World Health Organization

WHO POSITION PAPER ON MAMMOGRAPHY SCREENING

WHO position paper on mammography screening

World Health Organization

WHO Library Cataloguing-in-Publication Data

WHO position paper on mammography screening.

1.Mammography. 2.Early Detection of Cancer. 3.Breast Neoplasms – prevention and control. I.World Health Organization.

ISBN 978 92 4 150793 6 (NLM classification: WP 815)

Printed in Switzerland

Contents

Acknowledgements 5

Abbreviations 6

Executive Summary 7

Recommendations by age group and resource setting 11

1. Women aged 50–69 years 11
2. Women aged 40–49 years 12
3. Women aged 70–75 years 12

Introduction 13

Objectives, target audience and scope 15

Development process 16

Review groups 16
Management of conflict of interest 17
Decision-making 17
External review 18
Review of the evidence 18

Evidence and recommendations 20

Evidence of benefits and harms 20
Evidence on screening interval 22
General considerations 22

Recommendations by age and resource setting 27

1. Women aged 50–69 years 27
2. Women aged 40–49 years 29
3. Women aged 70–75 years 30

Research priorities 31

Dissemination and implementation 33

Useful web resources 34

References 35

List of contributors 37

Annex A - Evidence to recommendations template 39

Annex B - Evidence Summary
Benefits and harms of mammography screening: umbrella systematic review 47

Introduction 49

Methods and data source 50

Table 1. PICO question for the evidence review 51

Systematic reviews included 51
Available mammography randomized controlled trials 52
Areas of controversy 52

Outcome data sources 53
Breast cancer-specific mortality 53
Health-related quality of life, quality of life, disability-adjusted life years 54
Overtreatment 54
Masectomies 54
Overdiagnosis 54
False positive rate 55
All-cause mortality 55

Evidence profile 56
Table 2. Screening mammography (including data from contemporary observational studies) 56
Table 3. Systematic reviews of randomized controlled trials (RCTs) 59
Table 4. Systematic reviews of observational studies 60
Table 5. Systematic reviews evaluating psychological impact of mammography 63
Table 6. Mammography screening randomized controlled trials 64
Table 7. Convergence of relative estimates of breast cancer mortality across various reviews 65
Table 8. Evidence profile. Breast cancer mortality in randomized controlled trials (median follow-up of about 11 years) 66
Table 9a. Screening interval and breast cancer mortality: data from modelling 67
Table 9b. Screening interval and breast cancer mortality: data from randomized controlled trials 67
Table 10. Evidence profile for treatment type (proportion of women receiving treatment) 68
Table 11. Evidence profile. Mastectomy data from randomized controlled trials (median follow-up of about 11 years) 69
Table 12a. Evidence profile. False positive rate per single screening round in North America 70
Table 12b. Evidence profile. False positive rate per four screening mammographies over 11 years, in North America 70
Table 13. Evidence profile. All-cause mortality (median follow-up of about 11 years), data from randomized trials 71

References 72

Search strategy 75

PRISMA flow diagram and checklist 77

Acknowledgements

The *World Health Organization position paper on mammography screening* was produced under the overall direction of Dr Oleg Chestnov, Assistant Director-General, Noncommunicable Diseases and Mental Health, and Dr Shanthi Mendis, Director a. i., Management of Noncommunicable Diseases.

Dr Cecilia Sepúlveda, Senior Advisor, Cancer Control WHO, coordinated the overall publication and provided editorial input.

Professor Hassan Murad, Knowledge and Evaluation Unit of the Center for the Science of Healthcare Delivery and the Center for Translational Services Science Activities, Mayo Clinic, USA, conducted the systematic review.

Professor Carl Heneghan of the Centre for Evidence-Based Medicine, Nuffield Department of of Primary Care Health Sciences, University of Oxford, United Kingdom was co-chair of the Guideline Development Group meeting on recommendations.

Dr Gerald Gartlehner, GRADE Methodologist and Head of Department for Evidence-based Medicine and Clinical Epidemiology at the Danube University, Krems, Austria, was co-chair of the Guideline Development Group meeting on recommendations. He also provided assistance with the GRADE methodology and evidence during the recommendations process.

David Bramley was rapporteur and writer of these guidelines.

Dr Susan Norris of the WHO Guideline Review Committee secretariat, provided valuable support and guidance throughout the process of guideline development.

External experts and technical staff at WHO headquarters and regional offices, provided valuable input by contributing to and reviewing the drafts (see list of contributors).

Abbreviations

DALY	disability -adjusted life-year
ERG	External Review Group
GDG	Guideline Development Group
Global NCD Action Plan 2013–2020	Global Action Plan for Prevention and Control of Noncommunicable Diseases 2013–2020
GRADE	Grading of Recommendations Assessment, Development and Evaluation
GRC	Guideline Review Committee secretariat
IARC	International Agency for Research on Cancer
NCD	noncommunicable disease
RCT	randomized control trial
WHO	World Health Organization

Executive Summary

The *WHO position paper on mammography screening* and the *Guidelines for referral of suspected breast cancer at primary health care in low-resource settings* (WHO, 2013) are part of a broader set of breast cancer guidelines that will be developed in the coming years. These comprehensive guidelines will include primary prevention, diagnosis, treatment, rehabilitation and palliative care, as well as other screening modalities that could work in less affluent countries when evidence becomes available

So far the only breast cancer screening method that has proved to be effective in organized population-based programmes is mammography screening. However, reports of the benefits and harms of mammography screening differ widely in the context and intensity of screening examined, as well as in the interpretation of the available evidence. There is also uncertainty about the appropriate age groups for screening and the steps that should be taken by responsible authorities to commission and implement breast cancer screening programmes of appropriate quality. World Health Organization (WHO) Member States, particularly upper-middle-income countries that are implementing or planning to implement breast cancer screening programmes, are increasingly requesting guidance from WHO with regard to mammography screening.

The primary objectives of this guideline are: (i) to provide policy-makers, health-care managers, and health-care providers with clear, objective and independent guidance on the balance between benefits and harms of mammography screening in women of different age groups; and (ii) to disseminate the recommendations based on this guidance among policy-makers, health-care providers, health-care managers, women and the general public in order to promote informed decisions in this area.

The population addressed by this guideline comprises asymptomatic women at average risk for breast cancer in different age groups (40–49 years, 50–69 years, and 70 years and above). The scope of the guideline does not include women with an elevated risk for breast cancer independent of age. The questions addressed are the following:

- In women aged 40–49 years, 50–69 years and 70–75 years, asymptomatic and at average risk for breast cancer, what is the balance of benefits and harms in those offered mammography screening compared to those not offered screening?
- What is the effect of the screening interval on the balance of benefits and harms?

The Guideline Development Group (GDG) emphasized the importance of evaluating mammography screening in settings with organized population-based cancer screening programmes as defined by WHO (2007): high standard programmes that target all the population at risk in a given geographical area with high specific cancer burden, with everyone who takes part being offered the same level of screening, diagnosis and treatment service. Organized programmes include an administrative structure responsible for implementation, quality assurance and evaluation of the entire screening process. These programmes identify and individually invite each eligible woman to attend each round of screening. The organized, population-based approach to programme implementation is recommended because it provides an operational framework conducive to effective management of performance and continuous improvement of the screening process and outcomes (von Karsa et al., 2013). In addition, certain conditions need to be met in order to implement a successful organized mammography screening programme (Box 1).

Box 1
Organized, population-based breast cancer screening programmes
Key criteria for successful programme implementation

- Demonstrated feasibility, cost-effectiveness and affordability of the screening process in the respective setting through pilot studies and modelling.
- Coordination of all activities, including planning, feasibility testing, piloting and gradual rollout of the programme across a country or region, by an autonomous management team responsible for service delivery, quality assurance, and evaluation.
- A well-developed, equitable, health system with cancer control planning integrated into the national noncommuncable disease (NCD) prevention and control strategy and with balanced, objective information of women about the benefits and harms of mammography screening.
- Validated protocols for all steps in the screening process, including identification and individual invitation of all eligible women to attend screening, performing the screening test, diagnosis, treatment and palliative care.
- Adherence to comprehensive, evidence-based guidelines for quality assurance of the entire screening process, including standards and protocols for professional and technical quality assurance; and that are regularly updated based on current evidence.
- Quality assurance and information systems covering the entire screening process, including call and recall of participants for follow-up of abnormalities detected in screening, and for monitoring and evaluating programme performance at each step in the screening process.

- Regular monitoring, evaluation and reporting of programme performance and impact based on national or international standards that include process and outcome indicators and also cover women's safety and satisfaction.
- Sufficient organizational and financial resources to ensure the sustainability of all programme components, including the requisite equipment, infrastructure and workforce, and the capacity for training, reporting and national and international exchange of experience.

Sources: WHO, 2007; von Karsa et al., 2013.

These guidelines were developed according to the WHO process for guideline development. They include recommendations for different age groups and resource settings and are based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) tool and the GDG considerations.

The GDG decided to base its recommendations on systematic reviews of randomized controlled trials (RCTs) as well as systematic reviews of observational studies. The GDG was concerned about the applicability of trial results, because older trials no longer reflect current practice and might provide wrong estimates about some of the benefits and harms. Outcomes of interest, selected by the GDG on the basis of importance for decision-making, were: breast cancer-specific mortality, disability-adjusted life-years (DALYs) gained and health-related quality of life (rated as critical); plus all-cause mortality, overtreatment, reduction in mastectomies, overdiagnosis and cumulative false-positives (rated as important). Reliable quantitative data were not identified for DALYs, health-related quality of life or overtreatment.

Following the GRADE methodology, and in the context of well-organized, population-based programmes, the overall quality of evidence was graded as moderate or low across different age groups, and was graded as low for the screening interval. There is evidence across all age groups that organized, population-based mammography screening programmes can reduce breast cancer mortality by around 20% in the group of women invited to attend screening versus the uninvited group. In general, the expected benefit in women actually participating in screening is higher, but there appears to be a narrow balance between benefits and harms, particularly in younger and older women. There is uncertainty about the magnitude of the harms – particularly overdiagnosis and overtreatment. To date, the best trade-off seems to be provided by screening every two years.[1]

1. The complete Evidence Report in Annex B.

The GDG was concerned that the net benefit might be tilted towards harms if screening programmes are opportunistic,[1] not population-based, or lack the necessary quality control mechanisms. Irrespective of the social setting and the screening method used, all population-based cancer screening programmes must be well organized in order to obtain net benefits. Opportunistic screening or screening that is not well organized run the risk of causing more harm than good and should not be implemented in any setting (WHO, 2007; von Karsa et al., 2013). In addition, the GDG emphasized that access to objective, evidence-based information about the benefits and harms of breast cancer screening is crucial for women.

Cost-effectiveness analysis carried out by WHO and partners in various middle-income countries using the WHO CHOICE methodology showed that mammography screening was not cost-effective for a lower-middle-income country. In contrast, it was cost-effective for various upper-middle-income countries (Zelle et al., 2012, 2013; Niëns et al.). However, regional differences within countries were not taken into account. Furthermore, organized mammography screening programmes may not be feasible for nationwide implementation in the short or medium term in these countries due to fragmented health systems with uneven or limited capacity, resulting in lack of universal access to adequate diagnosis and treatment of symptomatic breast disease. Regional programmes may be an option in populations with an appropriate burden of breast cancer if sufficient resources are provided to implement and sustain an organized population-based screening programme.

Limited resource settings, where the majority of women with breast cancer are diagnosed in late stages and mammography screening is not cost-effective or feasible, should focus available resources on early diagnosis by ensuring universal access of women with symptomatic lesions to prompt and effective diagnosis and treatment (WHO, 2013). Low-cost screening approaches such as clinical breast examination, which seems to be a promising approach for these settings, could be implemented when the necessary evidence from ongoing studies becomes available (Sankaranarayanan et al., 2011; WHO, 2013).

Because the available evidence on the benefits and harms of mammography screening programmes in the different age groups comes only from higher-income countries and there is a greater level of uncertainty regarding the effects of these programmes in limited resource settings, the GDG decided to provide stratified recommendations by age group and resource setting.

---

1. Opportunistic screening is the unsystematic application of screening tests in routine health services (WHO, 2007).

Recommendations by age group and resource setting

1. Women aged 50–69 years

1.1 Well-resourced settings

In well-resourced settings, WHO recommends[1] organized, population-based mammography screening programmes for women aged 50–69 years if the conditions for implementing an organized programme specified in this guide[2] are met by the health-care system, and if shared decision-making strategies are implemented so that women's decisions are consistent with their values and preferences. **(Strong recommendation based on moderate quality evidence)**

WHO suggests a screening interval of two years. **(Conditional recommendation based on low quality evidence)**

1.2 Limited resource settings with relatively strong health systems

In limited resource settings with relatively strong health systems, WHO suggests[3] considering an organized, population-based mammography screening programme for women aged 50–69 years only if the conditions for implementing an organized programme specified in this guide[4] are met by the health-care system, and if shared decision-making strategies are implemented so that women's decisions are consistent with their values and preferences. **(Conditional recommendation based on moderate quality evidence)**

WHO suggests a screening interval of two years. **(Conditional recommendation based on low quality evidence)**

1.3 Limited resource settings with weak health systems

In limited resource settings with weak health systems, where the majority of women with breast cancer are diagnosed in late stages and mammography screening is not cost-effective and feasible, early diagnosis of breast cancer through universal access of women with symptomatic lesions to prompt and effective diagnosis and treatment should be high

1. According to GRADE, "recommend" is used when there is a strong recommendation.
2. See Box 1, page 8.
3. According to GRADE, "suggest" is used when there is a conditional recommendation.
4. See Box 1, page 8.

on the public health agenda (WHO, 2013). Clinical breast examination, a low-cost screening method, seems to be a promising approach for these settings and could be implemented when the necessary evidence from ongoing studies becomes available (Sankaranarayanan et al., 2011).

2. Women aged 40–49 years

2.1 Well-resourced settings

In well-resourced settings, WHO suggests an organized, population-based screening programme for women aged 40–49 years only if such programme is conducted in the context of rigorous research and monitoring and evaluation, if the conditions for implementing an organized programme specified in this guide[1] are met and if shared decision-making strategies are implemented so that women's decisions are consistent with their values and preferences. **(Conditional recommendation based on moderate quality evidence)**

2.2 Limited resource settings with weak or relatively strong health systems

In limited resource settings with weak or relatively strong health systems, WHO recommends against the implementation of population-based screening programmes for women aged 40–49 years. **(Strong recommendation based on moderate quality evidence)**

3. Women aged 70–75 years

3.1 Well-resourced settings

In well-resourced settings, WHO suggests an organized, population-based screening programme for women aged 70–75 years only if such programme is conducted in the context of rigorous research, if the conditions for implementing an organized programme specified in this guide[2] are met by the health-care system, and shared decision-making strategies are implemented so that women's decisions are consistent with their values and preferences. **(Conditional recommendation based on low quality evidence)**

3.2 Limited resource settings with weak or relatively strong health systems

In limited resource settings with weak or relatively strong health systems, WHO recommends against the implementation of population-based screening programmes for women aged 70–75 years. (Strong recommendation based on low quality evidence)

---

1. See Box 1, page 8.
2. See Box 1, page 8.

Introduction

The World Health Organization (WHO) Global Action Plan for Prevention and Control of Noncommunicable Diseases 2013–2020 (Global NCD Action Plan 2013–2020), which was endorsed by the Sixty-sixth World Health Assembly in resolution WHA66.10 in May 2013, calls on WHO to provide technical guidance to countries for the integration into their health systems of cost-effective interventions against major noncommunicable diseases (NCDs).[1] This includes the early detection of cancer.

Early detection of cancer comprises two strategies: screening and early diagnosis. Screening involves the systematic application of a screening test for a specific cancer in an asymptomatic population in order to detect and treat cancer or pre-cancers before they become a threat to the well-being of the individual or the community.

Early diagnosis is based on improved public and professional awareness (particularly at the primary health care level) of signs and symptoms associated with cancer, improved health-care-seeking behaviour, prompt clinical assessment and early referral of suspected cancer cases, such that appropriate diagnostic investigations and treatment can be rapidly instituted leading to improved mortality outcomes (WHO, 2007, 2013).

The *WHO position paper on mammography screening* and the *Guidelines for referral of suspected breast cancer at primary health care in low-resource settings* (WHO, 2013) are part of a broader set of breast cancer guidelines that will be developed in the coming years. These comprehensive guidelines will include primary prevention, diagnosis, treatment, rehabilitation and palliative care, as well as other screening modalities (e.g. clinical breast examination) that could work in less affluent countries when evidence becomes available (Sankaranarayanan et al., 2011).

Breast cancer is the leading cancer in women worldwide in both developed and developing countries. In resource-constrained settings with very limited health system capacity and lack of early-detection programmes, the majority of women with breast cancer are diagnosed in the late stages and the overall five-year survival rate is very low, with a range of 10–40%. On the other hand, the five-year survival rate for early localized breast cancer exceeds 80% in settings where early detection and basic treatment are available and accessible (Ferlay et al. 2010; Sankaranarayanan, Swaminathan and Lucas, 2011).

1. The Global NCD Action Plan 2013–2020 can be accessed at http://www.who.int/nmh/events/ncd_action_plan/en/index.html.

In resource-constrained settings, early diagnosis of breast cancer (as defined above) is a very appropriate and affordable strategy for early detection. It can complement screening strategies where these are justifiable, available and feasible (WHO, 2013). Screening is a much more complex public health undertaking than early diagnosis and is usually cost-effective and justified when the disease burden is relatively high, an adequate health system capacity has been achieved and when the quality of the entire multidisciplinary screening process is assured (WHO, 2007; von Karsa et al., 2013).

So far the only breast cancer screening method that has proved to be effective is mammography screening. Appendix 2 of the Global NCD Action Plan 2013–2020 provides a menu of policy options and cost-effective interventions for prevention and control of major NCDs, including population-based breast cancer mammography screening linked with timely and good-quality diagnosis and treatment services. The action plan's recommendation on mammography screening is based on the cancer prevention handbook on breast cancer screening published by the International Agency for Research on Cancer (IARC, 2002).

WHO Member States, particularly upper-middle-income countries that are implementing or planning to implement breast cancer screening programmes, are increasingly requesting guidance from WHO with regard to mammography screening. Furthermore, reports of the benefits and harms of mammography screening differ widely in the context and intensity of screening examined, as well as in the interpretation of the available evidence. There is also uncertainty about the adequate age groups for screening and the steps that should be taken by responsible authorities to commission and implement breast cancer screening programmes of appropriate quality.

Consequently, the *WHO position paper on mammography screening* responds to an urgent need of Member States and seeks to provide policy-makers, patients and health-care providers with clear, objective, independent and up-to-date guidance on the benefits and harms of mammography screening. The present guideline is focused on organized, population-based mammography screening programmes, essential for ensuring quality of screening services.

Objectives, target audience and scope

The primary objectives of this guideline are: (i) to provide policy-makers, health-care managers, and health-care providers with clear, objective and independent guidance on the balance between benefits and harms of mammography screening in women of different age groups; and (ii) to disseminate the recommendations based on this guidance among policy-makers, health-care providers, health-care managers, women and the general public in order to promote informed decisions in this area.

The primary target audiences of the guideline are policy-makers, health-care managers and health-care providers. The secondary target audiences are adult women and the public, in general, who need to be informed in a clear and constructive way of the WHO position on this topic.

The population addressed by this guideline comprises women at average risk for breast cancer in different age groups (40–49 years, 50–69 years, and 70 years and above). The scope of the guideline does not include women with breast symptoms or a palpable mass, or women with an elevated risk for breast cancer due to factors other than age (such as genetic mutations, personal history of invasive breast cancer, ductal carcinoma in situ, lobular carcinoma in situ or history of breast radiation). The questions addressed are the following:

- In women of different age groups (40–49 years, 50–69 years, and 70 years and above), what is the balance of benefits and harms in those offered mammography screening compared to those not offered screening?
- What is the screening interval (annually versus biannually) that provides the best balance between benefits and harms for women at average risk of breast cancer in different age groups?

Development process

Review groups[1]

WHO Steering Group: Members of the Steering Group were WHO staff members working in areas related to this topic at WHO headquarters and regional offices. The Steering Group contributed to the planning and oversight of the process of guideline development, reviewed the research questions, advised on the establishment of the Guideline Development Group (GDG) and the External Review Group (ERG), ensured that the process was carried out with objectivity and independence, and will provide the necessary support to mobilize resources for the dissemination, country adaptation and implementation of the guideline.

GDG: Members of the GDG were invited in their individual capacities. They represented different disciplines and diverse socioeconomic and geographical settings. The GDG was involved in the development of the guideline and the central task of the members was to produce evidence-based recommendations, taking into account diverse values and preferences. Individuals with very strong and passionate views on the subject were excluded from the GDG. The methodologist of the group was selected from the list provided by the Guideline Review Committee (GRC) secretariat and has not worked or published on mammography screening.

Chairs of the GDG: In consultation with the GRC secretariat it was decided to have two co-chairs for the GDG: the methodologist to facilitate discussions on methodological issues, and a GDG member with experience in assessment and management of screening programmes to guide discussions on content issues during the decision-making process. Both co-chairs were selected on the basis of their expertise and capacity in leading group discussions in a professional and unbiased manner.

ERG: Members of this group represented different geographical regions. The members were invited to review the completed draft of the guideline and were advised that the recommendations already agreed by the GDG could not be changed. Members included experts and stakeholders who had an interest in the topic and were likely to appraise the output from different scientific or philosophical perspectives, but who would eventually support the implementation of the recommendations.

1. See List of Contributors.

Management of conflict of interest

The "declaration of interests" form was collected for the methodologists and for all members of the GDG and ERG. Three experts declared some interest. The WHO legal office was consulted and no impediment was found for the full participation of these individuals in the GDG.

It was not possible to avoid having some panel members with an "intellectual" conflict of interest because of the critical need to include certain areas of expertise (such as radiology and experience in the management of breast cancer screening programmes). Consequently, there was careful management of intellectual interest in order to ensure the development of valid guidelines. This included the following:

- appointment of co-chairs with independent views on the topic, who had not published or conducted research on mammography screening and who were not in charge of managing mammography screening programmes or any of their components;
- limiting members with relevant intellectual conflict interest to a distinct minority of the panel;
- publicly disclosing the relevant conflicts of interest of panel members during the GDG consensus meeting;
- asking panel members to vote independently and anonymously on the recommendations at the GDG consensus meeting;
- requesting input on the draft recommendations on an individual basis;
- considering all guideline documents as strictly confidential during the development process;
- initial drafting and subsequent editing of the recommendations by a core group composed of the co-chairs and WHO secretariat with objective and independent views on the topic and without intellectual conflicts of interest.

Decision-making

Members of the WHO Steering Group, with the support of the guideline methodologists, drafted the scope of the guideline, refined the PICO (patient, intervention, comparison, outcome) questions and identified possible outcomes. The GDG agreed on the scoping document, selected the critical and important outcomes, provided input on the Evidence Report, and decided on the direction and strength of the recommendations during the GDG consensus meeting. An evidence-to-recommendations decision tool, which was adapted from a template provided by the GRC

secretariat, was used to guide the decision-making process.[1] The recommendations were agreed by consensus. This means that recommendations were accepted when the majority of the group members agreed with them and there was no major objection to acceptance.

After the meeting, the revised Evidence Report, the draft meeting report and the draft recommendations were circulated to the entire group. One GDG member who attended the consensus meeting did not agree later on with the recommendations and stated he/she could not co-author the document. Another member, who did not attend the consensus meeting, declared that the agreed recommendations were in conflict with the position of a regional patients' organization and, therefore, he/she could not endorse them. These two individuals had their names removed from the list of members of the GDG.

The discussion section elaborates the reasons behind possible mismatches between the quality of evidence and the strength of a recommendation.

External review

The final draft guideline, including the recommendations, was sent to the seven ERG members for review. Members of the ERG were advised that it was not possible to modify the recommendations already agreed by consensus by the GDG. One reviewer decided at this stage to abstain from participating without providing a reason. Two reviewers agreed with some of the recommendations but disagreed with others. Four reviewers agreed, in general, with the format and content of the guidelines. All six reviewers provided further input on methodological and research issues, as well as on the justification for the recommendations. Modifications were incorporated into the final document as appropriate and so long as they did not imply changing the agreed recommendations.

Review of the evidence[2]

An independent consultant was contracted to review and synthesize the evidence. Another independent consultant, the GDG methodologist, facilitated the selection of outcomes by the GDG and reviewed the evidence profiles.

1. See the template in the Annex A.
2. See the complete Evidence Report in Annex B.

The review followed an umbrella design.[1] Eligible studies were systematic reviews or evidence synthesis reports that evaluated mammography screening outcomes of interest in women at average risk, regardless of the study location or the language of the report. Databases were searched from the inception of each database up to December 2012 for relevant studies published in any language. Additional references were identified by contacting experts and reviewing bibliographies of identified studies.

The most relevant reviews were chosen on the grounds that they were: (i) the most comprehensive (summarizing the largest number of studies); (ii) the most recent; and (iii) the highest quality, as measured by the AMSTAR tool for assessing the methodological quality of systematic reviews. Systematic review selection, appraisal and data extraction were performed by a single methodologist considering the availability of multiple high-quality systematic reviews.

Outcomes of interest selected by the GDG after three rounds of voting and on the basis of importance for decision-making, were:

- **rated as critical:** breast cancer-specific mortality; disability-adjusted life-years (DALYs) gained; and health-related quality of life;
- **rated as important:** all-cause mortality; overtreatment; reduction in mastectomies; overdiagnosis; and cumulative false-positives.

The body of evidence used in the available systematic reviews and existing guidelines on mammography screening mainly comprised seven randomized controlled trials (RCTs) enrolling 600 000 women.

The quality of evidence was rated according to the GRADE framework (Grading of Recommendations Assessment, Development and Evaluation). With GRADE, the quality of evidence is rated down for increased risk of bias, indirectness, imprecision, publication bias and inconsistency; and is rated up for a large effect size, dose–response effect, and when all plausible confounding is considered to strengthen the association. The small number of existing trials and the heterogeneity of the observational studies did not allow for a formal statistical evaluation of publication bias; hence, this was not used to rate down the evidence although it may have existed. When existing GRADE evidence profiles were found, these were reviewed across multiple sources and their data were verified before being adapted for this report. When GRADE profiles were unavailable, they were created de novo.

1. Review of existing systematic reviews.

Evidence and recommendations

In women aged 40–49 years, 50–69 years and 70–75 years, asymptomatic and at average risk for breast cancer, what is the balance of benefits and harms in those offered mammography screening compared to those not offered screening? What is the effect of the screening interval on the balance of benefits and harms?

The GDG emphasized the importance of evaluating mammography screening in settings with organized, population-based cancer screening programmes (see the description under the General Considerations section).

The GDG decided to base its recommendations on systematic reviews of RCTs as well as on systematic reviews of observational studies. Although the reviewed trials generally have a lower risk of bias and confounding than observational studies, the GDG was concerned about the applicability of trial results. For example, older trials no longer reflect current practice and might provide wrong estimates about benefits and harms. These concerns particularly pertained to outcomes influenced by surgical practice such as rates of mastectomies and overall recall rates of women positive to mammography screening. Therefore, it was agreed to consider also the results from systematic reviews of observational studies that focused on organized, population-based screening programmes. Issues of limited applicability of RCTs have to be weighed against the higher risk that observational studies may present overestimated findings because of bias and confounding.

Evidence of benefits and harms

This section summarizes the main findings of the Evidence Report.[1] Evidence profiles were developed for all outcomes that were considered critical or important for decision-making by the GDG where data were available to provide reliable quantitative estimates. These outcomes were: (i) breast cancer mortality (rated as critical); (ii) mastectomies, overdiagnosis, cumulative false-positive rates; and (iii) all-cause mortality (rated as important). Reliable quantitative data were not identified for DALYs and health-related quality of life (rated as critical) or for overtreatment (rated as important).

In general across all ages and methods, on the basis of RCTs, most relative estimates of *breast cancer-specific mortality* are around a relative

1. See the complete Evidence Report, GRADE tables and references in Annex B.

risk of 0.80 in the group invited to attend screening, suggesting a 20% relative risk reduction with mammography screening programmes at 11 years of follow-up. However, longer follow-up has demonstrated a larger magnitude of risk reduction, thus suggesting that the full impact of mammography screening may be seen only after 20 years or more. Larger estimates were reported by observational studies, but these results may be overestimated due to risk of bias and confounding.

Data were not available to provide reliable quantitative estimates for rating the quality of evidence for *health-related quality of life*. Systematic reviews that focused mainly on *anxiety and psychological distress* were analysed. The main findings showed that mammography screening does not appear to create anxiety in women who are given a clear result after a mammogram. However, women who require further investigations following screening experience significant short-term anxiety. Women who received false-positive results on mammography screening had higher, though not apparently pathologically elevated, levels of distress and anxiety and thought more about breast cancer than did those with normal results.

Trial results showing increased *mastectomy rates* associated with screening are likely to be invalid at the present time because of the radical change in mastectomy practice since the RCTs were carried out. Most recent observational studies show results to the contrary.

The association between mammography screening and *overdiagnosis* has been demonstrated consistently across studies and is likely to be supported by high-quality evidence. However, there is significant uncertainty about the magnitude of overdiagnosis in the different age groups, particularly in younger and older women. The estimates vary greatly (from 0% to 54%) according to the method used, the source of the data and the definition of overdiagnosis. Thus, the evidence based on the current available data is low. Two recent reviews estimated that for every one or two overdiagnosed cases, at least one death due to breast cancer was avoided, a balance between benefit and harm considered to be appropriate (Marmot et al., 2012; Paci et al. 2012).

False-positive rates are common to all age groups, although they tend to be higher in younger age groups. The most precise estimates were available from studies that evaluated large registries and national databases in Europe with characteristics consistent with organized, population-based screening programmes.

Data for *all-cause mortality* were derived from RCTs with a median follow-up of about 11 years for women aged 39–49 and 50–69 years. However, there is concern about the accuracy/reliability of estimates due

to the small relative contribution of breast cancer mortality to all-cause mortality and the short duration of follow-up of the available trials.

Evidence on screening interval

Evidence of the effect of the *screening interval* on breast cancer-specific mortality was obtained from data from RCTs and modelling. Screening intervals in the RCTs ranged from 12 to 33 months over a median of 11 years and suggested no difference in breast cancer mortality for screening intervals less than 24 months compared to those of 24 months and longer. However, in view of the short follow-up period and other issues relating to the risk of bias or indirectness, the inference from this result is limited.

Modelling studies and further analysis of trials showed results that varied according to assumptions and trade-offs. Results from modelling showed that screening every two years seems to provide the best trade-off between benefits and harms. Screening biennially from age 50 years to 69 years achieved a median 16% reduction in breast cancer deaths compared to no screening. Biennial screening at age 40 years versus 50 years reduced mortality by an additional 3%, but it consumed more resources and yielded more false-positive results. Biennial screening after the age of 69 years yielded some additional mortality reduction in all models, but overdiagnosis increased substantially at older ages.

Following the GRADE methodology, and in the context of well-organized population-based programmes, the overall quality of evidence was graded as moderate or low across different age groups, and was graded as low for the screening interval.

General considerations

There is evidence across all age groups that organized population-based mammography screening programmes can reduce breast cancer mortality by around 20% in the group invited to participate in screening versus the uninvited group. In general, there appears to be a narrow balance of benefits compared with harms, particularly in younger and older women. There is uncertainty about the magnitude of the harms – particularly overdiagnosis and overtreatment. In addition, the best trade-off seems to be provided by screening every two years.

All-cause mortality was rated as an important outcome for decision-making. However, in view of the limitations of the available data the GDG did not consider these data to be sufficiently accurate or reliable to influence the recommendations.

The GDG was concerned that the net benefit might be tilted towards harms if screening programmes are opportunistic,[1] not population-based, or lack the necessary quality control mechanisms. For example, there is evidence that inequalities are generally reduced in countries with well-established, organized, population-based cancer screening programmes compared to those with opportunistic screening programmes (Palència et al., 2010).

Organized population-based cancer screening programmes, as defined by WHO (2007), share certain characteristics – i.e. they are of high standard, target all the population at risk in a given geographical area with high specific cancer burden and everyone who takes part is offered the same level of screening, diagnosis and treatment services. These programmes include an administrative structure responsible for implementation, quality assurance and evaluation of the entire screening process that includes information and invitation of the eligible women, performing the screening examination, and subsequent diagnosis and treatment of lesions detected through screening. Population-based screening programmes identify and individually invite each person in the eligible population to attend each round of screening so that each person in the eligible population has an equal chance of benefiting from screening. The organized, population-based approach to programme implementation is recommended because it provides an operational framework conducive to effective management of performance and continuous improvement of the screening process and outcomes. This is achieved, for example, through linkage of screening registry data with data in population and cancer registries, for optimization of invitation to screening and for evaluation of screening performance and impact (von Karsa et al., 2013).

In addition, the following conditions need to be met in order to implement an organized screening programme (WHO, 2007; von Karsa et al., 2013):

- demonstrated feasibility, cost-effectiveness and affordability of the screening process in the respective setting through pilot studies and modelling;
- coordination of all activities, including planning, feasibility testing, piloting and gradual rollout of the programme across a country or region, by an autonomous management team responsible for service delivery, quality assurance and evaluation;
- a well-developed, equitable health system with cancer control planning integrated into the national NCD prevention and control strategy and

1. Opportunistic screening is the unsystematic application of screening tests in routine health services (WHO 2007).

with balanced, objective information of women about the benefits and harms of mammography screening;

- validated protocols for all steps in the screening process, including identification and individual invitation of all eligible women to attend screening, performing the screening test, diagnosis, treatment and palliative care;
- adherence to comprehensive, evidence-based guidelines for quality assurance of the entire screening process, that are regularly updated based on current evidence and include standards and protocols for professional and technical quality assurance;
- quality assurance and information systems covering the entire screening process, including call and recall of participants for follow-up of abnormalities detected in screening, and for monitoring and evaluating programme performance at each step in the screening process;[1]
- regular monitoring, evaluation and reporting of programme performance and impact based on national or international standards that include process and outcome indicators and also cover women's safety and satisfaction;
- sufficient organizational and financial resources to ensure the sustainability of all programme, components, including the requisite equipment, infrastructure and workforce, and the capacity for training, reporting and national and international exchange of experience.

Cost-effectiveness analysis carried out by WHO and partners in various middle-income countries using the WHO CHOICE methodology[2] showed that mammography screening was not cost-effective for a lower-middle-income country such as Ghana (Zelle et al., 2012). In contrast, mammography screening was cost-effective for upper-middle-income countries such as Costa Rica, Mexico (Niëns et al., 2013) and Peru (Zelle et al., 2013). However, regional differences within countries should also be taken into account. Furthermore, organized mammography screening programmes may not be feasible for nationwide implementation in the short or medium term in these countries due to fragmented health systems with uneven or limited capacity resulting in lack of universal access to adequate diagnosis and treatment of symptomatic breast disease. Regional programmes may

1. Although the main intention of mammography screening programmes is to contribute to early detection and curative treatment, a proportion of women will still be identified in late stages of breast cancer. Therefore, palliative care should be made available to women diagnosed with late-stage cancer or who have progressive disease, but do not respond to curative treatment.
2. See the description of WHO CHOICE methodology at http://www.who.int/choice/description/en/.

be an option in populations with an appropriate burden of breast cancer if sufficient resources are provided to implement and sustain an organized population-based screening programme.

Because the evidence available on the benefits and harms of mammography screening programmes in the different age groups comes only from higher-income countries, and since there is a greater level of uncertainty regarding the effects of these programmes in limited resource settings, the GDG decided to provide stratified recommendations by age group and resource setting:

- **well-resourced settings** are settings with very strong health systems that, in general, have an existing capacity that allows them to develop and sustain organized population-based mammography screening programmes (e.g. most high-income countries).
- **limited resource settings with relatively strong health systems** are settings in which the existing capacity has the potential to gradually develop and sustain cost-effective, organized, population-based mammography screening programmes (e.g. a number of upper-middle-income countries).
- **limited resource settings with weak health systems** are settings with very limited capacity where mammography screening is not cost-effective, feasible and affordable (e.g. low-income and lower-middle-income countries).

Irrespective of the social setting and the screening method used, all population-based cancer screening programmes must be well organized in order to obtain net benefits. An organized, population-based screening programme with high coverage may reduce inequities by ensuring that all women, including those from lower socioeconomic groups, receive prompt diagnosis and treatment. Screening programmes that are not population-based, or programmes that are not well organized, run the risk of causing more harm than good and should not be implemented in any setting (WHO, 2007; von Karsa and Arrossi 2013).

Values may vary greatly across different groups, cultures and settings, particularly with respect to adverse events of mammography screening such as overdiagnosis, false-positives and psychosocial effects. For example, while overdiagnosis can be of great concern for policy-makers and programme managers it can be less of a problem for women. Qualitative research in the United Kingdom showed that overdiagnosis is viewed as less personally relevant by women than the possibility of underdiagnosis (Waller et al., 2013). Moreover, harms resulting from mammography screening (such as the false-positive rate, overdiagnosis or psychological distress) may be given a different value in settings where the majority of

women are diagnosed late and a screening programme is being introduced de novo.

Irrespective of the type of setting, access to objective, evidence-based information about the benefits and harms of breast cancer screening is crucial for women. Culturally tailored strategies are needed to address women in different settings, and particularly in underserved populations with low participation rates. However, pursuing high attendance rates for screening in a population-based programme should never take priority over informed decisions based on evidence and individual values and preferences.

Recommendations by age and resource setting

1. Women aged 50–69 years

1.1 Well-resourced settings

In well-resourced settings, WHO recommends[1] organized, population-based mammography screening programmes for women aged 50–69 years if the conditions for implementing an organized programme specified in this guide are met by the health-care system, and if shared decision-making strategies are implemented so that women's decisions are consistent with their values and preferences. **(Strong recommendation based on moderate quality evidence)**

WHO suggests a screening interval of two years. **(Conditional recommendation based on low quality evidence)**

Justification: Where feasible and affordable, organized mammography screening programmes represent so far the only population-based strategy that can reduce breast cancer mortality in women aged 50–69 years in well-resourced settings. While the balance between benefits and harms appears to be in favour of benefits, there is uncertainty as to the magnitude of the harms – particularly overdiagnosis and overtreatment. Breast cancer mortality is apparently decreasing in higher-income countries that have implemented mammography screening programmes, with the reduction probably due to both early detection and effective diagnosis and treatment. In addition, an organized screening programme, as opposed to an opportunistic screening programme, is able to ensure more efficient use of resources and equitable access to screening and management services.

Screening every two years seems to provide the best trade-off between benefits and harms. Further research is required to evaluate the effect of screening intervals.

Implementation: Because of the uncertainties regarding the magnitude of harms, the GDG emphasized the importance of implementing and maintaining well-organized population-based screening programmes and stressed that access to objective, evidence-based information about the benefits and harms of breast cancer screening is crucial for women.

---

1. According to GRADE, "recommend" is used when there is a strong recommendation.

1.2 Limited resource settings with relatively strong health systems

In limited resource settings with relatively strong health systems, WHO suggests[1] considering an organized, population-based mammography screening programme for women aged 50–69 years only if the conditions for implementing an organized programme specified in this guide are met by the health-care system, and if shared decision-making strategies are implemented so that women's decisions are consistent with their values and preferences. **(Conditional recommendation based on moderate quality evidence)**

WHO suggests a screening interval of two years. **(Conditional recommendation based on low quality evidence)**

Justification: There is no direct evidence that mammography screening programmes are effective in limited resource settings with weak or relatively strong health systems. However, in many such settings breast cancer has become an important public health problem (with high incidence and mortality rates) that justifies an early-detection programme being put in place. Organized mammography screening programmes for women aged 50–69 years could be a viable option in some limited resource settings with relatively strong health systems (e.g. various upper-middle-income countries), provided the WHO conditions for an organized, population-based programme are fulfilled. Taking into consideration the experience from higher-income countries described in many observational studies, only organized, population-based screening programmes with comprehensive quality control systems can provide the best balance between benefits and harms and can ensure equitable services.

Screening every two years seems to provide the best trade-off between benefits and harms in higher-resource settings and this may also apply to limited resource settings with relatively strong health systems.

Implementation: In limited resource settings with relatively strong health systems, even if the conditions for establishing population-based mammography screening programmes exist, nationwide implementation can be very challenging because of the complexity of such programmes. Higher-income countries have taken almost 10 years to fully implement well-organized population-based programmes. Therefore, it may be advisable to start the process in a pilot geographical area, implement activities in a stepwise manner, monitor and evaluate progress and establish a mechanism for gradual expansion of the programme (WHO, 2007). Because there appears to be a narrow balance of benefits compared with harms, the GDG emphasized that access to objective evidence-based

1. According to GRADE, "suggest" is used when there is a conditional recommendation.

information about the benefits and harms of breast cancer screening is crucial for women.

1.3 Limited resource settings with weak health systems

In limited resource settings with weak health systems, where the majority of women with breast cancer are diagnosed in late stages and mammography screening is not cost-effective and feasible, early diagnosis of breast cancer through universal access of women with symptomatic lesions to prompt and effective diagnosis and treatment should be high on the public health agenda (WHO, 2013). Clinical breast examination, a low-cost screening method, seems to be a promising approach for these settings and could be implemented when the necessary evidence from ongoing studies becomes available (Sankaranarayanan et al., 2011; WHO, 2013).

2. Women aged 40–49 years

2.1 Well-resourced settings

In well-resourced settings, WHO suggests an organized, population-based screening programme for women aged 40–49 years only if such programme is conducted in the context of rigorous research, and monitoring and evaluation, if the conditions for implementing an organized programme specified in this guide are met by the health-care system, and if shared decision-making strategies are implemented so that women's decisions are consistent with their values and preferences. **(Conditional recommendation based on moderate quality evidence)**

Justification: On the basis of the limited evidence available, there is uncertainty as to the balance between benefits and harms of mammography screening programmes in women aged 40–49 years. The reduction in breast cancer mortality is proven in RCTs; however, due to the much lower incidence rate of breast cancer in this age group and the somewhat lower sensitivity of mammography, the absolute benefits are small. On the other hand, harms – particularly in terms of cumulative false-positive rates – seem to be high. There is also uncertainty about the optimal screening interval. Therefore, there is a need for research in this age group.

Implementation: In well-resourced settings, when implementing mammography screening programmes for women aged 40–49 years in the context of rigorous research, it is important to ensure that organized population-based screening programmes are already well established for women aged 50–69 years.

2.2 Limited resource settings with weak or relatively strong health systems

In limited resource settings with weak or relatively strong health systems, WHO recommends against the implementation of population-based screening programmes for women aged 40–49 years. **(Strong recommendation based on moderate quality evidence)**

Justification: Because the limited evidence of mammography screening programmes for women aged 40–49 years comes only from higher-income countries, there is a greater level of uncertainty about the effects of these programmes in limited resource settings. Furthermore, although in such settings the proportion of women aged 40–49 years presenting with breast cancer may be relatively high (mainly due to demographic factors), the absolute risk of developing breast cancer in this age group is low compared to the risk in women over age 50 (GLOBOCAN, 2008, GLOBOCAN 2012). In limited resource settings, health investments should be made in interventions that promise a greater net benefit.

In limited resource settings, where the majority of women with breast cancer are diagnosed in late stages, and mammography screening is not cost-effective and feasible, early diagnosis of breast cancer through universal access of women with symptomatic lesions to prompt and effective diagnosis and treatment should be high on the public health agenda (WHO, 2013). Clinical breast examination, a low-cost screening method, seems to be a promising approach for these settings and could be implemented when the necessary evidence from ongoing studies becomes available (Sankaranarayanan et al., 2011; WHO, 2013)

3. Women aged 70–75 years

3.1 Well-resourced settings

In well-resourced settings, WHO suggests an organized, population-based screening programme for women aged 70–75 years only if the programme is conducted in the context of rigorous research, the conditions for implementing an organized programme specified in this guide are met by the health-care system, and if shared decision-making strategies are implemented so that women's decisions are consistent with their values and preferences. **(Conditional recommendation based on low quality evidence)**

Justification: There is uncertainty regarding the balance between benefits and harms of mammography screening programmes for women aged 70–75 years because of the limited and low level of evidence available. While existing data indicate an effect that is comparable to the effect in women aged 50–69 years, harms – particularly in terms of overdiagnosis

and overtreatment – seem to be very high. Therefore, there is a great need for research in this area.

Implementation: In well-resourced settings, when implementing mammography screening programmes for women aged 70–75 years in the context of rigorous research, it is important to ensure that there are already well-established population-based screening programmes for women aged 50–69 years.

3.2 Limited resource settings with weak or relatively strong health systems

In limited resource settings with weak or relatively strong health systems, WHO recommends against the implementation of population-based screening programmes for women aged 70–75 years. **(Strong recommendation based on low quality evidence)**

Justification: Because the scarce evidence available on mammography screening programmes for women aged 70–75 years comes only from higher-income countries, there is a greater level of uncertainty about the effects of these programmes in limited resource settings. Moreover, the GDG expressed the view that resources should be allocated to interventions with a clear net benefit. In limited resource settings generally, there are many other competing problems and a significant proportion of premature deaths correspond to avoidable causes for which there are cost-effective and feasible interventions.[1]

Research priorities

Well-resourced settings:

- evaluation of overdiagnosis, overtreatment, health-related quality of life issues, and the optimal screening interval of mammography screening programmes for women aged 50–69 years;
- evaluation of benefits and harms of mammography screening programmes for women aged 40–49 years, and the optimal screening interval; evaluation of the socioeconomic impact of expanding the mammography screening programme to this younger age group;
- evaluation of benefits and harms of mammography screening programmes for women over age 70, and the optimal screening interval; evaluation of the socioeconomic impact of expanding the mammography screening programme to this older age group.

1. As described, for example, in the Global NCD Action Plan 2013–2020 at http://www.who.int/nmh/events/ncd_action_plan/en/index.html.

Limited resource settings with relatively strong health systems:

- evaluation of benefits and harms of mammography screening programmes for women aged 50–69 years, including evaluation of the optimal screening interval;
- implementation research to test the feasibility of implementing nationwide organized mammography screening programmes for women aged 50–69 years.

Limited-resourced settings with relatively strong or weak health systems:

- evaluation of alternative breast cancer early-detection approaches that can work in limited resource settings, including validation of the protocols of low-cost screening methods.

Dissemination and implementation

The WHO guideline document, as well as the Evidence Report, will be published online (www.who.int/cancer). An official launch will be held and the recommendations will be widely disseminated to WHO regional and country offices, partners, governments, nongovernmental organizations, technical agencies and other stakeholders. A summary of the guideline will be published in a peer review journal. Clear and simple messages targeting women and the general public will be produced and posted on the Internet.

Mammography screening is known to be cost-effective, feasible and affordable mainly in countries where there is good health infrastructure and all the components for an early-detection programme are in place – including quality assurance systems and adequate, accessible diagnostic and treatment facilities, and palliative care. Therefore, this guideline can be implemented mainly in higher-income and upper-middle-income countries.

In collaboration with partners, WHO can support implementation activities by providing practical tools and direct technical assistance when needed. Multicountry demonstration projects of organized population-based screening programmes can be implemented, particularly in upper-middle-income countries that are planning to develop effective programmes.

WHO and partners will work with Member States to evaluate the impact of the guideline by coordinating efforts and providing advice and practical support. In this regard, tools and information systems will be developed to assess the impact of the guide. This will initially include assessment of performance indicators such as dissemination of the guidelines and adoption of the guideline recommendations within broad health policies and programmes and in the context of national cancer control programmes. Countries that have fully established screening programmes or that are developing demonstration programmes will be advised to include process indicators (such as compliance with and timeliness of screening, diagnosis and treatment, and quality assurance schemes) and outcome indicators (including stage distribution at diagnosis, survival, breast cancer mortality, rates of interval cancer, changes in end-users' knowledge and understanding of the benefits and harms of mammography screening, and economic consequences).

This guideline will be updated within five years as it is intended to evolve in response to new knowledge, evidence-based information, national needs and experience.

Useful web resources

- National Cancer Control Programmes
  http://www.who.int/cancer/nccp/en/
- How to plan and implement effective cancer control programmes
  http://www.who.int/cancer/modules/en/index.html
- IARC Screening Group: Breast Cancer
  http://screening.iarc.fr/breastindex.php
- European Reference Organization for Quality Assured Breast Screening and Diagnosis Services
  http://www.euref.org/
- Globocan 2008
  http://www.iarc.fr/en/media-centre/iarcnews/2010/globocan2008.php
- Globocan 2012
  http://globocan.iarc.fr/Default.aspx

References

Ferlay J et al. (2010). Cancer incidence in five continents, vol. I–IX. IARC CancerBase No. 9. Lyon: IARC (http://ci5.iarc.fr, accessed 8 February 2013).

IARC. (2002). Breast cancer screening. IARC handbook of cancer prevention, vol. 7. Lyon: IARC.

The Independent UK Panel on Breast Cancer Screening (2012), Cancer Research UK and the Department of Health The benefits and harms of breast cancer screening: an independent review *The Lancet*, Volume 380, Issue 9855, pp. 1778–1786 (http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(12)61611-0/abstract).

Niëns LM et al. (in press). Cost-effectiveness of breast cancer control strategies in Central America: the cases of Costa Rica and Mexico. *PlosOne*.

Paci E; EUROSCREEN Working Group (2012). Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen*. 19(Suppl. 1):5–13. (http://www.kreftregisteret.no/Global/Paci%20publication%20list.pdf, accessed 14.5.14)

Palència LEA et al. (2010). Socio-economic inequalities in breast and cervical cancer screening practices in Europe: influence of the type of screening program. *Int J Epidemiol*. 39(3):757–65.

Sankaranarayanan R et al. (2011). Clinical breast examination: preliminary results from a randomized controlled trial in India. *J Natl Cancer Inst*. 103:1476–80.

Sankaranarayanan R, Swaminathan R, Lucas E (2011). Cancer survival in Africa, Asia, the Caribbean and Central America (SurvCan). *IARC Scientific Publications*, vol. 162. Lyon: IARC.

von Karsa L, Arrossi S. (2013) *Development and implementation of guidelines for quality assurance in breast cancer screening – the European experience*. Salud pública de méxico. 55:318–28 (http://bvs.insp.mx/rsp/_files/File/2013/vol%2055%20No%203%20Mayo%20Junio/10Development.pdf, accessed 14.5.15).

Waller J et al. (2013). Women's responses to information about overdiagnosis in the UK breast cancer screening programme: a qualitative study. *BMJ Open*. 3:e002703. doi:10.1136/bmjopen-2013-002703.

WHO (2007). *Cancer control knowledge into action. WHO guide for effective programmes. Early detection*. Geneva: WHO.

WHO (2013). *Guidelines for referral of suspected breast and cervical cancer at primary health care in low resource settings*. Geneva: WHO.

Zelle SG et al. (2012). Costs, effects and cost-effectiveness of breast cancer control in Ghana. *Trop Med Int Health*. 17(8):1031–43.

Zelle SG et al. (2013). Cost-effectiveness analysis of breast cancer control interventions in Peru. *PlosOne*. doi: 10.1371/journal.pone.0082575 (http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0082575, accessed 14.5.2014).

List of contributors

Members of the Guideline Development Group

Silvina Arrosi
Cancer Prevention Programme
Ministry of Health
Buenos Aires, Argentina

Bettina Borisch
Institute of Preventive and Social Medicine
University of Geneva
Geneva, Switzerland

Jean-Francois Delaloye
Department of Gynecology and Obstetrics
Vaud University Hospital
Lausanne, Switzerland

Merce Marzo
Primary Health Care Research Unit
Catalan Institute of Health
Barcelona, Spain

Gerald Gartlehner (Co Chair)
Evidence-based Medicine and Clinical Epidemiology
Danube University
Krems, Austria

Carl Heneghan (Co Chair)
Centre for Evidence-Based Medicine
Nuffield Department of Primary Care Health Sciences
University of Oxford
United Kingdom

Sylvia Heywang-Köbrunner
National Reference Center Mammography
Munich, Germany

Julietta Patnick
NHS Cancer Screening Programmes
Directorate of Health and Wellbeing
Public Health England
United Kingdom

Rengaswamy Sankaranarayanan
Early Detection and Prevention Section
International Agency for Research on Cancer
Lyon, France

Nereo Segnan
Department of Cancer Screening
Reference Center for Cancer Epidemiology and Prevention
Piemonte, Italy

Mohamed Shaalan
Prevention and Early Detection Unit
National Cancer Institute
Cairo, Egypt

Surendra Shastri
WHO Collaborating Center for Cancer Prevention and Early Detection
Tata Memorial Hospital
Mumbai, India

Lawrence von Karsa
Quality Assurance Group
Early Detection and Prevention Section
International Agency for Research on Cancer
Lyon, France

Members of the External Review Group

Rajendra Badwe
Cancer Center, Tata Memorial Hospital
Mumbai, India

Ian Magrath
International Network for Cancer Treatment and Research
Brussels, Belgium

Raul Murillo
National Cancer Insititute
Bogotá, Colombia

Kofi Nyarko
Noncommunicable Diseases
Ministry of Health
Accra, Ghana

Yu Lin Qiao
Department of Cancer Epidemiology
Cancer Institute
Beijing, China

Cheng Har Yip
University Malaya Medical Center
Kuala Lumpur, Malaysia

Evidence Report prepared by

M. Hassan Murad
Knowledge and Evaluation Unit (KER), Center for the Science of Healthcare Delivery (CSHCD) and Center for Translational Services Science Activities (CTSA), Mayo Clinic Rochester, MN, USA

Members of the WHO Steering Group

Nathalie Broutet
World Health Organization
Geneva, Switzerland

Gauden Galea
World Health Organization
European Region
Copenhagen, Denmark

Silvana Luciani
Regional Offices for the Americas
Washington, DC, USA

Shanthi Mendis
World Health Organization
Geneva, Switzerland

Annex A

Evidence to recommendations template

Evidence-to-recommendations template, adapted from Health system and public health evidence to recommendations framework (Version 2)[1]

PICO Question:	
Problem: Background: Comparison: Setting:	Background:

	Criteria	Judgements						Research evidence	Additional information
	Are a large number of people affected?	No	Probably No	Uncertain	Probably Yes	Yes	Varies		
		☐	☐	☐	☐	☐	☐		

	Criteria	Judgements						Research evidence	Additional information
Benefits & harms of the options	Are the desirable anticipated effects large?	No	Probably No	Uncertain	Probably Yes	Yes	Varies		
		☐	☐	☐	☐	☐	☐		
	Are the undesirable anticipated effects small?	No	Probably No	Uncertain	Probably No	Yes	Varies		
		☐	☐	☐	☐	☐	☐		

1. http://www.decide-collaboration.eu/WP5/Strategies/Framework.

	Criteria	Judgements					Research evidence	Additional information
	What is the overall certainty of this evidence?	No included studies	Very low	Low	Moderate	High		
		☐	☐	☐	☐	☐		

	Criteria	Judgements					Research evidence	Additional information
Values	How certain is the relative importance of the desirable and undesirable outcomes?	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	No important uncertainty or variability	No known undesirable outcomes		
		☐	☐	☐	☐	☐		
	Are the desirable effects large relative to undesirable effects?	No	Probably No	Uncertain	Probably No	Yes	Varies	
		☐	☐	☐	☐	☐	☐	

	Criteria	Judgements						Research evidence	Additional information
Resource use	Are the resources required small?	No	Probably No	Uncertain	Probably Yes	Yes	Varies		
		☐	☐	☐	☐	☐	☐		

	Criteria	Judgements						Research evidence	Additional information
	Is the incremental cost small relative to the net benefits?	No	Probably No	Uncertain	Probably Yes	Yes	Varies		
		☐	☐	☐	☐	☐	☐		
Equity	What would be the impact on health inequities?	No	Probably No	Uncertain	Probably Yes	Yes	Varies		
		☐	☐	☐	☐	☐	☐		
Acceptability	Is the option acceptable to key stakeholders?	No	Probably No	Uncertain	Probably Yes	Yes	Varies		
		☐	☐	☐	☐	☐	☐		
Feasibility	Is the option feasible to implement?	No	Probably No	Uncertain	Probably Yes	Yes	Varies		
		☐	☐	☐	☐	☐	☐		

<table>
<tr><td>Balance of consequences</td><td>Undesirable consequences clearly outweigh desirable consequences in most settings</td><td>Undesirable consequences probably outweigh desirable consequences in most settings</td><td>The balance between desirable and undesirable consequences is closely balanced or uncertain</td><td>Desirable consequences probably outweigh undesirable consequences in most settings</td><td>Desirable consequences clearly outweigh undesirable consequences in most settings</td></tr>
<tr><td></td><td>☐</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td></tr>
<tr><td rowspan="2">Type of recommendation</td><td colspan="2">☐ Strong recommendation</td><td colspan="2">Conditional (weak) recommendation
☐ Only in the context of rigorous research
☐ Only with targeted monitoring and evaluation
☐ Only in specific contexts</td><td>We recommend the option</td></tr>
<tr><td colspan="2">☐ Strong recommendation against</td><td colspan="2">Conditional (weak)
☐ Recommendation against</td><td>☐</td></tr>
<tr><td>Recommendation (text)</td><td colspan="5"></td></tr>
<tr><td>Justification</td><td colspan="5"></td></tr>
<tr><td>Implementation considerations</td><td colspan="5"></td></tr>
<tr><td>Monitoring and evaluation</td><td colspan="5"></td></tr>
<tr><td>Research priorities</td><td colspan="5"></td></tr>
</table>

Implications of a strong recommendation	Implications of a weak recommendation
■ **Patients:** Most people in this situation would want the recommended course of action and only a small proportion would not ■ **Clinicians:** Most patients should receive the recommended course of action ■ **Policy makers:** The recommendation can be adapted as a policy in most situations	■ **Patients:** The majority of people in this situation would want the recommended course of action, but many would not ■ **Clinicians:** Be prepared to help patients to make a decision that is consistent with their own values/decision aids and shared decision making ■ **Policy makers:** There is a need for substantial debate and involvement of stakeholders

Definitions for ratings of the certainty of the evidence (GRADE)

Ratings	Implications of a strong recommendation	Implications of a weak recommendation
⊕⊕⊕⊕ High	This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different* is low.	This evidence provides a very good basis for making a decision about whether to implement the intervention. Impact evaluation and monitoring are unlikely to be needed if it is implemented.
⊕⊕⊕◯ Moderate	This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different* is moderate.	This evidence provides a good basis for making a decision about whether to implement the intervention. Monitoring of the impact is likely to be needed and impact evaluation may be warranted if it is implemented.
⊕⊕◯◯ Low	This research provides some indication of the likely effect. However, the likelihood that it will be substantially different* is high.	This evidence provides some basis for making a decision about whether to implement the intervention. Impact evaluation is likely to be warranted if it is implemented.
⊕◯◯◯ Very low	This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different* is very high.	This evidence does not provide a good basis for making a decision about whether to implement the intervention. Impact evaluation is very likely to be warranted if it is implemented.

* Substantially different: large enough difference that it might have an effect on a decision.

Annex B

Evidence Summary

Benefits and harms of mammography screening: umbrella systematic review

Introduction

Breast cancer is one of the most common cancers in women and one of the leading causes of death worldwide. Mammography screening seems to be the most effective method for early detection of breast cancer. Nevertheless, there is significant controversy about the balance of benefits and harms of mammography screening.

This systematic review aims to appraise and summarize the best available evidence about mammography screening to inform the development of WHO guidance on this issue in order to help policy-makers, patients and health-care providers with the process of decision-making. The guideline is focused on optimal screening methods that are service-based with a high participation rate and established follow-up and quality assurance strategies.

Methods and data source

This systematic review follows an umbrella design (overview of reviews). Eligible studies were systematic reviews or evidence synthesis reports that evaluated screening mammography outcomes of interest in women at average risk, regardless of the study location or the language of the report. Databases were searched from each database's inception and up to December 2012 for studies published in any language. The databases included Embase (1988 to 2013 Week 03), MEDLINE(R) In-process & other non-indexed citations and Ovid MEDLINE(R) (1946 to present), EBM Reviews and Cochrane Database of Systematic Reviews (2005 to December 2012). The search strategy was designed and conducted by an experienced librarian with input from the WHO methodologist. Controlled vocabulary supplemented with keywords was used to search for systematic reviews of outcomes of mammography screening. Additional references were identified by contacting experts and reviewing bibliographies of identified studies. The strategy used is described in the Appendix.

The most relevant reviews were chosen on the basis of: 1) the most comprehensive (summarizing the largest number of studies); 2) the most recent (published in the last 5 years for systematic reviews of randomized trials and observational studies and in the last 10 years for systematic reviews of psychological impact and quality of life); and 3) the highest quality, as measured by the AMSTAR tool for assessing the quality of systematic reviews. If two systematic reviews summarized the same trials, we chose the one with the higher AMSTAR score. In general, systematic reviews summarizing randomized trials had higher quality (score >10) whereas those summarizing observational studies had moderate scores (5-6) or did not describe details sufficient for quality assessment. Systematic review selection, appraisal and data extraction were performed by a single methodologist considering the availability of multiple high-quality systematic reviews. Searches for newer individual studies were not performed because there haven't been new randomized trials of mammography since 1991 nor they are expected to be conducted, and because the systematic reviews of the observational studies are very recent. Outcomes of interest were determined a priori by a WHO Expert Panel on the basis of importance for decision-making through a voting and consensus process.

The PICO question for this evidence review and the associated WHO statement are described in Table 1.

Table 1. PICO question for the evidence review

Patients	Intervention	Comparison	Outcomes	Study design	Other criteria
Women at average risk for breast cancer ≥ 40 years of age	Screening mammography (film or digital)	No screening	■ breast cancer-specific mortality ■ disability-adjusted life years (DALYs) gained ■ health-related quality of life (HrQoL) ■ overtreatment ■ reduction in mastectomies ■ overdiagnosis ■ cumulative false positives ■ all-cause mortality	Comparative (randomized or observational)	■ Data on well-resourced settings versus other settings, or on the screening interval, will be analysed if available ■ Focus on screening programmes with a high participation rate, established follow-up procedures and quality assurance strategies

The quality of evidence was rated using the GRADE framework (Grading of Recommendations Assessment, Development and Evaluation). The quality of evidence is rated down for increased risk of bias, indirectness, imprecision, publication bias and inconsistency; and is rated up for a large effect size, dose–response effect, and when all plausible confounding is considered to strengthen the association. (*1*) The small number of existing trials and the heterogeneity of the observational studies did not allow for a formal statistical evaluation of publication bias; hence, this is not used to rate down the evidence although it may have existed. When existing GRADE evidence profiles were found, these were reviewed across multiple sources and their data were verified before being adapted for this report. When GRADE profiles were unavailable, they were created de novo. Organized population-based cancer screening programmes were defined as those that target all the population at risk in a given geographical area with a high specific cancer burden and that offer the same level of screening, diagnosis and treatment services to all participants; and include quality control strategies that assure high quality of screening, assessment and therapy as well as adequate follow-up.

Systematic reviews included

The search yielded 229 citations of which 14 systematic reviews were selected (four summarized randomized trial data, eight summarized observational studies, and two summarized outcomes relevant to mammography-associated anxiety and quality of life). Reviews are described in Tables 3–5. Two additional systematic reviews were evaluated and their conclusions are presented in the results section but were not used

in the evidence tables because their data, list of included studies and conclusions greatly overlapped with other reviews.

Available mammography randomized controlled trials

Systematic reviews excluded three trials (two small ones in which mammography was used in combination with other interventions and one with a design that was inadequate for detecting the screening effect). One additional trial (Edinburgh, 1978) had a high risk of bias and was excluded by all the selected systematic reviews although its relative effect was consistent with the overall meta-analytical estimate for breast cancer mortality at 7 and 13 years – RR 0.84 [0.61, 1.17] and 0.86 [0.70, 1.05], respectively. Thus, the body of evidence used in the available systematic reviews and existing guidelines on mammography screening mainly included seven randomized controlled trials enrolling 600 000 women (Table 6). The risk of bias in these trials remains a controversial topic among experts and is likely to be moderate overall (table 6). The trials suffer indirectness to contemporary practice since the care of breast cancer (particularly surgical practice and mastectomy) has significantly changed over the last 30–40 years. Furthermore, the trials provide indirect evidence to screening mammography in settings with high participation rates, a service-based approach, and quality control strategies that assure adequate follow-up.

Areas of controversy

1. Randomized trials are old and the treatment of abnormal screening findings has changed (particularly for mastectomy), affecting the available estimates.
2. Digital mammography results may differ from film mammography.
3. The risk of bias in the trials has been quite controversial (see table 6). Trial data could not reliably exclude the effect of length time bias and bias due to evaluation during the implementation phase.
4. The adherence to mammography and follow up procedures varies across the RCTs and are lower than what is currently instituted in many countries.
5. Length of trials (average 11 years) may not be sufficient underestimating screening benefit.
6. Current observational evidence is more contemporary and described in settings with high adherence, follow up and quality measures, but is subject to biases inherent by design.

Outcome data sources

Breast cancer-specific mortality

The eight available randomized controlled trials were pooled in meta-analysis in numerous systematic reviews with varying methods (random effects, fixed effects, Bayesian methods) and stratifications (based on age, risk of bias, etc.). In general, across all ages and methods, most mortality relative estimates converge around RR of 0.80, suggesting a 20% relative risk reduction (RRR) with mammography (Tables 7 and 8) at 11 years of follow-up (however, longer follow-up has demonstrated larger magnitudes of risk reduction, suggesting that the full impact of screening may require more than 20 years). (*2*) Relative and absolute estimates derived from the randomized trials are for multiple screening rounds (median 4 rounds) over 11 years of follow-up, and are not modelled.

Heterogeneity was low (I^2<50%) in all analyses, suggesting minimal impact of the analysis model on the estimates. Clearly larger estimates of RRR were reported by observational studies. The baseline risk estimated from observational studies was also higher than that of the trials, leading to a larger absolute effect and a smaller number needed to screen. There is no known risk stratification or prediction model for women with average risk; hence, providing multiple baseline risks would not be helpful to clinicians, guideline developers or users (i.e. stratification would not be implementable).

Additional data on the screening interval from modelling and from the trials are provided in Tables 9a and 9b. Screening intervals in the randomized trials ranged from 12 to 33 months (over a median of 11 years). The confidence intervals for RR of breast cancer mortality for screening < 24 month vs. ≥24 months greatly overlap (Table 9b) (*3*) suggesting no statistically significant interaction (*4*) (i.e., no difference). However, considering the shorter duration of the median follow-up period and other issues relating to risk of bias or indirectness, the inference about the difference in breast cancer mortality between the two screening intervals is limited. Modelling studies and further analysis of trials showed results that varied on the basis of assumptions and trade-offs. (*5*,*6*)

Two additional systematic reviews evaluated breast cancer mortality. Magnes et al (*7*) pooled the randomized trials of screening mammography in women aged 39–49 years and reported a relative risk (RR) estimate of 0.83 (0.72–0.97) for breast cancer mortality. The results were consistent with previously published systematic reviews and the same trials were included in this review. Nickson et al (*8*) presented a case controlled study

from Western Australia and performed a meta-analysis of published case controlled studies. They reported odds ratios (ORs) for breast cancer mortality of 0.48 (0.38-0.59) and 0.51 (0.46-0.55); respectively. These estimates were consistent with other systematic reviews of observational studies in which the reduction in mortality was clearly higher in case controlled studies compared with cohort studies.

Health-related quality of life, quality of life, disability-adjusted life years

Data were not available to provide reliable quantitative estimates for rating the quality of evidence. Systematic reviews that focused mainly on anxiety and psychological distress associated with mammography are summarized (Table 5). Other systematic reviews described lower quality of life associated with higher stage breast cancer. (*9*,*10*)

Overtreatment

Data were not available to provide reliable quantitative estimates for rating the quality of evidence. Data were available only to provide estimates for the effect of mammography screening on the risk of receiving chemotherapy, radiotherapy and hormonal therapy (i.e. the proportion of women receiving treatment, not overtreatment) (Table 10). However, a large proportion of such treatments may be appropriate and may not be considered overtreatment.

Mastectomies

Data for mastectomy were derived from observational studies in view of the radical change in mastectomy practice since the randomized controlled trials were published. Evidence from randomized controlled trials (Table 11) showing an increased mastectomy rate associated with screening is likely to be invalid at the present time, and most recent studies in fact show the contrary.

Overdiagnosis

Numerous systematic reviews focused on overdiagnosis. Reports of the original randomized controlled trials, observational studies and modelling studies also provided their own estimates. Estimates of overdiagnosis varied greatly (from 0% to 54%) (*11*) according to the method used (i.e. incidence-based or modelling), (*12*) the source of the data (randomized controlled trials or observational studies) and the definition of overdiagnosis (at least four definitions in which excess cancers are divided

according to different denominators). (*13*) Overdiagnosis rates derived from modelling studies and population-based studies were markedly lower than the unadjusted rates from the randomized controlled trials. The adjustment for lead time and underlying trends considerably lowered estimates (6.5% of the expected incidence in the absence of screening). (*11*) One example of many other population-based studies is a well conducted and recent study that evaluated the cancer registry of British Columbia, Canada, for women aged 30–89 years between 1970 and 2009. Estimates of overdiagnosis from cumulative cancer rates among women between the ages of 40 and 89 years were: 1) when comparing participants in the screening programme with nonparticipants, 5.4% for invasive disease and 17.3% for invasive disease +DCIS, and 2) when comparing observed and predicted population rates, -0.7% for invasive disease and 6.7% for invasive disease +DCIS.14 The extent of overdiagnosis of invasive cancer was modest and primarily occurred among women aged over 60 years. (*14*) Further variations and nuances in estimating overdiagnosis rates relate to using in the analysis screened subjects versus those invited to be screened, using only invasive cancer versus all cancers (including in situ cancers in analysis), and using incidence rate versus cumulative incidence approaches. Although the association between breast cancer screening and overdiagnosis has been demonstrated consistently in all studies and is likely to be supported by high quality evidence, there is significant uncertainty about the quantitative estimates in the different age groups; thus this evidence is of low to very low quality due to its indirectness.

False positive rate

Data were available from several sources. However, the most precise estimates were available from studies evaluating large registries and national databases in Europe (*15*) with characteristics consistent with service screening. Data from North America (Table 12, USA (*16*) and Canada (*17*)) were also considered although it was unclear if such programmes had similarly organized screening practices.

All-cause mortality

In general, the observational studies of service-based screening focused on breast cancer mortality and not on all-cause mortality. The data presented for all-cause mortality (Table 13) were derived from the randomized trials.18 A concern about the accuracy/reliability of all-cause mortality estimates is the small relative contribution of breast cancer mortality to all-cause mortality as compared to various other potential sources of bias (self-selection; age distribution etc.). Another concern is the short follow-up of the available trials.

Evidence profile

Table 2. Screening mammography (including data from contemporary observational studies)

Breast cancer mortality							
Data source	Risk of bias	Inconsistency	Imprecision	Indirectness	Relative effect (95% CI)	Absolute effect per million	Quality of evidence
RCTs (details in Table 8) Average 4 screening rounds over 11 years	No serious concerns	No serious concerns	No serious concerns	Serious concerns*	0.79 (0.68–0.90)	1354 fewer (from 645 to 2064 fewer)	Moderate
Observational studies (Europe)	No serious concerns	No serious concerns	No serious concerns	No serious concerns	0.62 (0.56–0.69)**	4000† to 9000†† Number needed to invite 235† Number needed to screen 180†	Low

* Trials had short follow-up (11 years), low participation rate and practice has probably significantly changed to warrant rating down for indirectness.

** Data are from incidence-based studies in women actually screened. (*19*) The effect in women invited to screening is 0.75 (0.69–0.81).

† Modelling by the UK independent panel report (*13*) using relative risk from randomized trials (0.20) and UK observational data. The numbers of women needed to be invited/treated are for women screened 20 years starting at age 50.

†† Upper range is from Euroscreen modelling. (*20*)

Health-related quality of life

- No reliable estimates for the effect of screening mammography on health-related quality of life.
- Mammography is associated with short-term anxiety in women requiring further investigation. (*21*,*22*)
- Lower quality of life is associated with higher stage breast cancer. (*9*,*10*)

Disability-adjusted life years

- No reliable estimates for the effect of screening mammography on disability-adjusted life years.

Overtreatment

- No reliable estimates for the effect of screening mammography on overtreatment.

Mastectomies								
Data source	Risk of bias	Inconsistency	Imprecision	Indirectness	Age	Relative effect (95% CI)	Absolute effect per million	Quality of evidence
35 408 Norwegian women (pre- versus post-screening programme implementation) (*23*) ‡	Serious †	No serious concerns	No serious concerns	No serious concerns	40–49	0.65 (0.59–0.71)	381 less with screening	Very low
					50–69	**0.70 (0.66 –0.75)**	**494 less with screening**	
					70–79	0.59 (0.54–0.64)	834 less with screening	

Examples of other studies with consistent findings:

Reduction in mastectomy rates in Italy [1990 rate of 1.08/1000 (95% CI, 0.84 to 1.37) vs 1996 rate of 0.62/1000 (0.44 to 0.86)] ; [Mastectomy rates in screen detected (vs clinically detected) in Ireland were 32% vs 47%]; [Mastectomy rates in screen detected (vs clinically detected) in Australia were 41% vs 56%; P<0.05] and Germany [Mastectomy rates in 2000 of 32.6% vs in 2008: 19.6%]. (*24–28*)

† Reduction in mastectomy cannot be attributed only to screening but also to the change in surgical practice. Lead-time bias is present.
The Norwegian study was selected because it provided estimates for age groups consistent with those chosen a priori by the WHO panel. It was selected as a representative example of large observational studies (no meta-analysis was conducted). Similar trends of reduction in mastectomy rates were described in other observational studies. (*24–28*)

‡ Mastectomy rates are obtained from table 2 (page 6) and compare pre-screening rates (1993-2005) to fully implemented screening rates (2005-2008)

Overdiagnosis						
Data source	Risk of bias	Inconsistency	Imprecision	Indirectness	Rate	Quality of evidence
13 European population-based mammographic screening programmes (*11*)	No serious concerns	No serious concerns	Serious*	Serious**	Rate adjusted for breast cancer risk and lead time: 1–10%	Very low

* Estimates of overdiagnosis in younger and older women were quite disparate and may be better extrapolated from the middle age strata (50–70 years) in which somewhat more reliable estimates are available.

** There is very serious indirectness and imprecision due to the lack of clear definition, data source or methods used to estimate the magnitude of overdiagnosis quantitatively, or to adjust it (with rates from 0% to 54%). There is also indirectness due to inability to reliably stratify estimates by age as well as issues relating to digital versus film mammography. Markedly higher rates are reported in the randomized trials (although these estimates are likely to be outdated). The random-effect meta-analysis by the UK independent panel review (*13*) (with a definition of overdiagnosis as the proportion of cancers diagnosed during the screening period in women invited for screening) was 11% if the denominator included cancers diagnosed during long-term follow-up (as opposed to during screening). Some of the reported estimates of 1–10% included DCIS and some did not.

False positive rate						
Data source	Risk of bias	Inconsistency	Imprecision	Indirectness	Rate	Quality of evidence
Meta-analysis of 3 observational studies of 364 991 women (*15*)	No serious concerns	No serious concerns	No serious concerns	No serious concerns	■ Pooled rate is 19.7% ■ % Screening resulting in needle biopsy 2.2% (initial screening) and 1.1% (subsequent screening) ■ % Surgical interventions among women without breast cancer 0.19% (initial screening) and 0.07% (subsequent screening)	Low

Defined as the cumulative risk of being recalled for further assessment at least once during 10 biennial screens performed from age 50 to 51 until 68 to 69, among women without a diagnosis of breast cancer.

Table 3. Systematic reviews of randomized controlled trials (RCTs)

Author/agency	Quality (AMSTAR score)	Search date (month, year)	Systematic review appraised quality of evidence	Population
Fitzpatrick-Lewis, 2011, Canadian Task Force (*3*)	High (11/11)	October, 2010	GRADE	8 RCTs enrolling 600 000 average-risk women
Nelson HD, 2009, US Preventive Taskforce (*16*)	High (10/11)*	December, 2008	USPSTF method	8 RCTs enrolling 600 000 average-risk women
Gotzsche, 2013, Cochrane Collaboration (*18*)	High (11/11)	November, 2012	No	8 RCTs enrolling 600 000 average-risk women
The Independent United Kingdom Panel on Breast Cancer Screening, 2012 (*29*)	Low to moderate (NA)**	October, 2012	No	Relative estimates derived from 8 RCTs enrolling 600 000 average-risk women. Absolute estimate calculations used a baseline risk from United Kingdom registry data.

* It is unclear if grey literature/unpublished data were searched/included.

** The design is an umbrella literature review (overview) performed by a panel of experts; AMSTAR rating was not possible. Nevertheless, the included studies and the analysis were similar to other high-quality reviews. Furthermore, the selection criteria for the panel include no prior publication on breast cancer screening (considering the polarized views in the field), which seems an innovative method of reducing intellectual bias.

NA: not applicable.

Table 4. Systematic reviews of observational studies

Author/ agency	Quality (AMSTAR score)	Search date (month, year)	Systematic review appraised quality of evidence	Studies/ population	Outcomes
EUROSCREEN Studies (*11*,*20*,*30*)	Moderate* (6/11)	February, 2011	No	European women, likely average risk, 45 studies, trend studies (*17*), incidence-based studies (*20*) and case-control studies (*8*)	**Breast cancer mortality:** ■ Incidence-based studies: RR 0.75 (0.69–0.81) among invited women, and 0.62 (0.56–0.69) among screened women ■ Case-control studies: OR 0.69 (0.57–0.83), and 0.52 (0.42–0.65) adjusted for self-selection **Overdiagnosis:** ■ Unadjusted estimates ranged from 0% to 54% ■ Adjusted for lead time and underlying trends: 6.5% of the expected incidence in the absence of screening
Njor, 2012 (*31*) (part of EUROSCREEN)	Moderate¶ (6/11)	NR	No	European women, likely average risk, 16 studies	Breast cancer mortality associated with "service screening", which is screening in routine health care Based on the comparison group: 1. women not yet invited: RRs 0.76–0.81 2. historical data from the same region as well as from historical and current data from a region without screening: RRs 0.75–0.90 3. historical comparison group combined with data for nonparticipants: RRs 0.52–0.89 Study databases overlapped in Finnish and Swedish studies; adjustment for lead time was not optimal in all studies
Autier, 2011 (*32*)	Low to moderate** (5/11)	June, 2009	No	■ 8 studies from Australia, Italy, Netherlands, Norway, Switzerland and USA ■ Cancer registries from other countries ■ Screening coverage >60%	**Incidence trend of advanced breast cancer:** Age-adjusted annual percentage changes were stable or increasing in 10 areas, and had transient downward trends followed by increases back to prescreening rates in the remaining four areas. These trends were not supportive of a substantial role for screening on mortality.

Author/ agency	Quality (AMSTAR score)	Search date (month, year)	Systematic review appraised quality of evidence	Studies/ population	Outcomes
Breast Cancer Surveillance Consortium data (reported in Nelson HD, 2009, US Preventive Taskforce (*16*))	NA	NA	USPSTF method	600 830 women aged 40 years or older undergoing routine mammography screening from 2000 to 2005 ■ Results are based on a single screening round	**Rate of invasive breast cancer:** Lowest among women aged 40–49 years (2.7 per 1000 women per screening round) and increases with age. **Rate of DCIS:** Lowest among women aged 40–49 years (0.9 per 1000 women per screening round), increases for women aged 50–59 years (1.4 per 1000 women per screening round), and remains at approximately this level for older women. **False-positive mammography results:** Rate is highest among women aged 40–49 years (97.8 per 1000 women per screening round) and declines with each subsequent age decade. Common to all age groups. **False-negative mammography results:** Rate is lowest among women aged 40–49 years (1.0 per 1000 women per screening round) and increases slightly with subsequent age decades. **Sensitivity, recall rates, and cancer detection:** Rates increase as the months since previous mammography increase, whereas specificity decreases.
The Independent United Kingdom Panel on Breast Cancer Screening, 2012 (*29*)	Low to moderate (NA)***	October, 2012	No	21 case-control studies	**Breast cancer mortality:** Odds ratios ranged from 0.30 to 0.92 but were <0.60 in most studies

Author/ agency	Quality (AMSTAR score)	Search date (month, year)	Systematic review appraised quality of evidence	Studies/ population	Outcomes
Jorgensen, 2009 (*33*)	Moderate (8/11)‡	April, 2007	5 studies reporting incidence data covering at least seven years before screening and seven years after screening (Australia, Canada, United Kingdom and Europe) of publicly organized screening programmes	**Overdiagnosis:** Estimated rate 52% (46–58%) One in three breast cancers detected in a population offered organized screening is overdiagnosed.

* It was unclear if the review was done by duplicates. The review is restricted to the English language. There was no assessment of publication bias, evaluation of grey literature or clear description of excluded studies.

** It was unclear if the review was done by duplicates. There was no assessment of publication bias or evaluation of grey literature. Methods are not well described in general and the search strategy is not reported in sufficient detail.

*** The design is an umbrella literature review (overview) performed by a panel of experts; AMSTAR rating was not possible.

¶ The review searched only PubMed. The search date is not reported. There was no assessment of publication bias, evaluation of grey literature or clear description of excluded studies.

‡ The review searched only PubMed, with no clear evaluation of grey literature or of the quality of individual studies.

NR: not reported. NA: not applicable.

Table 5. Systematic reviews evaluating psychological impact of mammography

Systematic review	Studies	Main findings
Brett, 2005 (*21*)	54 studies from 13 countries, mostly published after 1990 (coinciding with routine mammographic screening implementation)	■ Mammographic screening does not appear to create anxiety in women who are given a clear result after a mammogram. ■ Women who require further investigations following screening experience significant short-term anxiety. ■ Factors associated with the adverse psychological impact of mammographic screening include: younger age, lower education, urban residence, manual occupation, having one or no children, waiting time between recall letter and recall appointment, pain during screening, and previous false-positive result.
Brewer, 2007 (*22*)	23 eligible studies (313 967 participants)	■ Women who received false-positive results on mammography screening had higher, but not apparently pathologically elevated, levels of distress and anxiety and thought more about breast cancer than did those with normal results.

Table 6. Mammography screening randomized controlled trials

	New York HIP	Malmö I and II	Swedish Two-county	Canada I and II	Stockholm	Göteborg	United Kingdom Age trial	Edinburgh*
Start date	1963	1976	1977	1980	1981	1982	1991	1978
Randomization method	Individual	Individual	Cluster	Individual	Day of birth	Day of birth	Individual	Cluster
Number of women (clusters)	62 000	60 076	133 065 (45)	89 835	60 800	52 222	160 921	54 654 (87)
Age (years)	40–64	45–69 and 43–49	38–75	40–49 and 50–59	39–65	39–59	39–41	45–64
Mammography								
Number of views	2	2 then 1 or 2	1	2	1	2 then 1	2 then 1	2 then 1
Screening interval (months)	12	18–24	24–33	12	24–28	18	12	24
Number of screening rounds	4	6–8	2–4	4–5	2	4–5	8–10	2–4
Duration of screening (years)	3	12	7	5	4	7	8	6
Attendance	65%	74%	85%	88%	82%	84%	81%	65%

Adapted from the evidence synthesis report by the Independent United Kingdom Panel on Breast Cancer Screening. (*13*)

* Excluded from the meta-analysis.

Risk of bias in mammography randomized controlled trials (RCTs)

1. Changes in standards for reporting trial methodology have affected the ability to judge adequately the risk of bias in most of these studies (despite attempts to contact authors by Cochrane reviewers and others).
2. The primary reasons for increased risk of bias (additional reasons exist and are summarized in several systematic reviews):
   - Göteborg, 1982: randomization ratios vary with age, biasing results towards exaggerated benefits; inadequate reporting of baseline data and other methodological information.
   - New York, 1963: poor description of methods as the trial started enrolment in 1963; probable lack of comparability of the two groups; unblinded outcome assessment.
   - Stockholm, 1981: probable dependence and overlap between the two subtrials; inadequate reporting; unclear blinded outcome assessment.
   - Two-county, 1977: conflicting information in various trial reports; probable noncomparable groups and unblinded outcome assessment.
   - Edinburgh, 1978: cluster randomization inadequate, with some clusters subsequently switched; variations in execution and poor reporting.
3. Meta-analyses demonstrated that, in general, trials at low risk of bias had smaller relative risk reduction of breast cancer mortality than those with increased susceptibility for bias. However, these estimates had greatly overlapping confidence intervals (e.g. mortality at 13 years was 0.90 [0.79, 1.02] in trials with low risk of bias versus 0.75 [0.67, 0.83] in trials with unclear or moderate risk of bias and versus 0.86 [0.70, 1.05] for the one trial excluded for high risk of bias).
4. The risk of bias designation (ie, adequately randomized vs. suboptimally randomized) is controversial and often disputed.

Table 7. Convergence of relative estimates of breast cancer mortality across various reviews

Systematic review	Overall RR (95% CI)
United Kingdom review	
13-year follow-up in trials reported in the Cochrane Review, random-e ects meta-analysis	0·80 (0·73–0·89)
Cochrane Review	
Fixed-effect meta-analysis of the above trials	0·81 (0·74–0·87)
Excluding women <50 years	0·77 (0·69–0·86)
Trials considered adequately randomized	0·90 (0·79–1·02)
Trials deemed suboptimally randomized	0·75 (0·67–0·83)
Overall considered by the authors as an average	0·85
USA Taskforce	
RR 0·86 (95% CI 0·75–0·99) for women aged 50–59 years, and RR 0·68 (0·54– 0·87) for those aged 60–69 years. These estimates have an inverse-variance weighted average RR of 0·81.	0·81
Canadian Task Force	
Overall effect	0·79 (0·68–0·90)
Duffy et al., 201234	
Overall effect	0·79 (0·73–0·86)

Source: United Kingdom Independent Panel Review. (*13*)

Table 8. Evidence profile: Breast cancer mortality in randomized controlled trials (median follow-up of about 11 years)

Cornelia Sepúlveda
World Health Organization
Geneva, Switzerland

Data source	Risk of bias	Inconsistency	Imprecision	Indirectness	Screening number of events (%)	Control number of events (%)	RR (95% CI)	Risk diff. per 1 million women (95% CI)	Quality of evidence
Ages 39–49 years									
8 RCTs	Not serious*	Not serious	Not serious	Serious††	n = 152 300 448 (0.29)	n = 195 919 625 (0.32)	0.85 (0.75–0.96)	474 fewer (from 115 to 792 fewer)	Moderate
Ages 50–69 years									
7 RCTs	Not serious*	Not serious	Not serious	Serious††	n = 135 068 639 (0.47)	n = 115 206 743 (0.64)	0.79 (0.68–0.90)	1354 fewer (from 645 to 2064 fewer)	Moderate
Ages 70–74 years									
2 RCTs	Not serious*	Not serious	Serious†	Serious††	n = 10 339 49 (0.47)	n = 7 307 50 (0.68)	0.68 (0.45–1.01)	2218 fewer (from 3734 fewer to 39 more)	Low

* We did not rate down for the risk of bias although there were some concerns. Sensitivity analysis based on trial quality does not show a significant change in conclusions. For ages 39–49 years (five quasi-randomized and three truly randomized trials). Blinding and allocation concealment were not clear for some of the studies.

† Although the sample size is large, the number of events is <300 and the confidence interval includes possible harm.

†† Considering the lower participation rate in screening and the fact that trials are outdated with significant changes to breast cancer treatment and mammography technique, evidence is indirect to contemporary service-based screening.

Notes:

Estimates of relative risk are based on a random-effects model, as reported by the Canadian Task Force Breast Cancer Screening. (*3*,*35*) These estimates are consistent with those from meta-analyses by a Cochrane systematic review (*36*), a systematic review from the Agency for Healthcare Research and Quality prepared for the United States Preventive Services Taskforce (USPSTF) (*37*), and an evidence synthesis report by the Independent United Kingdom Panel on Breast Cancer Screening. (*13*) The USPSTF divided the 50–69 years stratum further into two (50–59 years and 60–69 years) strata with overlapping CI of RRs of 0.86 (0.75–0.99) and 0.68 (0.54–0.87) respectively.

In all analyses, I-squared is less than 50% and the heterogeneity test is not significant.

Quality of evidence is not rated down for indirectness, although there is some concern about the age of the trials (start dates: 1963, 1976, 1977, 1980, 1981, 1982, 1991), as to whether film and digital mammography are interchangeable, and about the adequacy of the length of follow-up to ascertain screening benefits.

Assessment of publication bias is not statistically possible due to the small number of trials included.

The Swedish two-county trial is counted as two trials (Kopparberg and Östergötland).

Table 9a. Screening interval and breast cancer mortality: data from modelling

Description	Assumptions	Main findings
The six models were developed independently[†] within the Cancer Intervention and Surveillance Modeling Network of the National Cancer Institute	■ Assumes a cohort of women born in 1960 and followed from the age of 25 years through their entire lives. ■ All six models use a common set of age-specific variables for breast cancer incidence, mammography test characteristics, treatment algorithms and effects, and non-breast cancer competing causes of death.	■ The six models produced consistent rankings of screening strategies. ■ Screening biennially maintained an average of 81% of the benefit of annual screening with almost half the number of false-positive results. ■ Screening biennially from age 50 years to 69 years achieved a median 16.5% reduction in breast cancer deaths versus no screening. ■ Biennial screening at age 40 years (versus 50 years) reduced mortality by an additional 3%, consumed more resources, and yielded more false-positive results. ■ Biennial screening after age 69 years yielded some additional mortality reduction in all models, but overdiagnosis increased most substantially at older ages.

Data from Mandelblatt et al. (*5*)

[†] Erasmus Medical Center, Rotterdam, Netherlands; Georgetown University Medical Center, Washington, DC, USA and Albert Einstein College of Medicine, New York, NY, USA; M.D. Anderson Cancer Center, Houston, TX, USA; University of Wisconsin, Madison, WI, USA and Harvard Medical School, Boston, MA, USA; Stanford University, Palo Alto, CA, USA; Dana-Farber Cancer Institute, Boston, MA, USA. Two models include only invasive cancer and four also include DCIS.

Table 9b. Screening interval and breast cancer mortality: data from randomized controlled trials

	Screening interval < 24 months		Screening interval ≥ 24 months		
Age	Number of trials	RR (95% CI)	Number of trials	RR (95% CI)	Quality of evidence
39–49	5	0.82 (0.72–0.94)	3	1.04 (0.72–1.50)	Low[†]
50–69	4	0.86 (0.75–0.98)	3	0.67 (0.51–0.88)	Low[†]
≥ 70*	–	–	2	0.68 (0.45–1.01)	Low[†]
All ages	6	0.83 (0.76–0.92)	3	0.77 (0.58–1.03)	Low[†]

* No trial data.

[†] Concerns about indirectness, risk of bias and imprecision (only for the >24 month estimates)

Notes:

Range of screening intervals across RCTs was 12–33 months.

The Swedish two-county trial is counted as two trials (Kopparberg and Östergötland).

Table adapted from the Canadian Task Force guideline evidence summary. (*3*,*35*)

Table 10. Evidence profile for treatment type (proportion of women receiving treatment)

Data source	Risk of bias	Inconsistency	Imprecision	Indirectness	Screening number of events (%)	Control number of events (%)	RR (95% CI)	Risk diff. per 1 million women (95% CI)	Quality of evidence
Overtreatment**									
Women treated with radio-therapy	Serious‡‡	Not serious	Not serious	Very serious‡	260/21 242 (1.22%)	209/21 244 (0.98%)	1.24 [1.04, 1.49]	2361 more (from 394 more to 4821 more)	Very low
Women treated with chem-otherapy	Serious‡‡	Not serious	Not serious	Very serious‡	26/21 242 (0.12%)	41/21 244 (0.19%)	0.63 [0.39, 1.04]	714 fewer (from 1177 fewer to 77 more)	Very low
Women treated with hormonal therapy	Serious‡‡	Not serious	Not serious	Very serious‡	80/21 242 (0.38%)	99/21244 (0.47%)	0.81 [0.60, 1.08]	885 fewer (from 1864 fewer to 372 more)	Very low

* Data derived from the adequately randomized trials as reported in a Cochrane systematic review.36

** Reliable estimates of overtreatment are not available. Data on treatment with radiotherapy, chemotherapy and hormonal therapy are provided as a surrogate.

† Despite the nonsignificance of the relative effect, evidence is not rated down for imprecision because of the large number of events and the large sample size.

‡ Evidence is downgraded due to severe indirectness as overtreatment estimates are not available and trials are outdated (change in treatment pattern and practice).

‡‡ Bias is likely as treatment is mostly affected by treatment patterns and provider decisions and is not solely attributable to screening.

Table 11. Evidence profile. Mastectomy data from randomized controlled trials (median follow-up of about 11 years)

Data source	Risk of bias	Inconsistency	Imprecision	Indirectness	Screening number of events (%)	Control number of events (%)	RR (95% CI)	Risk diff. per 1 million women (95% CI)	Quality of evidence
Age 40–49 years									
1 RCT	Serious ‡‡	Not serious	Not serious†	Very serious‡	183/25 214 (0.73%)	157/25 216 (0.62%)	1.17 (0.94, 1.44)	1995 more (from 819 more to 3297 more)	Very low
Age 45–69 years									
7 RCTs	Serious ‡‡	Not serious	Not serious	Very serious**‡	621/40 953 (1.52%)	515/40 938	1.21 (1.07–1.36)	2646 more (from 882 more to 4536 more)	Very low

* Two other quasi-RCTs are not included in this evidence profile; the relative effect produced by these studies is similar, i.e. RR 1.21 (1.06 to 1.38).

† Despite the nonsignificance of the relative effect, evidence is not rated down for imprecision because of the large number of events and large sample size.

‡ The practice and context of mastectomy has significantly changed since the time of the trials.

‡‡ Bias is likely as treatment is mostly affected by treatment patterns and provider decisions and is not solely attributable to screening.

** Evidence can be considered indirect when applied to specific age groups. Estimates for age groups do overlap, however, suggesting no significant difference in relative effect (Canada 1 RCT age 40–49 years, RR = 1.17 [0.94, 1.44]; Canada 2 RCT age 50–59 years, RR = 1.12 [0.91, 1.37]; Malmo 2 RCT age 45–69 years, RR = 1.25 [1.09, 1.44]).

Notes:

Data were unavailable to provide reliable estimates for the predefined age groups (39–49 years, 60–69 years, >70 years).

Data adapted from the Canadian Task Force Breast Cancer Screening evidence summary (*3*,*35*) and from a Cochrane systematic review. (*36*)

Table 12a. Evidence profile. False positive rate per single screening round in North America

(Observed false positive rates in North America differ from those demonstrated in European programmes; see Hofvind et al)

Breast cancer mortality						
Data source	Risk of bias	Inconsistency	Imprecision	Indirectness	False positive rate per 1 million women screened	Quality of evidence
Age 39–49 years						
Observational study BCSC‡	Not serious	Not serious	Not serious	Not serious	98 000	Low
Age 50–69 years (*13*)						
Observational study BCSC‡	Not serious	Not serious	Not serious	Not serious	92 000	Low
Age 70–74 years						
Observational study BCSC‡	Not serious	Not serious	Not serious	Not serious	69 000	Low

† Data source: BCSC: Breast Cancer Surveillance Consortium. (*16*)

Table 12b. Evidence profile. False positive rate per four screening mammographies over 11 years, in North America

(Observed false positive rates in North America differ from those demonstrated in European programmes; see Hofvind et al)

Breast cancer mortality						
Data source	Risk of bias	Inconsistency	Imprecision	Indirectness	False positive rate per 1 million women screened	Quality of evidence
Age 39–49 years						
Observational study, Canada¶	Not serious	Not serious	Not serious	Not serious	330 000	Low
Age 50–69 years (*13*)						
Observational study, Canada¶	Not serious	Not serious	Not serious	Not serious	280 000	Low
Age 70–74 years						
Observational study, Canada¶	Not serious	Not serious	Not serious	Not serious	210 000	Low

¶ Data source: Organized Breast Cancer Screening Programs in Canada - Report on Program Performance in 2005 and 2006. (*17*)

Table 13. Evidence profile. All-cause mortality (median follow-up of about 11 years), data from randomized trials

Data source	Risk of bias	Inconsistency	Imprecision	Indirectness	Screening number of events (%)	Control number of events (%)	RR (95% CI)	Risk diff. per 1 million women (95% CI)	Quality of evidence
Age 39–49 years									
2 RCTs	Not serious	Not serious	Not serious*	Very serious†	1 373/79 098 (1.74%)	2 388/132 172 (1.81%)	RR 0.97 (0.91–1.04)	484 fewer (from 1 615 fewer to 726 more)	Low
Age 50–69 years									
1 RCT	Not serious	Not serious	Not serious*	Very serious†	734/19 711 (3.72%)	690/19 694 (3.50%)	RR 1.06 (0.96–1.18)	2 204 more (from 1 408 fewer to 6 201 more)	Low

† Rated down twice due to indirectness due to the fact that trials are old, and had low participation rate and shorter follow-up than what is expected for all-cause mortality benefit.

* Confidence interval includes benefit and harm. However, the sample size is very large and the number of events is high, suggesting no imprecision of estimates.

Adapted and modified from the Canadian Task Force Breast Cancer Screening evidence summary. (*3*,*35*)

References

1. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. Apr 2011;64(4):401-406.
2. Tabar L, Vitak B, Chen TH, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology*. Sep 2011;260(3):658-663.
3. Fitzpatrick-Lewis D, Hodgson N, Ciliska D, Peirson L, Gauld M, Liu Y. Breast *Cancer Screening*. Canadian Task Force. 2011. http://canadiantaskforce.ca/wp-content/uploads/2012/09/Systematic-review.pdf?9d7bd4.
4. Altman DG, Bland JM. Interaction revisited: the difference between two estimates.[see comment]. *Bmj*. Jan 25 2003;326(7382):219.
5. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Intern Med*. Nov 17 2009;151(10):738-747.
6. The frequency of breast cancer screening: results from the UKCCCR Randomised Trial. United Kingdom Co-ordinating Committee on Cancer Research. *Eur J Cancer*. Jul 2002;38(11):1458-1464.
7. Magnus MC, Ping M, Shen MM, Bourgeois J, Magnus JH. Effectiveness of mammography screening in reducing breast cancer mortality in women aged 39-49 years: a meta-analysis. *J Womens Health* (Larchmt). Jun 2011;20(6):845-852.
8. Nickson C, Mason KE, English DR, Kavanagh AM. Mammographic screening and breast cancer mortality: a case-control study and meta-analysis. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. Sep 2012;21(9):1479-1488.
9. Howard-Anderson J, Ganz PA, Bower JE, Stanton AL. Quality of life, fertility concerns, and behavioral health outcomes in younger breast cancer survivors: a systematic review. *J Natl Cancer Inst*. Mar 7 2012;104(5):386-405.
10. Majewski JM, Lopes AD, Davoglio T, Leite JC. [Quality of life of women recovering from breast cancer after being subjected to mastectomies compared with those who had conservative surgery: a review of the literature]. *Cien Saude Colet*. Mar 2012;17(3):707-716.
11. Puliti D, Duffy SW, Miccinesi G, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen*. 2012;19 Suppl 1:42-56.
12. Etzioni R, Gulati R, Mallinger L, Mandelblatt J. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. *Ann Intern Med*. Jun 4 2013;158(11):831-838.
13. The benefits and harms of breast cancer screening: an independent review. *Lancet*. Nov 17 2012;380(9855):1778-1786.

14. Coldman A, Phillips N. Incidence of breast cancer and estimates of overdiagnosis after the initiation of a population-based mammography screening program. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. Jun 10 2013.

15. Hofvind S, Ponti A, Patnick J, et al. False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. *J Med Screen*. 2012;19 Suppl 1:57-66.

16. Nelson HD TK, Naik A, Bougatsos C, Chan B, Nygren P, Humphrey L. *Screening for Breast Cancer: Systematic Evidence Review Update for the U.S. Preventive Services Task Force. Evidence Review Update No. 74. AHRQ Publication No. 10-05142-EF-1*. Rockville, MD: Agency for Healthcare Research and Quality; 2009.

17. Public Health Agency of Canada. Organized breast cancer screening programs in Canada: report on program performance in 2005 and 2006. Ottawa, ON: Public Health Agency of Canada. Report No.37 HP32-1/2004E-PDF. Available at: http://phac-aspc.gc.ca/cd-mc/publications/cancer/obcsp-podcs05/index-eng.php. 2011.

18. Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev*. Jun 4 2013;6:CD001877.

19. Broeders M, Moss S, Nystrom L, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen*. 2012;19 Suppl 1:14-25.

20. Paci E. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen*. 2012;19 Suppl 1:5-13.

21. Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. *Psychooncology*. Nov 2005;14(11):917-938.

22. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. *Ann Intern Med*. Apr 3 2007;146(7):502-510.

23. Suhrke P, Maehlen J, Schlichting E, Jorgensen KJ, Gotzsche PC, Zahl PH. Effect of mammography screening on surgical treatment for breast cancer in Norway: comparative analysis of cancer registry data. *Bmj*. 2011;343:d4692.

24. Paci E, Duffy SW, Giorgi D, et al. Are breast cancer screening programmes increasing rates of mastectomy? Observational study. *Bmj*. Aug 24 2002;325(7361):418.

25. Zorzi M, Puliti D, Vettorazzi M, et al. Mastectomy rates are decreasing in the era of service screening: a population-based study in Italy (1997-2001). *Br J Cancer*. Nov 6 2006;95(9):1265-1268.

26. Samnakay N, Tinning J, Ives A, et al. Rates for mastectomy are lower in women attending a breast-screening programme. *ANZ J Surg*. Nov 2005;75(11):936-939.

27. Dillon MF, Hill AD, Quinn CM, et al. Surgical intervention in screen-detected patients versus symptomatic patients with breast cancer. *J Med Screen*. 2004;11(3):130-134.

28. Schrodi S, Braisch U, Schenkirsch G, et al. [Changes in Therapy for Breast Cancer Patients as a Result of Mammography Screening. An Analysis of Bavarian Cancer Registry Data from 2000 to 2008.]. *Gesundheitswesen*. Feb 12 2013.

29. The Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *A report jointly commissioned by Cancer Research UK and the Department of Health (England).* 2012.

30. Broeders M, Moss S, Nystrom L, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen*. 2012;19 Suppl 1:14-25.

31. Njor S, Nystrom L, Moss S, et al. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. *J Med Screen*. 2012;19 Suppl 1:33-41.

32. Autier P, Boniol M, Middleton R, et al. Advanced breast cancer incidence following population-based mammographic screening. *Annals of Oncology*. 2011;22 (8):1726-1735.

33. Jorgensen KJ, Gotzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *Bmj*. 2009;339:b2587.

34. Duffy S, Yen M-FA, Chen T, et al. Long-term benefits of breast screening. *Breast Cancer Management*. 2012;1(1):31-38.

35. The Canadian Task Force on Preventive Health Care. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. *Cmaj*. 22 Nov 2011;183 (17):1991-2001.

36. Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane Database Syst Rev*. 2011(4).

37. Nelson HD, Tyne K, Naik A, et al. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2009 Nov 17 2009;151(10):727-737.

Search strategy

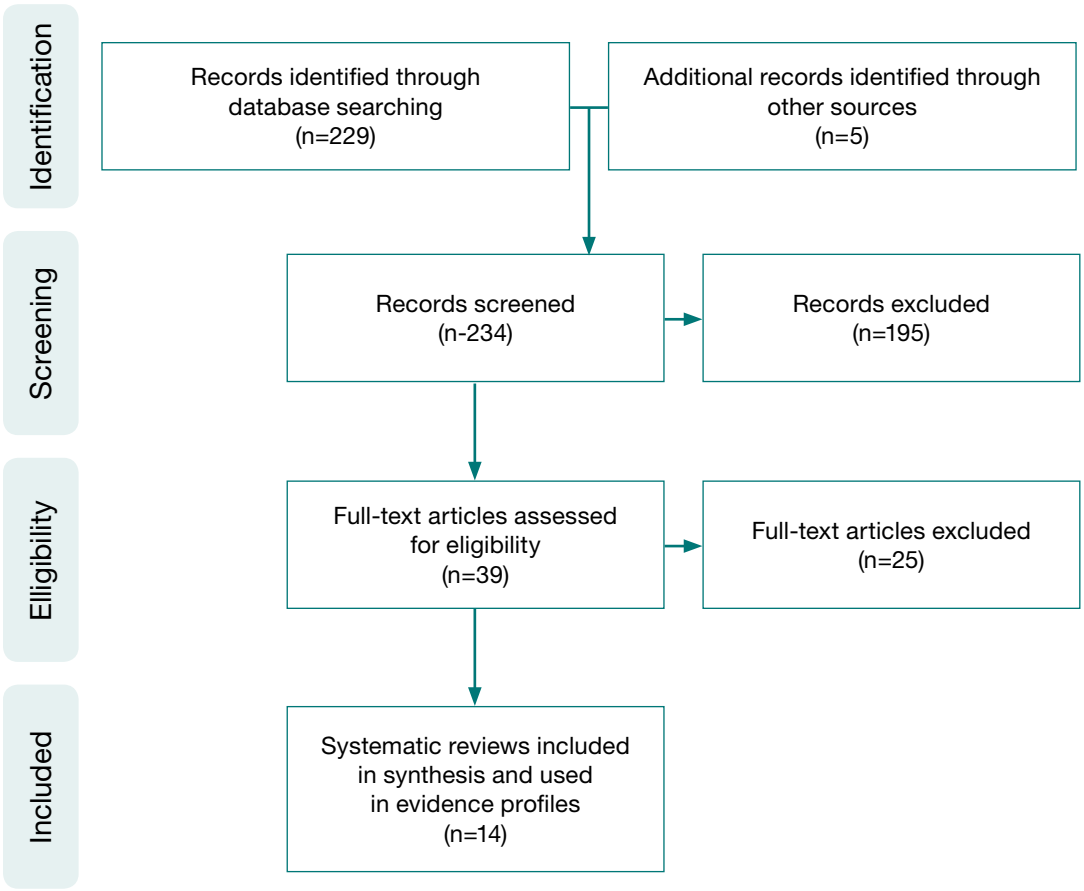
Database(s): Embase 1988 to 2013 Week 03, Ovid MEDLINE(R) In-process & other non-indexed citations and Ovid MEDLINE(R) 1946 to present, EBM Reviews – Cochrane Database of Systematic Reviews 2005 to December 2012

Search strategy:

#	Searches	Results
1	exp Mammography/	56475
2	(mammograph* or mammilloscop* or mastograph* or echomammograph* or galactograph* or scintimammograph* or xeromammograph*).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, kf, ps, rs, ui, tx, ct]	65662
3	1 or 2	65662
4	exp Mass Screening/	221308
5	anonymous testing.mp.	870
6	exp Physical Examination/	1058010
7	exp periodic medical examination/	1609
8	exp screening/	431518
9	exp Population Surveillance/	178654
10	exp DISEASE SURVEILLANCE/	7797
11	((periodic adj2 (examination* or checkup*)) or screening* or rescreening* or surveillance).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, kf, ps, rs, ui, tx, ct]	1080569
12	exp screening/	431518
13	or/4-12	2227271
14	3 and 13	29880
15	exp Treatment Outcome/	1461814
16	exp "Quality of Life"/	325704
17	exp "Costs and Cost Analysis"/	377175
18	exp "cost benefit analysis"/	112889
19	exp Program Evaluation/	1763886
20	exp Prognosis/	1333189
21	exp incidence/	389479
22	exp Prevalence/	500592
23	exp Survival Analysis/ or exp Survival/ or exp Survival Rate/	748016

#	Searches	Results
24	exp Mortality/	757729
25	exp risk/	1984633
26	exp morbidity/	502739
27	(outcome$ or effectiveness or efficacy or (quality adj2 life) or mortality or death or morbidity or cost or costs or economic* or financ* or survival or outcome* or prognosis or prognoses or incidence* or prevalence* or evaluat* or risk or risks).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, kf, ps, rs, ui, tx, ct]	12982369
28	exp death/	448107
29	or/15-28	13502502
30	14 and 29	22339
31	from 30 keep 22300-22339	40
32	exp "systematic review"/	56635
33	(systematic* adj3 review*).mp. [mp=ti, ot, ab, sh, hw, tn, dm, mf, dv, kw, nm, kf, ps, rs, an, ui, tx, ct]	136296
34	32 or 33	136296
35	30 and 34	328
36	limit 35 to (book or book series or editorial or erratum or letter or addresses or autobiography or bibliography or biography or comment or dictionary or directory or interactive tutorial or interview or lectures or legislation or news or newspaper article or patient education handout or periodical index or portraits or published erratum or video-audio media or webcasts) [Limit not valid in Embase,Ovid MEDLINE(R),Ovid MEDLINE(R) In-Process,CDSR; records were retained]	46
37	35 not 36	282
38	31 or 37	322
39	remove duplicates from 38	238
40	limit 39 to note [Limit not valid in Ovid MEDLINE(R),Ovid MEDLINE(R) In-Process,CDSR; records were retained]	139
41	from 40 keep 1-3	3
42	39 not 41	235

PRISMA flow diagram and checklist



Section/topic	#	Checklist item	Reported on page
Title			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1, 3
Abstract			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	NA
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
Methods			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g. web address) and, if available, provide registration information including registration number.	NA
Eligibility criteria	6	Specify study characteristics (e.g. PICOS, length of follow-up) and report characteristics (e.g. years considered, language, publication status) used as criteria for eligibility, giving rationale.	3, 4
Information sources	7	Describe all information sources (e.g. databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	32
Study selection	9	State the process for selecting studies (i.e. screening, eligibility, included in systematic review and, if applicable, included in the meta-analysis).	4
Data collection process	10	Describe method of data extraction from reports (e.g. piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g. PICOS, funding sources) and any assumptions and simplifications made.	4–6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	10 Tables 3 and 4
Summary measures	13	State the principal summary measures (e.g. risk ratio, difference in means).	Table 2
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I2) for each meta-analysis.	Table 2

World Health Organization

20 Avenue Appia
CH-1211 Geneva 27
Switzerland
www.who.int/

ISBN 978 92 4 150793 6