School of Epidemiology and Public Health

Knowledge Synthesis Group

Breast cancer screening:
Part A. An evidence report to inform an update of the Canadian Task Force on Preventive Health Care 2011 Guideline

October 2017 (Final)

Table 8 in the evidence report, which presents information on false positives and biopsies on false positives, was revised at the request of the Task Force guideline panel after the evidence report was finalized to present information on these outcomes over a median seven-year screening period. This change was requested because the median length of the screening period in the randomized controlled trials was 7 years. The revised table is available on the Task Force website at: www.canadiantaskforce.ca.

Principal Investigators:

David Moher, Julian Little

Centre Staff:

Pauline Barbeau, Adrienne Stevens, Andrew Beck, Becky Skidmore

Knowledge Synthesis Group, Ottawa Methods Centre

Ottawa Hospital Research Institute

Ottawa, Ontario, Canada

Clinical Experts and Methodological Collaborators:

Angel Arnaout, Muriel Brackstone, Alexandra Ginty, Amanda Hey, Brian Hutton, Beverley Shea

CTFPHC Working Group Chair:

Scott Klarenbach

PHAC Scientific Officer:

Nicki Sims-Jones, Susan Courage, Dana Reid

CTFPHC Working Group Members:

Brett Thombs, Harminder Singh, Gabriela Lewin, Marcello Tonelli, Guylene Thériault

Author Contribution
PB, AB, AS and JP drafted the protocol, BS developed the search strategy and provided text for the protocol and report. JL and DM contributed to discussion with the CTFPHC and PHAC on the scope for this work and critically reviewed protocol and report including all methodological considerations and the Centre's approach for integrating existing systematic reviews. PB and AB conducted screening, data extraction and quality assessment for the report. PB and AS produced the report, and all authors reviewed the report before it was finalized. The clinical and methodological experts provided expertise when needed and reviewed the report.

Acknowledgements
We would like to acknowledge the contribution of the following individuals for their clinical and methodological input as well as their peer review of the protocol and report:

Evidence Centre Additional Support Staff
Raymond Daniel (managing citations). Tara MacGregor (additional reviewer for full text screening for updated search)

CTFPHC Working Group
Scott Klarenbach (Chair), Brett Thombs, Harminder Singh, Gabriela Lewin, Marcello Tonelli, Guylene Thériault

PHAC Global Health & Guidelines Division Scientific Research Officers
Alejandra Jaramillo Garcia, Nicki Sims-Jones, Susan Courage, Dana Reid

External Reviewers
Anna M. Chiarelli, PhD, MHSc
Senior Scientist, Prevention and Cancer Control, Cancer Care Ontario
Provincial Lead Scientist, Ontario Breast Screening Program, Cancer Care Ontario
Professor, Dalla Lana School of Public Health, University of Toronto

Jennifer Payne, PhD
Associate Professor, Department of Diagnostic Radiology, Dalhousie University

Declaration of funding
Funding for the protocol and evidence update was provided by the Public Health Agency of Canada. This funding supported the collection of the data, data management, analyses, and writing of the protocol and the technical report.

Role of funder
The funder provided feedback on the protocol and draft evidence report, but was not involved in the study selection, data extraction, analyses, or interpretation of results and will not be involved in a decision to seek publication.

Registered Protocol: Part A- PROSPERO CRD# 42017051498

ABSTRACT

Background and Purpose:

The objectives of this evidence report were to synthesize up-to-date evidence on the benefits and harms of breast cancer screening for women who are ≥ 40 years old and not at high risk by conducting a modified overview of selected systematic reviews and an updated search for more recent primary studies. The findings will be used by the CTFPHC, along with additional considerations of feasibility, acceptability, affordability, and equity, to change or reaffirm previous recommendations.

Data Sources:

Eighteen pre-identified systematic reviews were considered for further assessment for inclusion for the overview of reviews. The search for primary studies was conducted from October 2014 to January 2017 for all screening modalities, except breast-self exam, which was from October 2010. For primary studies, we searched MEDLINE and the Cochrane Library and also supplemented with various grey literature sources. For false-positives, we relied on the Canadian Partnership against Cancer (CPAC) (cycle 2011-2012) report to calculate false-positive mammograms and biopsies on false positive data.

Study Selection:

The population of interest were women aged 40 or older who were not at high risk for breast cancer. The screening modalities of interest were the following compared to usual care/no screening: (i) mammography +/- clinical breast exam/breast self-exam; (ii) MRI +/- clinical breast exam/breast self-exam; (iii) ultrasound +/- clinical breast exam/breast self-exam; (iv): clinical breast exam alone; (v) breast self-exam alone. The outcomes of interest were breast cancer and all-cause mortality, overdiagnosis, and false positive mammograms and biopsies on false positives. Only randomized controlled trials (RCTs) were considered for mortality outcomes (breast cancer and all-cause). RCTs, controlled clinical trials and cohort studies were considered for the outcome of overdiagnosis.

For the overview, we had selected the SRs meeting specific methodological criteria, of moderate to high AMSTAR score, currency of the included evidence, and the best fit of data for each intervention-outcome pairing of relevance. Systematic reviews were evaluated by one reviewer and verified by a second person, and primary studies were reviewed by two independent reviewers. Disagreements were resolved by consensus.

Data Abstraction and Analysis:

For the overviews, information was extracted at face value as to how it was reported. This was done by a single reviewer. For the primary studies, data was extracted by a single reviewer and verified by a second reviewer. Disagreements were resolved by consensus. For the mortality outcomes, studies from the overviews were combined with the updated data in a meta-analysis, when feasible. For overdiagnosis and false-positive data, narrative syntheses were provided. The Cochrane Risk of Bias tool and the Newcastle-Ottawa scale was used for quality assessment. The rating of the quality of evidence was done using GRADE.

Results:

Three systematic reviews were included in the overviews: USPSTF 2016, CTFPHC 2011, and ACS 2014. The USPSTF 2016 review was used to inform the existing studies for breast-cancer and all-cause mortality, CTFPHC 2011 was used for breast self-examination, and all three provided evidence on overdiagnosis. For the updated search of primary studies, a total of three publications (reporting updated RCT data) were included, one of which reported on updated results for multiple trials. Across reviews and additional primary studies, a total of 20 reports addressing 16 unique studies (6 parallel group randomized, 4 cluster-randomized, 2 quasi-randomized and 4 cohort) were included, in addition to the CPAC report which informed the false positive calculations. Of the 16 studies, 14 addressed mammography and 2 breast self-exam.

Mammography was shown to decrease the risk of breast cancer mortality among cases identified during the study's screening period (short-case accrual- median follow-up 23 years), RR 0.85 (0.78-0.93). Decreases were also observed overall when including all cases identified during the follow-up period (long-case accrual-median follow-up 14 years), RR 0.82 (0.71-0.94) but with inconsistency in the effects across trials. No difference in all-cause mortality was found between groups; 95%CI (0.98-1.00). Due to issues of inconsistency of overdiagnosis data presentation and incomplete reporting, it was difficult to draw conclusions. False positive mammography and biopsy rates tended to be greater in women of younger age. In general, the quality of the body of evidence across outcomes was assessed as low or very low, indicating that the true effect may be substantially different from the estimates observed here. Of interest was to explore whether screening works differently in different age subgroups. Across outcomes (breast cancer mortality, all-cause mortality), it was unlikely that true subgroup differences were occurring according to age. No subgroup effects were occurring with other variables. No differences in breast cancer mortality or all-cause mortality were observed for breast self-examination, but this body of evidence was limited by indirectness, imprecision, and inconsistency.

Limitations:

Verification of select information and performing risk of bias assessments to inform GRADE assessments could not be completed in some instances. Limitations of this work include not searching EMBASE and relying on reviews' presentation of study information where verifications could not be sought; the latter was particularly true for overdiagnosis data, which may require a greater depth of investigation to understand the results. An evaluation to consider all relevant commentaries in relation to methodological and other issues to inform our assessments was not possible given constraints.

Conclusions:

A modified overview of reviews with an incorporated update was conducted on the benefits and harms of breast cancer screening modalities. Breast cancer mortality was reduced with the use of mammography compared with usual care for both short-case and long-case accrual, but true effects may be substantially different or very uncertain due to the low and very low quality of evidence, respectively. Interpretation of the available overdiagnosis evidence is limited. All-cause mortality data were not statistically different between groups and deemed of low quality. Outcomes in relation to breast self-examination compared with no screening were also not statistically significant, and the quality of evidence was low. Insufficient evidence exists regarding tomosynthesis, magnetic resonance imaging, ultrasound, and clinical breast examination. A number of considerations have been outlined

when interpreting these results for use in policy and practice. Implications for future research are also provided.

Contents

1. PURPOSE AND BACKGROUND

The burden and impact that breast cancer has on the Canadian population and the Canadian healthcare system are substantial. Breast cancer is the second-leading cause of cancer death in women(1). One in eight women are expected to develop breast cancer in their lifetime(2). In 2017, the projected age-standardized incidence rate of breast cancer among women in Canada is 130.3 cases per 100,000(3). The incidence rate in Canada has been stable since 2004, and variation in incidence between provinces in Canada is low. Death rates from breast cancer in women have been declining, from a peak of 41.7 per 100,000 in 1988 to a projected rate 23.2 per 100,000 in 2017(3). Across many countries, declines in mortality despite stable incidence rates suggests an improvement in breast cancer diagnostics and care(4). In Canada, it is speculated that this relationship may be mediated by both screening with mammography and the use of more effective therapies after surgery(5); with suggestions that the latter is the primary driver of improved outcomes(6). In Ontario, breast cancer was determined to be the fourth most costly for cancer care, and substantial personal economic burden may be experienced (7) (8).

A number of risk factors have been proposed: post-menopausal excess body weight, low levels of physical activity, early age at menarche, late age at first full term pregnancy (>30 years) or no pregnancy, late age at menopause, use of combination (estrogen and progestin together) hormone therapy and exposure to diethylstilboestrol (9-11). Relatively few cases occur because of ionizing radiation (from medical sources) (12), diagnostic x-rays and mammograms (13;14), and possibly type 2 diabetes (15). Risk factors commonly used to place women in a high-risk category include being a known carrier of the BRCA1 or BRCA2 gene mutation, being the first relative of a known carrier of the gene mutation, having a greater than 25% lifetime risk based on genetic testing (using the IBIS or BOADICEA tool), and having had chest radiation before the age of 30 or within the last 8 years (11;16-22). Breast density is an emerging risk factor not yet used to place women in the high-risk group, with a three- to five- time greater lifetime risk of developing breast cancer than women with mostly fatty breasts, regardless of other risk factors (23;24). About half of the incidence of breast cancer has been shown to occur in women ages 50-69 years, with 32% in women ages 70 and over and 17% in those under the age of 50(9).

Mammography is the primary method for breast cancer screening (25). Modalities include film mammography, which precede the more contemporary digital technology in terms of two-dimensional imaging, and tomosynthesis. Tomosynthesis has been introduced as an advancement, but has not been implemented for routine screening (often used as an adjunctive modality) (26-28). Tomosynthesis involves the use of computed tomography to create three-dimensional images without the overlapping of images of breast structures that is a limitation of two-dimensional technology (29). Clinical breast examination and breast self-examination have not been a part of mainstream recommendations for some time, but are known to still be employed in practice(30). Magnetic resonance imaging (MRI) uses magnetic fields to create displays of cross-sections of the breast and have been assessed as a supplementary test to mammography in high-risk women only (31); whether any studies exist in relation to its use in routine screening needs to be determined. Ultrasound has been suggested as supplementary to mammography, though not as a stand-alone test, and has been evaluated in high-risk women (25;32) (31).

Screening may benefit women by offering reassurance (if negative result) or by earlier and better treatment that may prevent or delay death from breast cancer. However, screen-positive results may be inaccurate (false-positives) or lead to overdiagnosis (screen-detected lesions that would not become clinically apparent during a woman's lifetime in the absence of screening and would not have caused any health implications), leading to psychological (e.g., now self-identifying as a cancer patient) or physical harms (e.g., overtreatment) (33) (34;35) (36) (37). Negative consequences related to biopsies or other invasive procedures, such as surgery, may also result following a false positive result (38).

Ideally, strategies to detect invasive breast cancer early, before a patient is symptomatic, allowing more tolerable and effective treatment that improves survival from breast cancer should be employed. Indeed, organized screening programs should potentially lead to a shift from late-stage to early-stage disease detection (39;40); however the occurrence or not of this shift in actual organized screening programs is debated (41;42). Nevertheless, the optimal screening modality, screening intervals, and age to start screening have not been definitively established. Increasing awareness of overdiagnosis and overtreatment have led to debates on the benefit-harm ratio of screening; the relative values and preferences of benefits and harms may impact decision making by women, and although adding complexity, have been integrated into some recommendations made by guideline panels.

Several groups have developed screening recommendations in recent years, with notable variation in guidance. The Canadian Task Force on Preventive Health Care (CTFPHC) is one of those groups and developed recommendations in 2011, which were updates from 2001 and 1994 (38). The purpose of this report is to synthesize evidence about breast cancer screening to inform an update of those recommendations. The CTFPHC guidelines are updated approximately every five years or as new evidence becomes available. This evidence report will seek new evidence syntheses, and assess the quality of evidence to date on the benefits and harms of screening determined by the CTFPHC as critical for decision-making. The CTFPHC will determine whether these findings, along with additional considerations of values and preferences, feasibility, acceptability, affordability, and equity, will change or reaffirm their 2011 recommendations.

1.1. Objectives

The objectives were to synthesize up-to-date evidence on the benefits and harms of breast cancer screening by conducting an overview of selected systematic reviews and an updated search for more recent primary studies. The findings will be used by the CTFPHC, along with additional considerations of feasibility, acceptability, affordability, and equity, to change or reaffirm previous recommendations.

1.2. Updated Analytic Framework

The analytical framework **(Figure 1)** and key questions were drawn from the 2011 review and adapted, as discussed below.

The population of interest remained as women aged 40 years or older who were not at high-risk for breast cancer: (a) without pre-existing or personal history of breast cancer and (b) not considered to

be at risk of breast cancer on the basis of extensive family history of breast or ovarian cancer, abnormal breast pathology, deleterious genetic mutations, or previously received radiation treatment to the chest for cancer (e.g., Hodgkin's). Studies of high-risk populations were excluded because of lack of applicability to our research questions. Although high breast density was not considered a high risk criterion, the CTFPHC considered study populations with 75% or more of women with known high tissue density to be outside of the scope of the target population for the guideline recommendations.

In its 2011 guideline, the CTFPHC included mammography (film, digital, and tomosynthesis), ultrasound, as well as magnetic resonance imaging in combination with clinical breast exam or self-breast exam, as relevant screening modalities for their guideline update. They had included outcomes which were rated by the CTFPHC working group as either critical (breast-cancer and all-cause mortality, false positive and false negative mammograms and biopsies on false positives), important (anxiety, distress, other psychological responses, radiation exposure, and overdiagnosis) and not important (pain during procedure).

Since its 2011 guideline, the CTFPHC had re-rated harm outcomes to better reflect their importance in patient decision-making. This evidence report only assessed outcomes considered by the CTFPHC working group as critical for decision making (i.e., rated as 7 or higher out of 9): breast-cancer and all-cause mortality, false-positive mammograms and biopsies on false-positives, as well as overdiagnosis.

2. METHODS

The CTFPHC requires its guidelines to be updated every 5 years. For this evidence synthesis, we conducted a modified overview of systematic reviews with an incorporated update. The modified review differed from a traditional overview of reviews since a systematic search for existing systematic reviews was not conducted; rather, a pre-specified list of known systematic reviews in the area were used as potential candidates(43). In addition, a search for more recent primary studies was also undertaken to update that literature. Further, it was our intention to identify the highest quality, fit-for-purpose systematic review to be used as a basis for our update, rather than providing an overall summary of the body of evidence for multiple reviews.

The methods were planned *a priori* and detailed in a publicly-available protocol (PROSPERO # 42017051498). Any changes from the protocol are provided herein. In the absence of an existing reporting guideline for overviews, we report this evidence synthesis according to the PRISMA statement **(Appendix 1)**.

2.1. RESEARCH QUESTIONS

Our research question of interest was:

For women ≥ 40 years of age and not at high-risk, what are the benefits and harms of the following screening modalities as compared with no screening or usual care:

(i) mammography (film and digital) with or without clinical breast examination or breast self-examination;
(ii) MRI with or without clinical breast examination or breast self-examination;
(iii) ultrasound with or without clinical breast examination or breast self-examination;
(iv) clinical breast examination;
(v) breast self-examination

2.2. INCLUSION/EXCLUSION CRITERIA

The selection criteria **(see Table 1)** were used to determine whether studies or systematic reviews containing studies were to be included or excluded.

2.3. LITERATURE SEARCH

To ensure the evidence needs of the CTFPHC were met in a timely manner, we used a modified overview of reviews approach, as described above. This approach involved identifying potentially relevant systematic reviews that would be assessed for quality of conduct, datedness, and fit-for-purpose according to the eligibility and methodological criteria, as a first step to updating the evidence from the 2011 CTFPHC guideline. Once the scope and datedness of the literature from those systematic reviews were mapped, we then conducted literature searches to identify more recently published or conducted primary studies.

2.3.1. Overview of Systematic Reviews

A total of 18 pre-identified, potentially relevant systematic reviews were considered for the overview of reviews portion of this report. The list **(outlined in Table 2**) included, but were not limited to reviews conducted for guidelines produced by the U.S Preventive Service Task Force (USPSTF)(44;45), the American Cancer Society (ACS)(46), and the Japanese National Cancer Center (NCC)(47). The

remaining potentially relevant systematic reviews were selected from the bibliography list from the USPSTF and ACS systematic review, since they were the most current and highest quality.

2.3.2. Primary Studies to Update Systematic Reviews

Given that we were aware of more recently conducted systematic reviews, we had anticipated that the updated search for primary studies would cover the timeframe of October 2014 to present for all screening modalities (search end date from the USPSTF 2016 review), except breast-self exam, which would be searched from as early as October 2010 (search end date from the CTFPHC 2011 review). The timeframes of these searches were to be confirmed once the selection process of systematic reviews for the overview component was completed.

An experienced medical information specialist developed and tested the strategies using an iterative process in consultation with the review team; the search strategies from the 2011 CTFPHC review were adapted for use in this update(48). Another senior information specialist peer reviewed the strategies prior to execution using the PRESS checklist(49). On January 4, 2017, we searched Ovid MEDLINE®, MEDLINE® Epub Ahead of Print and MEDLINE In-Process & Other Non-Indexed Citations, and the Cochrane Library on Wiley.

Strategies utilized a combination of controlled vocabulary (e.g., "Breast Neoplasms", "Early Detection of Cancer", "Mammography") and keywords (e.g., "breast cancer", "screen", "mammogram"). The strategy developed for MEDLINE was translated to the other databases. We applied standardized filters for study designs, including a highly sensitive search strategy for randomized controlled trials (RCTs). Search results were limited to the database entry dates October 2014 to the present, except for the breast self-examination search, which incorporated the period October 2010 to present. When possible, animal-only and opinion-pieces were removed. Specific details regarding the strategies appear in **Appendix 2**.

To update the grey literature search for unpublished literature (e.g., reports, theses, governmental publications) we consulted websites in the CADTH Grey Matters checklist. The CADTH checklist included national and international health technology assessment agencies, clinical practice guideline organizations, drug and device regulatory agencies, health economics resources, clinical trials registries, Canadian health prevalence and incidence databases, statistics, search engines, and databases. The clinical trial registries listed within the checklist included Canadian Cancer Trials, ClinicalTrials.gov, WHO International Clinical Trials Registry Platform, ISRCTN, CenterWatch, and Clinical Trials Registry India.

In addition to the checklist, we searched the following websites: British Columbia Cancer Agency, Cancer Care Ontario, Canadian Cancer Society, and World Conference on Breast Cancer. The searches utilized the following terms “breast cancer” OR “breast cancer screening,” and two reviewers screened results using the previously mentioned eligibility criteria.

2.4. SELECTION PROCESS

2.4.1. Overview of Systematic Reviews

In addition to the eligibility criteria above, the potentially relevant reviews (**see Tables 2 and 3**) had to satisfy the following methodological criteria to be considered a systematic review: (i) searched more than one database; (ii) reported their selection criteria; (iii) conducted quality or risk of bias

assessments on included studies; and (iv) provided a list and synthesis of included studies. Reviews were assessed by two independent reviewers. Thereafter, any remaining systematic reviews were subjected to AMSTAR assessments to inform quality of conduct. The thresholds for quality were 0-3 (low), 4-7 (moderate), 8-11 (high)(50;51).Any discrepancies were resolved through consensus. Finally, given our methodology, it was possible that multiple reviews could have been included to address the totality of the scope of interest. This final selection was determined based on quality score, completeness of reporting, and best fit for use of data for each intervention-outcome pair. This was assessed by a single reviewer and verified by a second reviewer.

2.4.2. Primary Studies to Update Systematic Reviews

Identified studies from the updated search were downloaded into a reference management software, and duplicates were removed. An online systematic review management software, DistillerSR (Evidence Partners, Ottawa, Canada)(52) was used to upload any remaining unique citations to be subjected to the eligibility criteria listed in **Table 1**. During the title and abstract screening phase, the liberal accelerated method was used: single reviewers assessed citations as relevant to move on to full text screening, whereas a second reviewer was needed to agree on exclusions. Two independent reviewers completed assessments of full-text articles, and any discrepancies were resolved either by consensus or a third reviewer. A pilot testing phase was conducted prior to commencing for both title/abstract (n=50) and full-text (n=10) study selection phases. Screening forms can be found in **Appendix 3**.

Articles not available electronically were ordered via interlibrary loan. If an ordered article, was not received within 30 days of our request, we considered the citation to be not available. Although citations in abstract form were not included, we attempted to locate the full text of these abstracts through searching author and title key words in Google and by contacting the authors (if the emails were in the abstract). Any full-texts identified were added to DistillerSR for screening.

2.5. Data Extraction and Management

2.5.1. Overview of Systematic Reviews

For the relevant systematic reviews, information was extracted at face value as to how it was synthesized and/or reported. We did not seek clarification from corresponding authors of the review. In the event that the evidence was presented which did not allow for clean extraction, the information was extracted and a commentary was provided. Single extractors collated study information (general characteristics and outcomes data) as reported in the included systematic reviews and verified some of this information, where critical and where time permitted, using the original study reports. Data extraction variables are those shown in (**Appendix 4**).

2.5.2. Primary Studies to Update Systematic Reviews

Extractions for any newly identified studies were recorded on a separate data extraction form, which was also pilot-tested. These were completed by one reviewer and verified by another. Any disagreements were resolved by consensus. In addition, we extracted (and verified) false positive data from the Canadian Partnership Against Cancer- Breast Cancer Screening in Canada (CPAC) Report (Cycle 2011-2012)(53). Authors (n=1) were contacted if there were data that were missing or needed clarification, and were given 2 weeks to respond.

2.6. Risk of Bias Assessment

All included studies (whether identified from systematic reviews or primary study searches), were assessed for the risk of bias using the Cochrane Risk of Bias tool (54) for RCTs and other controlled clinical trials **(Appendix 6)**, and the Newcastle-Ottawa scale for cohort studies(55) **(Appendix 7)**.

For the Cochrane risk of bias tool, under 'other biases', recruitment bias was considered. The general criteria used to determine an overall judgement for risk of bias per study was that all domains had to be 'low risk' for an overall judgement of low risk. If there were several 'unclears' and 'low risk', and perhaps a 'high' risk in some domains which are not considered to have serious implications (see explanation below), then it would have an overall judgement of moderate risk. Some domains were considered to have heavier weights for determining the overall judgement for risk of bias for a trial (ex: deficiencies in 'randomization and allocation concealment' were considered to have more serious implications compared to 'selective outcome reporting'.) If at least one domain (considered to have heavier weight) was considered to be 'high risk of bias', then the overall judgement was high risk.

For the Newcastle-Ottawa scale, age and hormone replacement therapy use were confounding variables regarded as important to be considered(56). Adjustment for lead time bias or allowing long-enough follow-up time to reduce lead time bias was considered important for overdiagnosis calculations. Therefore we modified the tool to add a specific question to address this. These assessments were conducted by one reviewer, and verified by another reviewer. Discrepancies were resolved by consensus. The assessments were used to inform the GRADE domain of study limitations.

The systematic reviews were not assessed for the risk of bias; AMSTAR (50;51)**(Appendix 5)** assessments to inform quality of conduct were conducted during the selection process, as described previously.

2.7. Analysis

Study characteristics of all included studies are presented in tables and were summarized narratively.

For the outcome of breast-cancer mortality, we conducted the main analysis according to whether authors analyzed only those cases identified during the screening period (short-case accrual) or all cases identified after the point of randomization and through the follow-up period (long-case accrual) **(Figure 2)**. In some studies, screening was provided for a brief period after the screening period. The short-case accrual method will include fewer cases because of the shorter follow-up timeframe within which to include cases. However, it would reduce contamination that would arise if control participants received screening after the intervention period of the trial; as such, would be expected to show an increased effectiveness of screening compared with long-case accrual methods (57). The long-case accrual method will include more cases because of the longer follow-up, but may underestimate the benefit of screening if present, because of the inclusion of cases diagnosed in both the control arm and the intervention arm after the intervention period where contamination may occur (57).

False positives mammograms and biopsies on false positives were calculated using the CPAC report (cycle 2011-2012)(53) following the same methods used in the CTFPHC 2011 systematic review. Data from this cycle were used to approximate a cohort of women in a breast cancer screening program. Initial and subsequent data for abnormal call rate (abnormal screening mammograms), invasive and in-

situ cancers, as well as non-malignant biopsy rates were used to calculate false positive data. As was done previously, the median follow-up time for breast-cancer mortality was determined from the trials included in the meta-analysis, and was used to estimate the number of screening rounds assuming a woman was screened every 2-3 years (ex: in the CTFPHC 2011, the median follow-up time from the RCTs was 11 years, assuming a woman gets screened every 2-3 years, we can anticipate 4 rounds of screening).

The values used to determine the numbers needed to screen (NNS) to prevent one death were calculated using the absolute effects as reported in the Summary of Findings (GRADE) tables. The false positive mammogram and biopsy data were then used to determine rates to correspond to the NNS.

In addition to the methods previously used by the CTFPHC 2011 review, we sought additional analyses to explore the data (see section 2.7.1.8). Instead of approximating a breast cancer screening cohort, we provided results cross-sectionally to represent the patterns occurring within a breast cancer screening program. The same variables were used to calculate false positive data.

2.7.1.1. Meta-Analysis

Where possible, effect estimates (as reported by review authors) were presented in a forest plot, as relative risk (RR) and corresponding 95% confidence intervals. If a meta-analysis was conducted in an existing review, we determined whether it was appropriate and feasible (an assessment of heterogeneity according to clinical and methodological characteristics) to update the pooled estimate with data from newly identified studies. Random effects models were used in these cases. Because of the lack of frequency and proportion data reported in systematic reviews, we undertook a generic inverse variance meta-analysis with the available risk ratio estimates and confidence intervals.

The existence of any unit of analysis errors were taken into consideration when evaluating the syntheses undertaken in systematic reviews. Unit of analysis errors can occur in studies that employ a cluster design (i.e., a clinical practice) and yet are analyzed at the individual level (i.e., patients), potentially leading to overly precise results and inappropriately contributing greater weight in a meta- analysis.

2.7.1.2. Statistical Heterogeneity

The Cochrane's Q (considered statistically significant at $p<0.10$) and I^2 statistic were used to assess the statistical heterogeneity of effect estimates among included studies. For the interpretation of I^2, a rough guide of low (0-25%), moderate (25-50%), substantial (50%-75%), and considerable (≥75%) was used (58;59).

2.7.1.3. Sub-Group Analyses

A priori-defined sub-group variables were defined below. The age sub-groups were updated, where possible, with information from newer studies. The results from sub-groups were interpreted cautiously.

- Age at entry (40-49 years, 50-59 years, 60-69 years, 70 years and older)
- Various ethnic populations, including indigenous populations
- Socioeconomic status (e.g., income, level of education; as assessed by study authors)
- Geographical location (rural vs. urban settings)

- Use of clinical breast examination/breast self-examination with mammography, MRI, or ultrasound
- Screening interval (≤12 months, 13 - <24 months, ≥24 months)
- Advancements in screening technology (film vs. digital mammography)

2.7.1.4. Sensitivity Analyses

Sensitivity analyses were planned to restrict analyses to those studies assessed as being of low risk of bias, as indicated in the Cochrane methods. Sensitivity analyses were not undertaken since there were no studies which had an overall judgement of low risk.

2.7.1.5. Small Study Effects

To assess for small study effects, a combination of graphical aids (i.e., funnel plot) and/or statistical tests (e.g., Egger regression test, Hedges-Olkin) were planned if at least 10 studies were available in any given analysis.

2.7.1.6. Meta-Analysis Software

Review Manager (RevMan) software version 5.3(60) was used to calculate effect estimates and conducting meta-analyses.

2.7.1.7. Rating the Quality of Evidence

For each critical outcome, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework(61) was used to assess the quality of the evidence.

Assessment of each of the GRADE domains (study limitations, consistency, precision, directness, reporting bias) were presented, where possible, based on the information provided in the reviews and newly identified studies. Absolute effects were calculated using baseline risks to facilitate reporting to GRADE.

2.7.1.8. Changes from Protocol

We originally intended to extract study information and data for face value from the relevant systematic reviews and not undertake *de novo* risk of bias assessments. However, due to the magnitude of missing information, the need for additional information, and the need to verify the critical information reported from the reviews, we consulted the original trial reports, where possible, and therefore, also extracted information to conduct risk of bias assessments. For overdiagnosis, a single reviewer extracted information which was missing for risk of bias assessments only, and this was verified by a second reviewer.

We modified our main analysis to consider both short-and long-case accrual. This was determined once study characteristics were identified and aligned with the approach taken with the USPSTF 2016 review. Due to incomplete reporting form the overviews of reviews, a generic inverse variance meta-analysis was used instead of an aggregate data approach.

We also modified our analysis for the age sub-groups. Based on the overview of systematic reviews methods, we relied on the existing data and analyses from the USPSTF 2016, which had presented their age groups as: 40-49, 50-59, 60-69, and 70-74.

Once we obtained and reviewed the CPAC 2011-2012 data, we performed calculations in addition to those undertaken in the 2011 CTFPHC systematic review to enhance our understanding of those data.

3. RESULTS

3.1. Overviews of Reviews

3.1.1. Selection and Characteristics Systematic Reviews to Form the Evidence Base

Three systematic reviews were included on the basis of their quality, relevance, and currency of included literature: CTFPHC 2011, USPSTF 2016; and ACS 2014. Methodological assessments are detailed in **Table 2 and 3**.

The USPSTF 2016 systematic review (AMSTAR 7, moderate quality) was selected to be used as the evidence base from which to supplement and update since it was considered the most recent, of the high-quality reviews that synthesized information for a majority of the screening modalities and outcomes of interest **(Table 4)**.

For studies addressing overdiagnosis and false positive data, the USPSTF 2016 review was used as the basis and supplemented with studies from ACS 2014 (AMSTAR 7, moderate quality) and CTFPHC 2011 (AMSTAR 10, high quality). The CTFPHC 2011 was used as the evidence base for breast self-exam and from which to update the evidence. A total of 16 unique studies were identified from these reviews.

Aligning with our objectives, the USPSTF 2016 systematic review assessed the effectiveness of various screening modalities (mammography [film, digital and tomosynthesis], MRI, ultrasound and clinical breast exam) for benefits (breast-cancer mortality and all-cause mortality) and harms (false positives, biopsies on false positives, and overdiagnosis) amongst women (≥40 years) not at high risk for breast cancer. The objective of the USPSTF 2016 systematic review was to update the previous iteration of the report (USPSTF 2009). The authors searched MEDLINE, the Cochrane Central Register of Controlled Trials, and the Cochrane Database of Systematic Reviews up to December 2014. For the mortality studies assessing the effectiveness of mammography, they identified updated follow-up data for 3 RCTs (The Canadian National Breast Screening Study 1&2, the Swedish Two County Trial, and the UK AGE trial). For false positives, the authors mainly used registry data from the Breast Cancer Screening Consortium. For overdiagnosis, no quantitative syntheses of the data were carried out given the extent of between-study heterogeneity in outcome definitions, and thus the authors instead prepared a narrative summary of study findings.

The authors did not find any studies assessing the effectiveness of digital, tomosynthesis, ultrasound, MRI, clinical breast exam or breast self-exam for any of the outcomes. GRADE assessments were not conducted.

The CTFPHC 2011 assessed the effectiveness of the following screening modalities for the same benefits and harms assessed in the USPSTF 2016: mammography (film & digital), MRI, clinical breast exam, and breast self-exam (but not ultrasound). The CTFPHC 2011 systematic review's main purpose was to determine whether findings from the CTFPHC 2001 systematic review had changed using the

results from the USPSTF 2009 systematic review as basis to update their results. The authors searched MEDLINE and the Cochrane Database of Systematic Reviews from December 2008 to October 2010. GRADE assessments were conducted on the benefit outcomes.

The ACS 2014 conducted a de-novo systematic review assessing the effectiveness of mammography (film & digital) and clinical breast exam (alone) on benefits (breast-cancer mortality) and harms (false positives, biopsies on false positives and overdiagnosis). The authors searched PubMED (up to March 6, 2014), CINAHL (up to September 10, 2013), and PsycINFO (up to September 10, 2013). GRADE assessments were conducted for all benefit and harm outcomes.

3.2. Results of the Search for Primary Studies

A total of 2,727 records were identified. After the removal of duplicates, a total of 1,980 records were uploaded into Distiller SR and examined for potential relevance (title/abstract screening). A total of 1,394 were excluded, resulting in 586 full-text articles that were assessed for eligibility.

Five hundred and eighty-three articles were excluded during full text screening with reasons (see **Appendix 9 and Figure 3**). Notably, 7 records were excluded: (i) outcomes did not match the pre-specified study design: cohort studies assessing breast-cancer mortality (n=4); and outcomes were not of interest (n=3). These 7 records were excluded to keep consistent with the PICOTs of the CTFPHC 2011 systematic review. An additional 2 RCTs were excluded since they did not present data in an appropriate format for use (62;63). A study by Bjurstam(62) for the Gothenburg trial was excluded since event rates were not provided (correspondence with author was unanswered); and updated data for the Gothenburg trial was already provided by the one of the included records(64). The other study by Narod(63) for the CNBSS 1 provided data up to the age of 60 assessed event rates for a certain follow-up period, which was not within the scope of this review.

Three articles (reporting updated RCT data) were included, one of which reported on updated results for multiple trials **(Figure 3)**. Studies from the above systematic reviews were then collated with the updated data. These studies represent the totality of the body of evidence that will be presented and analyzed in the remainder of this report: a total of 20 reports(64-83) addressing 16 unique studies are included, in addition to a grey literature report providing Canadian data on false positive screening results. Of the 16 studies, 14 addressed mammography and two addressed breast self-exam. No studies addressing the other modalities of interest were identified.

3.3. Characteristics of Included Studies

Study characteristics are shown in **Tables 5-7, 10-11**. Of the 16 included studies, there were six parallel-group randomized trials(64-69;72-74;76;81-84), four cluster randomized trials (70;71;76;84), two quasi-randomized trials(64;69), and four cohort studies(77-80) Date of trial initiation ranged from 1963 to 1989, while the conduct of cohort studies covered the 1986 to 2010 timespan. Seven studies were conducted in Sweden, two in Canada, and one study each for remaining countries (China, Denmark, Italy, Norway, Russia, United Kingdom, United States).

One of the few patient characteristics reported in all studies was age at entry and it varied across studies. Several studies evaluated participants around 40 years of age to around aged 65 or older, while five evaluated individuals within a discrete decade period or less (e.g., 39-41y, 40-49y, 60-69y). One study included women as young as 31 years of age, and we kept this study in as a carry-over inclusion from the CTFPHC 2011 review; it is unknown what proportion of the study population were under 40 years of age. No information was provided in the systematic reviews as to the ethnicity make-up of the study populations, or the proportion of women with high density of breast tissue across these studies. Socioeconomic status was provided for the CNBSS 1&2 trials; however, p-values were not provided.

Mammography screening

There were 10 trials (randomized and quasi-randomized) and four cohort studies. Most included studies addressed mammography screening with or without clinical breast exam. The ten trials were located and reported on breast cancer mortality, all-cause mortality, and/or overdiagnosis (Malmo I, Malmo II, CNBSS 1, CNBSS 2, AGE, HIP, Stockholm, Gothenburg, and the two Swedish Two County studies [Ostergotland and Kopparberg]). Updated data (in addition to the initially reported primary results, longer-term follow-up data were available) were located for seven of those trials (breast cancer mortality for Malmo I, Malmo II, Stockholm and Gothenburg; breast cancer and all-cause mortality for UK AGE trial; and overdiagnosis for the CNBSS 1 & 2 trial).

Across the 10 trials, sample size at randomisation ranged from around 18,000 to at least 160,000; sample sizes were not reported for the cohort studies. Mean follow-up ranged from <18 years to 30 years. Film mammography alone was specified in seven studies, and three studies specified film mammography with the addition of clinical breast exam. The comparator arm for all trials was usual care(85), which was further specified in a Cochrane review for breast cancer screening (85). The control arm in six of the ten RCTs received mammography screening at the end of the screening period. The duration of the screening period ranged from 3-12 years (median 7 years). The screening intervals ranged from 12 months to 33 months. Where reported, one to two views and readers were used, and the attendance rate ranged from 65% to 88%. All trials were conducted in urban settings.

Little information was provided on the four cohort studies reporting on overdiagnosis. Three of these studies reported on cancer screening programs that were implemented at a population level, so it might be reasonable to expect the available data to represent thousands of women per cohort. Screening interval was only reported in three studies and was 24 months. All remaining variables were not reported on.

In addition to the above studies, we obtained data for false-positive screening results in 2011-2012, as compiled by the Canadian Partnership against Cancer.

It should be noted that for the trials with two parts (Malmo I and II, CNBSS 1 and 2, and Swedish Two County- Kopparberg and Ostergotland), they were either treated as one dataset or two separate datasets, depending on how the authors reported the effect estimates.

Breast self-examination

Two trials were located that evaluated breast self-exam compared with no screening. Aside from age information, no other patient or study-level information was provided.

3.4. Risk of Bias of Included Studies

Tables 12-18 outlined the risk of bias assessments by domain (and supports for judgements) for the included primary studies. Overall, studies were deemed at moderate or high risk of bias for all outcomes.

Trials

For randomization, two studies were rated as high risk of bias for using allocation by date of birth rather than randomization (Gothenburg, Stockholm). Four studies did not provide adequate information to judge the method of randomization (Malmo I, Malmo II, HIP, Thomas) and were rated as unclear. The remaining studies reported an adequate method (CNBSS 1 & 2, AGE, Swedish Two County - Ostergotland & Kopparberg, Semiglazov) and were rated as low risk of bias.

Regarding allocation concealment, four trials were at high risk of bias due to the nature of their randomization method or lack of true randomization (Gothenberg, Stockholm, Swedish Two County - Ostergotland & Kopparberg). For CNBSS 1 & 2 (unclear risk of bias), all subjects received clinical breast-exams prior to being allocated to the study arms. There were more patients in the screening arm with invasive tumours (≥4 nodes) compared to the usual care arm at the first screen (17 vs. 5; p-value not reported in original publications). Information on concealment was not reported for remaining trials.

Blinding of participants and healthcare personnel was not possible due to the nature of the screening intervention. It is unknown whether knowledge of allocation might have changed care in a way that would have influenced outcomes (e.g., controls seeking mammography).

In relation to blinding of outcome assessors, all studies except the HIP trial were considered at low risk of bias as they used an independent endpoint committee, national registries, or blinding of assessors to determine mortality due to breast cancer. The HIP trial was deemed at high risk due to using a local endpoint committee that involved the study personnel. All studies were deemed at low risk in relation to all-cause mortality and overdiagnosis (specific to the trials). For overdiagnosis in relation to the cohort studies, blinding for outcome assessments were considered low risk, since they used record linkage to registries to obtain such data.

For attrition bias, the AGE trial was deemed at a low risk of bias due to little loss of follow-up and with reasons for dropout balanced between group. Conversely, the Thomas study was at high risk due to differential reasons for drop-out between groups. Information was not reporting for remaining trials.

For most studies, the lack of available protocols precluded a definitive assessment of selective outcome reporting. The exception was the AGE trial, where a protocol was identified, and outcomes were congruent.

In relation to other sources of bias, baseline imbalances were apparent in three studies. The Swedish Two County trials (treated as two separate datasets) had slightly older women in the mammography arm. In the HIP trial, there were more educated women in the mammography arm and more menopausal and women with breast lumps in the control arm. However, we were unable to verify

the above data based on the publications referenced in the overviews, and therefore the risk of bias remained 'unclear'. Four cluster randomized trials were included, of which two (Thomas, Semiglazov) were unclear for recruitment bias.

Cohorts

In relation to the selection of participants for the cohort, the four studies were deemed somewhat representative of the target population due to later age at entry and because it was unknown whether those at high risk for breast cancer were excluded. All studies drew their non-exposed cohort from the same community, except for Njor et al., 2013 who drew from a different region(80). Two studies used written self-report (through surveys) to ascertain mammography use (77;79). For the remaining two studies, one used secure records(78), and for the other, there was no description(80). Three studies noted that cancer was absent at the beginning of screening(77-79), whereas it was unclear for the remaining study(80). Out of a possible 4 stars for 'selection', 1 study scored 4 stars(78), 2 studies scored 3 stars(77;79), and the remaining study scored 1 star(80).

In relation to the comparability of the cohorts, age and hormone replacement therapy use were considered important confounding variables for stratification in study design or adjustment in the analysis (86). Only one study by Lund et al. addressed both confounding variables(77), while another adjusted for age(78). The two remaining studies did not consider these confounding factors(79;80). Therefore, out of a possible 2 stars, 1 study scored 2 stars(77), another scored 1 star(78), and 2 studies scored 0 stars(79;80).

Regarding aspects related to the assessment of outcomes, the four studies used record linkage to ascertain overdiagnosis. For these studies, it was difficult to ascertain whether all participants were followed for an adequate period of time; we used 10-15 years after randomization as a guide for assessment (86). Only one study adjusted for lead time bias(79). Adequacy of follow-up could not be assessed, since it was not clear whether the screening programs had ended. Out of a possible 4 stars, only 1 study scored 2 stars(79), whereas the remaining studies scored 1 star.

3.5. EFFECTS OF SCREENING

3.5.1. Mammography +/- Clinical Breast Exam vs. Usual Care

3.5.1.1. Breast-Cancer Mortality

Short-case accrual

Eight data sets (10 studies) were analyzed **(Evidence set 1)**; data from the two Canadian and two Swedish Two County studies were each combined in the primary study reports, thereby contributing one data set each in the analysis. The number of women diagnosed with breast cancer and followed up for outcomes were estimated to be in the thousands, although not all studies reported this data. Mammography with or without clinical breast examination reduced the risk of breast cancer mortality by 15% (95% CI 7% to 22%) compared with usual care for a median follow-up of 23 years. The absolute effect ranged from 31 to 91 fewer deaths per 100,000 in the screening arm depending on different

baseline risks (low to high: 205 to 606 per hundred thousand). The NNS to prevent one additional death attributed to breast cancer ranged from 3,226 (95%CI 2,222 to 6,667) to 1,099 (95%CI 746 to 2,326) based on low to high baseline risk **(Evidence set 1)**. Little statistical heterogeneity was detected among studies (I^2=10%, p=0.35), despite variation in study characteristics. GRADE assessment of low quality was given because the risk of bias across studies was deemed a very serious concern in this body of evidence.

Subgroup analysis by age. Studies by subgroup ranged from two to eight data sets; relatively fewer studies were available for the ≥70y subgroup **(Evidence set 1a).** The number of women included for analyses were unknown. The analysis of 40-49y included a mix of within-study subgroup data and studies whose sole population was within that age range ('between-study' assessment). All studies in the 50-59 y subgroup, except one, were within-study data. The remaining subgroups were within-study data only. A test for subgroup differences was not statistically significant (I^2=0%, p=0.44,). The GRADE assessments for each subgroup were rated at low or very low quality. The validity of subgroup effects is plausible but judged unlikely **(Evidence set 1a)**.

Other subgroup analyses. Subgroup analyses (use of clinical breast exam, screening modality and screening interval) are shown in **(Appendix 11)**. None were statistically significant for subgroup differences and confidence intervals overlapped among subgroups. The validity of subgroup effects may lack credibility **(Appendix 11)**; any differences among subgroups are likely spurious **(Appendix 16)**. We did not undertake remaining planned subgroup analyses because information was not reported (ethnicity, socioeconomic status, breast density) or was not applicable (all studies conducted in an urban setting, same comparator).

Long-case accrual

Six data sets (7 studies) were analyzed **(Evidence set 2)**. Data from the two Canadian studies were combined in the primary study report, thereby contributing one data set in the analysis. As above, the number of breast cancer cases was estimated to be in the tens of thousands. Mammography with or without clinical breast examination reduced the risk of breast cancer mortality by 18% (95% CI 6% to 29%) compared with usual care for a median follow-up of 14 years. The absolute effect ranged from 87 to 202 fewer deaths per 100,000 in the screening arm depending on different baseline risks (low to high: 482 to 1,125 per hundred thousand). The NNS to prevent one additional death attributed to breast cancer ranged from 1,149 (95%CI 714 to 3,448) to 493 (95%CI 306 to 1,471) according to different baseline risks **(Evidence set 2).** Substantial statistical heterogeneity was present across trials (I^2=68%, p=0.009). Where subgroup analyses were undertaken, none of the variables appear to account for the heterogeneity, which persisted in at least one subgroup per analysis **(refer to Appendix 12)**.. Relative to other studies in the analysis, the CNSS and HIP trials had the longest follow-up and most uncertain effects as per the extent that those data cross the null. The Kopparberg trial showed stronger effects than other trials. Another consideration is that, if events in the analyses are substantial enough, the test for heterogeneity may be overpowered and detecting differences that may not be clinically important. GRADE assessment of very low quality was given because the risk of bias and inconsistency across studies were each deemed very serious concerns in this body of evidence.

Subgroup analysis by age. The number of studies contributing to each subgroup ranged from two (≥70 years) to six (40-49 y), with the number of cases in the tens of thousands. The subgroups followed the same trend of representation of within-study and between-study data for the different age subgroups as

the short-case accrual analysis. As with short-case accrual, an examination of age subgroups in long-case accrual studies showed that there may not be differences among those subgroups (I^2=48%, p=0.12); **Evidence set 2a**), with the same trends of magnitude of effects across the subgroups. Subgroup data came from studies at a moderate (serious limitations) or high risk (very serious limitations) of bias; GRADE assessments were low or very low quality. The validity of subgroup effects for age is plausible but judged unlikely (**Evidence set 2a**).

Other subgroup analyses. These analyses are shown in **Appendix 12.** None were statistically significant for subgroup differences and confidence intervals overlapped among subgroups. The validity of subgroup effects may lack credibility; any differences among subgroups are likely spurious. We did not undertake remaining planned subgroup analyses for the same reasons as outlined above.

3.5.1.2. All-Cause Mortality

Eight data sets (9 studies) were analyzed **(Evidence set 3)**. Data from the two Canadian studies were combined in the primary study report, thereby contributing one data set in the analysis. As above, the number of breast cancer cases was approximated to be in the tens of thousands. No statistical differences were observed between mammography with or without clinical breast examination and usual care for all-cause mortality (RR 0.99, 95%CI 0.98 to 1.00) for a median follow-up of 16 years. No statistical heterogeneity was detected across trials (I^2=0%, p=0.59), despite variability in characteristics. GRADE assessment of low quality of evidence was given due to risk of bias issues considered of very serious concern.

Subgroup analysis by age. When examined by age in subgroups, no subgroup differences were detected (I^2=0%, p=0.72). The risk of bias across studies was deemed a very serious concern in this body of evidence. GRADE assessment of low quality was given. The validity of subgroup effects is plausible but judged unlikely **(Evidence set 3a)**.

Other subgroup analyses. No subgroup differences were detected for other subgroup variables **(Appendix 13).** The validity of subgroup effects may lack credibility **(Appendix 13**); any differences among subgroups are likely spurious. As above, we did not undertake remaining planned subgroup analyses (e.g., ethnicity, urban/rural setting) for the same reasons as outlined above.

3.5.1.3. Overdiagnosis

Three randomized controlled trials (Malmo I, CNBSS 1 & 2, Swedish Two County – Kopparberg) and four cohort studies reported on overdiagnosis **(Tables 6 and 7)**. As reported by the authors of the systematic reviews from which they were extracted, almost all studies were deemed as calculating 'excess incidence', but descriptions reveal that different calculations were undertaken **(Tables 6 and 7).** Information for both invasive and *in situ* cancers as well as at least invasive cancers were reported separately for most studies.

Table 6 provides data for the RCTs, which were reported differently among studies. The Kopparberg study reports that the cumulative incidence of diagnoses with mammography was not statistically different from usual care for invasive plus *in situ* cancers (relative risk [RR] 1.00, 95% CI 0.92 to 1.08), invasive cancers alone (RR 0.99, 95% CI 0.92 to 1.07) and *in situ* cancers alone (RR 1.17, 95% CI 0.88 to 1.55). Data from the two Canadian studies were combined and stratified by age; authors

reported the percentage of excess diagnoses in the mammography group but without an accompanying measure of dispersion (range of certainty) **(Table 6)**. The Malmo I trial also provides percentage data, but was reported as 'the percentage of incidence in the control group', inferring that the incidence of diagnoses is less in the mammography group. The risk of bias across those studies was deemed high. A GRADE rating of very low quality was given.

Overdiagnosis estimate calculations and reporting also varied among the four cohort studies **(Table 7)**. Rates and type of summary data differed across the four studies that reported invasive plus *in situ* data with confidence intervals, and although results were not statistically significant, these data represent different methodological approaches in the calculations. A few studies also reported on invasive only diagnoses. The risk of bias of those studies would be deemed a very serious concern across studies. GRADE rating of very low quality was given.

Not reported in the reviews' assessment of these studies were whether adjustments were made for lead time bias and what adjustments were made for possible differences in risk factors for the breast cancer groups being compared, particularly for the cohort studies. Overall, this body of evidence is of very low quality.

3.5.1.4. False-Positives and Biopsies on False-Positives

The false positive data were separated into short-case and long-case accrual. Using the 2011-2012 CPAC data and, short-case accrual, we approximated a cohort of women in a breast cancer screening program over 8 cycles (assuming a screening interval of every 2-3 years), representing a median follow-up of 23 years. For long-case, women would be screened over 5 cycles, representing a median follow-up of 14 years (**Table 8**). The discrepancy in the median follow-ups was reflective of the studies included in the meta-analysis. Our update identified more current short-case follow-up data (up to 30 years) for the Swedish Trials (Malmo I, Malmo II, and Stockholm); but not for long-case accrual. For the purposes of interpretation, it was assumed that the number of screens was equivalent to the number of women screened.

Weighted estimates for false-positive mammography screens and biopsies on false-positives across all screening ages (all ages), for short-case and long-case accrual are shown in **Table 8**. To allow for comparison with the CTFPHC 2011 review, we also provided these estimates according to age decades; the number of false positive screens and biopsies (per 1,000 women screened) appeared to be greatest in younger age groups..

Similar to false positive data, the NNS to prevent one additional death attributed to breast cancer would be reflective of the RR estimates and show overlap across the age subgroups.

For false positive data, 1 of 2 alternatives for cross-sectional assessment may be considered:

First, the 'initial' screens (abnormal call rates, biopsies, cancer detection) were consistently higher for all age groups compared to 'subsequent' screens. Knowing that first screens would identify prevalent cancers in women, this could provide support that 'initial' screens were actually picking up prevalent cases as part of the organized breast cancer screening programmes. The data were, therefore, calculated to present false positive rates cross-sectionally for initial and subsequent screens as separate quality indicators over a 1-year period (**see Table 9a**). Overall, the same patterns were observed, such

that false positives and biopsies on false positives per 1,000 women screened for 1 year decreased with increasing age.

The second alternative is to cast doubt that all initial screens were women's first screen. From the 2012 data, we are aware that only 54% of women had a screening mammogram as part of an organized program(87). Therefore, a certain proportion of data might be actually missing data (those who did not participate) or they were apart of organized screening, but were not captured by the database, and therefore, it was not their true first screen. With this uncertainty, we then calculated weighted average of the initial and subsequent screening data to gauge overall false positive rates cross-sectionally over a 1-year period (**Table 9b**). As was seen with the analysis above, per 1,000 women screened over a 1 year period, the number of false positives and biopsies on false positives displayed the same trend.

3.5.2. Breast Self-Exam vs. No Screening

3.5.2.1. Breast-Cancer Mortality

One cluster-randomized trial addressed this outcome. The Thomas et al., 2002 study evaluated women aged 31-65 years (see **Table 10**). In the CTFPHC 2011 review, these trial data were originally classified as 'all-cause mortality'; on closer inspection, we have reclassified under 'breast-cancer mortality' data as a result of an error during extraction (75). When comparing breast self-examination with no screening, groups were not statistically different for breast cancer mortality (RR 1.03, 95% CI 0.81-1.31), and duration of follow-up was not reported (75). The risk of bias was considered moderate, due to several domains not having the available information and deficiencies in loss to follow-up **(Table 17)**. The GRADE assessment was considered low-quality evidence **(Evidence Set 5)** due to serious concerns for both indirectness and imprecision.

3.5.2.2. All-Cause Mortality

Both cluster randomized trials contributed data for this outcome. All-cause mortality data for the Thomas et al., 2002 were extracted from the original trial report. Although we planned to use information extracted from the CTFPHC 2011 systematic review for the Semiglazov study, there was confusion as to whether the data used came from the correctly cited publication when comparing with information extracted from other systematic reviews as multiple versions of this report exist (Semiglazov 2003(76), Semiglazov 1992(88), Semiglazov 1999(89)). We elected to use the estimates from Semiglazov 2003 publication as reported in the USPSTF 2009 SR, as this was the most recent version of the trial, and these data were not reported in the USPSTF 2016 review. When pooling both studies, there was no statistical difference between groups (RR 0.96, 95% CI 0.81-1.12) and substantial heterogeneity between findings (I^2=67%, p=0.08). Follow-up was not reported in those studies. When considering the overall judgement of risk of bias, both studies were considered moderate risk **(Table 18).** The GRADE assessment for this outcome was considered low **(Evidence Set 6)** due to serious concerns with indirectness and inconsistency.

4. DISCUSSION

Summary of Main Results and Quality of the Evidence Ratings

The majority of evidence in relation to breast cancer screening for average risk women aged 40 years and older is in relation to the use of mammography with or without clinical breast exam. One grey literature report provided Canadian data for false positive screens and biopsies. Few trials had evaluated breast self-examination, and none were identified in relation to the use of mammography with or without breast self-exam, MRI, ultrasound, or clinical breast exam alone.

Mammography

Mammography was shown to decrease the risk of breast cancer mortality among cases identified during the study's screening period (short-case accrual). Decreases were also observed overall when including all cases identified during the follow-up period (long-case accrual), but with inconsistency in the effects across trials. No difference in all-cause mortality was found between groups. Due to issues of inconsistency of overdiagnosis data presentation and incomplete reporting, it is difficult to draw conclusions. False positive mammography and biopsy rates tended to be greater in younger compared with older age groups.

In general, the quality of the body of evidence across outcomes was assessed as low or very low, indicating that the true effect (direction unknown) may be substantially different from the estimates observed here.

Methodological limitations of the evidence pose uncertainty in the validity of the findings. Of importance are selection bias concerns: the method of or uncertainty in the method of randomisation and/or concealment of the allocation sequence across trials. For cohort studies, selection, confounding, and outcome assessment issues were identified.

The observed inconsistency in results for long-case accrual breast cancer mortality could not be explained by subgroup analyses. It is unknown how results might differ if the remaining trials (Malmo I, Malmo II, Stockholm) could have been included in the analysis, had they reported the long-case data. The quality of the evidence for overdiagnosis was downgraded to very low due to various concerns within the GRADE domains. Poor reporting and inability to undertake a detailed analysis inhibited the ability to make any definitive conclusions.

Imprecision was not of concern given the presumed large sample sizes and corresponding number of events. An insufficient number of studies precluded the ability to perform formal tests for small study effects.

Subgroups by age

Of interest was to explore whether screening works differently in different age subgroups. Across outcomes (breast cancer mortality, all-cause mortality), it is unlikely that true subgroup differences are occurring according to age. No subgroup effects were occurring with other variables.

Breast Self-Examination

No differences in breast cancer mortality or all-cause mortality were observed between breast self-exam, but this body of evidence is limited by indirectness (inclusion of women <40 years of age) and imprecision due to few included studies (breast cancer mortality) or inconsistency of effects (all-cause mortality).

Overall Completeness and Applicability of the Mammography Evidence

Several considerations may influence how the results summarized above are interpreted. Firstly, we are aware of commentaries that have discussed the potential merits and drawbacks of the included trials in relation to the methods that were assumed or understood to have been used (56;90;91). Varied interpretations of some trials exist. Given that these trials are old, poorly reported, and that additional information to clarify uncertainties may not be possible to attain, a final or 'true' assessment of the risk of bias may not be possible and may lead to differing assessments across systematic reviews. Certainly, such differing assessments are apparent when comparing this report to that of the CTFPHC 2011 systematic review.

We know from the literature that population-based breast cancer screening programmes were implemented in the late 1980s and early 1990s in the countries represented by the included trials (92;93) **(Appendix 18)**. The timing of these programs and the proportion of the respective populations covered by these programmes would infer that the women in these studies would have been receiving routine screening during the follow-up period (with either overlap with the screening study period or some lag before the implementation of a screening program). For the mammography group, this would affect the fidelity of the intervention in attempting to attribute the results of effects of screening (and, accordingly, variation across trials) to the characteristics of the delivery of mammography in those trials (e.g., duration of screening during the study period, the interval between mammographic screens). In turn, this would mean 'contamination' of the usual care group with the receipt of mammography. These issues would be particularly true for long-case accrual analyses where breast cancer cases identified during the follow-up period are included and tracked for mortality. For short-case accrual, however, only cases identified during the screening period (and possibly for a short time thereafter) are followed.

Determining the balance of these effects for long-case accrual may be difficult when considering that, at some point during follow-up, the identification of breast cancer cases in the control group should catch-up to the screening group in relation to lead time (and hence reflected in breast cancer incidence rates with time). Marmot et al suggest that 10-15 years after randomization would provide the most reliable estimate of the effect of screening on breast cancer mortality, after which a diluting effect of the control group may occur(56). This is a potential explanation for why heterogeneity might be occurring in the long-case accrual analyses in relation to findings of the older trials (Canadian and UK) nearing and crossing the null. Considering the findings in relation to the duration of follow-up was not identified *a priori* when planning the methodological approach for this modified overview of reviews, but was considered in other systematic reviews (56;85). For short-case accrual, given that screening in the control group at the end of the study period took place in over half of the trials, at least prevalent cancers would be identified and followed. However, this may not completely account for lead time if remaining screen-negative women were not followed. An additional consideration for the short-case accrual analyses would be the generalizability to current-day screening programs that would provide mammography screening for a longer duration.

Hanley et al. propose that, after a period of time (shorter in trials than in population-based screening programs), one would expect diminishing returns of the effects of screening following its cessation(94;95). The difficulty in assessing this from available data is that if these women are subsequently enrolled in population-based screening programmes, there would be no discontinuation of screening until they reach the end of eligibility (70 to 74 year range, depending on the programme). Hanley et al. further discuss undertaking time-specific breast cancer mortality rates instead of cumulative mortality estimates(94;95). Through their calculations and simulations, they suggest that

underestimation of the meta-analytic results may be occurring. They propose a method that uses year-specific breast cancer deaths; calculations by incident densities according to age may be helpful in understanding the findings.

The NNS values provided in the results were the best estimate based on the available data. NNS is the most widely used term; however, it is possible that these values represent NNI (numbers needed to invite). It is probable that the NNS values are influenced by contamination in the control group and the attrition rate after being invited; these factors may influence whether the NNS values are reflective of active screening.

Due to inconsistent presentation and the poor reporting of data (including understanding how lead time bias was adjusted for and any adjustment in the possible differences in risk factors between comparisons groups), insight from the existing literature was limited. If no overdiagnosis was occurring, though, the cumulative incidence of breast cancer cases would equalize between groups once a period equal to the lead time has passed(56). Marmot indicates that 5-10 years would be adequate follow-up after delivery of study screening has stopped(56). Zahl et al. lean towards 5 years for adequate follow-up time and further caution that including cancers that emerge following the end of the screening period for both screened and unscreened groups underestimates the amount of overdiagnosis occurring with screening(96). Further, both Marmot and Zahl note the importance of the control group not receiving screening to facilitate this comparison to avoid overestimation of the results, which may be difficult to consider for the included mammography trials in light of the implementation of population-base screening programs(56;96).

An aspect of the process of care not addressed in this modified overview is in relation to treatment. Treatment options and their effectiveness would have improved over the course of these trials in improving patient care. Presumably, however, treatment options to women in the comparison groups would have been the same after the screening period, since this timeframe was brief, and follow-up for all trials were long.

The number of women in these trials with high breast density is unknown. Higher tissue density has been shown to increase the risk of breast cancer(24;97), thought to be due to the masking of cancers during screening thus increasing identification between screens(24). However, women will not know they have high density tissue until they are screened or have undergone imaging for another reason. Given the age of the existing trials, it is highly unlikely that, if collected, these data would be accessible; however, for ongoing and future trials, quantifying this subpopulation within studies and stratifying their results might provide information on screening effectiveness, false-positive rates, and how they were managed (e.g., different screening frequency or use of adjunct modalities). It is unknown how much literature currently exists in relation to this population; though we excluded studies where 75% or more of the women had high breast density. We encountered two RCTs (98;99) and two cohort studies (100;101) (from search update) with a population with greater than 75% high breast density, and they were excluded since they either did not address a screening modality of interest: negative mammogram and subsequent MRI (98;99) or addressed a screening modality of interest, but not a comparator of interest: no comparator or not reported (98;100;101).

Based on the age of the trials it is likely that, if not reported, all included trials used film mammography. Digital mammography is its more contemporary counterpart, the diagnostic accuracy of which has been compared with film and found it to be similar (102;103). It is therefore reasonable to

consider the results of this review to be applicable to digital mammography. We did not find any studies that compared 3-dimensional mammography (tomosynthesis) with usual care or no screening; evidence suggests that tomosynthesis improves cancer detection rates (which may or may not affect mortality), reduces recall, has an acceptable harms profile, and may be cost-saving relative to 2-dimensional (conventional) digital mammography(104). However, due to its novelty, its use as a primary screening modality may not be freely available. From our updated search, it was apparent that tomosynthesis usually, but not always, was being included in combination with other technologies as a supplemental or adjunctive screening modality; comparators were also usually digital or film mammography (which was an exclusion criteria)(26-28).

Inadequate reporting in these trials affected our assessment of the internal validity of studies, what effect estimate data that were available for use, and our ability to gauge anticipated absolute effects for GRADE tables. Although some reports of those trials were published before reporting guidelines such as CONSORT, STROBE, and CONSORT extension for cluster randomized trials were developed, we would encourage authors publishing new studies or follow-up data for known studies to use these guidelines when writing their manuscripts.

Some of the above considerations may be applicable when considering all-cause mortality; however, the breast cancer cases are relatively few in the context of all causes of mortality.

Overall Completeness and Applicability of the Breast Self-Examination Evidence

Few studies were located that addressed breast self-examination, and no differences were observed between groups for breast cancer and all-cause mortality. Poor reporting of trial information impeded an understanding of study characteristics and assessments of risk of bias. For breast cancer mortality, the inclusion of women younger than the population of interest (proportion unknown) and the extent of imprecision reduced the confidence in the findings. For all-cause mortality, inclusion of younger women in one study and the heterogeneity of the findings decreased the quality of the evidence.

Comparison with other reviews

Several other reviews have addressed breast cancer screening. As with other reviews, we did not include the Edinburgh trial in light of important methodological issues that would affect the interpretation of findings(57;105).

The USPSTF 2016 was a main source of information for our review. Their results are similar overall to our review, but with some notable differences: their main analyses were according to age, their long-case accrual was a mix of short- and long- case accrual (an assessment of the 'longest' case accrual), and they did not undertake GRADE assessments. It should be noted that the USPSTF 2016 systematic review was the first to separate their analyses into short- and long-case accrual, and they used data from the long-case accrual analysis to inform their recommendations. Justifications for doing so were not provided.

The previous CTFPHC review did not consider short- versus long-case accrual and combined ages 50 to 69 as an aggregate group. Our final GRADE assessments also differ with those in the 2011 review, mainly in relation to the downgrading of study limitations.

The ACS 2014 did not provide a meta-analysis incorporating any updated study data identified, with their reason being that they did not feel that this would appreciably change any of the existing effect estimates. It was difficult to compare their GRADE assessments, since complete justification for judgements were not provided; however their GRADE assessments were similar to those outlined by the CTFPHC 2011.

In regards to false positive data from the modified overviews exercise, the same pattern was also observed from other, similar organizations and cohort studies **(see Appendix 14)**.

Strengths and limitations of the modified overview of reviews

We were able to undertake verification of select information and perform risk of bias assessments to inform GRADE assessments to meet the timelines of the CTFPHC guidelines working group. Limitations of this work include not searching EMBASE and relying on reviews' presentation of study information where verifications could not be sought; the latter was particularly true for overdiagnosis data, which may require a greater depth of investigation to understand the results. An evaluation to consider all relevant commentaries in relation to methodological and other issues to inform our assessments was not possible given constraints.

Implications for Research

As indicated above, the current literature is fraught with issues to consider in understanding the applicability of the evidence. Individual-patient data meta-analyses may be ideal approach to overcome some of those issues, but likely impractical given the age of the trials and that the needed information to elucidate uncertainties either wasn't collected or may not be accessible. Consideration of Hanley's suggested approach to calculate time-specific breast cancer mortality rates could be given. Overdiagnosis is an aspect of this research that needs further and systematic assessment. de Gelder et al. outline seven approaches to estimating overdiagnosis, which depend on choice of numerator and denominator(106). Again, individual patient data from the included studies to test the relevant calculation options would be ideal.

The AgeX cluster randomized trial in the UK is underway to address screening in average risk women ages 47-49 years and 70 years and older. As screening is routinely offered to women ages 50-70 years every three years in the UK, this trial will offer one additional screening to those <50 years of age and three triennial screenings to those over 70 years of age. Most of the 80 National Health Service breast screening centres in England will participate. Results are not anticipated until 2026. Potentially relevant, ongoing RCTs were identified from our grey literature search and may prove informative for any subsequent updates of this evidence report **(Appendix 15)**.

We did not retrieve studies evaluating tomosynthesis, MRI, and ultrasound; these modalities are usually supplementary or offered as adjunctive techniques in the evaluation of high-risk women or those

with higher breast density. Future research initiatives may focus on high risk patient groups, in which potential benefits would be expected to outweigh the harms associated with screening (number of false positives and overdiagnosis), more clearly than in screening of average risk women. Personalized risk assessment (based on risk factors in addition to family history) is also emerging as an area of intense research activity. This approach involves completion of a risk assessment tool prior to screening in order to select an appropriate screening approach, and after confirmation of a positive biopsy to tailor treatment course to fulfill a woman's unique needs(107). These approaches will require rigorous evaluation of clinical validity and clinical utility.

Implications for Practice and Policy

The implications for practice for those developing policy or guideline recommendations will need to consider the observed benefits of reduced breast cancer mortality in a body of evidence that for which the aforementioned considerations may need to be made. Those results then need to be balanced against the risk and extent of overdiagnosis, for which the information to date is incomplete, and false positive data.

5. CONCLUSIONS

A modified overview of reviews with an incorporated update was conducted on the benefits and harms of breast cancer screening modalities. Breast cancer mortality was reduced with the use of mammography compared with usual care for both short-case (median follow-up 23 years) and long-case accrual (median follow-up 14 years), but true effects may be substantially different or very uncertain due to the low and very low quality of evidence, respectively. Interpretation of the available overdiagnosis evidence is limited. All-cause mortality data were not statistically different between groups and deemed of low quality.

Outcomes in relation to breast self-examination compared with no screening were also not statistically significant, and the quality of evidence was low. Evidence was non-existent regarding tomosynthesis, magnetic resonance imaging, ultrasound, and clinical breast examination for the purposes of this report.

A number of considerations have been outlined when interpreting these results for use in policy and practice. Considerations for future research are provided.

References

(1) DeSantis C, Siegel R, Bandi P, Jemal A. Breast cancer statistics, 2011. CA: a cancer journal for clinicians 2011;61(6):408-18.

(2) Canadian Cancer Society. Canadian Cancer Statistics. Special Topic: Pancreatic Cancer. 2017. http://www.cancer.ca/~/media/cancer.ca/CW/publications/Canadian%20Cancer%20Statistics/Canadian-Cancer-Statistics-2017-EN.pdf

(3) Canadian Cancer Society. Canadian Cancer Statistics. 2017. http://www.cancer.ca/~/media/cancer.ca/CW/cancer%20information/cancer%20101/Canadian%20cancer%20statistics/Canadian-Cancer-Statistics-2017-EN.pdf?la=en

(4) Kleibl Z, Kristensen VN. Women at high risk of breast cancer: Molecular characteristics, clinical presentation and management. The Breast 2016;28:136-44.

(5) Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics. 2015. http://www.cancer.ca/~/media/cancer.ca/CW/cancer%20information/cancer%20101/Canadian%20cancer%20statistics/Canadian-Cancer-Statistics-2015-EN.pdf?la=en

(6) Welch HG, Prorok PC, O'Malley J, Kramer BS. Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. NEJM 2016;375:1438-47.

(7) Canadian Breast Cancer Network. Breast Cancer: Economic Impact and Labour Force Re-entry. 2010. http://www.cbcn.ca/index.php?id=2912&lang=en&pageaction=content.page

(8) de Oliveira C, Bremner KE, Pataky R, Gunraj N, Chan K, Peacock S, et al. Understanding the costs of cancer care before and after diagnosis for the 21 most common cancers in Ontario: a population-based descriptive study. CMAJ open 2013;1(1):E1-E8.

(9) Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics. 2016. http://www.cancer.ca/~/media/cancer.ca/CW/cancer%20information/cancer%20101/Canadian%20cancer%20statistics/Canadian-Cancer-Statistics-2016-EN.pdf?la=en

(10) American Cancer Society. Breast Cancer Facts & Figures 2013-2014. 2013. http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-042725.pdf

(11) Cancer Research UK. Breast cancer risk factors. 2016. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/risk-factors

(12) Parkin DM, Darby SC. 12. Cancers in 2010 attributable to ionising radiation exposure in the UK. British journal of cancer 2011;105:S57-S65.

(13) de Gonzalez AB, Darby S. Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. The lancet 2004;363(9406):345-51.

(14) de Gonzalez AB, Curtis RE, Gilbert E, Berg CD, Smith SA, Stovall M, et al. Second solid cancers after radiotherapy for breast cancer in SEER cancer registries. British journal of cancer 2010;102(1):220-6.

(15) Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JP. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. Bmj 2015;350:g7607.

(16) Cancer Care Ontario. Ontario Breast Screening Program (OBSP). Guideline Summary. 2015. https://www.cancercare.on.ca/common/pages/UserFile.aspx?fileId=349883

(17) Kleibl Z, Kristensen VN. Women at high risk of breast cancer: Molecular characteristics, clinical presentation and management. The Breast 2016;28:136-44.

(18) Garber JE, Offit K. Hereditary cancer predisposition syndromes. Journal of Clinical Oncology 2005;23(2):276-92.

(19) American Cancer Society. What are the risk factors for breast cancer? 2016. https://www.cancer.org/cancer/breast-cancer/risk-and-prevention.html

(20) Nelson HD, Zakher B, Cantor A, Fu R, Griffin J, O'Meara ES, et al. Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. Annals of internal medicine 2012;156(9):635-48.

(21) Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M. Genetic susceptibility to breast cancer. Molecular oncology 2010;4(3):174-91.

(22) Antoniou A, Pharoah PDP, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. The American Journal of Human Genetics 2003;72(5):1117-30.

(23) McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiology Biomarkers & Prevention 2006;15(6):1159-69.

(24) Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. New England journal of medicine 2007;356(3):227-36.

(25) Nothacker M, Duda V, Hahn M, Warm M, Degenhardt F, Madjar H, et al. Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. A systematic review. BMC cancer 2009;9(1):335.

(26) Lang K, Nergarden M, Andersson I, Rosso A, Zackrisson S. False positives in breast cancer screening with one-view breast tomosynthesis: An analysis of findings leading to recall, work-up and biopsy rates in the Malmo Breast Tomosynthesis Screening Trial. European radiology 2016;26(11):3899-907.

(27) Houssami N. Digital breast tomosynthesis (3D-mammography) screening: data and implications for population screening. 2015. Taylor & Francis.

(28) Jong R. Tomosynthesis Mammography Imaging Screening Trial (TMISTLead-in). 2017. https://clinicaltrials.gov/ct2/show/NCT02616432

(29) Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. Jama 2014;311(24):2499-507.

(30) Provencher L, Hogue JC, Desbiens C, Poirier B, Poirier E, Boudreau D, et al. Is clinical breast examination important for breast cancer detection? Current Oncology 2016;23(4):e332.

(31) Lord SJ, Lei W, Craft P, Cawson JN, Morris I, Walleser S, et al. A systematic review of the effectiveness of magnetic resonance imaging (MRI) as an addition to mammography and ultrasound in screening young women at high risk of breast cancer. European journal of cancer 2007;43(13):1905-17.

(32) Gartlehner G, Thaler K, Chapman A, Kaminski-Hartenthaler A, Berzaczy D, Van Noord MG, et al. Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk. The cochrane library 2013.

(33) Nelson HD, Tyne K, Naik A, Bougatsos C, Chan B, Nygren P, et al. Screening for Breast Cancer: Systematic Evidence Review Update for the U.S. Preventive Services Task Force. Evidence Review Update No. 74. 2009.

(34) Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. The cochrane library 2009.

(35) Salz T, Richman AR, Brewer NT. Meta–analyses of the effect of false–positive mammograms on generic and specific psychosocial outcomes. Psycho-Oncology 2010;19(10):1026-34.

(36) Brodersen J, Thorsen H, Kreiner S. Validation of a Condition–Specific Measure for Women Having an Abnormal Screening Mammography. Value in health 2007;10(4):294-304.

(37) Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. The Annals of Family Medicine 2013;11(2):106-15.

(38) Canadian Task Force on Preventive Health Care. Recommendations on screening for breast cancer in average-risk women aged 40ΓÇô74 years. Canadian Medical Association Journal 2011;183(17):1991-2001.

(39) Helvie MA, Chang JT, Hendrick RE, Banerjee M. Reduction in late-stage breast cancer incidence in the mammography era: Implications for overdiagnosis of invasive cancer. Cancer 2014;120(17):2649-56.

(40) Williams JH, Carter SM, Rychetnik L. Organised cervical screening 45 years on: How consistent are organised screening practices? European journal of cancer 2014;50(17):3029-38.

(41) Harding C, Pompei F, Burmistrov D, Welch HG, Abebe R, Wilson R. Breast cancer screening, incidence, and mortality across US counties. JAMA internal medicine 2015;175(9):1483-9.

(42) Jorgensen KJ, Gotzsche PC, Kalager M, Zahl PH. Breast Cancer Screening in Denmark A Cohort Study of Tumor Size and OverdiagnosisBreast Cancer Screening in Denmark. Annals of internal medicine 2017;166(5):313-23.

(43) Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC medical research methodology 2011;11(1):15.

(44) Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 US Preventive Services Task Force recommendation. Annals of internal medicine 2016;164(4):244-55.

(45) Siu AL. Screening for breast cancer: US Preventive Services Task Force recommendation statement. Annals of internal medicine 2016;164(4):279-96.

(46) Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YCT, et al. Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. Jama 2015;314(15):1599-614.

(47) Hamashima C, Hamashima C, Hattori M, Honjo S, Kasahara Y, Katayama T, et al. The Japanese Guidelines for Breast Cancer Screening. Japanese journal of clinical oncology 2016;hyw008.

(48) Robinson KA, Chou R, Berkman ND, Newberry SJ, Fu R, Hartling L, et al. Integrating Bodies of Evidence: Existing Systematic Reviews and Primary Studies BTI - Methods Guide for Effectiveness and Comparative Effectiveness Reviews. 2015.

(49) McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. Journal of clinical epidemiology 2016.

(50) Grimshaw J, Hill S, Lowe D, Mayhew A, Fiander M, Pratt M, et al. Rx for Change- Methods for Development. CADTH . 2011.
https://www.cadth.ca/rx-change

(51) Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC medical research methodology 2007;7(1):1.

(52) DistillerSR [computer program]. Ottawa: Evidence Partners; 2011.

(53) Canadian Partnership Against Cancer. Breast Cancer Screening in Canada. Monitoring & Evaluation of Quality Indicators. 2017.

(54) Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions. 5 ed. Wiley Online Library; 2008.

(55) Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. Newcastle-Ottawa Scale. Hospital Research Institute Ottawa (http://www.ohri.ca/programs/clinical_epidemiology/nosgen pdf 2007.

(56) Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. Br J Cancer 2013;108(11):2205-40.

(57) Nelson HD, Cantor A, Humphrey L, Fu R, Pappas M, Daeges M, et al. Screening for Breast Cancer: A Systematic Review to Update the 2009 U.S. Preventive Service Task Force Recommendation. 2016.

(58) Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj 2003;327(7414):557-60.

(59) Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. Bmj 2011;343:d4002.

(60) Review Manager [computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre; 2014.

(61) Oxman AD, GRADE Working Group. Grading quality of evidence and strength of recommendations. Bmj 2004;328(19):1490-4.

(62) Bjurstam NG, Bjorneld LM, Duffy SW. Updated results of the Gothenburg trial of mammographic screening. Cancer 2016;122(12):1832-5.

(63) Narod SA, Sun P, Wall C, Baines C, Miller AB. Impact of screening mammography on mortality from breast cancer before age 60 in women 40 to 49 years of age. Current oncology (Toronto, Ont) 2014;21(5):217-21.

(64) Nystrom L, Bjurstam N, Jonsson H+, Zackrisson S, Frisell J. Reduced breast cancer mortality after 20+ years of follow-up in the Swedish randomized controlled mammography trials in Malmo, Stockholm, and Goteborg. Journal of Medical Screening 2016;0969141316648987.

(65) Bjurstam N, Bjorneld L, Warwick J, Sala E, Duffy SW, Nystrom L, et al. The Gothenburg breast screening trial. Cancer 2003;97(10):2387-96.

(66) Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. 2014.

(67) Moss SM, Wale C, Smith R, Evans A, Cuckle H, Duffy SW. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. The Lancet Oncology 2015;16(9):1123-32.

(68) Moss S, Johns L, Cuckle H, Bobrow L, Evans A, Kutt L, et al. Randomised controlled trial of mammographic screening in young women: updated results on incidence and mortality. National Cancer Research Institute . 2009.

(69) Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. The Lancet 2002;359(9310):909-19.

(70) Tabar L, Fagerberg G, Chen H, Duffy SW, Smart CR, Gad A, et al. Efficacy of breast cancer screening by age. New results swedish two–county trial. Cancer 1995;75(10):2507-17.

(71) Tabar L, Vitak B, Chen TH-H, Yen AM-F, Cohen A, Tot T, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011;260(3):658-63.

(72) Shapiro S, Venet W, Strax P, et al. Current results of the breast cancer screening randomized trial: the health insurance plan (HIP) of greater New York study. Hans Huber; 1988. p. 3-15.

(73) Habbema JDF, Oortmarssen GJ, van Putten DJ, Lubbe JT, Maas PJ. Age-specific reduction in breast cancer mortality by screening: an analysis of the results of the Health Insurance Plan of Greater New York study. JNCI: Journal of the National Cancer Institute 1986;77(2):317-20.

(74) ARON JL, PROROK PC. An analysis of the mortality effect in a breast cancer screening study. International journal of epidemiology 1986;15(1):36-43.

(75) Thomas DB, Gao DL, Ray RM, Wang WW, Allison CJ, Chen FL, et al. Randomized trial of breast self-examination in Shanghai: final results. JNCI: Journal of the National Cancer Institute 2002;94(19):1445-57.

(76) Semiglazov VF, Manikhas AG, Moiseyenko VM, Protsenko SA, Kharikova RS, Seleznev I, et al. [Results of a prospective randomized investigation [Russia (St. Petersburg)/WHO] to evaluate the signficance of self-examination for the early detection of breast cancer]. Vopr Onkol 2003;49(4):434-41.

(77) Lund E, Mode N, Waaseth M, Thalabard JC. Overdiagnosis of breast cancer in the Norwegian Breast Cancer Screening Program estimated by the Norwegian Women and Cancer cohort study. BMC cancer 2013;13(1):614.

(78) Puliti D, Miccinesi G, Zappa M, Manneschi G, Crocetti E, Paci E. Balancing harms and benefits of service mammography screening programs: a cohort study. Breast Cancer Research 2012;14(1):R9.

(79) Hellquist BN, Duffy SW, Nystr+Âm L, Jonsson H+. Overdiagnosis in the population-based service screening programme with mammography for women aged 40 to 49 years in Sweden. Journal of Medical Screening 2012;19(1):14-9.

(80) Njor SH, Olsen AH, Blichert-Toft M, Schwartz W, Vejborg I, Lynge E. Overdiagnosis in screening mammography in Denmark: population based cohort study. Bmj 2013;346:f1064.

(81) Baines CJ, To T, Miller AB. Revised estimates of overdiagnosis from the Canadian National Breast Screening Study. Preventive Medicine 2016;90:66-71.

(82) Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP. Rate of over-diagnosis of breast cancer 15 years after end of Malmo mammographic screening trial: follow-up study. Bmj 2006;332(7543):689-92.

(83) Yen AM, Duffy SW, Chen TH, Chen L, Chiu SY, Fann JC, et al. Long-term incidence of breast cancer by trial arm in one county of the Swedish Two–County Trial of mammographic screening. Cancer 2012;118(23):5728-32.

(84) Thomas DB, Gao DL, Ray RM, Wang WW, Allison CJ, Chen FL, et al. Randomized trial of breast self-examination in Shanghai: final results. JNCI: Journal of the National Cancer Institute 2002;94(19):1445-57.

(85) Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. Thecochrane library 2013.

(86) Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M, et al. 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) October 2012. British journal of cancer 2013;108(11):2205.

(87) Canadian Partnership Against Cancer. Report from the Evaluation Indicators Working Group. Guidelines for Monitoring Breast Cancer Screening Porgram Performance, 3rd edition. 2013.

(88) Semiglazov VF, Moiseyenko VM, Bavli JL, Migmanova NS, Seleznyov NK, Popova RT, et al. The role of breast self-examination in early breast cancer detection (results of the 5-years USSR/WHO randomized study in Leningrad). European journal of epidemiology 1992;8(4):498-502.

(89) Semiglazov VF, Moiseenko VM, Manikhas AG, Protsenko SA, Kharikova RS, Popova RT, et al. Interim results of a prospective randomized study of self-examination for early detection of breast cancer (Russia/St. Petersburg/WHO). Voprosy onkologii 1999;45(3):265-71.

(90) Heywang-Kobrunner SH, Schreer I, Hacker A, Noftz MR, Katalinic A. Conclusions for mammography screening after 25-year follow-up of the Canadian National Breast Cancer Screening Study (CNBSS). European radiology 2016;26(2):342-50.

(91) Brackstone M, Latosinsky S, Saettler E, George R. CJS debate: Is mammography useful in average-risk screening for breast cancer? Canadian Journal of Surgery 2016;59(1):62.

(92) Dowling EC, Klabunde C, Patnick J, Ballard-Barbash R, International Cancer Screening Network (ICSN). Breast and cervical cancer screening programme implementation in 16 countries. Journal of Medical Screening 2010;17(3):139-46.

(93) Shapiro S, Coleman EA, Broeders M, Codd M, de Koning H, Fracheboud J, et al. Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. International journal of epidemiology 1998;27(5):735-42.

(94) Hanley JA. Analysis of mortality data from cancer screening studies: looking in the right window. Epidemiology 2005;16(6):786-90.

(95) Hanley JA, McGregor M, Liu Z, Strumpf EC, Dendukuri N. Measuring the mortality impact of breast cancer screening. Can J Public Health 2013;104(7):437-42.

(96) Zahl PH, Jorgensen KJ, Gotzsche PC. Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. British journal of cancer 2013;109(7):2014-9.

(97) Kerlikowske K, Shepherd J, Creasman J, Tice JA, Ziv E, Cummings SR. Are breast density and bone mineral density independent risk factors for breast cancer? Journal of the National Cancer Institute 2005;97(5):368-74.

(98) Samavat H, Dostal AM, Wang R, Bedell S, Emory TH, Ursin G, et al. The Minnesota Green Tea Trial (MGTT), a randomized controlled trial of the efficacy of green tea extract on biomarkers of breast cancer risk: study rationale, design, methods, and participant characteristics. Cancer causes & control 2015;26(10):1405-19.

(99) Emaus MJ, Bakker MF, Peeters PH, Loo CE, Mann RM, De Jong MD, et al. MR imaging as an additional screening modality for the detection of breast cancer in women aged 50ΓÇô75 years with extremely dense breasts: the DENSE trial study design. Radiology 2015;277(2):527-37.

(100) Goossens MC, De Brabander I, De Greve J, Vaes E, Van Ongeval C, Van Herck K, et al. Breast cancer risk is increased in the years following false-positive breast cancer screening. European Journal of Cancer Prevention 2017.

(101) Kerlikowske K, Zhu W, Tosteson AN, Sprague BL, Tice JA, Lehman CD, et al. Identifying women with dense breasts at high risk for interval cancer: a cohort study. Annals of internal medicine 2015;162(10):673-81.

(102) Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. the Digital Mammographic Imaging Screening Trial (DMIST) Investigators Group Diagnostiv Performance of Digital versus Film Mammography for Breast-Cancer. N Engl J Med 2005;353(17):1.

(103) Lewin JM, D'Orsi CJ, Hendrick RE, Moss LJ, Isaacs PK, Karellas A, et al. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. American Journal of Roentgenology 2002;179(3):671-7.

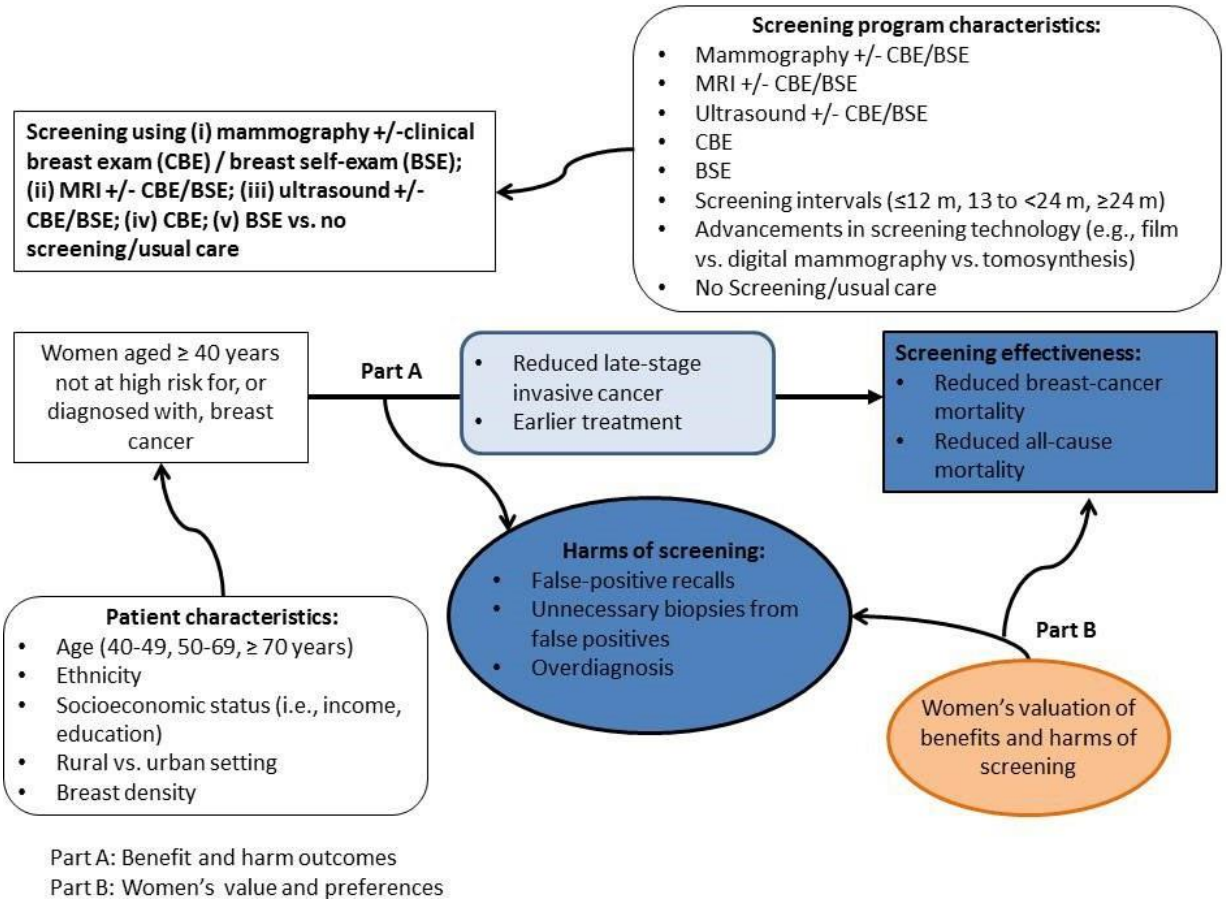
(104) Foerster V. Tomosynthesis (3D Mammography) for Breast Cancer Scrreening [Issues in emerging health technologies[Issue 135]. 2015. Canadian Agency for Drugs and Technologies in Health.

(105) Fitzpatrick-Lewis D, Hodgson N, Ciliska D, Peirson L, Gauld M, Liu YY, et al. Breast cancer screening. Calgary: Canadian Task Force on Preventive Health Care 2011.

(106) De Gelder R, Heijnsdijk EA, van Ravesteyn NT, Fracheboud J, Draisma G, De Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiologic reviews 2011;mxr009.

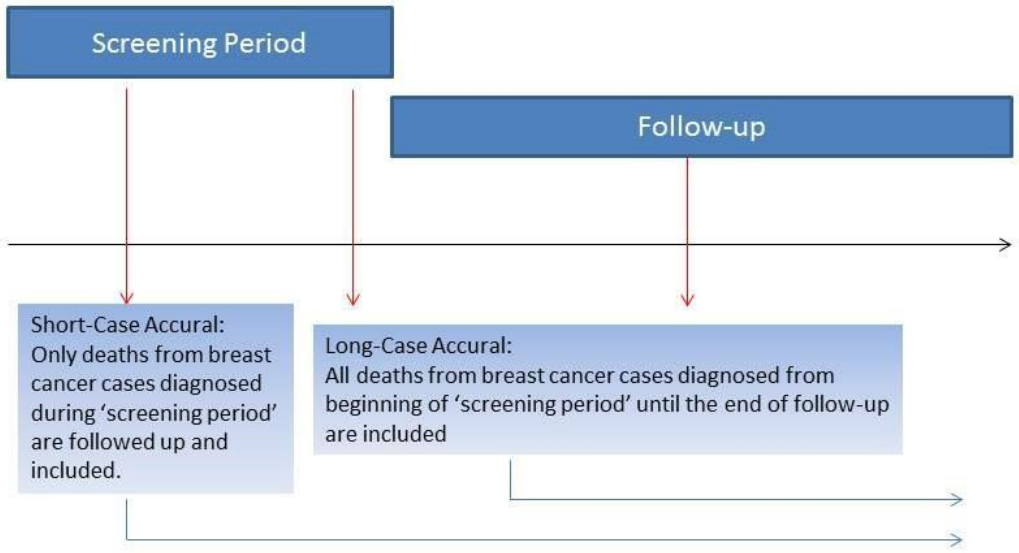
(107) Onega T, Beaber EF, Sprague BL, Barlow WE, Haas JS, Tosteson AN, et al. Breast cancer screening in an era of personalized regimens: A conceptual model and National Cancer Institute initiative for risk-based and preference–based approaches at a population level. Cancer 2014;120(19):2955-64.

Figure 1. Analytical Framework



Part A: Benefit and harm outcomes
Part B: Women's value and preferences

Figure 2. Short Case vs. Long Case Accrual

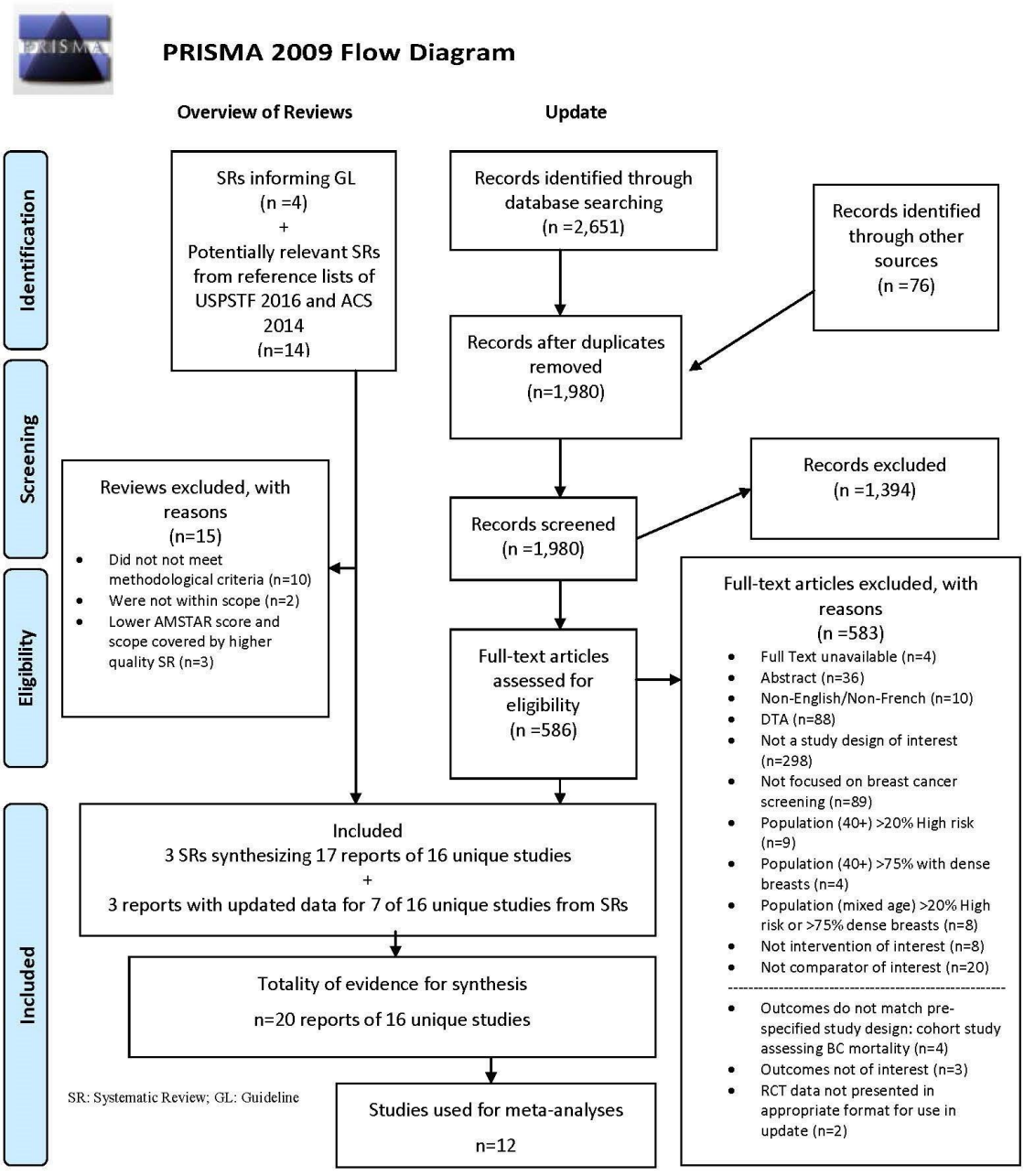


Explanation:

Short-case: Only the breast cancer cases diagnosed during the screening period (ex: 5 years) are considered; however, they are followed up to determine if they died as a result of breast cancer for an additional 10 years (for example), which would be the follow-up.

Long-case: Breast cancer cases diagnosed during the screening period (ex: 5 years) and the follow-up period (additional 10 years- regardless if they received screening with in the first 5 year screening period) are followed to determine if they died as a result of breast cancer.

Figure 3. PRISMA Flow Diagram



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

Table 1. Inclusion/Exclusion Criteria for Part A

Table 1. Inclusion/Exclusion Criteria for Part A		
	Inclusion	**Exclusion**
Population	Women aged ≥40 years of age and were not at high-risk for breast cancer ‡. Subgroups: age (40-49 years, 50-69 years, 70 years and older); ethnicity, including whether women were from indigenous populations; socioeconomic status; geographical location (rural vs. urban settings); breast density. ‡If the study population comprised of <75% of women with high breast density or the population comprised of <20% of women who were high risk, these studies were included. The definition of 'dense breasts' was defined in each included systematic review or primary study.	Cohorts comprised of 75-100% women with high breast density; men with breast cancer, women with pre-existing or personal history of breast cancer; women who were considered to be at high-risk for breast cancer on the basis of family history (in a first degree relative) of breast or ovarian cancer or other personal risk factors, such as abnormal breast pathology or BRCA1/BRCA2 genetic mutations, previously received radiation treatment to the chest (such as Hodgkin's) for cancer.
Intervention and comparator	(i) Mammography (film, digital, or tomosynthesis-3D mammography*) with or without clinical breast examination/breast self-examination vs. no screening/usual care (ii) MRI with or without clinical breast examination/breast self-examination vs. no screening/usual care (iii) Ultrasound* with or without clinical breast examination/breast self-examination vs. no screening/usual care (iv) clinical breast examination vs. no screening/usual care (v) breast self-examination vs. no screening/usual care Subgroups: screening interval (≤12 months, 13 - 24 months, >24 months); advancements in screening technology (comparison of film mammography, digital mammography or tomosynthesis); no screening vs. usual care, delayed screening (post-hoc) * Added modality since the CTFPHC 2011 review. Also (i) added +/- CBE/BSE to the following screening modalities: MRI, ultrasound; (ii) explicitly stated 'no screening/usual care' as comparator.	Combination modalities other than that already indicated.
Outcomes	• Breast cancer related mortality • All-cause mortality • Overdiagnosis* • False-positive results and consequences (e.g., FP recalls, FP recalls requiring unnecessary biopsies)	

	Overdiagnosis: screen-detected lesions that would not become clinically apparent during a woman's lifetime in the absence of screening and would not have caused any health implications. These are a subset of true positives where cancer is diagnosed, but would never have become clinically apparent otherwise. Different from false-positive as there is a diagnosis. **False-positive**: positive screening result, but cancer is not present. *Authors' calculations/or reporting of overdiagnosis were used (various iterations of the same data was possible due to differing methods of calculation from various authors). We did not undertake calculations for these outcomes. Note: Changes/modifications since CTFPHC 2011 review: -Included 'overdiagnosis' as a critical harm outcome.	
Timing	No limit	
Settings	Primary care or other settings generalizable to primary care, including referrals by primary care providers	Any setting where it could not be reasonably generalizable to a Canadian screening context
Databases	Medline, Cochrane Library	
Study designs	**Mortality outcomes*:** Randomized controlled trials, including cluster randomized controlled trials **Harms outcomes:** Overdiagnosis *: RCTs, non-RCTs, cohort studies, ecological studies (that were conducted according to a study design of interest) FP, FP consequences: As per the 2011 CTFPHC guideline, for these outcomes, we used the Canadian Breast Cancer Screening Database Registry data (Canadian Partnership Against Cancer, 2017). The 2011 approach was repeated with the use of standard tables. We also considered other Canadian studies captured in our search update and captured in the existing reviews. Subgroups: Truly randomised vs. quasi-randomised (post-hoc) *Existing systematic reviews reporting on outcomes using these study designs were extracted. If the synthesis included designs other than those of interest, a commentary was provided.	Cross-sectional studies, case series, case reports, controlled before-after, case-control studies, diagnostic test accuracy studies, modelling studies
Language	English and French	

Table 2. Determining Whether Reviews are True Systematic Reviews (Overview of Reviews)

Reference	1.More than one bibliographic database searched? (Y/N)	2.Reported selection criteria? (Y/N)	3.Quality assessment of included studies reported? (Y/N)	4.Provides a list and synthesis of included studies? (Y/N)	5. Is this an eligible SR? (Must satisfy all 4 criteria) (Y/N)
Broaeders 2012	No	Yes	No	Yes	No
Gotzsche 2013	No	Yes	Yes	Yes	No
Hofvind 2012	No	Yes	No	Yes	No
Iared 2011	Yes	Yes	No	Yes	No
Independent UK 2012	No	No	No	No	No
Magnus 2011	Yes	Yes	Yes	Yes	Yes(Include)
Marmot 2013	No	No	No	No	No
Paci 2012	No	No	No	No	No
Puliti 2012	No	No	No	Yes	No
Rashidian 2013	Yes	Yes	Yes	Yes	Yes * (Exclude)
Armstrong 2007	Yes	Yes	Yes	Yes	Yes (Include)
Nothacker 2009	Yes	Yes	Yes	Yes	Yes* (Exclude)
Harris 2011	No	Yes	Yes	Yes	No
Otto 2003	No	No	No	Unclear	No

*Excluded due to not being within scope of research question:

Rashidian 2013: cost-effectiveness

Nothacker 2009: Ultrasound as a subsequent screening modality after a negative mammogram.

The other 4 SRs listed in the protocol used to develop guidelines (USPSTF, CTFPHC, ACS and NCC) were not subjected to these criteria and were instead advanced to the next round of quality assessment (using AMSTAR).

Table 3. AMSTAR Assessment of Systematic Reviews (Overview of Reviews)

	USPSTF 2016	ACS 2014	CTFPHC 2011	NCC 2016	Magnus 2011	Armstrong 2007
1. Was an 'a priori' design provided? **(Y=Yes; N=No; CA=Can't answer; NA=not applicable)**	Yes	Yes	Yes	No	No	No
2. Was there duplicate study selection and data extraction? **(Y=Yes; N=No; CA=Can't answer; NA=not applicable)**	Yes	Yes	Yes	No	No	No
3. Was a comprehensive literature search performed? **(Y=Yes; N=No; CA=Can't answer; NA=not applicable)**	Yes	Yes	Yes	Yes	Yes	Yes
4. Was the status of publication (i.e., grey literature) used as an inclusion criterion? **(Y=Yes; N=No; CA=Can't answer; NA=not applicable)**	No	No	Yes	No	Yes	No
5. Was a list of studies (included and excluded) provided? **(Y=Yes; N=No; CA=Can't answer; NA=not applicable)**	Yes	Yes	Yes	No	No	No
6. Were the characteristics of the included studies provided?	Yes	Yes	Yes	Yes	Yes	Yes

(Y=Yes; N=No; CA=Can't answer; NA=not applicable)						
Y7. Was the scientific quality of the included studies assessed and documented? **(Y=Yes; N=No; CA=Can't answer; NA=not applicable)**	Yes	Yes	Yes	No	Yes	Yes
8. Was the scientific quality of the included studies used appropriately in formulating conclusions? **(Y=Yes; N=No; CA=Can't answer; NA=not applicable)**	Yes	Yes	Yes	No	Yes	Yes
9. Were the methods used to combine the findings of the studies appropriate? **(Y=Yes; N=No; CA=Can't answer; NA=not applicable)**	No	Not Applicable	Yes	No	Can't Answer	Not Applicable
10. Was the likelihood of publication bias assessed? **(Y=Yes; N=No; CA=Can't answer; NA=not applicable)**	No	No	Yes	No	No	No
11. Was the conflict of interest included? **(Y=Yes; N=No; CA=Can't answer; NA=not applicable)**	No	No	No	No	No	No
Final AMSTAR score out of 11 **0-3 (low); 4-7 (moderate); 8-11 (high)**	7 (Moderate)	7 (Moderate)	10 (High)	2 (Low)	5 (Moderate)	4 (Moderate)

Table 4. Designated Systematic Reviews to be used as Evidence from Overview

Table 2. Designated SRs to be used as evidence from overview.		
Screening Modality	**Outcome**	**SRs from Overview**
(i) Mammography +/- CBE/BSE vs. usual care/ no screening; (ii) MRI +/- CBE/BSE vs. usual care/ no screening; (iii) Ultrasound +/- CBE/BSE vs. usual care/no screening; (iv) CBE vs. usual care/ no screening	Breast-Cancer Mortality All-Cause Mortality	USPSTF 2016
	False Positives & False Positive Biopsies	USPSTF 2016, ACS 2014, CTFPHC 2011
	Overdiagnosis	USPSTF 2016, ACS 2014, CTFPHC 2011
BSE vs. usual care/ no screening	Breast-Cancer Mortality All-Cause Mortality False Positives & False Positive Biopsies Overdiagnosis	CTFPHC 2011

Table 5. Mammography +/- Clinical Breast Exam for Breast Cancer and All-Cause Mortality (Study Characteristics Table) (Overviews of Reviews & Updated Search)

	Malmo I	Malmo II	Stockholm	Gothenburg	CNBSS 1	CNBSS2	AGE	HIP	Swedish Two County (Ostergotland)	Swedish Two County (Kopparberg)
Year of study	1976	1978	1981	1982	1980	1980	1991	1963	1977	1978
Country (Rural/Urban)	Sweden (Urban)	Sweden (Urban)	Sweden (Urban)	Sweden (Urban)	Canada (Urban)	Canada (Urban)	UK (Urban)	USA (Urban)	Sweden (Urban)	Sweden (Urban)
Study Design	RCT	RCT	Quasi-RCT	Quasi-RCT	RCT	RCT	RCT	RCT	Cluster-RCT[A]	Cluster-RCT[A]
Age at Entry Total Randomized (N)	45-70 N=42,283	43-49 N=17,793	39-65 N=60,117	39-59 N=50,200	40-49 N=50,489	50-59 N=39,459	39-41 N=160,921	40-64 N=61,004[B]	40-74 N=75,894	40-74 N=57,171
Ethnicity	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
SES	NR	NR	NR	NR	Level of Education, Occupation (p-values not provided)	Level of Education, Occupation (p-values not provided)	NR	NR	NR	NR
% Breast Density	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Longest Follow-up Reported	30 yrs (mean)*	22 yrs (mean)*	25 yrs (mean)*	24 yrs (mean)*	21.9 yrs (mean)*	21.9 yrs (mean)*	17.7 yrs (median)*	18 yrs (mean)	25.7 yrs (mean)	25.7 yrs (mean)
Intervention (type) *n* randomized	M (Film) (n=21,088)	M (Film) (n=9,581)	M (Film) (n=39,139)	M (Film) (n=21,000)	M (Film) + CBE (n=25,246)	M (Film) + CBE (n=19,735)	M (NR) (n=53,914)	M (Film) + CBE (n=30,239)[B]	M (NR) (n=38,491)	M (NR) (n=38,589)
Comparator *n* randomized	UC (n=21,195)	UC (n=8,212)	UC (n=20,978)	UC (n=29,200)	UC (n=25,243)	CBE (n=19,724)	UC (n=107,007)	UC (n=30,765)[B]	UC (n=37,403)	UC (n=18,582)
Received screening at end of study	No	Yes	Yes	Yes	No	No	Yes	NR	Yes	Yes

period?										
# of views	2[C]	2	1	2[D]	2	2	2[E]	2	1	1
# of readers	2	2	1	1[F]	NR	NR	NR	NR	1	1
Screening Interval	18-24 mo.	18-24 mo.	28 mo.	18 mo.	12 mo.	12 mo.	12 mo.	12 mo.	24-33 mo. [G]	24-33 mo. [G]
Duration of Screening	12 yrs	12 yrs	4 yrs	7 yrs	4 yrs	4 yrs	8 yrs	3 yrs	7 yrs	7 yrs
Attendance Rate	74%	74%	82%	84%	88%	88%	81%	65%	85%	85%

M=Mammography; NR= Not Reported; CBE= Clinical Breast Exam UC= Usual Care; SES= Socioeconomic status

*Updated follow-up from studies identified in updated search.

[A] Geographic clusters within each county stratified by socioeconomic status.

[B] The number randomized is unclear. The number analyzed is presented.

[C] Starting at round 3: used either single or two views depending on breast density.

[D] Starting at round 2: used either single or two views depending on breast density.

[E] Starting at round 2: used single view.

[F] Starting at round 4: used two readers.

[G] 40-49 yrs: average 24 mo.; 50-59 yrs: average 33 mo.

Table 6. Mammography +/- Clinical Breast Exam for Overdiagnosis (Study Characteristics Table-RCTs) (Overviews of Reviews & Updated Search)

Randomized Controlled Trials (from Overview and Update) for Overdiagnosis			
	Malmo I (Zackrisson 2006)	**CNBSS 1 & 2 (Baines 2016)***	**Swedish Two County (Kopparberg) (Yen 2012)**
Year of study	1976	1980	1978
Country (Rural/Urban)	Sweden (Urban)	Canada (Urban)	Sweden (Urban)
Study Design	RCT	RCT	Cluster-RCT[A]
Age at Entry Total Randomized (N)	45-70 N=42,283 For analysis: 55-69 (controls never screened)	40-59 N=89,948	40-74 N=57,171
Ethnicity	NR	NR	NR
SES	NR	Level of education and Occupation (p-values not provided)	NR
% Breast Density	NR	NR	NR
Longest Follow-up Reported	25 years of follow up (including screening period)	20 years of follow up (including screening period)	29 years of follow up (including screening period)
Intervention (type) ***n* randomized**	M (Film) (n=21,088)	M (Film) + CBE (n=44,967)	M (NR) (n=38,589)
Comparator ***n* randomized**	UC (n=21,195)	UC/CBE (n=44,967	UC (n=18,582)
Received screening at end of study period?	No	No	Yes
# of views	2[B]	2	1
# of readers	2	NR	1
Screening Interval	18-24 mo.	12 mo.	24-33 mo. [D]
Duration of Screening	12 yrs	4 yrs	7 yrs
Attendance Rate	74%	88%	85%
Methodological Approach	Excess Incidence "Comparison of incidence in screened vs. unscreened." (USPSTF 2016)	"The numerator is the difference in numbers of cancers in the mammography arm less those in the control arm; and the denominator is the # of screen-detected cancers in the	Excess Incidence "Cumulative incidence in active screening vs. usual care groups."

		mammography arm" (Baines 2016)	
Overdiagnosis Estimate**	**Invasive + in situ:** 10% of incidence in control group **Invasive**: 7% **In situ**: 3%	**40-49:** ***5 years post-screen*** **Invasive + in situ:** 41% **Invasive**: 32% ***20 years post-screen*** **Invasive + in situ:** 55% **Invasive**: 48% **50-59:** ***5 years post-screen*** **Invasive + in situ:** 25% **Invasive**: 16% ***20 years post-screen*** **Invasive + in situ:** 16% **Invasive:** 5%	**Invasive + in situ:** RR 1.00 (0.92-1.08) **Invasive:** RR 0.99 (0.92-1.07) **In situ:** RR 1.17 (0.88-1.55)
Overall Risk of Bias Judgement	Moderate Risk	Moderate Risk	High Risk

M=Mammography; NR= Not Reported; CBE= Clinical Breast Exam UC= Usual Care; SES= Socioeconomic status

Note: Since all 3 RCTs were included in the outcomes (breast cancer mortality and all-cause mortality), we used the study characteristics and risk of bias extracted for those outcomes. Overdiagnosis characteristics were extracted as was reported in the overviews.

*Updated follow-up from studies identified in updated search. Revised estimates to replace previous publication of Miller 2014.

** Numerator and denominator could not be provided, since a majority of studies were from the overviews or reviews, and as per the methodology, data was extracted as reported by the reviews.

[A] Geographic clusters within each county stratified by socioeconomic status.

[B] Starting at round 3: used either single or two views depending on breast density.

[D] 40-49 yrs: average 24 mo.; 50-59 yrs: average 33 mo.

Table 7. Mammography +/- Clinical Breast Exam for Overdiagnosis (Study Characteristics Table-Cohort) (Overview of Reviews)

Cohort studies (from overview) for Overdiagnosis				
	Lund 2013	**Puliti 2012**	**Hellquist 2012**	**Njor 2013**
Years	2002-2010	1991-2009	1986-2005	1991-2009
Country	Norway (Norwegian Breast Screening Program)	Italy (Florentine Screening Program)	Sweden (SCRY cohort)	Denmark
Study Design	Cohort	Cohort	Cohort	Cohort
Age at Entry	52-79	60-69	40-49	56-70
% Breast Density	NR	NR	NR	NR
Longest Follow-up Reported	NR	NR	NR	NR
Intervention (type) ***n***	NR	NR	NR	NR
Comparator ***n***	NR	NR	NR	NR
# of views	NR	NR	NR	NR
# of readers	NR	NR	NR	NR
Screening Interval	24 mo.	24mo.	24 mo.	NR
Duration of Screening	NR	NR	NR	NR
Attendance Rate	NR	NR	NR	NR
Methodological Approach	Excess Incidence "Cases overdiagnosed/all diagnosed cancers." (ACS 2014)	Excess Incidence "Excess cases/cases observed in unscreened women (non-attenders)." (ACS 2014)	Excess Incidence "Cases overdiagnosed/ cases expected without screening." (ACS 2014)	Excess Incidence "Cumulative incidence in screened population vs. expected incidence in unscreened counties." (USPSTF 2016)
Overdiagnosis Estimate	**Invasive + in situ:** 22% (-0.9% to 64%)[A] **Invasive:** 7% (-0.8% to 45%)[A]	**Invasive + in situ**: 10% (-2% to 23%) **Invasive**: 5% (-7% to 18%)	**Invasive + in situ:** RR 1.01 (0.94-1.08) **Invasive**: RR 0.95 (0.88-1.01)	≥8 years of follow up: **Invasive + in situ:** Copenhagen: 3% (-14% to 25%) Funen: 0.7% (-9% to 12%)
Risk of Bias (New-Castle Ottawa Scale)	Selection (max 4 stars): *** Comparability (max 2 stars): **	Selection (max 4 stars): **** Comparability (max 2 stars): *	Selection (max 4 stars): *** Comparability (max 2 stars): No Stars	Selection (max 4 stars): * Comparability (max 2 stars): No Stars

	Outcome (max 4 stars): *	Outcome (max 4 stars): *	Outcome (max 4 stars): **	Outcome (max 4 stars): *
M=Mammography; NR= Not Reported (study characteristics were extracted as reported from the overviews); CBE= Clinical Breast Exam UC= Usual Care; SES= Socioeconomic status Note: Study characteristics were extracted as reported from the overviews. Quality assessment was re-assessed for this report as reporting was unclear between the existing reviews. ** Numerator and denominator could not be provided, since a majority of studies were from the overviews or reviews, and as per the methodology, data was extracted as reported by the reviews. [A] Calculated from inverse of OR (incidence in unscreened vs screened). Adjusted for age, parity, HRT, maternal history of breast cancer, BMI, education.				

Table 8. False Positive and Biopsy on False Positive Calculations (Breast Cancer Screening Cohort)

False Positives and Unnecessary Biopsies from an Estimated Cohort of Women in a Breast Screening Program[1]										
	All Ages*		40-49 years		50-59 years		60-69 years		70-74 years	
	Short-Case[2]	Long-Case[3]	Short-Case[2]	Long-Case[3]	Short-Case[2]	Long-Case[3]	Short-Case[2]	Long-Case[3]	Short-Case[2]	Long-Case[3]
Short-Case reflects a median follow-up of 23 years (8 cycles) Long-Case reflects a median follow-up of 14 years (5 cycles)										
Per 1,000 women screened										
FP Mammography	611	416	660	442	652	437	578	385	495	329
Biopsies on FP	79	54	90	64	80	55	76	51	68	45
Per one breast cancer death prevented										
NNS (95%CI)	1,471 (1,000 to 3,125) L: 3,226 (2,222 to 6,667) M: 1,724 (1,176 to 3,704) H: 1,099 (746 to 2,326)	840 (521 to 2,500) L: 1,149 (714 to 3,448) M: 725 (450 to 2,174) H: 493 (306 to 1,471)	2,000 (1,042 to -25,000) CI includes ∞	3,704 (1,667 to -10,111) CI includes ∞ L: 5,882 (2,564 to -14,286) CI includes ∞ M: 4,000 (1,754 to -10,000) CI includes ∞ H: 3, 030 (1,333 to -8,333) CI includes ∞	1,136 (625 to ∞) CI includes ∞ L: 2,041 (1,111 to ∞) M: 1,250 (690 to ∞) H:730 (402 to ∞)	962 (541 to 8,333) L: 1,220 (690 to 11,111) H: 847 (478 to 7,692)	541 (369 to 1,351)	452 (316 to 1,053)	885 (231 to -255) CI includes ∞	699 (299 to -662) CI includes ∞ L: 746 (321 to -714) CI includes ∞ H: 610 (262 to -585) CI includes ∞
FP Mammography	899	350	1,320	1,639	741	420	312	174	438	230
Biopsies on FP	115	45	180	242	91	53	41	23	60	31

L: Low baseline risk, M: Moderate baseline risk; H: High baseline risk

*The data is presented as the weighted average.

[1]The data is used to approximate a cohort of women entering the screening program. Although assumed, but not confirmed, the 'initial screen' in the CPAC report is the first screen documented in the database, and may not necessarily be the first 'true' screen of a woman. This is especially true for data originating from Alberta.

[2] Short-Case Accrual: data is estimated for 8 cycles over a median of 23 years, assuming women get screened every 2-3 years. The median of 23 years is reflective of the follow-up time of the studies included in the meta-analysis for short-case breast cancer mortality. Calculation: Initial + 7 (Subsequent).

[3] Long-Case Accrual: data is estimated for 5 cycles over a median of 14 years, assuming women get screened every 2-3 years. The median of 14 years is reflective of the follow-up time of the studies included in the meta-analysis for long-case breast cancer mortality. Calculation: Initial + 4 (Subsequent).

Table 9a. False Positive and Biopsy on False Positive Calculations (Cross-Sectional Analysis-Initial & Subsequent)

False Positives and Unnecessary Biopsies from a cross-sectional population using Breast Cancer Screening Program Data (Cycle 2011-2012)[1]																				
	All Ages*				40-49 years				50-59 years				60-69 years				70-74 years			
	Short-Case		Long-Case		Short-Case		Long-Case		Short-Case		Long-Case		Short-Case		Long-Case		Short-Case		Long-Case	
	I	S	I	S	I	S	I	S	I	S	I	S	I	S	I	S	I	S	I	S
Per 1,000 women screened																				
FP Mammography	145	67	145	67	148	73	148	73	151	73	151	73	128	64	128	64	109	55	109	55
Biopsies on FP	21	8	21	8	28	9	28	9	21	9	21	9	18	8	18	8	15	8	15	8

*The data is presented as the weighted average.

[1]This analysis was undertaken, under the assumption that the initial screen was likely the woman's first screen. This was demonstrated by the higher FP rates under initial vs. subsequent, which leads us to believe that a true first screen was taking place since the higher FP rates were likely reflective of prevalent cases being identified.

I=Initial; S=subsequent

Table 9b. False Positive and Biopsy on False Positive Calculations (Weighted Cross-Sectional Analysis)

False Positives and Unnecessary Biopsies from a cross-sectional population using Breast Cancer Screening Program Data (Cycle 2011-2012)[1]										
	All Ages (Weighted*)		40-49 years (Weighted*)		50-59 years (Weighted*)		60-69 years (Weighted*)		70-74 years (Weighted*)	
	Short-Case	Long-Case	Short-Case	Long-Case	Short-Case	Long-Case	Short-Case	Long-Case	Short-Case	Long-Case
Per 1,000 women screened										
FP Mammography	80	80	92	92	90	90	69	69	58	58
Biopsies on FP	11	11	14	14	12	12	9	9	7	7

[1]This analysis was undertaken, under the assumption that we cannot be certain that the woman's 'initial' screen is in fact her 'true' first screen.

*The data is presented as the weighted average with initial + subsequent combined (weighted by the total population receiving initial and subsequent screens per age group).

Table 10. Breast Self-Exam for Breast-Cancer Mortality (Study Characteristics Table) (Overview of Reviews)

Randomized Controlled Trials (from overview) for Breast Cancer Mortality	
	Thomas 2002
Year of study	1989
Country **(Rural/Urban)**	Shanghai (urban)
Study Design	Cluster-RCT
Age at Entry Total Randomized (N)	31-65
Ethnicity	NR
SES	NR
% Breast Density	NR
Longest Follow-up Reported	NR
Intervention (type) ***n* randomized**	BSE n=NR
Comparator ***n* randomized**	No Screening n=NR
Received screening at end of study period?	NR
# of views	NR
# of readers	NR
Screening Interval	NR

Duration of Screening	NR
Attendance Rate	NR

NR= Not Reported; SES= Socioeconomic status

Note: Results for Thomas were reported in the CTPFHC 2011 under 'All-Cause Mortality'; however, upon verification, this data was actually classified under 'Breast-Cancer Mortality'.

Study characteristics were extracted as was reported in the overviews.

Table 11. Breast Self-Exam for All-Cause Mortality (Study Characteristics Table) (Overview of Reviews)

Randomized Controlled Trials (from overview) for All-Cause Mortality		
	Thomas 2002	**Semiglazov 2003**
Year of study	1989	1985
Country **(Rural/Urban)**	Shanghai, China (Urban)	St. Petersburg, Russia (Urban)
Study Design	Cluster-RCT	Cluster-RCT
Age at Entry Total Randomized (N)	31-65	40-64
Ethnicity	NR	NR
SES	NR	NR
% Breast Density	NR	NR
Longest Follow-up Reported	NR	NR
Intervention (type) ***n* randomized**	BSE n=NR	BSE n=NR
Comparator ***n* randomized**	No Screening n=NR	No Screening n=NR
Received screening at end of study period?	NR	NR
# of views	NR	NR
# of readers	NR	NR
Screening Interval	NR	NR
Duration of Screening	NR	NR

Attendance Rate	NR	NR

NR= Not Reported; SES= Socioeconomic status

Note: Results for Thomas were not reported in the USPSTF 2009 and CTFPHC 2011; however, upon verification, the data was provided in the publication. Results initially in the CTPFHC 2011 were from Semiglazov 1999 (and it was unclear whether it is for breast-cancer or all-cause mortality). USPSTF 2009 used Semiglazov 2003, and the CTPFHC 2011 mentions including Semiglazov 2003, but instead uses results form Semiglazov 1992 (which data actually comes from Semiglazov 1999). To remain consistent and clear, we will use the estimates from Semiglazov 2003 from the USPSTF 2009 review.

Study characteristics were extracted as was reported in the overviews.

Table 12. Mammography +/- Clinical Breast Exam for Breast-Cancer Mortality- Short-Case Accrual (Risk of Bias) (Overview of Reviews and Updated Search)

Study	Randomisation	Allocation Concealment	Blinding (patients/personnel)	Blinding (outcomes)	Incomplete Outcome Assessments	Selective Outcome Reporting	Other	Overall Judgement
Short-Case Accrual (All Ages)								
Gothenburg (Nystrom 2016)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
CNBSS 1 & 2 (Miller 2014)	Low Risk (randomised by prepared allocation lists-block stratified)	Unclear (all patients received CBE before being allocated to groups. 17 vs. 5 women in intervention vs. control had tumours with >4 nodes at initial screen (p-value: not reported)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (we are aware of a protocol, but this was not referenced in the any citations from the overviews or reviews or updated search)	Low Risk	Moderate
AGE (Moss 2015)	Low Risk (randomisation software)	Unclear (not described)	Unclear	Low Risk (independent endpoint committee, national registries)	Low Risk (<10% losses in each group, balanced reasons)	High Risk (protocol does not list follow-up assessment timing)	Low Risk	Moderate
Malmo I (Nystrom 2016)	Unclear (not described) NB: More stringent criteria for randomisation employed compared to previous CTPFHC	Unclear (not described)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	Moderate

	2011. It was determined that the descriptions for the Malmo trials were not sufficiently detailed.							
Malmo II (Nystro 2016)	Unclear (not described) NB: More stringent criteria for randomisation employed compared to previous CTPFHC 2011. It was determined that the descriptions for the Malmo trials were not sufficiently detailed.	Unclear (not described)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	Moderate
Stockholm (Nystrom 2016)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
Swedish Two County (Tabar 2011)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear(Screening women slightly older than controls)- However, could not verify from publication cited in USPSTF 2002	High
HIP (Shapiro)	Unclear (not described)	Unclear (not described)	Unclear	High Risk (Local endpoint committee)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used	Unclear (no protocol)	Unclear(More menopausal women in control and women with previous breast lumps vs.	High

					in analysis)		screening group, More educated women in the screening group) However, could not verify from publication cited in USPSTF 2002	
Short-Case Accrual (Stratified by Age)								
40-49 years								
AGE (Moss 2015)	Low Risk (randomisation software)	Unclear (not described)	Unclear	Low Risk (independent endpoint committee, national registries)	Low Risk (<10% losses in each group, balanced reasons)	High Risk (protocol does not list follow-up assessment timing)	Low Risk	Moderate
Malmo I (Nystrom 2002)	Unclear (not described) NB: More stringent criteria for randomisation employed compared to previous CTPFHC 2011. It was determined that the descriptions for the Malmo trials were not sufficiently detailed.	Unclear (not described)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	Moderate
Gothenburg (Nystrom 2016)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
Malmo II (Nystrom 2016)	Unclear (not described)	Unclear (not described)	Unclear	Low Risk (independent endpoint	Unclear (No consort flow diagram, unclear	Unclear (no protocol)	Low Risk	Moderate

	NB: More stringent criteria for randomisation employed compared to previous CTPFHC 2011. It was determined that the descriptions for the Malmo trials were not sufficiently detailed.			committee, national registries)	number of patients lost to follow-up or used in analysis)			
Stockholm (Nystrom 2016)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
Swedish Two County (Tabar 2011)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
CNBSS-1 (Miller 2014)	Low Risk (randomised by prepared allocation lists-block stratified)	Unclear (all patients received CBE before being allocated to groups. 17 vs. 5 women in intervention vs. control had tumours with >4 nodes at initial screen (p-value: not reported)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (we are aware of a protocol, but this was not referenced in the any citations from the overviews or reviews or updated search)	Low Risk	Moderate
HIP (Shapiro 1988)	Unclear (not described)	Unclear (not described)	Unclear	High Risk (Local endpoint committee)	Unclear (No consort flow diagram, unclear number of patients lost to	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could	High

					follow-up or used in analysis)		not verify from publication cited in USPSTF 2002	
50-59 years								
Gothenburg (Nystrom 2016)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
Malmo I (Nystrom 2016)	Unclear (not described) NB: More stringent criteria for randomisation employed compared to previous CTPFHC 2011. It was determined that the descriptions for the Malmo trials were not sufficiently detailed.	Unclear (not described)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	Moderate
Stockholm (Nystrom 2016)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
Swedish Two County (Tabar 2011)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High

CNBSS-2 (Miller 2014)	Low Risk (randomised by prepared allocation lists-block stratified)	Unclear (all patients received CBE before being allocated to groups.	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (we are aware of a protocol, but this was not referenced in the any citations from the overviews or reviews or updated search)	Low Risk	Moderate
HIP (Shapiro 1988)	Unclear (not described)	Unclear (not described)	Unclear	High Risk (Local endpoint committee)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
60-69 years								
Stockholm (Nystrom 2002)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
Malmo I (Nystrom 2016)	Unclear (not described) NB: More stringent criteria for randomisation employed compared to previous CTPFHC 2011. It was determined that the descriptions for the Malmo trials were not sufficiently detailed.	Unclear (not described)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	Moderate

Swedish Two County (Tabar 2011)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
HIP (Shapiro 1988)	Unclear (not described)	Unclear (not described)	Unclear	High Risk (Local endpoint committee)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
70-74 years								
Malmo I (Nystrom 2002)	Unclear (not described) NB: More stringent criteria for randomisation employed compared to previous CTPFHC 2011. It was determined that the descriptions for the Malmo trials were not sufficiently detailed.	Unclear (not described)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	Moderate
Swedish Two County (Tabar 2011)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High

Table 13. Mammography +/- Clinical Breast Exam for Breast-Cancer Mortality- Long-Case Accrual (Risk of Bias) (Overview of Reviews and Updated Search)

Study	Randomization	Allocation Concealment	Blinding (patients/personnel)	Blinding (outcomes)	Incomplete Outcome Assessments	Selective Outcome Reporting	Other	Overall Judgement
Long-Case Accrual (All Ages)								
Gothenburg (Bjurstam 2003)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
CNBSS 1 & 2 (Miller 2014)	Low Risk (randomised by prepared allocation lists-block stratified)	Unclear (all patients received CBE before being allocated to groups. 17 vs. 5 women in intervention vs. control had tumours with >4 nodes at initial screen (p-value: not reported)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (we are aware of a protocol, but this was not referenced in the any citations from the overviews or reviews or updated search)	Low Risk	Unclear
AGE (Moss 2015)	Low Risk (randomisation software)	Unclear (not described)	Unclear	Low Risk (independent endpoint committee, national registries)	Low Risk (<10% losses in each group, balanced reasons)	High Risk (protocol does not list follow-up assessment timing)	Low Risk	Moderate
Swedish Two County-Kopparberg (Tabar 1995)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	High (Screning owmen slightly older than controls)	High
Swedish Two County-Ostergotland (Tabar	Low Risk (flipping a coin)	High Risk (no indication it was a	Unclear	Low Risk (independent endpoint	Unclear (No consort flow diagram, unclear	Unclear (no protocol)	High (Screning owmen	High

1995)		concealed process)		committee, national registries)	number of patients lost to follow-up or used in analysis)		slightly older than controls)	
HIP (Habbema 1986)	Unclear (not described)	Unclear (not described)	Unclear	High Risk (Local endpoint committee)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear(More menopausal women in control and women with previous breast lumps vs. screening group, More educated women in the screening group) However, could not verify from publication cited in USPSTF 2002	High
40-49 years								
Gothenburg (Bjurstam 2003)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
AGE (Moss 2009)	Low Risk (randomisation software)	Unclear (not described)	Unclear	Low Risk (independent endpoint committee, national registries)	Low Risk (<10% losses in each group, balanced reasons)	High Risk (protocol does not list follow-up assessment timing)	Low Risk	Moderate
Swedish Two County-Kopparberg (Tabar 1995)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population).	High

				registries)	in analysis)		However, could not verify from publication cited in USPSTF 2002	
Swedish Two County-Ostergotland (Tabar 1995)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
CNBSS 1 (Miller 2014)	Low Risk (randomised by prepared allocation lists-block stratified)	Unclear (all patients received CBE before being allocated to groups. 17 vs. 5 women in intervention vs. control had tumours with >4 nodes at initial screen (p-value: not reported)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (we are aware of a protocol, but this was not referenced in the any citations from the overviews or reviews or updated search)	Low Risk	Moderate
HIP (Habbema 1986)	Unclear (not described)	Unclear (not described)	Unclear	High Risk (Local endpoint committee)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
50-59 years								

Gothenburg (Bjurstam 2003)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
Swedish Two County-Kopparberg (Tabar 1995)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
Swedish Two County-Ostergotland (Tabar 1995)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
CNBSS 2 (Miller 2014)	Low Risk (randomised by prepared allocation lists-block stratified)	Unclear (all patients received CBE before being allocated to groups.	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (we are aware of a protocol, but this was not referenced in the any citations from the overviews or reviews or updated search)	Low Risk	Moderate
HIP (Habbema 1986)	Unclear (not	Unclear (not	Unclear	High Risk (Local endpoint	Unclear (No consort flow	Unclear (no	Unclear (Baseline	High

	described)	described)		committee)	diagram, unclear number of patients lost to follow-up or used in analysis)	protocol)	imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	
60-69 years								
Swedish Two County-Kopparberg (Tabar 1995)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
Swedish Two County-Ostergotland (Tabar 1995)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
HIP (Habbema 1986)	Unclear (not described)	Unclear (not described)	Unclear	High Risk (Local endpoint committee)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from	High

							publication cited in USPSTF 2002	
70-74 years								
Swedish Two County-Kopparberg (Tabar 1995)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
Swedish Two County-Ostergotland (Tabar 1995)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (independent endpoint committee, national registries)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High

Table 14. Mammography +/- Clinical Breast Exam for All-Cause Mortality (Risk of Bias) (Overview of Reviews and Updated Search)

Study	Randomization	Allocation Concealment	Blinding (patients/personnel)	Blinding (outcomes)	Incomplete Outcome Assessments	Selective Outcome Reporting	Other	Overall Judgement
All Ages								
Gothenburg (Nystrom 2002)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
CNBSS 1 & 2 (Miller 2014)	Low Risk (randomised by prepared allocation lists-block stratified)	Unclear (all patients received CBE before being allocated to groups. 17 vs. 5 women in intervention vs. control had tumours with >4 nodes at initial screen (p-value: not reported)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (we are aware of a protocol, but this was not referenced in the any citations from the overviews or reviews or updated search)	Low Risk	Moderate
Swedish Two County-Ostergotland (Nystrom 2002)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	High (Screning women slightly older than controls)	High
Stockholm (Nystrom 2002)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
HIP (Aron and Prorak	Unclear (not	Unclear (not	Unclear	Low Risk (all-cause death is	Unclear (No consort flow	Unclear (no	Unclear(More menopausal	Moderate

1986)	described)	described)		objective)	diagram, unclear number of patients lost to follow-up or used in analysis)	protocol)	women in control and women with previous breast lumps vs. screening group, More educated women in the screening group) However, could not verify from publication cited in USPSTF 2002	
Malmo II (Nystrom 2002)	Unclear (not described) NB: More stringent criteria for randomisation employed compared to previous CTPFHC 2011. It was determined that the descriptions for the Malmo trials were not sufficiently detailed.	Unclear (not described)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	Moderate
Malmo I (Nystrom 2002)	Unclear (not described) NB: More stringent criteria for randomisation employed compared to previous CTPFHC 2011. It was determined that the descriptions for the Malmo	Unclear (not described)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	Moderate

	trials were not sufficiently detailed.							
AGE (Moss 2015)	Low Risk (randomisation software)	Unclear (not described)	Unclear	Low Risk (all-cause death is objective)	Low Risk (<10% losses in each group, balanced reasons)	High Risk (protocol does not list follow-up assessment timing)	Low Risk	Moderate
40-49 years								
AGE (Moss 2015)	Low Risk (randomisation software)	Unclear (not described)	Unclear	Low Risk (all-cause death is objective)	Low Risk (<10% losses in each group, balanced reasons)	High Risk (protocol does not list follow-up assessment timing)	Low Risk	Moderate
Malmo II (Nystrom 2002)	Unclear (not described) NB: More stringent criteria for randomisation employed compared to previous CTPFHC 2011. It was determined that the descriptions for the Malmo trials were not sufficiently detailed.	Unclear (not described)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	Moderate
Swedish Two County-Ostergotland (Tabar 1989)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High

Swedish Two County-Kopparberg (Tabar 1989)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
Stockholm (Frisell 1997)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to randomisation)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
CNBSS 1 (Miller 2002)	Low Risk (randomised by prepared allocation lists-block stratified)	Unclear (all patients received CBE before being allocated to groups. 17 vs. 5 women in intervention vs. control had tumours with >4 nodes at initial screen (p-value: not reported)	Unclear	Low Risk (all-cause death is objective)	Low Risk (<10% losses in each group, balanced reasons)	Unclear (we are aware of a protocol, but this was not referenced in the any citations from the overviews or reviews or updated search)	Low Risk	Moderate
Gothenburg (Bjurstam 1997)	High Risk (randomised by date of birth)	High Risk (could not be concealed due to poor randomisation process)	Unclear	Low Risk (all-cause death is objective)s)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	High
50-59 years								
Swedish Two County-Ostergotland (Tabar 1989)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole	High

					follow-up or used in analysis)		population). However, could not verify from publication cited in USPSTF 2002	
Swedish Two County-Kopparberg (Tabar 1989)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
CNBSS 2 (Miller 2000)	Low Risk (randomised by prepared allocation lists-block stratified)	Unclear (all patients received CBE before being allocated to groups.	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (we are aware of a protocol, but this was not referenced in the any citations from the overviews or reviews or updated search)	Low Risk	Moderate
60-69 years								
Swedish Two County-Ostergotland (Tabar 1989)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High

Swedish Two County-Kopparberg (Tabar 1989)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
70-74 years								
Swedish Two County-Ostergotland (Tabar 1989)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High
Swedish Two County-Kopparberg (Tabar 1989)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear (Baseline imbalances pertain to whole population). However, could not verify from publication cited in USPSTF 2002	High

Table 15. Mammography +/- Clinical Breast Exam for Overdiagnosis (Risk of Bias-RCTs) (Overview of Reviews and Updated Search)

Study	Randomization	Allocation Concealment	Blinding (patients/personnel)	Blinding (outcomes)	Incomplete Outcome Assessments	Selective Outcome Reporting	Other	Overall Judgement
All Ages								
Malmo I (Zackrisson 2006)	Unclear (not described) NB: More stringent criteria for randomisation employed compared to previous CTPFHC 2011. It was determined that the descriptions for the Malmo trials were not sufficiently detailed.	Unclear (not described)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Low Risk	Moderate
CNBSS 1 & 2 (Baines 2016)	Low Risk (randomised by prepared allocation lists-block stratified)	Unclear (all patients received CBE before being allocated to groups. 17 vs. 5 women in intervention vs. control had tumours with >4 nodes at initial screen (p-value: not reported)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (we are aware of a protocol, but this was not referenced in the any citations from the overviews or reviews or updated search)	Low Risk	Moderate
Swedish Two County (Kopparberg) (Yen 2012)	Low Risk (flipping a coin)	High Risk (no indication it was a concealed process)	Unclear	Low Risk (all-cause death is objective)	Unclear (No consort flow diagram, unclear number of patients lost to follow-up or used in analysis)	Unclear (no protocol)	Unclear(Screening women slightly older than controls)-However, could not verify from publication cited	High

							in USPSTF 2002)	

Table 16. Mammography +/- Clinical Breast Exam for Overdiagnosis (Risk ofBias- Cohort) (Overview of Reviews)

	Selection. Max 4 stars (****)				Comparability. Max 2 stars (**)	Outcomes. Max 4 stars (****)			
Author, Year	1.Representativeness of the exposed cohort	2. Selection of the non-exposed cohort	3. Ascertainment of exposures	4. Demonstration that outcome of interest was not present at start of study	1. Comparability of cohorts on the basis of the design or analysis (controls for age and hormone replacement therapy use)	1. Assessment of outcome	2. Was follow-up long enough for outcomes to occur	3. Did the authors adjust for lead time bias in the analysis (or was follow up long enough to reduce lead time bias)?§	4. Adequacy of follow up of cohorts
Lund 2013 Selection: *** Compar: ** Outcome: *	Somewhat rep. of target population * (includes high risk population-maternal history and HRT use. Represent. of ages 50+)	Drawn from same community as exposed cohort *	Written self-report (Surveyed mammogram use)	Yes*	Controls for age and HRT use **	Record linkage*	No/Unclear (there was no end of screening period, and therefore follow-up is unclear)	No/Unclear	No statement
Puliti 2012 Selection: **** Compar: * Outcome: *	Somewhat rep. of target population * (ages 50-69. No mention of excluding high risk)	Drawn from same community as exposed cohort *	Secure record *	Yes*	Controls for age (or does a sub-group)*	Record linkage*	No/Unclear (there was no end of screening period, and therefore follow-up is unclear)	No/Unclear (only for 60-69 at entry)	No statement
Hellquist 2012 Selection: ***	Somewhat rep. of target population * (ages 40-49). No mention of	Drawn from same community as exposed cohort *	Written self-report (Survey to the screening program)	Yes*	None	Record linkage*	No/Unclear (there was no end of screening period, and therefore follow-up is unclear)	Yes*	No statement

Compar: No stars Outcome: **	excluding high risk)								
Njor 2013 Selection: ** Compar: No stars Outcome: *	Somewhat rep. of target population * (ages 50-69). No mention of excluding high risk)	Drawn from same community as exposed cohort *	No description	No/Unclear	None	Record linkage*	No/Unclear (there was no end of screening period, and therefore follow-up is unclear)	No/Unclear	No statement
§Added to address the nature of the topic.									

Table 17. Breast Self-Exam for Breast-Cancer Mortality (Risk of Bias) (Overview of Reviews)

Study	Randomization	Allocation Concealment	Blinding (patients/personnel)	Blinding (outcomes)	Incomplete Outcome Assessments	Selective Outcome Reporting	Other	Overall Judgement
All Ages								
Thomas 2002	Unclear	Unclear	Unclear	Low Risk	High Risk	Unclear	Low Risk	Moderate

Table 18. Breast Self-Exam for All-Cause Mortality (Risk of Bias) (Overview of Reviews)

Study	Randomization	Allocation Concealment	Blinding (patients/personnel)	Blinding (outcomes)	Incomplete Outcome Assessments	Selective Outcome Reporting	Other	Overall Judgement
All Ages								
Thomas 2002	Unclear	Unclear	Unclear	Low Risk	High Risk	Unclear	Low Risk	Moderate
Semiglazov 2003	Low Risk	Unclear	Unclear	Low Risk	Unclear	Unclear	Low Risk	Moderate

Evidence Set 1- Mammography +/- Clinical Breast Exam- Breast-Cancer Mortality (Short-Case Accrual) (All Ages)

Part A- GRADE Evidence Profile Table – Breast Cancer Mortality (Short-Case Accrual) (All Ages)

Included Studies:

1: Gothenburg (Nystrom 2016) 2: Age (Moss 2015) 3. Swedish Two County (Kopparberg & Ostergotland) (Tabar 2011)

4: Malmo I (Nystrom 2016) 5: Malmo II (Nystrom 2016) 6: CNBSS 1 & 2 (Miller 2014)

7: Stockholm (Nystrom 2016) 8: HIP (Shapiro 1988)

Mammography +/- CBE compared to Usual Care											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mammography +/- CBE	Usual Care	Relative (95% CI)	Absolute (95% CI)	
Main Analysis: Breast-Cancer Mortality (All Ages)											
8 # R: Unclear # A: Unclear Range of follow-up (yrs): 17.7 to 29.0	randomised trials [a]	very serious [b]	not serious [c]	not serious [d]	not serious [e]	none [f]	Unavailable*	0.2% ‡	RR 0.85 (0.78 to 0.93)	**31 fewer per 100,000** (from 14 fewer to 45 fewer) ‡	⊕⊕◯◯ LOW
								0.4% ‡		**58 fewer per 100,000** (from 27 fewer to 85 fewer) ‡	
								0.6% ‡		**91 fewer per 100,000** (from 42 fewer to 133 fewer) ‡	

CI: Confidence interval; **RR:** Risk ratio; # R: Number randomised; # A: Number analyzed
*Complete data was not available. Numerators and/or denominators were either unclear or not reported for all included studies.
‡The baseline risk (in the control group) may not have been representative of all included studies because the numerators and/or denominators were either unclear or not reported. Values provided reflect those studies which provided complete data.

Explanations

a. Two studies considered quasi-randomised (Stockholm & Gothenburg)
b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas.
c. Heterogeneity may be low (I^2=10%); (p-value=0.35)
d. Studies seemed relevant to the PICO being addressed. Breast density was not addressed. For some studies, the control group received screening after the screening period. Studies reported in Nystrom 2016 included one round of screening in the control group as part of the short-case accrual calculation. Therefore, the study estimates may be underestimated (a larger benefit from the intervention may be possible).
e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs excludes the null, and does not cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
f. Cannot assess publication bias (insufficient number of trials)

Part A- GRADE Summary of Findings Table – Breast Cancer Mortality (Short-Case Accrual) (All Ages)

Mammography +/- CBE compared to Usual Care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with Usual Care (Assumed Risk)	**Risk with Mammography +/- CBE (Corresponding Risk)**				
Main Analysis: Breast-Cancer Mortality (All Ages) # Randomised : Unclear # Analyzed: Unclear Range of follow-up (yrs): 17.7 to 29.0	**Low**		RR 0.85 (0.78 to 0.93)	Unavailable (8 RCTs) [a]	⨁⨁◯◯ LOW [b,c,d,e,f]	Short-Case Accrual: Only deaths from breast cancer cases diagnosed during the screening period are included in the analysis. NNS (Low): 3,226 (2,222 to 6,667) NNS (Moderate): 1,724 (1,176 to 3,704) NNS (High): 1,099 (746 to 2,326)
	205 per 100,000 ‡	**174 per 100,000** (160 to 190)				
	Moderate					
	386 per 100,000 ‡	**328 per 100,000** (301 to 359)				
	High					
	606 per 100,000 ‡	**515 per 100,000** (472 to 563)				

***The corresponding risk** (and its 95% confidence interval) is based on **the assumed risk** in the comparison group and the **relative effect** of the intervention (and its 95% CI).
‡The baseline risk (in the control group) may not have been representative of all included studies because the numerators and/or denominators were either unclear or not reported. Values provided reflect those studies which provided complete data.
CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

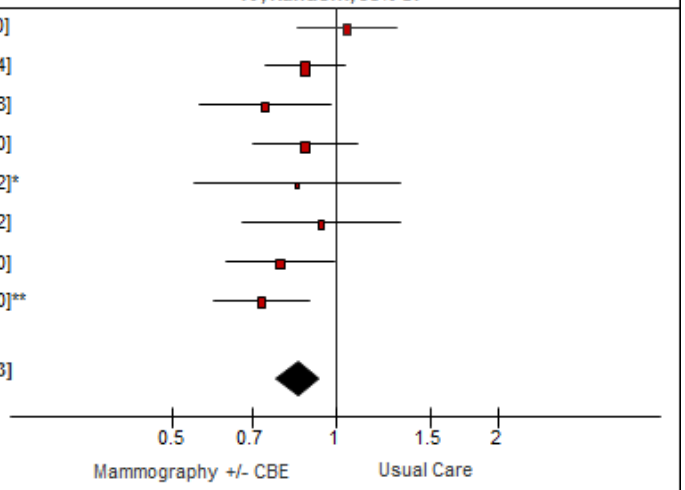
a. Two studies considered quasi-randomised (Stockholm & Gothenburg)
b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas.
c. Heterogeneity may be low (I^2=10%); (p-value=0.35)
d. Studies seemed relevant to the PICO being addressed. Breast density was not addressed. For some studies, the control group received screening after the screening period. Studies reported in Nystrom 2016 included one round of screening in the control group as part of the short-case accrual calculation. Therefore, the study estimates may be underestimated (a larger benefit from the intervention may be possible).
e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs excludes the null, and does not cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
f. Cannot assess publication bias (insufficient number of trials)

Part A- Forest Plot – Breast Cancer Mortality (Short-Case Accrual) (All Ages)

Reference	Study	Age (at entry)	Mean Follow-up (yrs)	log [RR]	SE	Weight	Risk Ratio [95%CI]
Miller 2014	CNBSS 1 & 2	40-69	21.9	0.05	0.11	15.3%	1.05 [0.85, 1.30]
Moss 2015	AGE	39-41	17.7[a]	-0.13	0.09	22.2%	0.88 [0.74, 1.04]
Nystrom 2016	Gothenburg	40-59	24.0	-0.30	0.14	9.3%	0.74 [0.56, 0.98]
Nystrom 2016	Malmo I	45-70	30.0	-0.13	0.12	13.7%	0.88 [0.70, 1.10]
Nystrom 2016	Malmo II	43-49	22.0	-0.16	0.22	4.0%	0.85 [0.55, 1.32]*
Nystrom 2016	Stockholm	40-65	25.0	-0.06	0.17	6.5%	0.94 [0.67, 1.32]
Shapiro 1988	HIP	40-64	18.0	-0.24	0.12	12.9%	0.79 [0.63, 1.00]
Tabar 2011	Swedish Two County	40-74	29.0[b]	-0.31	0.10	16.2%	0.73 [0.59, 0.90]**
Total (95% CI)						**100.0%**	**0.85 [0.78, 0.93]**

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 7.80$, df = 7 (P = 0.35); $I^2 = 10\%$

Test for overall effect: Z = 3.49 (P = 0.0005)



[a]Median; [b]Time since randomization; *Adjusted for Age; ** Adjusted for clustering

Evidence Set 1a- Mammography +/- Clinical Breast Exam- Breast-Cancer Mortality (Short-Case Accrual) (Stratified by Age)

Part A- GRADE Evidence Profile Table – Breast Cancer Mortality (Short-Case Accrual) (Stratified by Age)

Included Studies:

1: Gothenburg (Nystrom 2016) 2: Age (Moss 2015) 3. Swedish Two County (Kopparberg & Ostergotland) (Tabar 2011)

4: Malmo I (Nystrom 2016) 5: Malmo I (Nystrom 2002) 6: Malmo II (Nystrom 2016)

7: CNBSS 1 (Miller 2014) 8: CNBSS 2 (Miller 2014) 9: HIP (Shapiro 1988)

10: Stockholm (Nystrom 2016) 11: Stockholm (Nystrom 2002)

Mammography +/- CBE compared to Usual Care											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mammography +/- CBE	Usual Care	Relative (95% CI)	Absolute (95% CI)	
Sub-Group: Breast-Cancer Mortality (40-49 years)											
8 [1-3, 5-7, 9-10] # R: Unclear # A: Unclear Range of follow-up (yrs): 17.7 to 25.7	randomised trials [a]	very serious [b]	not serious [c]	not serious [d]	not serious [e]	none [f]	Unavailable*	412/106,956 (0.4%) ‡	RR 0.87 (0.75 to 1.01)	**50 fewer per 100,000** (from 4 more to 96 fewer) ‡	⊕⊕◯◯ LOW
Sub-Group: Breast-Cancer Mortality (50-59 years)											
6 [1,3,4,8-10] # R: Unclear	randomised trials [a]	very serious [b]	serious [g]	not serious [d]	serious [h]	none [f]	Unavailable*	0.3% §	RR 0.84 (0.71 to 1.00)	**49 fewer per 100,000** (from 0 fewer to 90 fewer) §	⊕◯◯◯ VERY LOW
								0.5% §		**80 fewer per 100,000** (from 0 fewer to 145 fewer) §	

Mammography +/- CBE compared to Usual Care											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mammography +/- CBE	Usual Care	Relative (95% CI)	Absolute (95% CI)	
# A: Unclear Range of follow-up (yrs): 18.0 to 30.0								0.9% §		**137 fewer per 100,000** (from 0 fewer to 249 fewer) §	
Sub-Group: Breast-Cancer Mortality (60-69 years)											
4 [3,4,9,11] # R: Unclear # A: Unclear Range of follow-up (yrs): 13.1 to 30.0	randomised trials [i]	very serious [b]	not serious [j]	not serious [d]	not serious [k]	none [f]	Unavailable*	48/7,806 (0.6%)§	RR 0.70 (0.56 to 0.88)	**184 fewer per 100,000** (from 74 fewer to 271 fewer) §	⊕⊕◯◯ LOW
Sub-Group: Breast-Cancer Mortality (70-74 years)											
2 [3,5] # R: 18,233 # A: Unclear Range of follow-up (yrs): 13.2 to 13.6	randomised trials	very serious [b]	not serious [l]	not serious [d]	serious [m]	none [f]	Unavailable*	3/291 (1.0%) §	RR 0.89 (0.58 to 1.38)	**113 fewer per 100,000** (from 392 fewer to 433 more) §	⊕◯◯◯ VERY LOW

CI: Confidence interval; **RR:** Risk ratio; # R: Number randomised; # A: Number analyzed
*Complete data was not available. Numerators and/or denominators were either unclear or not reported for all included studies.

‡The baseline risk (in the control group) may not have been representative of all included studies because the numerators and/or denominators were either unclear or not reported. Values provided reflect those studies which provided complete data.
§ The baseline risk (in the control group) may not representative of all included studies because the numerators and/or denominators were either unclear or not reported for all studies. The number randomised for those publications where event rates were given were used to estimate baseline risk. The numbers randomised were initially not used to estimate baseline risk, since an independent investigation found that we could not be confident that the numbers randomised were used to calculate RR.

Explanations

a. Two studies considered quasi-randomised (Stockholm & Gothenburg)
b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas
c. Heterogeneity may be low (I^2=22%); (p-value =0.06)
d. Studies seemed relevant to the PICO being addressed. Breast density was not addressed. For some studies, the control group received screening after the screening period. Studies reported in Nystrom 2002 and Nystrom 2016 included one round of screening in the control group as part of the short-case accrual calculation. Therefore, the study estimates may be underestimated (a larger benefit from the intervention may be possible).
e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs include the null, but do not cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
f. Cannot assess publication bias (insufficient number of trials)
g. Heterogeneity may be moderate (I^2=26%); (p-value=0.24)
h. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs include the null, and do cross appreciable benefit (RR 0.75) .Given the large sample sizes, an optimal sample size calculation was not warranted.
f. Cannot assess publication bias (insufficient number of trials)
i. One study considered quasi- randomised (Stockholm)
j. Heterogeneity may be low (I^2=0%); (p-value=0.48)
k. (i) The total population is large (>2000); and (ii) the 95%CIs do not include the null but do cross appreciable benefit (RR 0.75) , Given the large sample sizes, an optimal sample size calculation was not warranted.
l. Heterogeneity may be low (I^2=0%); (p-value=0.92)
m. (i) The total population is large (>2000), and (ii) the 95%CIs include the null and do cross appreciable benefit (RR 0.75) .Given the large sample sizes, an optimal sample size calculation was not warranted.

Part A- GRADE Summary of Findings Table – Breast Cancer Mortality (Short-Case Accrual) (Stratified by Age)

Mammography +/- CBE compared to Usual Care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with Usual Care (Assumed Risk)	**Risk with Mammography +/- CBE (Corresponding Risk)**				
Sub-Group: Breast-Cancer Mortality (40-49 years) # Randomised: Unclear # Analyzed: Unclear Range of follow-up (yrs): 17.7 to 25.7	385 per 100,000 ‡	**335 per 100,000** (289 to 389)	RR 0.87 (0.75 to 1.01)	Unavailable (8 RCTs) [a]	⊕⊕◯◯ LOW [b,c,d,e,f]	Short-Case Accrual: Only deaths from breast cancer cases diagnosed during the screening period are included in the analysis. NNS: 2,000 (1,042 to -25,000) *CIs include ∞
Sub-Group: Breast-Cancer Mortality (50-59 years) # Randomised: Unclear # Analyzed: Unclear Range of follow-up (yrs): 18.0 to 30.0	**Low**		RR 0.84 (0.71 to 1.00)	Unavailable (6 RCTs) [a]	⊕◯◯◯ VERY LOW [b,d,f,g,h]	Short-Case Accrual: Only deaths from breast cancer cases diagnosed during the screening period are included in the analysis. NNS (Low): 2,041 (1,111 to ∞) NNS (Moderate): 1,250 (690 to ∞) NNS (High): 730 (402 to ∞)
	309 per 100,000 §	**260 per 100,000** (219 to 309)				
	Moderate					
	500 per 100,000 §	**420 per 100,000** (355 to 500)				
	High					
	858 per 100,000 §	**721 per 100,000** (609 to 858)				
Sub-Group: Breast-Cancer Mortality (60-69 years) # Randomised: Unclear # Analyzed: Unclear Range of follow-up (yrs): 13.1 to 30.0	615 per 100,000 §	**430 per 100,000** (344 to 541)	RR 0.70 (0.56 to 0.88)	Unavailable (4 RCTs) [i]	⊕⊕◯◯ LOW [b,d,f,j,k]	Short-Case Accrual: Only deaths from breast cancer cases diagnosed during the screening period are included in the analysis. NNS: 541 (369 to 1,351)
Sub-Group: Breast-Cancer Mortality (70-74 years) # Randomised: 18,233 # Analyzed: Unclear Range of follow-up (yrs): 13.2 to 13.6	1,031 per 100,000 §	**918 per 100,000** (598 to 1,423)	RR 0.89 (0.58 to 1.38)	Unavailable (2 RCTs)	⊕◯◯◯ VERY LOW [b,d,f,l,m]	Short-Case Accrual: Only deaths from breast cancer cases diagnosed during the screening period are included in the analysis. NNS: 885 (231 to -255) *CIs include ∞

Mammography +/- CBE compared to Usual Care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with Usual Care (Assumed Risk)	**Risk with Mammography +/- CBE (Corresponding Risk)**				

***The corresponding risk** (and its 95% confidence interval) is based on **the assumed risk** in the comparison group and the **relative effect** of the intervention (and its 95% CI).

‡The baseline risk (in the control group) was not representative of all included studies. Numerators and/or denominators were either unclear or not reported for some studies. Values provided reflect those studies which provided complete data.

§ The baseline risk (in the control group) was not representative of all included studies. Numerators and/or denominators were either unclear or not reported for all studies. The number randomised for those publications where event rates were given were used to estimate baseline risk. The numbers randomised were initially not used to estimate baseline risk, since an independent investigation found that we could not be confident that the numbers randomised were used to calculate RR.

CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Two studies considered quasi-randomised (Stockholm & Gothenburg)

b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas

c. Heterogeneity may be low (I^2=22%); (p-value =0.06)

d. Studies seemed relevant to the PICO being addressed. Breast density was not addressed. For some studies, the control group received screening after the screening period. Studies reported in Nystrom 2002 and Nystrom 2016 included one round of screening in the control group as part of the short-case accrual calculation. Therefore, the study estimates may be underestimated (a larger benefit from the intervention may be possible).

e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs include the null, but do not cross appreciable benefit (RR 0.75) . Given the large sample sizes, an optimal sample size calculation was not warranted.

f. Cannot assess publication bias (insufficient number of trials)

g. Heterogeneity may be moderate (I^2=26%); (p-value=0.24)

h. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs include the null, and do cross appreciable benefit (RR 0.75) .Given the large sample sizes, an optimal sample size calculation was not warranted.

f. Cannot assess publication bias (insufficient number of trials)

i. One study considered quasi- randomised (Stockholm)

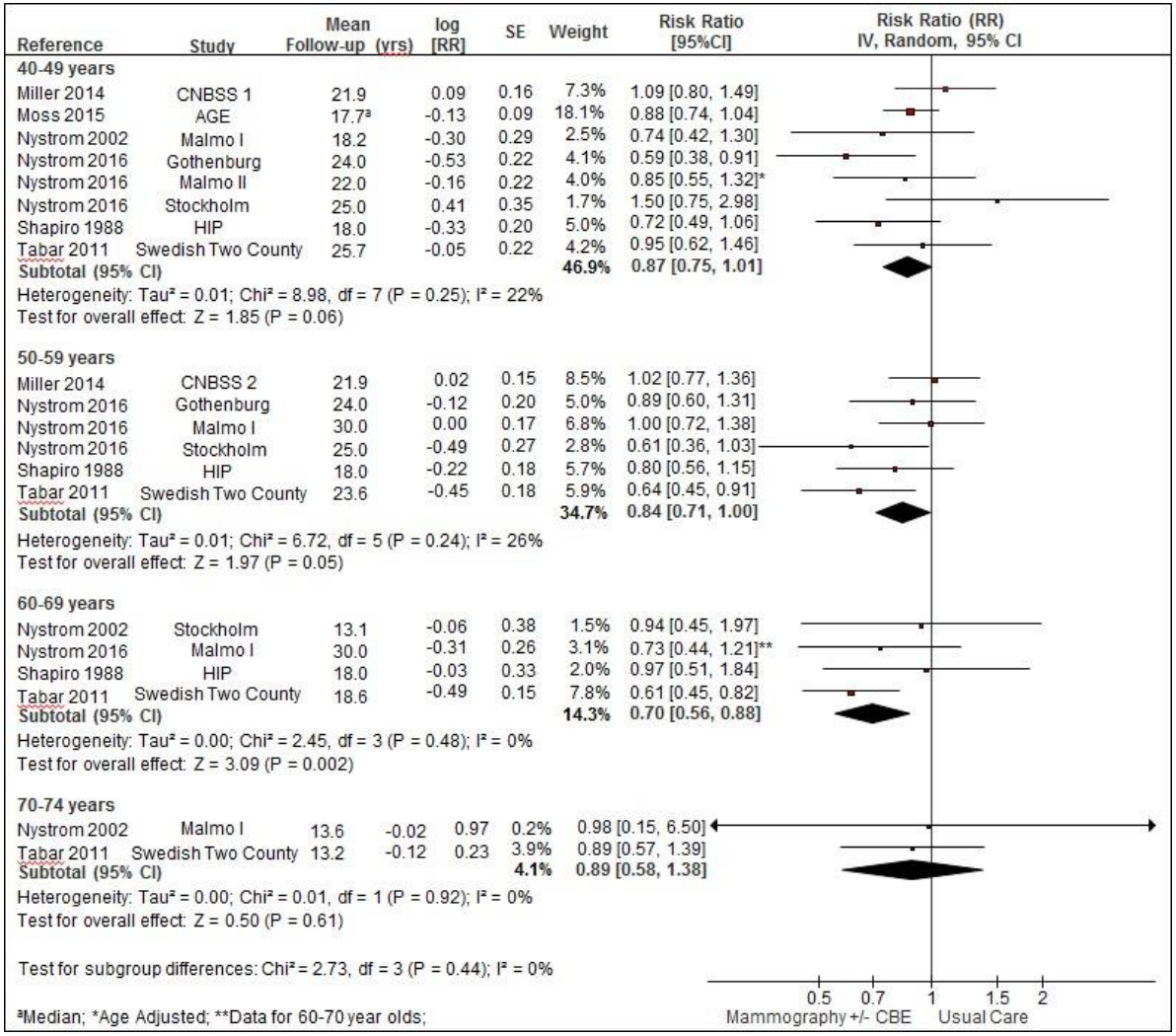
j. Heterogeneity may be low (I^2=0%); (p-value=0.48)

k. (i) The total population is large (>2000); and (ii) the 95%CIs do not include the null but do cross appreciable benefit (RR 0.75) , Given the large sample sizes, an optimal sample size calculation was not warranted.

l. Heterogeneity may be low (I^2=0%); (p-value=0.92)

m. (i) The total population is large (>2000), and (ii) the 95%CIs include the null and do cross appreciable benefit (RR 0.75) .Given the large sample sizes, an optimal sample size calculation was not warranted.

Part A- Forest Plot – Breast Cancer Mortality (Short-Case Accrual) (Stratified by Age)



Evidence Set 2- Mammography +/- Clinical Breast Exam – Breast-Cancer Mortality (Long-Case Accrual) (All Ages)

Part A- GRADE Evidence Profile Table – Breast Cancer Mortality (Long-Case Accrual) (All Ages)

Included Studies:

1: Gothenburg (Bjurstam 2003) 2: CNBSS 1 & 2 (Miller 2014) 3: AGE (Moss 2015)

4: Swedish Two County (Kopparberg) (Tabar 1995) 5: Swedish Two County (Ostergotland) (Tabar 1995) 6: HIP (Habbema 1986)

Note: Long-case accrual unavailable for the following studies: Malmo I, Malmo II, and Stockholm

Mammography +/- CBE compared to Usual Care											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mammography +/- CBE	Usual Care	Relative (95% CI)	Absolute (95% CI)	
Main Analysis: Breast-Cancer Mortality (All Ages)											
6 # R: Unclear # A: 460,389 Range of follow-up (yrs): 12.5 to 21.9	randomised trials [a]	very serious [b]	very serious [c]	not serious [d]	not serious [e]	none [f]	1,256/207,939 (0.6%)	0.5%	RR 0.82 (0.71 to 0.94)	**87 fewer per 100,000** (from 29 fewer to 140 fewer)	⊕◯◯◯ VERY LOW
								0.8%		**138 fewer per 100,000** (from 46 fewer to 222 fewer)	
								1.1%		**202 fewer per 100,000** (from 67 fewer to 326 fewer)	

CI: Confidence interval; **RR:** Risk ratio; # R: Number randomised; # A: Number analyzed

Explanations

a. One study considered quasi-randomised (Gothenburg)

b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas.

c. Heterogeneity may be substantial (I^2=68%); (p-value=0.009)

d. Studies are relevant to the PICO being addressed. Breast density was not addressed. For some studies, the control group received screening after the screening period. Gothenburg, AGE, and Swedish Two County -usual care arm received screening at end of study period.

e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs exclude the null, but do cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
f. Cannot assess publication bias (insufficient number of trials)

Part A- GRADE Summary of Findings Table – Breast Cancer Mortality (Long-Case Accrual) (All Ages)

Mammography +/- CBE compared to Usual Care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with Usual Care (Assumed Risk)	**Risk with Mammography +/- CBE (Corresponding Risk)**				
Main Analysis: Breast-Cancer Mortality (All Ages) # Randomised: Unclear # Analyzed: 460,389 Range of follow-up (yrs): 12.5 to 21.9	**Low**		RR 0.82 (0.71 to 0.94)	460,389 (6 RCTs) [a]	⨁◯◯◯ VERY LOW [b,c,d,e,f]	Long-Case Accrual: All deaths from breast cancer cases diagnosed during the screening period and follow-up period are included in the analysis. NNS (Low): 1,149 (714 to 3,448) NNS (Moderate): 725 (450 to 2,174) NNS (High): 493 (306 to 1,471)
	482 per 100,000	**395 per 100,000** (342 to 453)				
	Moderate					
	767 per 100,000	**629 per 100,000** (545 to 721)				
	High					
	1,125 per 100,000	**922 per 100,000** (798 to 1,057)				

***The corresponding risk** (and its 95% confidence interval) is based on **the assumed risk** in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio

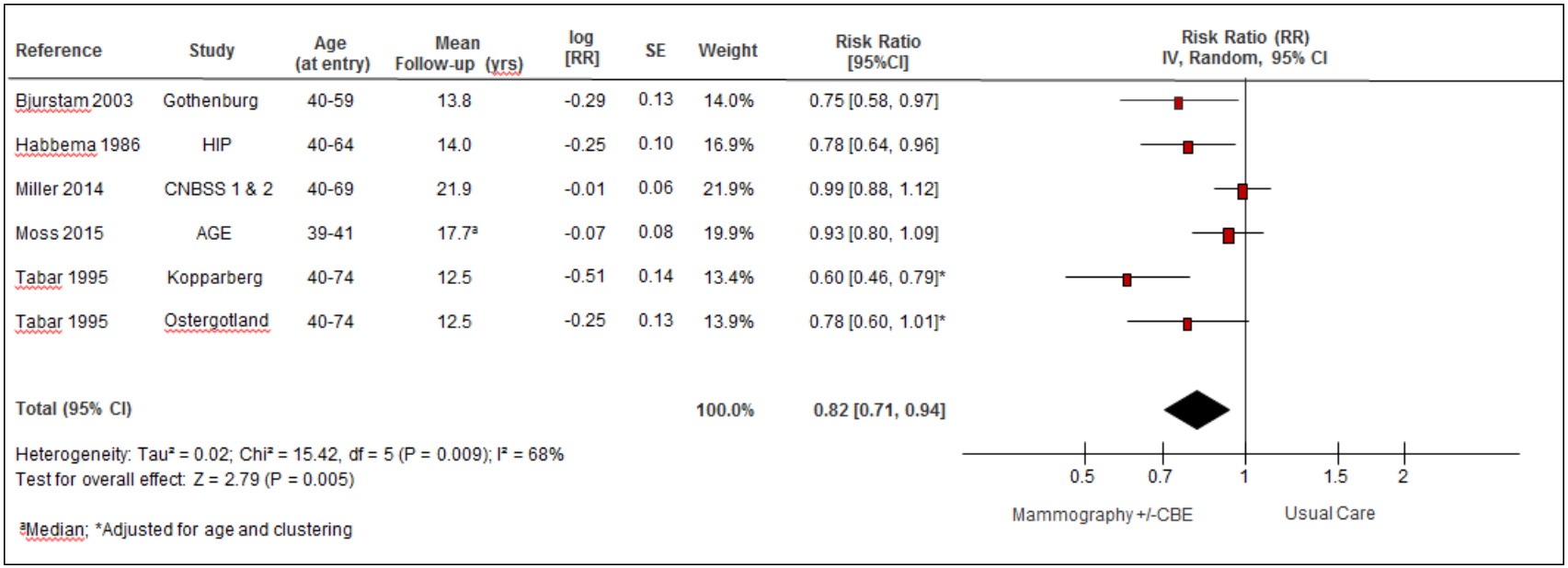
GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. One study considered quasi-randomised (Gothenburg)
b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas.
c. Heterogeneity may be substantial (I^2=68%); (p-value=0.009)

d. Studies are relevant to the PICO being addressed. Breast density was not addressed. For some studies, the control group received screening after the screening period. Gothenburg, AGE, and Swedish Two County -usual care arm received screening at end of study period.
e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs exclude the null, but do cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
f. Cannot assess publication bias (insufficient number of trials)

Part A- Forest Plot – Breast Cancer Mortality (Long-Case Accrual) (All Ages)



Evidence Set 2a- Mammography +/- Clinical Breast Exam- Breast-Cancer Mortality (Long-Case Accrual) (Stratified by Age)

Part A- GRADE Evidence Profile Table – Breast Cancer Mortality (Long-Case Accrual) (Stratified by Age)

Included Studies:

1: Gothenburg (Bjurstam 2003) 2: Age (Moss 2015) 3. Swedish Two County (Kopparberg) (Tabar 1995)

4: Swedish Two County (Ostergotland) (Tabar 1995) 5: CNBSS 1 (Miller 2014) 6: CNBSS 2 (Miller 2014) 7: HIP (Habbema 1986)

Note: Long-case accrual unavailable for the following studies: Malmo I, Malmo II, and Stockholm

Mammography +/- CBE compared to Usual Care											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mammography +/- CBE	Usual Care	Relative (95% CI)	Absolute (95% CI)	
Sub-Group: Breast-Cancer Mortality (40-49 years)											
6 [1-5, 7] # R: Unclear # A: Unclear Range of follow-up (yrs): 12.5 to 21.9	randomised trials [a]	very serious [b]	not serious [c]	not serious [d]	not serious [e]	none [f]	Unavailable*	0.2% ‡	RR 0.92 (0.82 to 1.03)	**17 fewer per 100,000** (from 7 more to 39 fewer) ‡	⊕⊕◯◯ LOW
								0.3% ‡		**25 fewer per 100,000** (from 10 more to 57 fewer) ‡	
								0.4% ‡		**33 fewer per 100,000** (from 12 more to 75 fewer) ‡	
Sub-Group: Breast-Cancer Mortality (50-59 years)											
5 [1, 3, 4, 6, 7]	randomised trials [a]	very serious [b]	serious [g]	not serious [d]	not serious [h]	none [f]	Unavailable*	0.5% ‡	RR 0.82 (0.68 to 0.98)	**82 fewer per 100,000** (from 9 fewer to 145 fewer) ‡	⊕◯◯◯

Mammography +/- CBE compared to Usual Care											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mammography +/- CBE	Usual Care	Relative (95% CI)	Absolute (95% CI)	
# R: Unclear # A: Unclear Range of follow-up (yrs): 12.5 to 21.9								0.7% ‡		**118 fewer per 100,000** (from 13 fewer to 209 fewer) ‡	VERY LOW
Sub-Group: Breast-Cancer Mortality (60-69 years)											
3 [3,4,7] # R: Unclear # A: Unclear Range of follow-up (yrs): 12.5 to 14.0	randomised trials	very serious [b]	not serious [i]	not serious [d]	not serious [j]	none [f]	Unavailable*	103/16,269 (0.6%)	RR 0.65 (0.50 to 0.85)	**222 fewer per 100,000** (from 95 fewer to 317 fewer) ‡	⊕⊕◯◯ LOW
Sub-Group: Breast-Cancer Mortality (70-74 years)											
2 [3,4]	randomised trials	very serious [b]	not serious [k]	not serious [d]	serious [l]	none [f]	49/10,339 (0.5%)	0.6%	RR 0.79 (0.51 to 1.22)	**134 fewer per 100,000** (from 140 more to 312 fewer)	⊕◯◯◯

Mammography +/- CBE compared to Usual Care											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mammography +/- CBE	Usual Care	Relative (95% CI)	Absolute (95% CI)	
# R: 17,646 # A: 17,646 Range of follow-up (yrs): 12.5								0.8%		**164 fewer per 100,000** (from 171 more to 382 fewer)	VERY LOW

CI: Confidence interval; **RR:** Risk ratio; # R: Number randomised; # A: Number analyzed

*Complete data was not available. Numerators and/or denominators were either unclear or not reported for all included studies.
‡The baseline risk (in the control group) may not be representative of all included studies because the numerators and/or denominators were either unclear or not reported. Values provided reflect those studies which provided complete data.

Explanations

a. One study considered quasi-randomised (Gothenburg)
b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas
c. Heterogeneity may be low (I^2=10%); (p-value=0.35)
d. Studies are relevant to the PICO being addressed. Breast density was not addressed. For some studies, the control group received screening after the screening period. Gothenburg, AGE, and Swedish Two County -usual care arm received screening at end of study period.
e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs include the null, but do not cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
f. Cannot assess publication bias (insufficient number of trials)
g. Heterogeneity may be moderate (I^2=38%); (p-value=0.17)
h. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs do not cross the null, but do cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
i. Heterogeneity may be low (I^2=0%); (p-value=0.57)
j. (i) The total population is large (>2000); and (ii) the 95%CIs do not cross the null, but do cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
k. Heterogeneity may be low (I^2=0%); (p-value=0.87)
l. (i) The total population is large (>2000); and (ii) the 95%CI crosses the null, and also crosses appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.

Part A- GRADE Summary of Findings Table – Breast Cancer Mortality (Long-Case Accrual) (Stratified by Age)

Mammography +/- CBE compared to Usual Care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with Usual Care (Assumed Risk)	**Risk with Mammography +/- CBE (Corresponding Risk)**				
Sub-Group: Breast-Cancer Mortality (40-49 years) # Randomised: Unclear # Analyzed: Unclear Range of follow-up (yrs): 12.5 to 21.9	**Low**		RR 0.92 (0.82 to 1.03)	Unavailable (6 RCTs) [a]	⊕⊕◯◯ LOW [b,c,d,e,f]	Long-Case Accrual: All deaths from breast cancer cases diagnosed during the screening period and follow-up period are included in the analysis. NNS (Low): 5,882 (2,564 to -14,286) *CIs include ∞ NNS (Moderate): 4,000 (1,754 to -10,000) *CIs include ∞ NNS (High): 3,030 (1,333 to -8,333) *CIs include ∞
	217 per 100,000 ‡	**200 per 100,000** (178 to 224)				
	Moderate					
	318 per 100,000 ‡	**293 per 100,000** (261 to 328)				
	High					
	415 per 100,000 ‡	**382 per 100,000** (340 to 427)				
Sub-Group: Breast-Cancer Mortality (50-59 years) # Randomised: Unclear # Analyzed: Unclear Range of follow-up (yrs): 12.5 to 21.9	**Low**		RR 0.82 (0.68 to 0.98)	Unavailable (5 RCTs) [a]	⊕◯◯◯ VERY LOW [b,d,f,g,h]	Long-Case Accrual: All deaths from breast cancer cases diagnosed during the screening period and follow-up period are included in the analysis. NNS (Low):1,220 (690 to 11,111) NNS (High): 847 (478 to 7,692)
	453 per 100,000 ‡	**371 per 100,000** (308 to 444)				
	High					
	654 per 100,000 ‡	**536 per 100,000** (445 to 641)				
Sub-Group: Breast-Cancer Mortality (60-69 years) # Randomised: Unclear # Analyzed: Unclear Range of follow-up (yrs): 12.5 to 14.0	633 per 100,000 ‡	**412 per 100,000** (317 to 538)	RR 0.65 (0.50 to 0.85)	Unavailable (3 RCTs)	⊕⊕◯◯ LOW [b,d,f,i,j]	Long-Case Accrual: All deaths from breast cancer cases diagnosed during the screening period and follow-up period are included in the analysis. NNS: 452 (316 to 1,053)
Sub-Group: Breast-Cancer	**Low**		RR 0.79	17,646	⊕◯◯◯	Long-Case Accrual: All deaths from breast cancer cases

Mammography +/- CBE compared to Usual Care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with Usual Care (Assumed Risk)	**Risk with Mammography +/- CBE (Corresponding Risk)**				
Mortality (70-74 years) # Randomised: 17,646 # Analyzed: 17,646 Range of follow-up (yrs): 12.5	637 per 100,000	**503 per 100,000** (325 to 777)	(0.51 to 1.22)	(2 RCTs)	VERY LOW [b,d,f,k,l]	diagnosed during the screening period and follow-up period are included in the analysis. NNS (Low): 746 (321 to -714) *CIs include ∞
	High					
	779 per 100,000	**615 per 100,000** (397 to 950)				NNS (High): 610 (262 to -585) *CIs include ∞

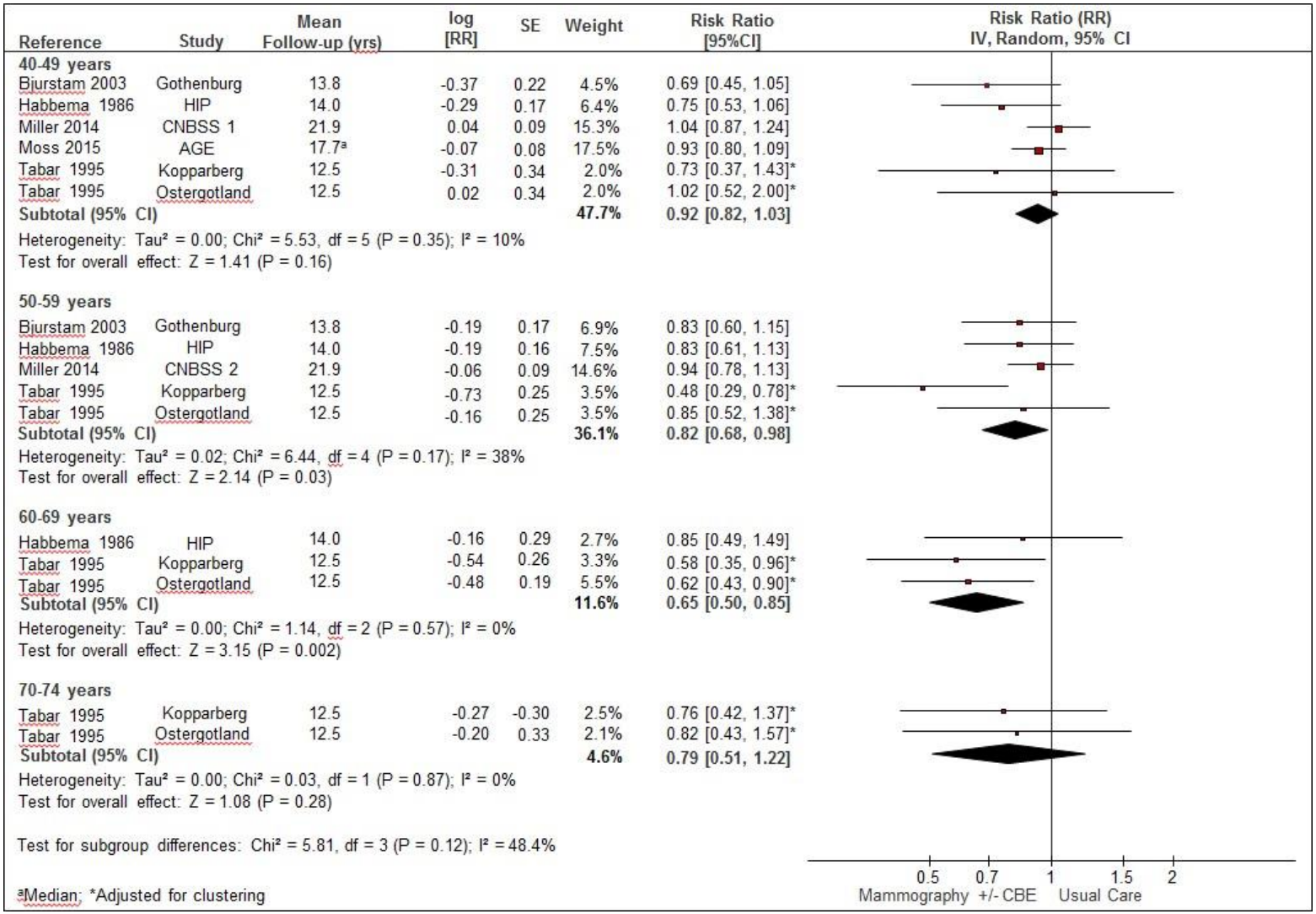
***The assumed risk** (and its 95% confidence interval) is based on **the assumed risk** in the comparison group and the **relative effect** of the intervention (and its 95% CI).
‡The baseline risk (in the control group) may not be representative of all included studies because the numerators and/or denominators were either unclear or not reported. Values provided reflect those studies which provided complete data.
CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. One study considered quasi-randomised (Gothenburg)
b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas
c. Heterogeneity may be low (I^2=10%); (p-value=0.35)
d. Studies are relevant to the PICO being addressed. Breast density was not addressed. For some studies, the control group received screening after the screening period. Gothenburg, AGE, and Swedish Two County -usual care arm received screening at end of study period.
e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs include the null, but do not cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
f. Cannot assess publication bias (insufficient number of trials)
g. Heterogeneity may be moderate (I^2=38%); (p-value=0.17)
h. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs do not cross the null, but do cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
i. Heterogeneity may be low (I^2=0%); (p-value=0.57)
j. (i) The total population is large (>2000); and (ii) the 95%CIs do not cross the null, but do cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
k. Heterogeneity may be low (I^2=0%); (p-value=0.87)
l. (i) The total population is large (>2000); and (ii) the 95%CI crosses the null, and also crosses appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.

Part A- Forest Plot – Breast Cancer Mortality (Long-Case Accrual) (Stratified by Age)



Evidence Set 3- Mammography +/- Clinical Breast Exam- All-Cause Mortality (All Ages)

Part A- GRADE Evidence Profile Table – All-Cause Mortality (All Ages)

Included Studies:

1: Gothenburg (Nystrom 2002) 2: CNBSS 1 & 2 (Miller 2014) 3. Swedish Two County- Ostergotland (Nystrom 2002)

4: Stockholm (Nystrom 2002) 5: HIP (Aron & Prorak 1986) 6: Malmo I (Nystrom 2002)

7: Malmo II (Nystrom 2002) 8: AGE (Moss 2015)

Mammography +/- CBE compared to Usual Care											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mammography +/- CBE	Usual Care	Relative (95% CI)	Absolute (95% CI)	
Main Analysis: All-Cause Mortality (All Ages)											
8 # R: Unclear # A: Unclear Range of follow-up (yrs): 9.1 to 25.0	randomised trials [a]	very serious [b]	not serious [c]	not serious [d]	not serious [e]	none [f]	Unavailable*	4.0% ‡	RR 0.99 (0.98 to 1.00)	**40 fewer per 100,000** (from 0 fewer to 81 fewer) ‡	⊕⊕◯◯ LOW
								6.9% ‡		**69 fewer per 100,000** (from 0 fewer to 138 fewer) ‡	
								10.4% ‡		**104 fewer per 100,000** (from 0 fewer to 209 fewer) ‡	

CI: Confidence interval; **RR:** Risk ratio; # R: Number randomised; # A: Number analyzed

*Complete data was not available. Numerators and/or denominators were either unclear or not reported for all included studies.

‡The baseline risk (in the control group) may not have been representative of all included studies because the numerators and/or denominators were either unclear or not reported. Values provided reflect those studies which provided complete data.

Explanations

a. Two studies considered quasi-randomised (Stockholm & Gothenburg)

b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas.

c. Heterogeneity may be low (I^2=0%); (p-value=0.59)

d. Studies are relevant to the PICO being addressed. Breast density was not addressed. For some studies, the control group received screening after the screening period. Studies reported in Nystrom 2002 included one round of screening in the control group

e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs include the null, but do not cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.

f. Cannot assess publication bias (insufficient number of trials)

Part A- GRADE Summary of Findings Table – All-Cause Mortality (All Ages)

Mammography +/- CBE compared to Usual Care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with Usual Care (Assumed Risk)	**Risk with Mammography +/- CBE (Corresponding Risk)**				
Main Analysis: All-Cause Mortality (All Ages) # Randomised: Unclear # Analyzed: Unclear Range of follow-up (yrs): 9.1 to 25.0	**Low**		RR 0.99 (0.98 to 1.00)	Unavailable (8 RCTs) [a]	⨁⨁◯◯ LOW [b,c,d,e,f]	NNS (Low): 2,500 (1,235 to ∞) NNS (Moderate): 1,449 (725 to ∞) NNS (High): 952 (478 to ∞)
	4,039 per 100,000 ‡	**3,999 per 100,000** (3,958 to 4,039)				
	Moderate					
	6,878 per 100,000 ‡	**6,809 per 100,000** (6,740 to 6,878)				
	High					
	10,439 per 100,000 ‡	**10,334 per 100,000** (10,230 to 10,439)				

***The corresponding risk** (and its 95% confidence interval) is based on **the assumed risk** in the comparison group and the **relative effect** of the intervention (and its 95% CI).
‡The baseline risk (in the control group) may not have been representative of all included studies because the numerators and/or denominators were either unclear or not reported. Values provided reflect those studies which provided complete data.
CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

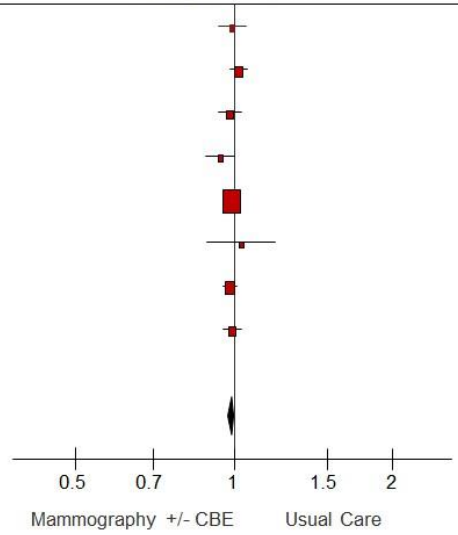
a. Two studies considered quasi-randomised (Stockholm & Gothenburg)
b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas.
c. Heterogeneity may be low (I^2=0%); (p-value=0.59)
d. Studies are relevant to the PICO being addressed. Breast density was not addressed. For some studies, the control group received screening after the screening period. Studies reported in Nystrom 2002 included one round of screening in the control group.
e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95% CIs include the null, but do not cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
f. Cannot assess publication bias (insufficient number of trials)

Part A- Forest Plot – All-Cause Mortality (All Ages)

Reference	Study	Age (at entry)	Mean Follow-up (yrs)	log [RR]	SE	Weight	Risk Ratio [95%CI]
Aron and Prorok 1986	HIP	40-59	10.0	-0.01	0.03	4.7%	0.99 [0.93, 1.05]
Miller 2014	CNBSS 1& 2	40-59	25.0	0.02	0.02	11.4%	1.02 [0.98, 1.06]
Moss 2015	AGE	39-41	17.7[a]	-0.02	0.03	6.7%	0.98 [0.93, 1.03]
Nystrom 2002	Gothenburg	40-59	13.2	-0.06	0.03	4.3%	0.94 [0.88, 1.00]
Nystrom 2002	Malmo I	45-70	19.2	-0.01	0.01	42.8%	0.99 [0.97, 1.01]
Nystrom 2002	Malmo II	43-49	9.1	0.03	0.08	0.8%	1.03 [0.89, 1.20]
Nystrom 2002	Ostergotland	40-74	17.2	-0.02	0.02	18.6%	0.98 [0.95, 1.01]
Nystrom 2002	Stockholm	40-64	14.7	-0.01	0.02	10.7%	0.99 [0.95, 1.03]
Total (95% CI)						**100.0%**	**0.99 [0.98, 1.00]**

Heterogeneity: Tau² = 0.00; Chi² = 5.57, df = 7 (P = 0.59); I² = 0%
Test for overall effect: Z = 1.65 (P = 0.10)

[a]Median



Evidence Set 3a- Mammography +/- Clinical Breast Exam- All-Cause Mortality (Stratified by Age)

Part A- GRADE Evidence Profile Table – All-Cause Mortality (Stratified by Age)

Included Studies:

1: Age (Moss 2015)
2. Malmo II (Nystrom 2002)
3: Swedish Two County (Kopparberg) (Tabar 1989)
4: Swedish Two County (Ostergotland) (Tabar 1989)
5: Stockholm (Frisell 1997)
6: CNBSS 1 (Miller 2002)
7: CNBSS 2 (Miller 2000)
8: Gothenburg (Bjurstam 1997)

Mammography +/- CBE compared to Usual Care											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mammography +/- CBE	Usual Care	Relative (95% CI)	Absolute (95% CI)	
Sub-Group: All-Cause Mortality (40-49 years)											
7 [1-6, 8] # R: 311,066 # A: Unclear Range of follow-up (yrs): 7.9 to 17.7	randomised trials [a]	very serious [b]	not serious [c]	not serious [d]	not serious [e]	none [f]	Unavailable*	0.2% ‡	RR 0.99 (0.95 to 1.03)	**2 fewer per 100,000** (from 5 more to 8 fewer) ‡	⊕⊕◯◯ LOW
								2.1% ‡		**22 fewer per 100,000** (from 65 more to 108 fewer) ‡	
								4.0% ‡		**40 fewer per 100,000** (from 121 more to 202 fewer) ‡	
Sub-Group: All-Cause Mortality (50-59 years)											
3 [3,4,7] # R:	randomised trials	very serious [b]	not serious [g]	not serious [d]	not serious [e]	none [f]	1,836/43,196 (4.3%)	3.5%	RR 1.02 (0.95 to 1.09)	**70 more per 100,000** (from 175 fewer to 315 more)	⊕⊕◯◯ LOW

Mammography +/- CBE compared to Usual Care											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mammography +/- CBE	Usual Care	Relative (95% CI)	Absolute (95% CI)	
79,749 # A: 79,695 Range of follow-up (yrs): 7.9 to 13.0								4.8%		**95 more per 100,000** (from 239 fewer to 429 more)	
Sub-Group: All-Cause Mortality (60-69 years)											
2 [3,4] # R: 39,681 # A: 39,681 Range of follow-up (yrs): 7.9	randomised trials	very serious [b]	not serious [h]	not serious [d]	not serious [e]	none [f]	2,899/23,412 (12.4%)	2,080/16,269 (12.8%)	RR 0.97 (0.92 to 1.02)	**384 fewer per 100,000** (from 256 more to 1,023 fewer)	⊕⊕◯◯ LOW
Sub-Group: All-Cause Mortality (70-74 years)											
2 [3,4] # R: 17,646 # A: 17,646 Range of follow-up (yrs): 7.9	randomised trials	very serious [b]	very serious [i]	not serious [d]	not serious [e]	none [f]	2,869/10,339 (27.7%)	26.1% 29.5%	RR 0.98 (0.87 to 1.11)	**523 fewer per 100,000** (from 2,877 more to 3,400 fewer) **590 fewer per 100,000** (from 3,243 more to 3,832 fewer)	⊕◯◯◯ VERY LOW

CI: Confidence interval; **RR:** Risk ratio; # R: Number randomised; # A: Number analyzed
*Complete data was not available. Numerators and/or denominators were either unclear or not reported for all included studies.
‡The baseline risk (in the control group) may not have been representative of all included studies because the numerators and/or denominators were either unclear or not reported. Values provided reflect those studies which provided complete data.

Explanations

a. Two studies considered quasi-randomised (Stockholm & Gothenburg)
b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas
c. Heterogeneity may be low (I^2=0%); (p-value=0.55)
d. Studies are relevant to the PICO being considered. Breast density was not addressed. For some studies, the control group received screening after the screening period. Studies reported in Nystrom 2002 included one round of screening in the control group
e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs include the null, but do not cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
f. Cannot assess publication bias (insufficient number of trials)
g. Heterogeneity may be low (I^2=0%); (p-value=0.60)
h. Heterogeneity may be low (I^2=0%); (p-value=0.59)
i. Heterogeneity may be substantial (I^2=74%); (p-value=0.05)

Part A- GRADE Summary of Findings Table – All-Cause Mortality (Stratified by Age)

Mammography +/- CBE compared to Usual Care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with Usual Care (Assumed Risk)	**Risk with Mammography +/- CBE (Corresponding Risk)**				
Sub-Group: All-Cause Mortality (40-49 years) # Randomised: 311,066 # Analyzed: Unclear Range of follow-up (yrs): 7.9 to 17.7	**Low**		RR 0.99 (0.95 to 1.03)	Unavailable (7 RCTs) [a]	⊕⊕◯◯ LOW [b,c,d,e,f]	NNS (Low): 50,000 (12,500 to -20,000) *Cis include ∞ NNS (Moderate): 4,762 (926 to -1,538) *Cis include ∞ NNS (High): 2,500 (495 to -826) *Cis include ∞
	169 per 100,000 ‡	**167 per 100,000** (161 to 174)				
	Moderate					
	2,150 per 100,000 ‡	**2,129 per 100,000** (2,042 to 2,215)				
	High					
	4,039 per 100,000 ‡	**3,999 per 100,000** (3,837 to 4,160)				
Sub-Group: All-Cause Mortality (50-59 years) # Randomised: 79,749 # Analyzed: 79,695 Range of follow-up (yrs): 7.9 to 13.0	**Low**		RR 1.02 (0.95 to 1.09)	79,695 (3 RCTs)	⊕⊕◯◯ LOW [b,d,e,f,g]	NNS (Low): -1,429 (571 to -317) *Cis include ∞ NNS (High): -1,053 (418 to -233) *Cis include ∞
	3,503 per 100,000	**3,573 per 100,000** (3,328 to 3,818)				
	High					
	4,770 per 100,000	**4,865 per 100,000** (4,531 to 5,199)				
Sub-Group: All-Cause Mortality (60-69 years) # Randomised: 39,681 # Analyzed: 39,681 Range of follow-up (yrs): 7.9	12,785 per 100,000	**12,401 per 100,000** (11,762 to 13,041)	RR 0.97 (0.92 to 1.02)	39,681 (2 RCTs)	⊕⊕◯◯ LOW [b,d,e,f,h]	NNS (Low): 260 (98 to -391) *Cis include ∞
Sub-Group: All-Cause Mortality (70-74 years) # Randomised: 17,646 # Analyzed: 17,646	**Low**		RR 0.98 (0.87 to 1.11)	17,646 (2 RCTs)	⊕◯◯◯ VERY LOW [b,d,e,f,i]	NNS (Low): 191 (29 to -35) *Cis include ∞ NNS (High): 169 (26 to -31) *Cis include ∞
	26,150 per 100,000	**25,627 per 100,000** (22,751 to 29,027)				
	High					

Mammography +/- CBE compared to Usual Care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with Usual Care (Assumed Risk)	**Risk with Mammography +/- CBE (Corresponding Risk)**				
Range of follow-up (yrs): 7.9	29,480 per 100,000	**28,890 per 100,000** (25,648 to 32,723)				

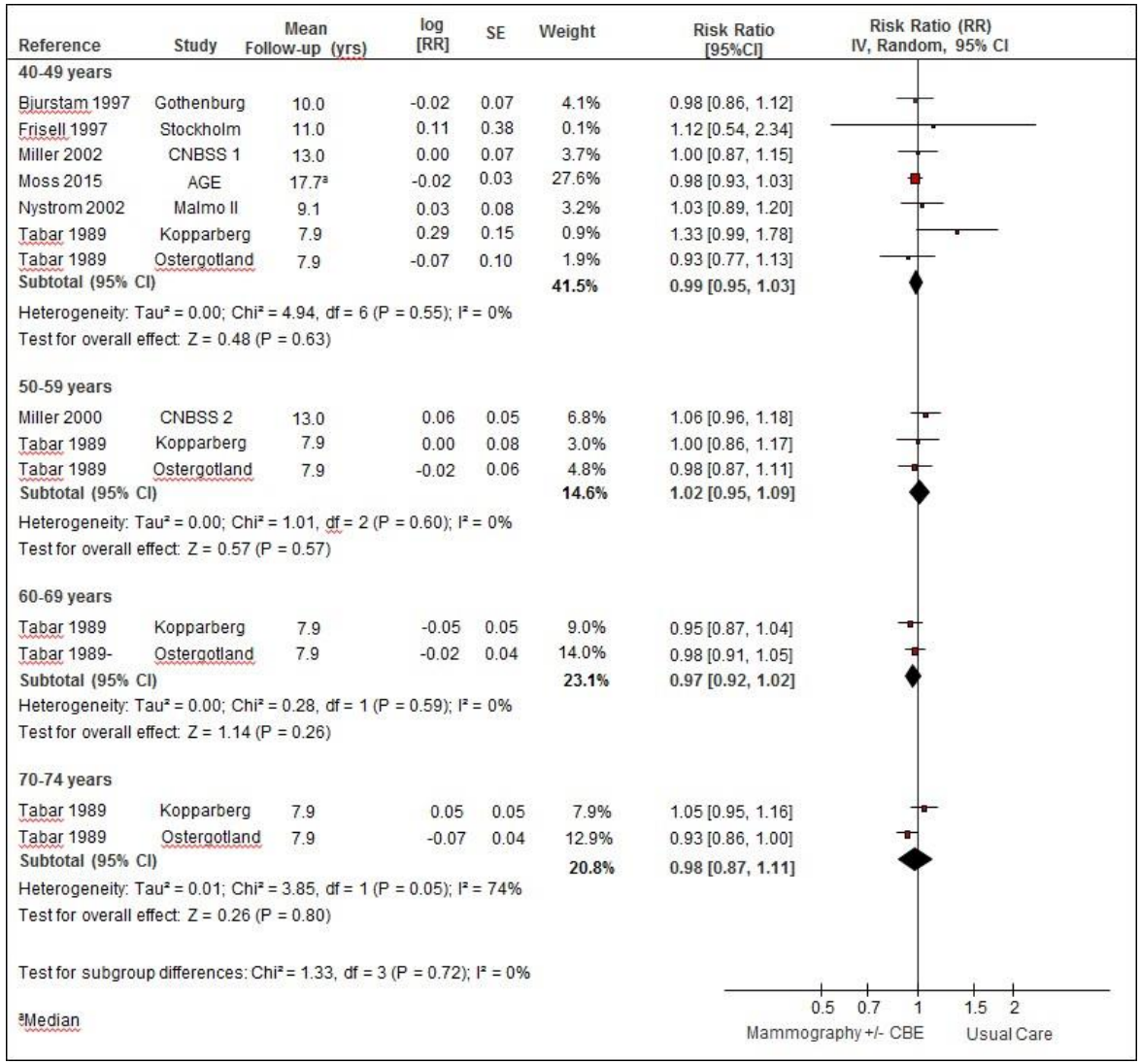
***The corresponding risk** (and its 95% confidence interval) is based on **the assumed risk** in the comparison group and the **relative effect** of the intervention (and its 95% CI).
‡The baseline risk (in the control group) may not have been representative of all included studies because the numerators and/or denominators were either unclear or not reported. Values provided reflect those studies which provided complete data. **CI:** Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Two studies considered quasi-randomised (Stockholm & Gothenburg)
b. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas
c. Heterogeneity may be low (I^2=0%); (p-value=0.55)
d. Studies are relevant to the PICO being considered. Breast density was not addressed. For some studies, the control group received screening after the screening period. Studies reported in Nystrom 2002 included one round of screening in the control group
e. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs include the null, but do not cross appreciable benefit (RR 0.75). Given the large sample sizes, an optimal sample size calculation was not warranted.
f. Cannot assess publication bias (insufficient number of trials)
g. Heterogeneity may be low (I^2=0%); (p-value=0.60)
h. Heterogeneity may be low (I^2=0%); (p-value=0.59)
i. Heterogeneity may be substantial (I^2=74%); (p-value=0.05)

Part A- Forest Plot – All-Cause Mortality (Stratified by Age)



Evidence Set 4a- Mammography +/- Clinical Breast Exam- Overdiagnosis (RCTs)

Part A- GRADE Evidence Profile Table – Overdiagnosis (All Ages) - Randomised Controlled Trials

Included Studies:

1: Malmo I (Zackrisson 2006) 2: Baines 2016 (CNBSS 1 & 2)* 3: Swedish Two County (Kopparberg) (Yen 2012)

Mammography +/- CBE vs. Usual Care											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mammography +/- CBE	Usual Care/No Screening	Relative (95% CI)	Absolute (95% CI)	
Invasive + In situ											
3 [1-3] Range of follow-up: 20 to 29 years	randomised trials	very serious [a]	very serious [b]	not serious [c]	serious [d]	none [e]	Results were reported differently between studies. One study reported no overdiagnosis [RR 1.00 (0.92-1.08)]. Another study reported 55% and 16% of overdiagnosis for 40-49, and 50-59, respectively. The remaining study reported that there was a 10% of incidence in the control group [g].				⊕◯◯◯ VERY LOW
Invasive only											
3 [1-3] Range of follow-up: 20 to 29 years	randomised trials	very serious [a]	very serious [b]	not serious [c]	serious [d]	none [e]	Results were reported differently between studies. One study reported no overdiagnosis [RR 0.99 (0.92-1.07)]. Another reported 48% and 5% overdiagnosis for 40-49, and 50-59, respectively. The remaining study reported 7% overdiagnosis [g].				⊕◯◯◯ VERY LOW
In situ only											
2 [1,3] Range of follow-up: 20 to 29 years	randomised trials	very serious [f]	very serious [b]	not serious [c]	serious [d]	none [e]	Results were reported differently between studies. One study noted that there was overdiagnosis in the screening arm (although not statistically significant) [RR 1.17 (0.88-1.55)]. The remaining study reported 3% overdiagnosis [g].				⊕◯◯◯ VERY LOW

CI: Confidence interval
*Updated follow-up from studies identified in updated search. Revised estimates to replace previous publication of Miller 2014.

Explanations

a. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas. Baseline imbalances apparent in two studies.
b. Reporting of estimates varied between studies. One cannot be confident that the same methodological approach was used.
c. Studies are relevant to the PICO being addressed.
d. Narrative analysis. Effect sizes were not provided consistently across studies.
e. Cannot assess publication bias (insufficient number of trials).
f. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas. Baseline imbalances apparent in one study.
g. Overdiagnosis characteristics were extracted as reported in the overviews.

Part A- GRADE Summary of Findings Table – Overdiagnosis (All Ages)- Randomised Controlled Trials

Mammography +/- CBE compared to Usual Care

Outcomes	**Anticipated absolute effects*** (95% CI) **Risk with Usual Care/No Screening (Assumed Risk)**	**Risk with Mammography +/- CBE (Corresponding Risk)**	Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
Invasive + In situ Range of follow-up: 20 to 29 years	Results were reported differently between studies. One study reported no overdiagnosis [RR 1.00 (0.92-1.08)]. Another study reported 55% and 16% of overdiagnosis for 40-49, and 50-59, respectively. The remaining study reported that there was a 10% of incidence in the control group [g].			(3 RCTs)	⊕◯◯◯ VERY LOW [a,b,c,d,e]	
Invasive only Range of follow-up: 20 to 29 years	Results were reported differently between studies. One study reported no overdiagnosis [RR 0.99 (0.92-1.07)]. Another reported 48% and 5% overdiagnosis for 40-49, and 50-59, respectively. The remaining study reported 7% overdiagnosis [g].			(3 RCTs)	⊕◯◯◯ VERY LOW [a,b,c,d,e]	
In situ only Range of follow-up: 20 to 29 years	Results were reported differently between studies. One study noted that there was overdiagnosis in the screening arm (although not statistically significant) [RR 1.17 (0.88-1.55)]. The remaining study reported 3% overdiagnosis [g].			(2 RCTs)	⊕◯◯◯ VERY LOW [b,c,d,e,f]	

***The corresponding risk** (and its 95% confidence interval) is based on **the assumed risk** in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas. Baseline imbalances apparent in two studies.
b. Reporting of estimates varied between studies. One cannot be confident that the same methodological approach was used.
c. Studies are relevant to the PICO being addressed.
d. Narrative analysis. Effect sizes were not provided consistently across studies.
e. Cannot assess publication bias (insufficient number of trials)
f. Randomisation and allocation concealment were either not reported or there were serious deficiencies in these areas. Baseline imbalances apparent in one study.
g. Overdiagnosis characteristics were extracted as reported in the overviews.

Evidence Set 4b- Mammography +/- Clinical Breast Exam - Overdiagnosis (Cohort)

Part A- GRADE Evidence Profile Table – Overdiagnosis (All Ages)- Cohort Studies

1: Lund 2013 2: Puliti 2012 3: Hellquist 2012

4: Njor 2013

Mammography +/- CBE vs. Usual Care/ No Screening								
Quality assessment							Impact	Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		
Invasive + In situ								
4 [1-4] Range of follow-up: NR	observational studies	very serious [a]	very serious [b]	serious [c]	serious [d]	none [e]	Results were reported differently between studies. One study reported no overdiagnosis [RR 1.01 (0.94-1.08)]. The other studies reported percentages: 22% (-0.9% to 64%); 10% (-2% to 23%); 3% (-14% to 25%- Copenhagen); and 0.7% (-9% to 12%- Funen) [f].	⊕◯◯◯ VERY LOW
Invasive only								
3 [1-3] Range of follow-up: NR	observational studies	very serious [a]	very serious [a]	serious [c]	serious [d]	none [e]	Results were reported differently between studies. One study reported no overdiagnosis [RR 0.95 (0.88-1.01)]. The other studies reported percentages: 7% (-0.8% to 45%) and 5% (-7% to 18%) [f].	⊕◯◯◯ VERY LOW
In situ only								
0	observational studies							-

CI: Confidence interval; **NR**: Not Reported

Explanations

a. A number of studies did not adjust for age and hormone therapy use (significant confounders) and also did not adjust or address for lead time bias.
b. Reporting of estimates varied between studies. One cannot be confident that the same methodological approach was used.
c. Some studies had either included high risk subjects in their cohort, or did not mention whether they were excluded.
d. Narrative analysis. Effect sizes were not provided consistently across studies.
e. Cannot assess publication bias (insufficient number of studies)
f. Overdiagnosis characteristics were extracted as reported in the overviews.

Part A- GRADE Summary of Findings Table – Overdiagnosis (All Ages)- Cohort Studies

Mammography +/- CBE compared to Usual Care

Outcomes	Impact	№ of participants (studies)	Quality of the evidence (GRADE)
Invasive + In situ	Results were reported differently between studies. One study reported no overdiagnosis [RR 1.01 (0.94-1.08)]. The other studies reported percentages: 22% (-0.9% to 64%); 10% (-2% to 23%); 3% (-14% to 25%- Copenhagen); and 0.7% (-9% to 12%- Funen) f.	(4 observational studies)	⊕◯◯◯ VERY LOW a,b,c,d,e
Invasive only	Results were reported differently between studies. One study reported no overdiagnosis [RR 0.95 (0.88-1.01)]. The other studies reported percentages: 7% (-0.8% to 45%) and 5% (-7% to 18%) f.	(3 observational studies)	⊕◯◯◯ VERY LOW a,c,d,e
In situ only		(0 observational studies)	-

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval**; NR:** Not reported

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. A number of studies did not adjust for age and hormone therapy use (significant confounders) and also did not adjust or address for lead time bias.
b. Reporting of estimates varied between studies. One cannot be confident that the same methodological approach was used.
c. Some studies had either included high risk subjects in their cohort, or did not mention whether they were excluded.
d. Narrative analysis. Effect sizes were not provided consistently across studies.
e. Cannot assess publication bias (insufficient number of studies)
f. Overdiagnosis characteristics were extracted as reported in the overviews.

Evidence Set 5- Breast Self Exam- Breast Cancer Mortality (All Ages)

Part A- GRADE Evidence Profile Table – Breast-Cancer Mortality (All Ages)

Included Studies:

1: Thomas 2002

BSE vs. No Screening											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	BSE	No Screening	Relative (95% CI)	Absolute (95% CI)	
Main Analysis: Breast-Cancer Mortality (All Ages)											
1 # R: Unclear # A: 266,064 Range of Follow-up: NR	randomised trials	not serious [a]	not serious [b]	serious [c]	serious [d]	none [e]	135/132,979 (0.1%)	131/133,085 (0.1%)	**RR 1.03** (0.81 to 1.31)	**3 more per 100,000** (from 19 fewer to 31 more)	⊕⊕◯◯ LOW

CI: Confidence interval; **RR:** Risk ratio; #R: # Randomised; #A: # Analyzed

Explanations

a. Randomisation and allocation concealment were not reported. Differences in losses to follow-up were of concern.
b. Only one study (cannot downgrade for inconsistency)
c. Study population included patients 31-65 years old, which does not completely correspond with our population of interest (40+ years)
d. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95% CIs include the null, and also cross appreciable harm (1.25)
e. Cannot assess publication bias (insufficient number of trials)

Part A- GRADE Summary of Findings Table – Breast-Cancer Mortality (All Ages)

BSE compared to No Screening

Outcomes	**Anticipated absolute effects*** (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with No Screening (Assumed Risk)	**Risk with BSE (Corresponding Risk)**				
Main Analysis: Breast-Cancer Mortality (All Ages) # Randomised: Unclear # Analyzed: 266,064 Range of Follow-up: NR	98 per 100,000	**101 per 100,000** (80 to 129)	**RR 1.03** (0.81 to 1.31)	266,064 (1 RCT)	⊕⊕◯◯ LOW a,b,c,d,e	NNS: 33,333 (5,348 to -3,279) *Cis include ∞

***The corresponding risk** (and its 95% confidence interval) is based on **the assumed risk** in the comparison group and the **relative effect** of the intervention (and its 95% CI).

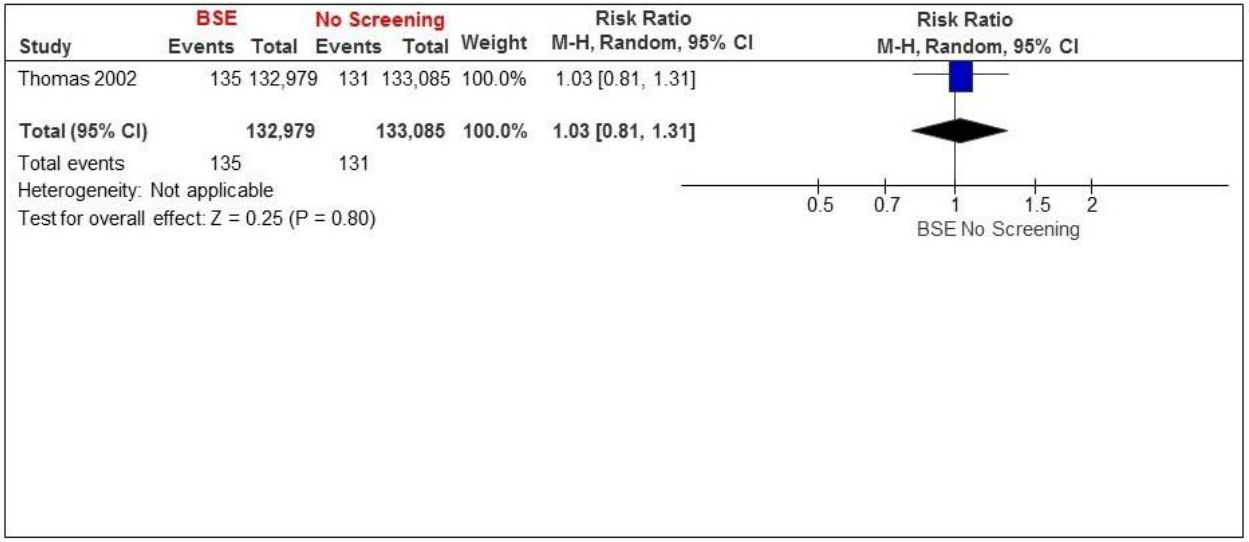
CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Randomisation and allocation concealment were not reported. Differences in losses to follow-up were of concern.
b. Only one study (cannot downgrade for inconsistency)
c. Study population included patients 31-65 years old, which does not completely correspond with our population of interest (40+ years)
d. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs include the null, and also cross appreciable harm (1.25)
e. Cannot assess publication bias (insufficient number of trials)

Part A- Forest Plot – Breast-Cancer Mortality (All Ages)



Evidence Set 6- Breast Self-Exam- All Cause Mortality (All Ages)

Part A- GRADE Evidence Profile Table – All-Cause Mortality (All Ages)

Included Studies:

1: Semiglazov 2003 2: Thomas 2002

BSE vs. No Screening											
Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	BSE	No Screening	Relative (95% CI)	Absolute (95% CI)	
All-Cause Mortality (All Ages)											
2 # R: Unclear # A: Unclear Range of follow-up: NR	randomised trials	not serious [a]	serious [b]	serious [c]	not serious [d]	none [e]	Unavailable*	5,939/133,085 (4.5%) ‡	**RR 0.96** (0.81 to 1.12)	**179 fewer per 100,000** (from 536 more to 848 fewer) ‡	⊕⊕◯◯ LOW

CI: Confidence interval; **RR:** Risk ratio; **# R**: # Randomised; **#A:** # Analyzed; **NR**: Not reported
*Complete data was not available. Numerators and/or denominators were either unclear or not reported for all included studies.
‡The baseline risk (in the control group) may not have been representative of all included studies because the numerators and/or denominators were either unclear or not reported. Values provided reflect those studies which provided complete data.

Explanations

a. Randomisation and allocation concealment were either low risk or unclear. Differences in losses to follow-up were of concern for one study.
b. Heterogeneity may be substantial (I^2=67%); (p-value=0.08)
c. Study population included patients 31-65 years old, which does not completely correspond with our population of interest (40+ years)
d. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs includes the null, but does not cross appreciable benefit (RR 0.75)
e. Cannot assess publication bias (insufficient number of trials)

Part A- GRADE Summary of Findings Table – All-Cause Mortality (All Ages)

BSE compared to No Screening

Outcomes	**Anticipated absolute effects*** (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with No Screening (Assumed Risk)	**Risk with BSE (Corresponding Risk)**				
All-Cause Mortality (All Ages) # Randomised: Unclear # Analyzed: Unclear Range of follow-up: NR	4,463 per 100,000 ‡	**4,284 per 100,000** (3,615 to 4,998)	**RR 0.96** (0.81 to 1.12)	Unavailable (2 RCTs)	⊕⊕◯◯ LOW [a,b,c,d,e]	NNS: 560 (118 to -187) *Cis include ∞

***The corresponding risk** (and its 95% confidence interval) is based **on the assumed risk** in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio; **NR**: Not reported

‡The baseline risk (in the control group) may not have been representative of all included studies because the numerators and/or denominators were either unclear or not reported. Values provided reflect those studies which provided complete data.

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Randomisation and allocation concealment were either low risk or unclear. Differences in losses to follow-up were of concern for one study.

b. Heterogeneity may be substantial (I^2=67%); (p-value=0.08)

c. Study population included patients 31-65 years old, which does not completely correspond with our population of interest (40+ years)

d. (i) The number of events and total population are large (>300 threshold for events); and (ii) the 95%CIs includes the null, but does not cross appreciable benefit (RR 0.75)

e. Cannot assess publication bias (insufficient number of trials)

Part A- Forest Plot – All-Cause Mortality (All Ages)