*). The principle of the technique is that the intra- and extracellular fluid compartments account for almost all the electrical conductance of the body, and whole body impedance is thus directly related to total body water (and inversely to percentage body fat), after standardizing for body size. Devices should give a reading of impedance directly, rather than only an estimate of percentage body fat derived from a built-in prediction formula. The instruments are relatively inexpensive (about US$ 500), and the measurement takes only a few minutes. The tetrapolar method minimizes the impedance of the skin contacts. Standardizing the placement of electrodes is important for the reliability of measurements. Electrodes are placed on the dorsal surfaces of the hand and foot, and the ventral surfaces of the wrist and ankle, usually on the right side of the body. The participant lies supine on a non-conducting surface with arms and legs slightly abducted.

Body impedance measurements are reliable both within individuals and between observers. Measurements are affected by hydration status, recent strenuous exercise and environmental temperature. Measurements of impedance 2–4 hours after a meal can be 3–4% lower than those made at other times. Instruments should be calibrated periodically using resistors.

Impedance measurements have been shown to improve the prediction of percentage body fat obtained by using BMI alone (*82*). However, it is not yet clear whether equations used to predict total body water from impedance measurements are valid when applied in different populations from those in which they were originally developed (*81*). Equations for prediction of body fat mass from body impedance should therefore be derived specifically for the population under study, using some more direct method of measurement of total body water or percentage body fat as a reference standard in a separate validation study.

Dual-energy X-ray absorptiometry (DEXA)

DEXA is one of the most accurate techniques available for measurement of body composition (coefficient of variation for fat mass about 2%), and can also yield mea-

sures of body fat distribution. An X-ray source that emits photons at two discrete peaks of energy is mounted beneath a table, opposite a scintillation detector above the table. The source and detector are passed across the body while attenuation of the beam is measured. The fat content of soft tissue is calculated from the ratio of attenuation of low-energy photons to that of high-energy photons. A scan takes approximately 20 minutes, and the radiation dose is less than that received during a long-distance air flight. The main disadvantage is the high cost of the equipment—about US$ 10 000. There are also some uncertainties about standardization of the machines and the built-in software used to calculate percentage body fat from the absorption measurements. At present the equipment is still too expensive for use outside specially-equipped centres.

Underwater weight

The standard reference method for measuring body fat mass is densitometry. Underwater weight and residual lung volume are measured simultaneously, and residual lung volume is subtracted from the body volume calculated from weight underwater. This measurement is too complex and time-consuming for large population surveys. It requires a special water tank and other equipment. Techniques for measurement of body volume that do not require underwater weighing may eventually be developed.

Measurement of cardiorespiratory fitness

Cardiorespiratory fitness is inversely associated with several cardiovascular risk factors (*32*). It is usually assessed by the maximal oxygen uptake (VO_2max) which can be attained by an individual. This can be measured directly in a maximal exercise test, in which the participant walks or runs on a treadmill the speed and slope of which are gradually increased according to a predefined protocol, while oxygen uptake is measured continuously. As the workload increases, the oxygen uptake rises then reaches a plateau as the participant's work relies on anaerobic metabolism.

Instead of being measured directly, VO_2max can be predicted from a sub-maximal exercise test, in which oxygen uptake is measured while the workload is increased steadily until heart rate reaches a specified proportion of the predicted maximal heart rate of a participant of this age. The VO_2 value is then estimated by extrapolation to the predicted maximal heart rate, or from a standard nomogram. In a modification of this procedure, oxygen uptake can be estimated from the work output of the participant, so that expensive measuring equipment uptake is not required. This estimate depends on assumptions about the efficiency of the participant (energy expended/work output) and the respiratory quotient. The estimate of efficiency is more accurate if a cycle ergometer is used rather than a treadmill, provided that participants are familiar with pedalling a cycle.

The advantages and limitations of submaximal versus maximal oxygen uptake measurements are discussed elsewhere (*83*, *84*). Maximal uptake tests yield highly reproducible estimates of VO_2max but must be undertaken under direct medical supervision; although deaths are extremely rare in apparently healthy participants undergoing a maximal exercise test, a serious medical complication rate of 6 per 70 000 tests has been reported. Another problem is that some individuals who are usually sedentary will be unable to achieve their maximal oxygen uptake or unable to complete the procedure, or will find it exhausting. Submaximal tests on healthy participants do not generally require direct medical supervision; in community surveys of 130 000 adults, no deaths or serious medical complications have been recorded. However, performance of this test does require well-trained personnel and clear protocols to ensure safety. Moreover, submaximal tests have poor reproducibility within individuals, although they may be useful in measuring the fitness of groups (*84*, *85*).

To equip a field laboratory for direct measurements of VO_2max will usually cost at least US$ 40 000, with oxygen analysis equipment being the largest single item of expenditure. Participants must be familiarized beforehand with the use of equipment such as mouthpieces and nose clips, as using these for the first time is likely to provoke tachycardia and increased oxygen consumption. In contrast, a simple submaximal fitness test, using a cycle ergometer and a heart rate monitor can be undertaken with equipment costing less than US$ 1000.

Where physical fitness measurements are to be included in a cardiovascular field survey, it is recommended that a simple submaximal protocol using a cycle ergometer and measuring the heart rate response to exercise be used. Detailed protocols are described elsewhere (*86*).

Before fitness testing, an initial questionnaire screening for pre-existing CVD should be carried out. Beta-blockade is an obvious contraindication to testing. The protocol is selected according to the participant's weight, age and activity status as determined by questionnaire. A constant-workload cycle ergometer, in which the resistance varies with pedal rate to ensure a constant work output, is easiest to use. The protocol begins with a warm-up phase in which the participant pedals on the ergometer with no load for 2 minutes. At regular intervals (1 or 2 minutes) the workload is increased. At the end of each interval the pulse rate is recorded. The test is terminated when heart rate reaches 86% of the predicted age-adjusted maximal value. Angina, cardiac arrhythmias, fall in blood pressure or signs of cardiorespiratory insufficiency are indications for stopping the test before this stage is reached. A plot of heart rate against workload will usually give an approximately straight line. Maximal oxygen uptake can be estimated by extrapolation to the predicted maximal heart rate or from equations previously validated against maximal exercise testing. Where predictive equations are unavailable for the population under study, indirect estimates of maximal oxygen uptake can be validated with a subsample of participants in a well-equipped laboratory (*87*). Protocols for treadmills are included in Appendix 34.

Electrocardiography

Recording of the electrocardiogram

The recommended procedure for recording a resting ECG, and the technical requirements for a suitable electrocardiograph, are described in detail in the reference manual for the Minnesota code (*88*). Although an automated, multichannel electrocardiograph is quickest, a single-channel nonautomated machine can be used if care is taken to label the tracings corresponding to each lead correctly. The participant should lie supine and flat on a couch wide enough to allow the arms to rest comfortably clear of the trunk. The position of the chest electrodes should be measured and marked as follows:

end-point on the mid-line of the sternum, level with the 6th interspace

— V1 4th interspace, right sternal edge
— V2 4th interspace, left sternal edge
— V6 at the level of the E point in the mid-axillary line
— V4 halfway from V6 to the E point
— V5 halfway between V6 and V4
— V3 halfway between V2 and V4

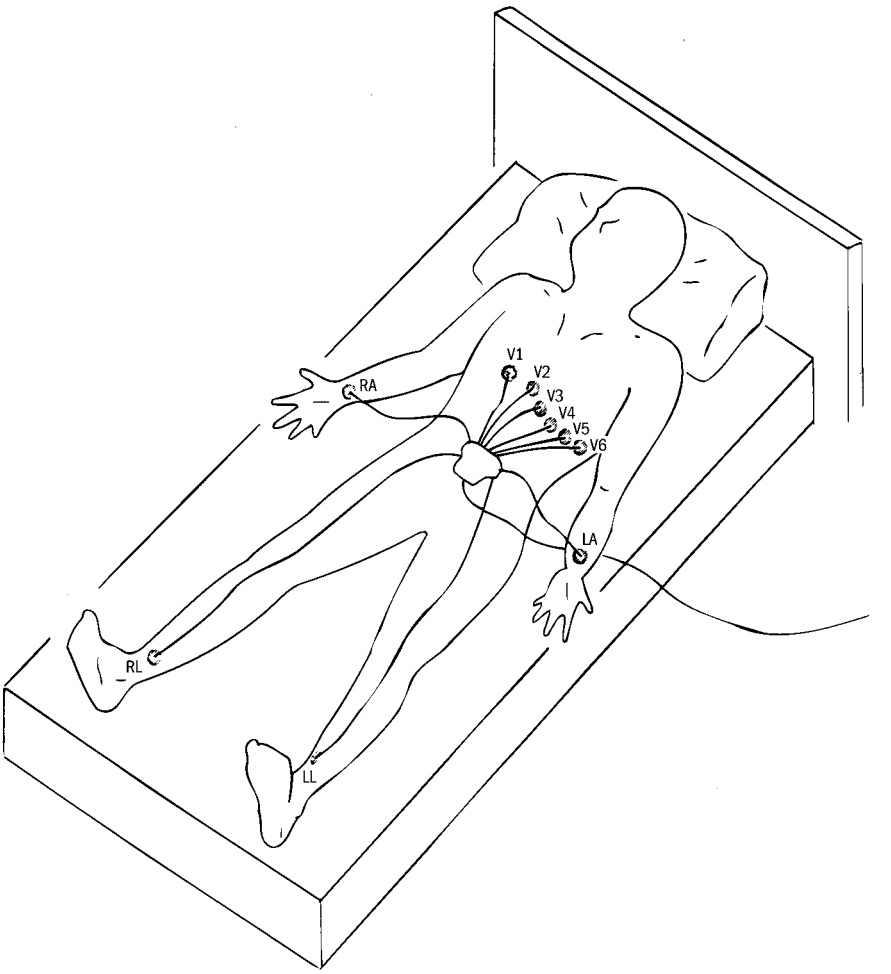
Each lead should be recorded for at least 5 seconds, so that a minimum of four complete beats are on the trace—this differs from clinical practice in which a recording period of 2.6 seconds is usual. No clinical details should be included on the recording strip, as coding should be performed blind—only the participant's study number, the date and the time should be included.

Coding of the ECG

Current computer programs for Minnesota coding are in development and show promise (*89*). Manual coding by two coders working in duplicate remains the reference standard (*88*). This is not a job for cardiologists, who are likely to find it difficult to follow a predetermined scoring scheme that does not allow them to use their clinical judgement. It is possible to train any observer who has a good basic technical education to undertake Minnesota coding within a few weeks. When the prevalence of abnormalities is low, as it usually is in a representative population sample of working age, experienced coders can code 10 tracings per hour.

Minnesota-coded major Q waves (codes 1–1 or 1–2) are recommended as the standard criterion for CHD in prevalence surveys on general population samples. Some measure of the validity of this criterion is available from comparison of ECGs recorded shortly before death with autopsy evidence of old myocardial infarction (*90*). It is possible that the specificity of Q waves is poorer in some regions of the

Fig. 9.1. Proper electrode placement of limb and chest leads



world where non-ischaemic cardiomyopathies are common, but there is no evidence of this from any study so far reported.

Use of more nonspecific criteria, including S-T and T wave changes, as a measure of prevalence in epidemiological surveys is to be discouraged. In women especially, these ECG signs have poor specificity for IHD.

The choice of major Q waves as the main criterion for prevalent CHD has implications for sample size. Typically, in a high-risk population such as the United

Kingdom, about 3% of adult men aged 40–59 years will be found to have major Q waves on ECG. This necessitates a sample of about 1000 men in this age group to obtain an estimate of prevalence with a standard error of 0.5%. A common error in prevalence surveys is to choose too small a sample, or too young an age group, to obtain a useful estimate of the prevalence of major Q waves.

The use of exercise ECGs in population studies is sometimes advocated as a solution to the poor sensitivity of the resting ECG. However the specificity of the exercise ECG is too poor for it to be useful. In a large study of patients referred for angiography, the specificity of the exercise ECG in identifying those with coronary artery disease at angiography was 74% (*91*). It can thus be estimated that if the prevalence of coronary artery disease in the group under study is 5%, the predictive value of a positive exercise ECG will be only 16%.

Ultrasound of the heart

Ultrasound has only limited use in the detection of IHD, but is widely used to measure hypertensive changes in the heart and to study peripheral arterial disease. New techniques are being rapidly developed, and it is possible that ultrasound may eventually find wider use in the non-invasive detection of coronary artery disease. Improved techniques for measurement of left ventricular function and cardiac output by ultrasound may make it possible to study the occurrence of heart failure in the general population. The use of these techniques in epidemiological studies is still at an early stage, and few validation studies are available.

Attention to the technical details of the procedure helps to maintain reproducibility (*92*). The left decubitus position is recommended as this improves imaging of the heart from the parasternal window. Wedges of firm foam made to support participants at specified angles are useful. Tissue phantoms are available that permit calibration of the equipment in both horizontal and lateral axes.

M-mode echocardiography

M-mode echocardiography transmits and receives ultrasound pulses on a single scan line, and displays the echoes on a graph of depth against time. M-mode echocardiography is used to measure cardiac dimensions and to time events within the heart. In cardiovascular field surveys its main application is in the measurement of hypertensive changes in the heart. Cardiomyopathy and valve regurgitation can also be detected.

Echocardiograms are obtained with the participant lying in a partial left decubitus position with the head at 30% from the horizontal. Left ventricular measurements are made using at end diastole in a parasternal short axis view at the mitral leaflet tips. The echocardiographic beam should be oriented to pass perpendicularly through

the interventricular septum and the posterolateral left ventricular wall. This can be defined by independent M-mode echocardiography, or under guidance of two-dimensional echocardiography. The correct level is that which shows the largest left ventricular internal dimension at end diastole.

Measurable images are defined by the presence of dominant lines with correct motion representing interfaces for at least 0.1 seconds (5 mm at standard recording speed). A skilled echocardiographer can obtain measurable images in about 90% of participants examined.

Three measurements are obtained at end diastole for:

— Interventricular septal thickness (I_d)
— Posterior wall thickness (P_d)
— Left ventricular internal diameter (L_d)

According to the American Society of Echocardiography (*93*), end diastolic measurements are made at the onset of the QRS complex of the ECG. All measurements are made from one leading edge to another leading edge. Left ventricular mass (M) in gram (g) is then calculated (*94*). These calculations correlate well with left ventricular mass measured at autopsy (*95*).

Criteria for left ventricular hypertrophy are usually based on the left ventricular mass index, defined as left ventricular mass divided by body surface area. An alternative index of cardiac hypertrophy is the relative wall thickness, defined as $2P/L$. Cut-off levels should be derived specifically for the population under study; for instance, one can define cutoffs at the 95th or 97th centiles for normotensive men and women.

Assessment of left ventricular function by M-mode echocardiography

If the ventricle is contracting symmetrically, as in uncomplicated hypertension, the fractional shortening of the left ventricular internal diameter between end diastole and end systole can be calculated as a substitute for left ventricular ejection fraction. Measurements of left atrial dimension, obtained at end systole, have been found to predict stroke (*96*).

Although normal ranges for these measurements have been derived for United States populations, it is emphasized that the validity of these normal ranges in other populations has not been established. Where possible, normal ranges and cut-offs that are specifically derived for the population under study are recommended.

Two-dimensional echocardiography

Two-dimensional echocardiography transmits and receives ultrasound pulses along multiple scan lines to generate a cross-sectional image of the heart in the chosen plane. Typically, about 120 scan lines are transmitted over a 90-degree arc 20–30 times/second. Where M-mode echocardiography is used, two-dimensional echocardiography is useful in guiding. Its use in evaluating left ventricular function in epidemiological surveys has not yet been established. Left ventricular end-systolic and end-diastolic volumes, ejection fraction, and load can be calculated from measurements obtained during two-dimensional echocardiography by a variety of methods (*97*). Segmental wall motion abnormalities can also be detected by two-dimensional echocardiography.

Regional scoring is more accurate than global measures of left ventricular function. Each segment must be imaged separately. Abnormal regional wall motion and absence of normal systolic thickening are signs of myocardial infarction, but the sensitivity and specificity for myocardial ischaemia of these echocardiographic findings have not been determined, and the validity of scores for wall motion abnormalities has not been established in populations outside the United States.

Doppler ultrasound

Doppler ultrasound derives information on the velocity of blood flow from the frequency shift that occurs when sound is reflected by a moving object. Continuous-wave Doppler uses two crystals, one transmitting and the other receiving, to generate a graph of velocity against time in which the density of points on the display is related to the number of red cells moving at that velocity. Continuous-wave Doppler can be used to estimate the severity of valve stenosis. Its main limitation is that the flow signal cannot be accurately localized. Pulsed Doppler uses a single crystal to transmit ultrasound in pulses 30–40 cycles long and then receive after a preset time delay. The time delay setting, combined with two-dimensional imaging, allows the region where flow velocities are measured to be accurately localized. Pulsed Doppler can be used to measure cardiac output, to describe the diastolic behaviour of the left ventricle, and to estimate effective valve orifice area.

Colour flow mapping is an automated two-dimensional version of pulsed Doppler in which flow velocity is measured at multiple depth settings and scan lines within the chosen plane to give a two-dimensional image in which the velocity of flow is represented by the colour of the image.

Measurements of transmitral and transaortic flow velocity during the cardiac cycle can be used to evaluate left ventricular function and to estimate stroke volume. Early diastolic filling of the left ventricle is assessed by measurement of the peak early diastolic transmitral flow velocity and the time–velocity integral of early diastolic flow.

Similarly, late diastolic filling is assessed by measurement of peak flow velocity and time–velocity integral during atrial systole. Doppler measurements of mitral inflow velocity have high reproducibility (*98*), and are an indirect measure of instantaneous pressure differences between the left atrium and left ventricle. The time–velocity integral for one pulse, multiplied by the cross-sectional area of the outflow tract, can be used to estimate the stroke volume. This method can be used in the left ventricular outflow tract or in the mitral valve. Estimates of cardiac output from Doppler measurements of transmitral blood flow have been shown to correlate well with thermodilution as the reference standard for measurement of cardiac output (*99*). Experience with the use of these methods in epidemiological surveys is increasing.

Ultrasound of peripheral arteries

B-mode ultrasound yields a two-dimensional image in which the brightness of the image represents the intensity of echoes from each point. Doppler ultrasound can be used to measure flow velocity. A combination of these two methods is duplex scanning, in which B-mode ultrasound is combined with colour imaging or spectral analysis of pulsed-wave Doppler to give an image in which information about arterial wall thickness and flow velocity are combined. The use of this technique to study peripheral arterial disease in epidemiological surveys is still expanding.

Ultrasound of the carotid arteries

B-mode ultrasound can be used in the carotid arteries to measure intima–media thickness as a marker of arteriosclerosis. These measurements have high specificity for atherosclerosis.

In the ARIC study, the largest population survey using carotid ultrasound so far reported, three segments are scanned: distal 1 cm straight portion of the common carotid, carotid bifurcation, and proximal first cm of the internal carotid. The common carotid is the easiest site to image, but appears to be less susceptible to disease than the carotid bifurcation and the internal carotid. In each segment, 11 measurements of the thickness of the far wall are attempted at 1-mm increments, and the intima–media thickness of each segment is estimated as the mean of these 11 measurements (Appendix 35).

Three scanning angles are used: anterior oblique, lateral and posterior oblique (*100*), and scans are performed bilaterally. Scans can be read at a central station, using computerized edge-tracking to reduce observer variation (*100*). The reliability of measurements of intimal–medial thickness in the common carotid is fairly high; the coefficient of variation for intimal–medial thickness (mean of both common carotids) has been reported to be 2.5% for duplicate observers using a computerized edge-tracking program, and 6% for duplicate measurements by different operators (*100*). Cut-off levels for defining atherosclerosis must be derived with respect to the popu-

lations under study, as few data are available for populations outside the United States.

Localized atherosclerotic plaques can also be detected in the carotid arteries by B-mode ultrasound (*101*, *102*) but the reproducibility of quantitative measurements of plaque is poor (*103*). Continuous-wave Doppler ultrasound can be used to detect reduction in the diameter of the carotid arteries (*104*). A reduction of the internal carotid artery diameter of 30% or more, corresponding to a frequency shift of 3.1 kHz, has been taken as evidence of carotid stenosis (*105*).

Ultrasound of the abdominal aorta

B-mode ultrasound can also be used to detect aortic aneurysms in population screening. Elderly men are the group at highest risk. The detection of small abdominal aortic aneurysms in apparently healthy men presents some ethical problems because the most appropriate management for this group has not yet been defined. Decisions about whether to include this in a protocol should take into account the results of controlled trials of screening that are now in progress.

Endothelium-dependent vasodilatation

Impairment of endothelium-dependent vasodilatation is associated with several cardiovascular risk factors, including smoking, raised plasma cholesterol and hypertension. Whether atherosclerosis causes endothelial dysfunction, or endothelial dysfunction causes atherosclerosis is not yet clear. Some of the protocols that have been developed to measure endothelium-dependent vasodilatation are simple and non-invasive enough to be used in cardiovascular field surveys. Dilatation of arteries in response to increased flow is dependent on intact and functional endothelium. In one widely-used protocol, reactive hyperaemia is induced in the forearm, and the change in brachial artery diameter in response to the increased flow is measured by B-mode ultrasound (*106*). A cuff around the forearm is inflated to 300 mmHg (40·0KPa) for 4–5 minutes, then deflated. The increased flow in the brachial artery is the result of dilatation. The brachial artery above the elbow is scanned in longitudinal section at rest, 45–60 seconds after cuff deflation, and again 10 minutes later. Mean vessel diameter is measured on four cardiac cycles, coincident with the R wave on the ECG. Flow-mediated vasodilatation is calculated as the mean arterial diameter during hyperaemia divided by the mean arterial diameter at rest. Measurements on the same individuals measured up to 1 month apart are highly reproducible (*106*). The measurement can be completed in 20 minutes, and is acceptable to both adults and children. At present, however, measurement of endothelium-dependent vasodilation is still new and its utility for field surveys intended to monitor risk factors in the population remains to be established.

Coronary calcification by electron beam computed tomography

Electron beam computed tomography (EBCT) is a new technique that allows calcification to be detected in the coronary arteries with high sensitivity and specificity. Autopsy studies have shown that EBCT accurately quantifies calcification in the coronary arteries and that calcification in the coronary arteries is always associated with atheromatous plaque (*107*). Coronary calcification is also correlated to traditional risk factors such as hypertension or hypercholesterolaemia. The prevalence of coronary calcification (defined as coronary calcium area >1 mm^2) in unselected population samples parallels the age–sex patterning of clinical coronary disease: in Caucasians aged 30–39 years in the USA the prevalence is 21–31% in men and 10–11% in women (*108, 109*), rising steeply with age (*109*). In young adults cardiovascular risk factors predict coronary calcification (*110*), and in older adults coronary calcification predicts major coronary events (*111*). Although EBCT may be of limited use in the clinical assessment of patients with known coronary disease, it has promise for studying the occurrence of early coronary atheroma in the population.

At present EBCT scanners are quite costly. In surveys, at least six participants can be scanned each hour. However, there are ethical problems associated with detection of coronary calcification in otherwise fit young adults. As yet no trial evidence is available to guide the clinical management of such individuals.

Collection and analysis of biological specimens

Blood sampling

Most cardiovascular field surveys include a venous blood sample for measurement of plasma cholesterol and other lipids. The protocol for blood sampling (non-fasting, fasting or post-load) will depend on the objectives of the study. Plasma total cholesterol is not altered by recent food consumption, and fasting specimens are thus not necessary if plasma cholesterol is the only variable to be measured. High-density lipoprotein (HDL) cholesterol falls slightly after a meal, but measurements of HDL cholesterol on non-fasting samples may still be useful (*112*). For measurement of triglyceride, glucose and insulin levels, samples should be taken in the fasting state or at a fixed time after challenge with carbohydrate. Asking subjects to fast overnight does not usually affect their willingness to participate if they are offered an appointment early in the morning at a field station near their home. A protocol for blood sampling, including recommendations for possible HIV/AIDS exposure, is shown in Appendix 36.

Plasma lipids

EDTA plasma is the preferred specimen to use for lipid analyses, as lipoproteins are more stable in EDTA plasma than in serum (*113*). EDTA plasma for HDL determinations can be kept for up to 48 hours at 4 °C before the HDL-containing supernatant is separated.

Cholesterol and triglyceride are usually measured by enzymatic methods. These methods give high reproducibility within batches, but results are likely to differ between laboratories even when the same equipment and reagents are used, and the levels tend to drift with time. The standard quality control and calibration procedures recommended by manufacturers of the assay kits are usually sufficient to detect a drift of the assay large enough to be important for clinical measures (say 5%) but not to detect smaller drifts which may be important in epidemiological surveys, especially if an objective of the survey is to monitor changes in the plasma lipids of the population. Portable cholesterol meters are even less satisfactory for monitoring and research purposes. For plasma cholesterol it is essential that the laboratory measurements are standardized against a reference method.

The essential components of a chemical reference method are a definitive method and reference materials that have long-term stability and can be used to calibrate any assay that is in use. In the USA, a national reference system for cholesterol has been organized through the National Institute for Standards and Technology (NIST) (*114*). The NIST maintains a definitive isotope-dilution mass spectrometer method and distributes reference materials. The Abell–Kendall reference method, maintained by the Centers for Disease Control and Prevention, gives similar results and is widely used. Tracing the laboratory results back to the reference system depends upon assay of reference materials in the laboratory. A problem with standardizing cholesterol measurements is matrix effects—interactions between the effects of sample preparation (fresh, frozen or lyophilized) and assay (reference method or routine enzymatic method) (*115*).

Apo B-containing lipoproteins (VLDL, IDL and LDL) can be separated from whole plasma by precipitation, leaving a supernatant in which HDL is the only lipoprotein. Several precipitation techniques are available; heparin–manganese chloride is the most widely used, but has relatively low precision. Phosphotungstate precipitation has higher precision, but gives results that are biased upwards in comparison with the reference method of ultracentrifugation. Where the objective is to rank individuals rather than to record absolute values for comparison with other surveys, this is not a serious disadvantage. Differences between methods of HDL measurement are more serious where clinical decisions are to be made on the basis of measurements of LDL cholesterol calculated from total cholesterol, HDL cholesterol and triglyceride by the Friedewald method, i.e. $C_{LDL} = C_{plasma} - C_{HDL} - TG/5$ (*116*).

No national reference standards are currently available for HDL cholesterol or plasma triglyceride research measurements.

The value of including HDL subfractions in population surveys is not yet clearly established. Direct measurements of cholesterol and triglyceride in LDL, VLDL and IDL can be made only by ultracentrifugation, which is complex and expensive.

Lipoprotein (a) can be measured by commercial kits at a modest cost. The interpretation of differences in lipoprotein (a) levels between populations is difficult unless apo (a) phenotype is also measured, as most of the variation in lipoprotein (a) levels is attributable to the isoforms of apo (a) that are present, which are in turn determined by the alleles present at the apo (a) genetic locus.

Other lipoprotein measurements (LDL subfractions, IDL, apolipoprotein A-I and A-II) have not consistently been shown to improve the prediction of IHD or to add information of public health importance. Some progress is being made with the development of standardized reference materials (*117*).

Post-prandial lipids have rarely been measured in epidemiological surveys. One reason for this has been that the standard fat tolerance test, in which hourly blood samples are taken for 12 hours following a fat meal, is too time-consuming for population surveys. Recent studies, however, suggest that a single measurement of plasma triglyceride at 8–9 hours after a fat meal discriminates as well as the full protocol between those with and those without, coronary artery disease (*118*). This is feasible to include in large surveys. Measurements of postprandial apo B-100 and apo B-55, in addition to plasma triglyceride, can help to distinguish between endogenous and exogenous triglyceride in plasma.

Glucose tolerance testing and insulin sensitivity

Measurement of glucose tolerance is especially important in many non-European populations in whom the prevalence of non-insulin-dependent diabetes is high and the excess risk associated with glucose intolerance accounts for a high proportion of cases of cardiovascular disease. To measure the prevalence of diabetes, it is essential to obtain a minimum of one blood sample 2 hours after a 75-g oral glucose load which has itself been consumed after a fast. This allows participants to be categorized as normal glucose tolerance, impaired glucose tolerance or diabetic. Usually surveys that include a blood sample at 2 hours also obtain a fasting sample, but where resources are limited a 2-hour sample is sufficient to measure the prevalence of diabetes.

The standard procedure for an oral glucose tolerance test (*119*) requires a fast of at least 14 hours. A shorter period of fasting may affect the fasting glucose value but probably does not seriously affect the 2-hour post-load value. The participant should

be instructed to refrain from smoking on the day of examination until the test has been completed, and to have nothing to eat or drink except water on the day of the test. After the fasting blood sample has been taken, the participant is given a glucose drink containing the equivalent of 75 g anhydrous dextrose and instructed to drink it steadily over 5 minutes. A second blood sample is taken 2 hours later. Note that 75 g anhydrous dextrose is equivalent to more than 75 g dextrose monohydrate, the form in which glucose powder is supplied.

Whole blood should be collected in tubes containing fluoride, which inhibits enzymatic glucose oxidation. If glucose determinations are to be made on plasma rather than whole blood, plasma should be separated within 20 minutes of taking the blood sample, as plasma glucose values fall when whole blood is kept for an hour or more, even at 4 °C. Once plasma has been separated, the glucose level in the fluoride plasma should be stable if the specimen is frozen.

Insulin levels and insulin resistance

Resistance to insulin-mediated glucose uptake is believed to have a primary role in the etiology of non-insulin-dependent diabetes, and also in disturbances of plasma lipids such as raised triglyceride levels and low HDL cholesterol. Thus measurements of insulin resistance may be desirable in populations in whom both cardiovascular disease and non-insulin-dependent diabetes are common.

Fasting, 1-hour, and 2-hour insulin levels which can easily be obtained if a glucose tolerance test is being undertaken are crude proxy measures of insulin resistance. The sum of these three insulin values can be used as a summary measure. Some investigators have used the product of fasting insulin and glucose values as an index of insulin resistance ("homeostasis model assessment") (*120*).

Measurements of insulin resistance by more direct techniques, such as the frequently-sampled intravenous glucose tolerance test or the euglycaemia hyperinsulinaemic clamp, are too invasive, costly and time-consuming for routine population surveys. One method that has been successfully used in studies of more than 100 participants is the short insulin tolerance test. For this test, participants attend after an overnight fast. Insulin (0.05 U/kg body weight) is injected intravenously. "Arterialized" blood samples for glucose measurement are obtained from a dorsal vein of a warmed hand every 2 minutes for 16 minutes (*121*). The test is terminated by the administration of intravenous glucose at 15 minutes. The rate of fall of blood glucose (on a logarithmic scale) is taken as a measure of insulin sensitivity. The short insulin tolerance test can be performed quickly and correlates well with measurements of insulin resistance obtained by more invasive methods. Hypoglycaemia within 15 minutes of insulin injection is uncommon, and the procedure has a good safety record provided that venous access is maintained to allow intravenous glucose to be given promptly.

A wide variety of measures are performed on routine blood samples. The remaining issues are correct sampling handling, and accurate and stable measurement techniques.

Methods in genetic studies

There are a number of unique issues that should be borne in mind in the planning of genetic epidemiological studies. These are outlined in the following paragraphs.

Consent forms

A key issue in any genetic study is that consent for the examination of DNA must be obtained in writing, and should specify the general area of proposed testing. Without such consent, results, no matter how remarkable, will be unpublishable. Publishers will demand consent for genetic analyses because this is an increasingly sensitive area that has seized the public's concern (Appendix 1).

Family studies

Family studies require the utmost tact and diplomacy. There are surprises in most family cupboards. Contact with relatives to validate reported episodes of disease can be difficult. It is therefore wise to have well-trained staff who are experienced in such procedures.

Genetic material

There are many sources of DNA, ranging from buccal mucosal cells through material from skin biopsies, hair and post mortem specimens. White blood cells are the most accessible and can be extracted from whole blood or stored as buffy coats removed from the top of the column of red blood cells after centrifugation.

Storage

DNA should be prepared immediately. Buffy coats can be kept for some years at −70 °C. A good rule for low-temperature storage is that there is access to another appliance, preferably on a separate power supply, in case of failure of the first one. Buccal mucosal cells can be stored simply and cheaply.

Shipping

Samples of buffy coats are best air-freighted batched in a plastic bag placed on a block of dry ice at the bottom of a large polystyrene box. More dry ice is then packed around the sample, taking care that the ice fragments are large so they will take longer to evaporate. A polystyrene lid is applied and well taped on. After addressing, choose a reliable air freight company. Fax or telephone the laboratory of destination with

the date and expected time of arrival of the consignment, the flight number, and the airway bill number.

Other biological samples

Blood cells can be kept after separation from plasma, frozen, and used for DNA analyses. Care should taken to avoid removing the buffy coat layer, which lies between the plasma and red cell layers, when the supernatant is removed by pipette.

Buccal brushings are an alternative source of DNA for genetic studies—the recommended method uses two or three cotton swabs. Participants can collect their own brushings and return them by post.

Urine collected over 24 hours should include a marker such as para-amino benzoic acid (PABA) where possible, so that the completeness of urine collection can be checked. Three 80-mg capsules are given to be taken three times a day with meals. Single 24-hour collections that contain 85% or more of the marker can be assumed to be complete. Measurements on 24-hour urine include electrolyte excretion, hormones and their metabolites, and albumin. For measurements of albumin excretion, a timed overnight urine collection is an acceptable alternative to 24-hour collection. An example of a urine collection protocol is shown in Appendix 37.

Adipose tissue biopsies of about 35 mg of tissue can be obtained from the upper outer quadrant of the buttock or from the supra-iliac region. The least invasive method is to use a needle of about 1.5 mm diameter, attached to a syringe or vacuum tube assembly via a connecting tube (*122*). A skinfold is raised and the needle is inserted into the subcutaneous tissue space. When suction is applied, adipose tissue collects in the connecting tube (*122*). The evacuated tube is disconnected before the needle is removed from the tissue. This method causes slightly more discomfort than a venepuncture, but much less than use of a conventional biopsy needle.

Analysis of the component fatty acids is carried out by gas–liquid chromatography of the methyl esters. The reliability of duplicate measurements taken from different sites should be high.

Storage of biological specimens

For long-term storage of specimens, freezing at −70 °C or lower is usually required. Precautions against breakdown of freezers or power loss must be taken. Freezers should be sited in a hospital or other installation that has back-up power supplies. Regular maintenance of the freezer should be arranged, and an emergency back-up freezer should be available in case of breakdowns. The freezer should be fitted with a temperature alarm linked to the telephone switchboard so that the research group can be alerted if a breakdown or loss of power occurs.

References

1. Rose GA. Chest pain questionnaire. *Millbank Memorial Fund Quarterly*, 1965, 43: 32–39.

2. Bass EB, Follansbee WP, Orchard TJ. Comparison of a supplemented Rose questionnaire to exercise thallium testing in men and women. *Journal of Clinical Epidemiology*, 1989, 42:385–394.

3. Garber CE, Carleton RA, Heller GV. Comparison of Rose questionnaire angina to exercise thallium scintigraphy: different findings in males and females. *Journal of Clinical Epidemiology*, 1992, 45:715–720.

4. Lessa I, Bastos CAG. Epidemiology of cerebrovascular accidents in the city of Salvador, Bahia, Brazil. *Bulletin of the Pan America Health Organization*, 1983, 17:292–303.

5. Osuntokun BO et al. Incidence of stroke in an African city: results from the Stroke Registry at Ibadan, Nigeria, 1973–1975. *Stroke*, 1979, 10:205–207.

6. Murray CJ. Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bulletin of the World Health Organization*, 1994, 72:429–445.

7. Murray C. Rethinking DALYs. In Murray C, Lopez A, eds. *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020.* Geneva, World Health Organization, 1996: 1–98.

8. Ware JEJ, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, 1992, 30:473–483.

9. Croog SH et al. The effects of antihypertensive therapy on the quality of life. *New England Journal of Medicine*, 1986, 314:1657–1664.

10. Kaplan GA, Keil JE. Socioeconomic factors and cardiovascular disease: a review of the literature. *Circulation*, 1993, 88:1973–1998.

11. Office of Population Censuses and Surveys. *Classification of occupations.* London, HMSO, 1980.

12. Iribarren C et al. Twelve-year trends in cardiovascular disease risk factors in the Minnesota Heart Survey: are socioeconomic differences widening? *Archives of Internal Medicine*, 1997, 157:873–881.

13. Luepker RV et al. Socioeconomic status and coronary heart disease risk factor trends: the Minnesota Heart Survey. *Circulation*, 1993, 269:505–510.

14. Reed T et al. Family history of cancer related to cholesterol level in young adults. *Genetic Epidemiology*, 1986, 3:63–71.

15. Hunt SC, Williams RR, Barlow GIL. A comparison of positive family history definitions for defining risk of future disease *Journal of Chronic Disorders*, 1986, 39: 809–821.

16. Kee F et al. Reliability of reported family history of myocardial infarction. *British Medical Journal*, 1993, 307:1528–1530.

17. Barker DJP. The fetal and infant origins of adult disease. *British Medical Journal*, 1990, 301:111.

18. Pechacek TF et al. Measurement of adolescent smoking behavior: rationale and methods. *Journal of Behavioral Medicine*, 1984, 7:123–140.

19. Sobell LC et al. Do quantity-frequency data underestimate drinking-related health risks? *American Journal of Public Health*, 1982, 72:823–828.

20. Romelsjo A, Leifman H, Nystrom S. A comparative study of two methods for the measurement of alcohol consumption in the general population. *International Journal of Epidemiology*, 1995, 24:929–936.

21. Girela E et al. Comparison of the CAGE questionnaire versus some biochemical markers in the diagnosis of alcoholism. *Alcohol and Alcoholism*, 1994, 29:337–343.

22. Duffy JC. Questionnaire measurement of drinking behaviour in sample surveys. *Journal of Official Statistics*, 1990, 1:229–234.

23. Bernadt MW et al. Comparison of questionnaire and laboratory tests in the detection of excessive drinking and alcoholism. *Lancet*, 1982, 1:325–328.

24. Berkman LF, Syme SL. Social networks, host resistance and mortality: a nine-year follow-up of Alameda County residents. *American Journal of Epidemiology*, 1979, 109: 186–204.

25. Rosenman RH et al. Multivariate prediction of coronary heart disease during 8.6 year follow-up in Western Collaborative Group Study. *American Journal of Cardiology*, 1976, 37:903–909.

26. Ragland DR, Brand RJ. Coronary heart disease mortality in the Western Collaborative Group Study: follow-up experience of 22 years. *American Journal of Epidemiology*, 1988, 127:462–476.

27. Farmer RD, Harvey PG. Minor psychiatric disturbance in young adults. The use of the general health questionnaire in the estimation of the prevalence of non-psychotic disturbance in different groups. *Social Science and Medicine*, 1976, 9:469–474.

28. Barefoot JC, Dahlstrom WG, Williams RB. Hostility, CHD incidence, and total mortality: a 25-year follow-up study of 255 physicians. *Psychosomatic Medicine*, 1983, 45:59–63.

29. Marmot MG, Syme SL. Acculturation and coronary heart disease in Japanese Americans. *American Journal of Epidemiology*, 1976, 104:225–246.

30. Wenger HA, Bell GJ. The interactions of intensity, frequency and duration of exercise training in altering cardiorespiratory fitness. *Sports Medicine*, 1986, 3:346–356.

31. Passmore R, Durnin JVGA. Human energy expenditure. *Physiology Review*, 1965, 35:801–840.

32. United States Department of Health and Human Services. Physical activity and health: a report of the Surgeon General. Atlanta, GA, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, 1996.

33. Siconolfi SF et al. Self-reported physical activity compared with maximal oxygen uptake. *American Journal of Epidemiology*, 1985, 122:101–105.

34. Tucker LA, Bagwell M. Television viewing and obesity in adult females. *American Journal of Public Health*, 1991, 81:908–911.

35. Rissel CE. Overweight and television watching. *Australian Journal of Public Health*, 1991, 15:147–150.

36. Richardson MT et al. Ability of the Caltrac® accelerometer to assess daily physical activity levels. *Journal of Cardiopulmonary Rehabilitation*, 1995, 15:107–113.

37. Ceesay SM et al. The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. *British Journal of Nutrition*, 1989, 61:175–186.

38. Marr JW. Individual dietary surveys: purposes and methods. *World Review of Nutrition and Dietetics*, 1971, 13:105–164.

39. Bingham SA et al. Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24h recalls, food-frequency questionnaires and estimated-diet records. *British Journal of Nutrition*, 1994, 11:341–355.

40. Bingham SA. Limitations of the various methods for collecting dietary intake data. *Annals of Nutrition and Metabolism*, 1991, 35:117–127.

41. Bingham SA. The dietary assessment of individuals; methods, accuracy, new techniques and recommendations. *Nutrition Abstracts and Reviews (Series A)*, 1987, 57:706–742.

42. Black AE et al. Measurements of total energy expenditure provide insights into the validity of dietary measurements of energy intake. *Journal of the American Dietetic Association*, 1993, 93:572–579.

43. Bingham SA. The use of 24-h urine samples and energy expenditure to validate dietary assessments. *American Journal of Clinical Nutrition*, 1994, 59:227S–231S.

44. Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. *Human Nutrition—Clinical Nutrition*, 1985, 39:5–41.

45. Stunkard AJ, Messick S. The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *Journal of Psychomatic Research*, 1985, 29:71–83.

46. Schakel SF, Sievert YA, Buzzard IM. Sources of data for developing and maintaining a nutrient database. *Journal of the American Dietetic Association*, 1988, 88:1268–1271.

47. Feunekes GIJ et al. Relative and biomarker-based validity of a food frequency questionnaire estimating intake of fats and cholesterol. *American Journal of Clinical Nutrition*, 1993, 58:489–496.

48. Balogh M, Kahn HA, Medalie JH. Random repeat 24-hour dietary recalls. *The American Journal of Clinical Nutrition*, 1971, 24:304–310.

49. Liu K et al. Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. *Journal of Chronic Diseases*, 1978, 31:399–418.

50. Willett WC et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *American Journal of Epidemiology*, 1985, 122:51–65.

51. Sempos CT. Invited commentary: some limitations of semiquantitative food frequency questionnaires. *American Journal of Epidemiology*, 1992, 135:1127–1132.

52. Rimm EB et al. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *American Journal of Epidemiology*, 1992, 135:1114–1126.

53. Feskanich D et al. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *Journal of the American Dietetic Association*, 1993, 93:790–796.

54. Kushi LH. Gaps in epidemiologic research methods: design considerations for studies that use food frequency questionnaires. *American Journal of Clinical Nutrition*, 1994, 59:180S–184S.

55. Bingham SA et al. Validation of weighed records and other methods of dietary assessment using the 24h urine nitrogen technique and other biological markers. *British Journal of Nutrition*, 1995, 6:531–550.

56. Lindroos AK, Lissner L, Sjostrom L. Validity and reproducibility of a self-administered dietary questionnaire in obese and non-obese subjects. *European Journal of Clinical Nutrition*, 1993, 47:461–481.

57. James WP, Bingham SA, Cole TJ. Epidemiological assessment of dietary intake. *Nutrition and Cancer*, 1981, 2:203–212.

58. Crockett SJ et al. Parent education in youth-directed nutrition interventions. *Preventive Medicine*, 1989, 18:475–491.

59. Jones PR, Wilkinson S, Davies PS. A revision of body surface area estimations. *European Journal of Applied Physiology and Occupational Physiology*, 1985, 53:376–379.

60. Seidell JC et al. Assessment of intra-abdominal and subcutaneous abdominal fat: relation between anthropometry and computed tomography. *American Journal of Clinical Nutrition*, 1987, 45:7–13.

61. Tornaghi G et al. Anthropometric or ultrasonic measurements comparative study. *International Journal of Obesity*, 1994, 18:771–775.

62. Armellini F et al. Measurements of intra-abdominal fat by ultrasound and computed tomography: predictive equations in women. *Basic Life Sciences*, 1993, 60:75–77.

63. Helsing E. *Measuring obesity: classification and description of anthropometric data.* (Report EUR/HFA 16) Copenhagen, WHO Regional Office for Europe, 1988.

64. Roldan CA, Shively BK, Crawford MH. Value of the cardiovascular physical examination for detecting valvular heart disease in asymptomatic subjects. *American Journal of Cardiology*, 1996, 77:1327–1331.

65. Wagener HP, Keith NM. Diffuse arteriolar disease with hypertension and associated retinal lesions. *Medicine*, 1939, 18:317.

66. Sharp PS et al. Hypertensive retinopathy in Afro-Caribbeans and Europeans: prevalence and risk factor relationships. *Hypertension*, 1995, 25:1322–1325.

67. Petrie JC et al. Recommendations on blood pressure measurement. *British Medical Journal* (*Clinical Research Edition*), 1986, 293:611–615.

68. Prineas RJ. *Blood pressure sounds: their measurement and meaning.* Minneapolis, MN, Gamma Medical Products Corporation, 1977.

69. White WB et al. National standard for measurement of resting and ambulatory blood pressures with automated sphygmomanometers. *Hypertension*, 1993, 21:504–509.

70. de Gaudemaris R et al. The random-zero versus the standard mercury sphygmomanometer: a systematic blood pressure difference. *American Journal of Epidemiology*, 1985, 121:282–290.

71. O'Brien E et al. Inaccuracy of the Hawksley random-zero sphygmomanometer. *Lancet*, 1990, 336:1465–1468.

72. Mackie A, Whincup P, McKinnon M. Does the Hawksley random-zero sphygmomanometer underestimate blood pressure, and by how much? *Journal of Human Hypertension*, 1995, 9:337–343.

73. Morrison C. *Scottish MONICA risk factor survey methods* (video tapes). Glasgow, Scotland. MONICA Project.

74. Walma EP et al. Accuracy of an oscillometric automatic blood pressure device: the Omron HEM403C. *Journal of Human Hypertension*, 1995, 9:169–174.

75. Reeves RA. A review of the stability of ambulatory blood pressure: implications for diagnosis of hypertension. *Clinical and Investigative Medicine*, 1991, 14:251–255.

76. Laing S, Greenhalgh RM. The detection and progression of asymptomatic peripheral arterial disease. *British Journal of Surgery*, 1983, 70:628–630.

77. Lukaski HC. Methods for the assessment of human body composition, traditional and new. *American Journal of Clinical Nutrition*, 1987, 46:537–556.

78. Jensen MD. Research techniques for body composition assessment. *Journal of the American Dietetic Association*, 1992, 92:454–460.

79. van der Kooy K, Seidell JC. Techniques for the measurement of visceral fat: a practical guide. *International Journal of Obesity*, 1993, 17:187–196.

80. Grimley Evans J, Prior IAM. Indices of obesity derived from height and weight in two Polynesian populations. *British Journal of Preventive Social Medicine*, 1969, 23: 56–59.

81. Houtkooper LB et al. Why bioelectrical impedance analysis should be used for estimating adiposity. *American Journal of Clinical Nutrition*, 1996, 64:436S–448S.

82. Roubenoff R. Applications of bioelectrical impedance analysis for body composition to epidemiologic studies. *American Journal of Clinical Nutrition*, 1996, 64: 459S–462S.

83. Shephard RJ. Tests of maximal oxygen intake: a critical review. *Sports Medicine* 1984, 1:99–124.

84. Davies CTM. Limitations to the prediction of maximum oxygen intake from cardiac frequency measurements. *Journal of Applied Physiology*, 1968, 24:700–706.

85. Glassford RC et al. Comparison of maximal oxygen uptake values determined by predicted and actual methods. *Journal of Applied Physiology*, 1965, 20:509–513.

86. American College of Sports Medicine. *Guidelines for exercise testing and prescription.* Philadelphia, Lea & Febiger, 1991.

87. Evaluation of physical performance on the basis of tests. In: Åstrand PO, Rodahl K, eds. *Textbook of work physiology: physiological bases of exercise.* New York, McGraw-Hill, 1986:354–390.

88. Prineas RJ, Crow RS, Blackburn H. *The Minnesota Code: manual of electrocardiographic findings. Standards and procedures for measurement and classification.* Littleton, M.A. Wright, 1982.

89. Kors JA et al. Validation of a new computer program for Minnesota coding. *Journal of Electrocardiology*, 1996, 29:83–88.

90. Uusitupa M et al. Sensitivity and specificity of Minnesota Code Q-QS abnormalities in the diagnosis of myocardial infarction verified at autopsy. *American Heart Journal*, 1983, 106:753–757.

91. Weiner DA et al. Exercise stress testing. Correlations among history of angina, ST-segment response and prevalence of coronary-artery disease in the coronary artery surgery study (CASS). *New England Journal of Medicine*, 1979, 301:230–235.

92. Devereux RB, Liebson PR, Horan MJ. Recommendations concerning use of echocardiography in hypertension and general population research. *Hypertension*, 1987, 9:97–104.

93. Sahn DJ et al. Recommendations regarding quantitation in M-mode electrocardiography: results of a survey of echocardiographic measurements. *Circulation*, 1978, 58:1072–1083.

94. Devereux RB, Reichek N. Echocardiographic determination of left ventricular mass in man: anatomic validation of the method. *Circulation*, 1977, 55:613–618.

95. Devereux RB et al. Echocardiographic assessment of left ventricular hypertrophy: comparison to necropsy findings. *American Journal of Cardiology*, 1986, 57:450–458.

96. Benjamin EJ et al. Left atrial size and the risk of stroke and death. The Framingham Heart Study. *Circulation*, 1995, 92:835–841.

97. Devereux RB. Evaluation of cardiac structure and function by echocardiography and other noninvasive techniques. *Hypertension*, 1990, 92:1479–1491.

98. Galderisi M et al. Intra- and interobserver reproducibility of Doppler-assessed indexes of left ventricular diastolic function in a population-base study (the Framingham Heart Study). *American Journal of Cardiology*, 1992, 70:1341–1346.

99. Hoit BD et al. Calculating cardiac output from transmitral volume flow using Doppler and M-mode echocardiography. *American Journal of Cardiology*, 1988, 62: 131–135.

100. Adams MR et al. Carotid intima-media thickness is only weakly correlated with the extent and severity of coronary artery disease. *Circulation*, 1995, 92:2127–2134.

101. Howard G et al. Carotid artery intimal-medial thickness distribution in general populations as evaluated by B-mode ultrasound. *Stroke*, 1993, 24:1297–1304.

102. Li R et al. B-mode-detected carotid artery plaque in a general population. *Stroke*, 1994, 25:2377–2383.

103. Salonen JT, Salonen R. Ultrasonographically assessed carotid morphology and the risk of coronary heart disease. *Arteriosclerotic Thrombolysis*, 1991, 11:1245–1249.

104. Jungquist G et al. Spectral analysis with digital presentation of C-W Doppler signals in the evaluation of carotid stenosis. *Clinical Physiology*, 1986, 6:319–327.

105. Jungquist G, Arborelius MJ, Lindell SE. Features of carotid artery flow velocity in healthy subjects and consequences for evaluating stenosis in the internal carotid artery. *Ultrasound in Medicine and Biology*, 1989, 15:305–310.

106. Sorensen KE et al. Non-invasive measurement of human endothelium dependent arterial responses: accuracy and reproducibility. *British Heart Journal*, 1995, 74:247–253.

107. Rumberger JA et al. Coronary artery calcium area by electron-beam computed tomography and coronary atherosclerotic plaque area: a histopathologic correlative study. *Circulation*, 1995, 92:2157–2162.

108. Kaufmann RB et al. Quantity of coronary artery calcium detected by electron beam computed tomography in asymptomatic subjects and angiographically studied patients. *Mayo Clinic Proceedings*, 1995, 70:223–232.

109. Devries S et al. Influence of age and gender on the presence of coronary calcium detected by ultrafast computed tomography. *Journal of the American College of Cardiology*, 1995, 25:76–82.

110. Mahoney LT et al. Coronary risk factors measured in childhood and young adult life are associated with coronary artery calcification in young adults: the Muscatine Study. *Journal of the American College of Cardiology*, 1996, 27:277–284.

111. Arad Y et al. Predictive value of electron beam computed tomography of the coronary arteries. 19-month follow-up of 1173 asymptomatic subjects. *Circulation*, 1996, 93:1951–1953.

112. Folsom A et al. Lipid concentrations in serum and EDTA-treated plasma, from fasting and nonfasting normal persons, with particular regard to HDL cholesterol. *Clinical Chemistry*, 1983, 29:2025–2028.

113. Hainline A et al. Accuracy and comparability of long-term measurements of cholesterol. *Clinical Chemistry*, 1986, 32:611–615.

114. Cooper GR, Myers GL. Reference system for cholesterol measurements. *Scandinavian Journal of Clinical and Laboratory Investigation*, 1990, 198:27–31.

115. Waymack PP, Miller WG, Myers GL. Assay instrument-dependent matrix effects in standardization of cholesterol measurements. *Clinical Chemistry*, 1993, 39:2058–2062.

116. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clinical Chemistry*, 1972, 18:499–502.

117. Marcovina SM et al. International Federation of Clinical Chemistry standardization project for measurements of apolipoproteins A-I and B. III. Comparability of apolipoprotein A-I values by use of international reference material. *Clinical Chemistry*, 1993, 39:773–781.

118. Patsch JR et al. Relation of triglyceride metabolism and coronary artery disease: studies in the postprandial state. *Arteriosclerotic Thrombolysis*, 1992, 12:1336–1345.

119. *Diabetes mellitus. Report of a WHO Study Group.* Geneva, World Health Organization, 1986 (WHO Technical Report Series, No. 727).

120. Matthews DR et al. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*, 1985, 28:412–419.

121. Gelding SV et al. Validation of the low dose short insulin tolerance test for evaluation of insulin sensitivity. *Clinical Endocrinology*, 1994, 40:611–615.

122. Beynen AC, Katan MB. Rapid sampling and long-term storage of subcutaneous adipose-tissue biopsies for determination of fatty acid composition. *American Journal of Clinical Nutrition*, 1985, 42:317–322.

GLOSSARY

The following glossary defines certain epidemiological and statistical terms as used in this monograph. The numbers in parentheses following some of the definitions indicate pages on which additional information will be found.

cluster sampling. A method of sampling that uses groups or clusters (e.g., households, schools, or villages) instead of individuals as the initial sampling units.

cohort. A group of persons, defined by a common characteristic (usually year of birth), whose experience over a period of time is studied.

confidence limits. Limits placed on both sides of an estimated quantity to provide a specified probability that the true value is included.

cross-sectional study. A study of the characteristics of a population at a particular point in time.

enriched sample. A stratified sample in which the sampling fraction is disproportionately high in a particular stratum.

epidemiology. The study of disease in defined populations.

frequency distribution. A table showing the numbers (or proportions) of individuals falling in each of a series of defined classes.

incidence rate. The rate of occurrence of new cases (or manifestations) in a defined population.

longitudinal (prospective) study. A study of a population over a period of time, starting with the present.

population. A group of individuals defined by some common characteristic (e.g., occupation).

power of a statistical test. The probability that the test will reject a false alternative hypothesis.

predictive (diagnostic) value. The probability of disease (or occurrence of some event), given a positive test result.

prevalence rate. A ratio expressing the frequency with which a specified characteristic occurs in a defined population at a particular point in time.

CARDIOVASCULAR SURVEY METHODS

random error. An unbiased and unpredictable form of error whose mean, in repeated measurements, usually tends towards zero. Also called random variation.

random sampling. A method of sampling in which each person or unit in a population has an equal chance of being selected. Also called random allocation.

repeatability. The extent of agreement between repeated measurements-i.e., the sum of subject and observer variation. Also called precision or reliability.

sampling frame. A list that identifies all persons in a population at a particular point in time.

secular change. A change in the characteristics of a population over a period of time.

sensitivity. The extent to which a method gives results free from false negatives-i.e., the fewer false-negative results, the greater the sensitivity.

specificity. The extent to which a method gives results free from false positives-i.e., the fewer false-positive results, the greater the specificity.

standard deviation. A statistical measure of the variability or scatter of individual observations. In a series of observations that follow a normal (i.e., Gaussian) distribution, about 95% of the observations fall within two standard deviations on either side of the mean. (The standard deviation is an inappropriate summary of variability if the observations show a major departure from a normal distribution; in this case, the use of percentiles may be more suitable.) The standard deviation of the difference between duplicate readings is a measure of the variability of disagreement between such readings.

standard error of the difference between two means. A measure of the sampling variability of estimates of the mean difference between two groups. When there is no true difference between the groups, their observed means will differ, by chance alone, by more than twice the standard error of the difference in about 1 in 20 pairs of randomly selected samples.

standard error of the mean. A measure of the sampling variability of estimates of the mean-i.e., the amount of variation to be expected between estimates of the mean in a series of random samples (of equal size) from the same population. In about 95% of the samples, the observed mean will fall within two standard errors of the true mean.

stratified sampling. A method of sampling in which the population is divided into strata of known size on the basis of some known characteristic (e.g., age or social class); separate random samples are then drawn from each stratum.

systematic error. A biased form of measurement error or disagreement-i.e., one whose mean, on repeated measurement, tends to be other than zero. validity. The extent to which a method provides a true assessment of that which it purports to measure.

variable. An attribute that is measured. Measurement may be either qualitative (involving assignment of individuals to a class) or quantitative (involving numerical description).

variance. The square of the standard deviation.

variation. The tendency of measurements of a given attribute to differ. Such variation may be either (1) observer variation, the tendency of repeated measurements of a given attribute, made by one or more observers, to differ for reasons other than subject variation, or (2) subject variation, of which there are two types: (a) between-subject variation, which is the tendency of a given attribute to show true differences from one individual to another, and (b) within-subject variation, which is the tendency of the attribute to show true change during the interval between consecutive measurements in the same individual.

LIST OF APPENDICES