Review

J Med Screen
2021, Vol. 28(4) 389–404

DOI: 10.1177/0969141321993866
journals.sagepub.com/home/msc


Benefits and harms of breast cancer mammography screening for women at average risk of breast cancer: A systematic review for the European Commission Initiative on Breast Cancer

Carlos Canelo-Aybar[1,2], Diogenes S Ferreira[2,3], Mónica Ballesteros[2], Margarita Posso[2,4], Nadia Montero[2], Ivan Solà[2], Zuleika Saz-Parkinson[5], Donata Lerda[5], Paolo G Rossi[6], Stephen W Duffy[7], Markus Follmann[8], Axel Gräwingholt[9] and Pablo Alonso-Coello[1,2]

Abstract

Objectives: Mammography screening is generally accepted in women aged 50–69, but the balance between benefits and harms remains controversial in other age groups. This study systematically reviews these effects to inform the European Breast Cancer Guidelines.

Methods: We searched PubMed, EMBASE and Cochrane Library for randomised clinical trials (RCTs) or systematic reviews of observational studies in the absence of RCTs comparing invitation to mammography screening to no invitation in women at average breast cancer (BC) risk. We extracted data for mortality, BC stage, mastectomy rate, chemotherapy provision, overdiagnosis and false-positive-related adverse effects. We performed a pooled analysis of relative risks, applying an inverse-variance random-effects model for three age groups (<50, 50–69 and 70–74). GRADE (Grading of Recommendations Assessment, Development and Evaluation) was used to assess the certainty of evidence.

Results: We identified 10 RCTs including 616,641 women aged 38–75. Mammography reduced BC mortality in women aged 50–69 (relative risk (RR) 0.77, 95%CI (confidence interval) 0.66–0.90, *high certainty*) and 70–74 (RR 0.77, 95%CI 0.54–1.09, *high certainty*), with smaller reductions in under 50s (RR 0.88, 95%CI 0.76–1.02, *moderate certainty*). Mammography reduced stage IIA+ in women 50–69 (RR 0.80, 95%CI 0.64–1.00, *very low certainty*) but resulted in an overdiagnosis probability of 23% (95%CI 18–27%) and 17% (95%CI 15–20%) in under 50s and 50–69, respectively (*moderate certainty*). Mammography was associated with 2.9% increased risk of invasive procedures with benign outcomes (*low certainty*).

Conclusions: For women 50–69, high certainty evidence that mammography screening reduces BC mortality risk would support policymakers formulating strong recommendations. In other age groups, where the net balance of effects is less clear, conditional recommendations will be more likely, together with shared decision-making.

Keywords

Guidelines, breast cancer, mass screening, mammography

Date received: 8 July 2020; accepted: 31 December 2020

[1]CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[2]Department of Clinical Epidemiology and Public Health, Iberoamerican Cochrane Centre, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain
[3]Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
[4]Department of Epidemiology and Evaluation, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
[5]European Commission, Joint Research Centre (JRC), Ispra, VA, Italy
[6]Epidemiology Unit, AUSL – IRCCS di Reggio Emilia, RE, Italy
[7]Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK
[8]German Cancer Society, Berlin, Germany
[9]Radiologie am Theater, Paderborn, Germany

Corresponding authors:
Carlos Canelo-Aybar, Iberoamerican Cochrane Centre, Biomedical Research Institute Sant Pau (IIB Sant Pau – CIBERESP), Sant Antonio María Claret 167, 08025 Barcelona, Spain.
Email: ccanelo@santpau.cat

Zuleika Saz-Parkinson, European Commission, Joint Research Centre (JRC), Via E. Fermi, 2749. TP127, I-21027 Ispra (VA), Italy.
Email: zuleika.saz-parkinson@ec.europa.eu

Introduction

Breast cancer (BC) is the second most common malignancy in the world.[1] In the European Union, 404,920 women were diagnosed with BC and 98,755 died during 2018.[2] Over the last 20 years, BC mortality has decreased due to improvements in treatment, services delivery and implementation of population screening. However, the role of population screening has been under debate over the last three decades due to conflicting systematic reviews and recommendations.[3,4]

Randomised clinical trials (RCTs) carried out during the 1970s and 1980s showed that mammography screening is associated with a reduction in BC mortality.[5] However, screening a healthy population is also associated with undesirable effects such as recalling women with a false-positive result for additional imaging.[6] Overdiagnosis (BC cases that would not have clinically surfaced in the absence of screening) is another downside of screening, and its magnitude is controversial.[7]

Numerous organisations have issued screening recommendations. The WHO recommended in favour of screening starting at 40 years of age in well-resourced settings.[8] The Canadian Task Force recommended screening only in women over 50, due to the risk of overdiagnosis and unnecessary biopsies in younger women.[9] The American Cancer Society, including evidence from RCTs and observational and modelling studies, recommended initiating annual screening at 45.[10]

In 2015, the (ECIBC) was launched to develop the European Guidelines on Breast Cancer Screening and Diagnosis. This systematic review informed the recommendations about mammography screening for early detection of BC in asymptomatic women at average risk. During the guideline development,[11] the Guidelines Development Group (GDG) made detailed considerations on the evidence of effects as well as values and preferences, equity, acceptability and feasibility. Readers are welcome to refer to these considerations in the published recommendations and on the ECIBC website (https://healthcare-quality.jrc.ec.europa.eu/european-breast-cancer-guidelines/screening-ages-and-frequencies).[12,13]

Methods

Structured question and outcome prioritisation

The clinical question prioritised by the GDG, 'Which is the optimal age range in which to carry out screening for breast cancer?', followed the Population, Intervention, Comparison and Outcomes format (Box 1). Three sub-populations were pre-defined: women under 50, 50–69 and 70–74 years old at the moment of invitation to screening.

During the development of the recommendations,[11] the GDG decided to split the sub-group of women under 50 into two sub-groups: 40–44 and 45–49. The outcomes were prioritised by the GDG using a 1–9 scale as suggested by the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach.[14]

Data sources and searches

We searched MEDLINE (April 2016), EMBASE (April 2016) and CENTRAL (March 2016) databases using pre-defined algorithms for both systematic reviews and individual studies. We adapted the search terms to each database (see Supplemental material 1). We also reviewed lists of references of the included studies, and members of the GDG were consulted about potentially missing studies.

During June 2018, we performed a new search in MEDLINE as part of the ECIBC's guideline updating process. The results were assessed by the GDG and, as no relevant studies were identified that could potentially

Box 1 Structured clinical question.

Population	Intervention	Comparison	Outcomes
Women who are at average risk of breast cancer: • Under 50 y • 50–69 y • 70–74 y	Invitation to mammography screening	No invitation to mammography screening	Critical • Breast cancer mortality • Overdiagnosis Important • Stage of breast cancer • Other cause mortality • Rate of mastectomies • Provision of chemotherapy: provision of either adjuvant, neoadjuvant or both • Psychological effects: include anxiety caused or relieved by screening, anxiety caused by assessment of suspicious screening findings, longer length of life qualified by longer periods of life spent with a diagnosis of breast cancer, treatment side effects including psychosocial effects of body image following surgery • False positive related adverse effects: psychological distress • False positive related adverse effects: biopsies and surgeries

change the recommendations, they decided to not update the review. In November 2019, the GDG met and considered that, to the best of their knowledge, there were no relevant new publications.

Study selection

We included RCTs of women at average risk of BC (without family history of BC or inherited changes of BRCA1 and BRCA2 genes), comparing invitation to mammography screening versus no invitation. If no RCTs were identified, we included systematic reviews of observational studies. For overdiagnosis, we included only trials in which, after completing the study phase, neither women in the control nor the intervention group were offered mammography screening. We excluded studies conducted outside the context of screening programmes or not published in English.

Initially, at the title and abstract level, two reviewers after calibration assessed the eligibility of the references retrieved. Two reviewers independently reviewed the full text of the selected references. Discrepancies were solved either by consensus or with the help of a third reviewer.

Data extraction and risk of bias assessment

Details of the study design, population, follow-up and results were extracted by one reviewer and confirmed by a second reviewer. If needed, we requested additional data from authors of the included studies. We assessed the risk of bias (RoB) of RCTs using the Cochrane RoB Assessment tool[15], and systematic reviews with the AMSTAR (A Measurement Tool to Assess Systematic Reviews) checklist (see Supplemental material 2).[16]

Data analysis

To estimate the effect of mammography screening on BC mortality, we used two methods. The 'short case accrual' method includes only BC deaths among BC cases diagnosed during the screening intervention phase.[17,18] The 'long case accrual' method considers all BC deaths irrespective of the date of diagnosis, accrual time being equivalent to the follow-up of the study.[17,18]

We estimated overdiagnosis as the difference in the cumulative number of BC in the groups invited and not invited to screening, expressed: (a) as a percentage of the number of cancers in the screening group (population perspective) or (b) as a percentage of the cancers diagnosed during the screening phase of the trial in the invited to screening group (individual perspective). We pooled data as relative risks (RR) using a random-effects model (Review Manager v5.3). We assessed the presence of heterogeneity among studies using the Cochrane chi-square test and the I^2 statistic. Additionally, we provided subgroup analysis based on the risk of bias assessment and a post-hoc sensitivity analysis excluding RCTs with a substantial concern for breaking the concealment at randomisation.

To estimate risk differences, we used the baseline risk from the control arms of the RCTs; we also provide estimates using baseline risks proposed by the GDG members considering European population surveillance data. All results are expressed as by 100,000 women invited to mammography screening.

Certainty of the evidence

We rated the certainty of the evidence for each outcome taking into consideration the standard GRADE domains,[19,20] described in the evidence profiles (see Supplemental material 4).

Results

Search results

From 2393 unique citations, we selected 57 to be appraised as full text. At this stage, we excluded seven reviews, and 13 observational studies of mammography screening reporting outcomes available from RCTs. Additionally, we excluded the Edinburgh trial because of important baseline differences between the screening and control groups, suggesting suboptimal randomisation (see Supplemental material 5).[21,22]

We included 30 publications from nine RCTs: the Health Insurance Plan (HIP) of Greater New York trial,[5,23–26] the Canadian Breast Cancer Screening Study (CNBSS-1 and CNBSS-2),[27–32] the United Kingdom Age trial,[33–35] the Stockholm trial,[36–38] the Malmö Mammographic Screening Trial (MMST I and MMST II),[39–41] the Göteborg trial,[42,43] the Swedish Two-County trial (Östergötland and Kopparberg counties),[44–49] one publication that reported results for the five Swedish mammography trials,[18] and updated results of the UK Age Trial[50] and the Göteborg Trial (Table 1).[51] We also obtained additional age-stratified results for BC mortality from the authors of the CNBSS trial.

Four systematic reviews of observational studies fulfilled the eligibility criteria (Figure 1).[52–55] Brett et al.[53] assessed the adverse psychological impact of mammography screening in the general population. Salz et al.[55] examined the effects of false-positive mammogram results. Bond et al.[52] evaluated the psychological effects of false-positive screening mammograms in the UK, and one review assessed the cumulative risk of false-positive results leading to an invasive procedure (needle biopsy or surgery).[54]

BC mortality

Eight trials included 348,478 women less than 50,[24,26,28,29,36,37,39,40,43,50,51] six trials 249,930 women aged 50–69,[18,23,24,26,27,29,36,37,43,47,48,51] and two trials 18,233 women aged 70–74.[18,39,40,47,48] The trial time ranged from 3.5 to 18.8 years, the median short case accrual follow-up time from 9.1 to 24 years, and the median long case accrual follow-up time from 13 to 21.9 years, depending on the age strata (Table 1).

Table 1. Overview of included randomised clinical trials.

	HIP	Malmö I/Malmö II	Swedish two-county (*Ostergotland/Kopparberg*)	CNBSS-1/CNBSS-2	Stockholm	Göteborg	UK age
Start date	1963	1976	1977	1980	1981	1982	1991
Randomisation (if cluster, type)	Individual	Individual	Cluster (geographical area)	Individual[a]	Cluster (day of birth)	Individual, cluster (day of birth)	Individual
Number of women[b]	60,490	42,283 and 17,793	133,065	50,430/39,405	60,117	52,222	160,836
Age at entry (years)	40–64	45–70/43–49	38–75	40–49/50–59	39–65	39–59	39–41
Intervention							
Number of views	2	2 then 1 or 2/2[c]	1	2	1	2 (round 1) then 1 or 2[d]	2 then 1
Number of readers	2	2	1	1	1	1 (round 1-3), and 2 (round 4-5)	ND
Screening interval (months)	12	18-24	24,33[e]	12	24–28	18	12
Number of screening rounds	4	6–8/17	2–4	5/4–5	2	4–5[f]	4-6
Trial time (y)[g]	3.5	5.8/18.8	7	5	4.4	6.7	9
Last follow-up (years)[h]	18	19.2/9.1	20	21.9	17.4	24	17
Attendance, first round (%)	67	74/75–80	89	86/87	82	84	68
Controls invited to screening at end of intervention[i]	ND	No/yes[j]	Yes (after 7 y)	No	Yes (after 4 y)	Yes (after 7 y)	Yes (after 10 y)

ND: not described.

[a]There was a clinical breast examination before randomization, and the personnel in charge of allocation had access to this information.

[b]Numbers of participants included in the analysis, for some trials these numbers vary in different publications.

[c]Depending on parenchymal pattern.

[d]Depending on the density of breast.

[e]Women 38–49: 24 months; Women 50–74: 33 months.

[f]Women born 1923–1932: 4 rounds; women born 1933–1944: 5 rounds.

[g]Number represents median time of women spent in the active intervention phase of the study; otherwise, number of years from start of randomization to last screening in the intervention arm.

[h]Follow-up might differ for specific outcomes and accrual methods (i.e. short vs. long case). Number represents median time when available otherwise reported time from entry.

[i]Mean years after randomization.

[j]Applies to the cohort 1908–1922 (55–69 years at entry).

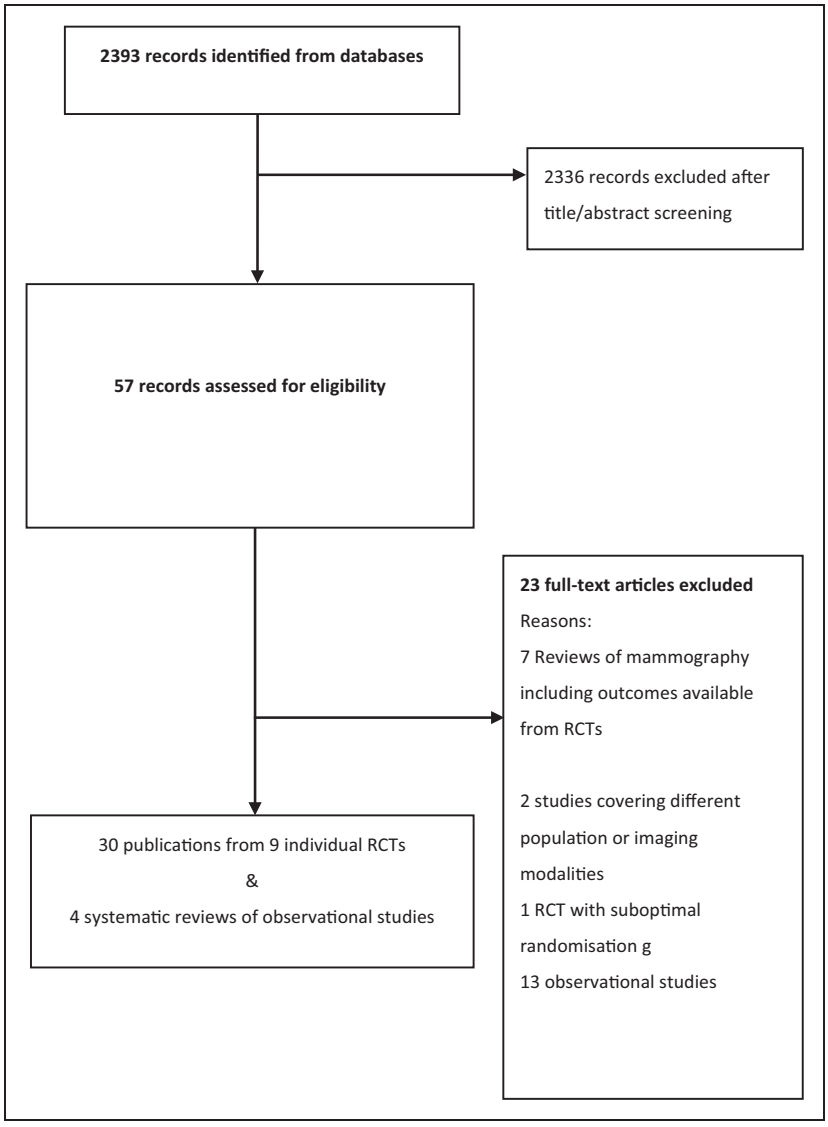

Figure 1. PRISMA flowchart.

In women under 50, invitation to mammography screening probably reduces the risk of BC mortality (RR 0.88; 95% CI (confidence interval) 0.76–1.02; $I^2 = 20\%$; short case accrual) (*moderate certainty*).[18,24,26,28,50,51] Comparable results were obtained using a long case accrual follow-up time (RR 0.92; 95%CI 0.83–1.02; $I^2 = 6\%$)[18,24,28,43,45,50,51] (Figure 2). In women aged between 50 and 69, mammography screening reduced the risk both for short (RR 0.77; 95%CI 0.66–0.90; $I^2 = 49\%$)[18,24,26,27,47,51] and long case accrual time[18,24,27,43,45,51] *(high certainty)* (Figure 3). For women aged 70–74, the Malmö I reported short case accrual follow-up and the Swedish Two-County reported long accrual follow-up time; mammography screening reduced the risk of BC mortality (RR 0.77; 95%CI 0.54–1.09; $I^2 = 0\%$)[18,47] *(high certainty)* (Figure 4).

The risk difference in BC mortality for women aged 50–69 was 138 fewer deaths per 100,000 women invited to screening (95%CI 204 fewer to 60 fewer) using short accrual follow time, and 175 fewer deaths per 100,000 women invited to screening (95%CI 251 fewer to 91 fewer) using long accrual time (Table 2). Sensitivity analysis including only RCTs at low RoB yielded similar results.

Other cause mortality

In women under 50, mammography screening may make no difference to other-cause mortality, but the evidence is uncertain (RR 1.04; 95%CI 0.95 to 1.15; $I^2 = 62\%$) *(very low certainty)*.[18,28,42,44,50] Two trials were included in the 50–69 group[27,44] and one trial in the 70–74 group;[44] in these age strata, mammography screening may also result in no difference (Table 2) *(low certainty)* (Supplemental material 3: Figures S5, S8 and S9).

Advanced BC

We defined advanced stage as either stage II or greater, tumour size $\geqslant 20$ mm or $\geqslant 1$ positive lymph node, which is consistent with stage IIA disease or higher. Additionally, we used a second definition of advanced disease as

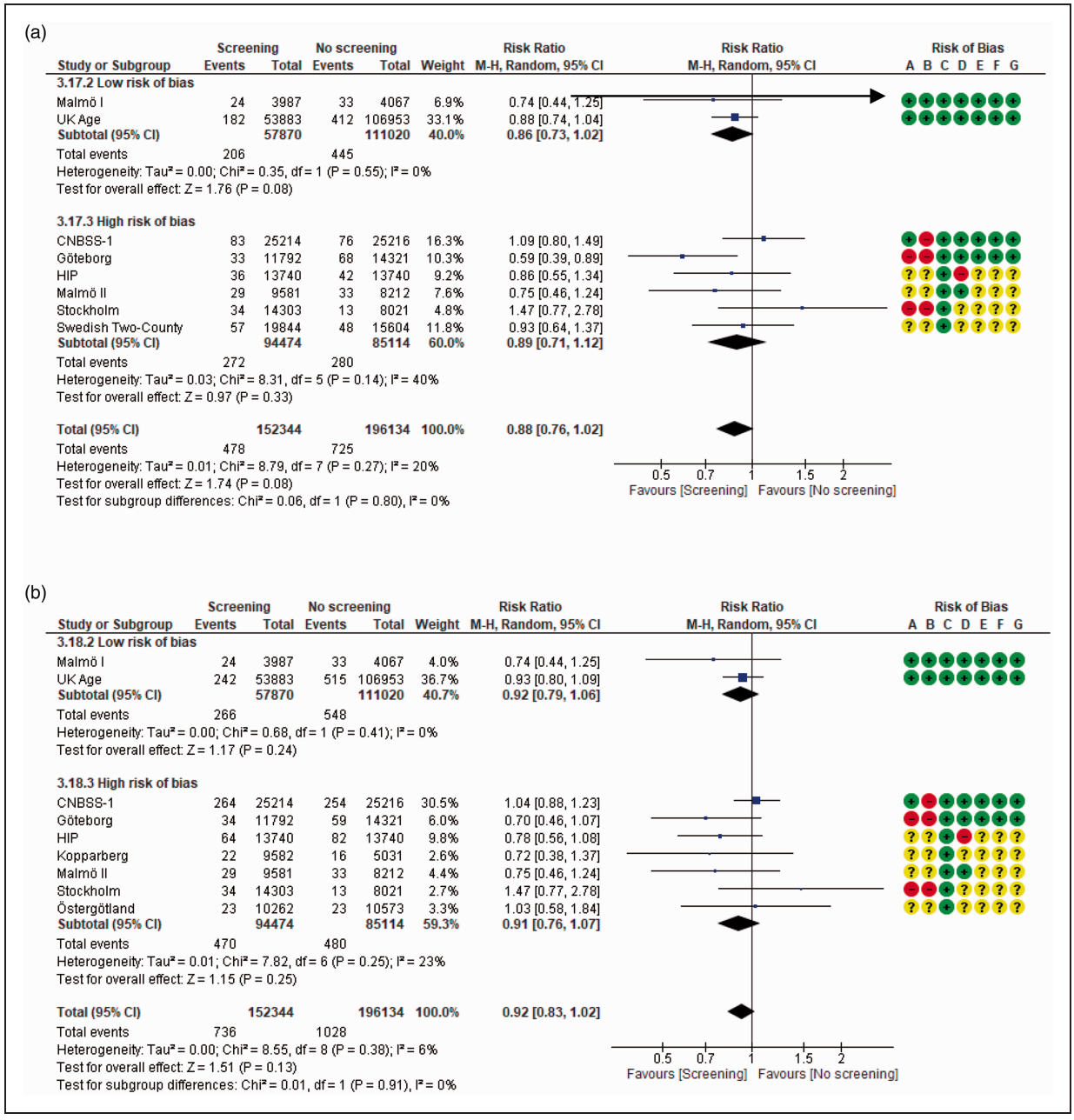

Figure 2. Effect on breast cancer mortality of mammography screening (women under 50 years of age): (a) short case accrual, mean follow up across studies 16.8 years; (b) longest case accrual, mean follow-up across studies 15.2 years. Risk of bias legend: (A) Random sequence generation, (B) allocation concealment, (C) blinding of participants and personnel, (D) blinding of outcome assessment, (E) incomplete outcome data and (F) selective reporting, (G) other bias.

regional or metastatic or tumour size ⩾ 40 mm, equivalent to stage III or higher.

Using the stage IIA or higher definition, in women under 50, mammography screening may reduce the risk of advanced disease, but the evidence is uncertain (RR 0.88; 95%CI 0.78 to 0.99; $I^2 = 0\%$) *(very low certainty)*.[23,24,28,34,43,45,50,51] In women aged 50–69, the effect size was similar (RR 0.80; 95%CI 0.64–1.00, $I^2 = 70\%$) *(very low certainty)*.[23,24,27,43,45,51] One trial including older women (aged 50–74) showed that mammography screening may reduce the risk of advanced disease (RR 0.64; 95%CI 0.55–0.73),[45] equivalent to 385 fewer cases (95%CI 482 fewer to 289 fewer) (Table 2) *(low certainty)* (Supplemental material 3: Figures S1 and S6).

Using the stage III or higher definition, in women under 50, screening may make little difference to the risk of advanced disease (RR 0.98; 95%CI: 0.74–1.29; $I^2 = 0\%$) *(low certainty)*.[23,24,28,34,45,50] In women aged between 50 and 69, mammography screening may reduce the risk of advanced disease (RR 0.62, 95%CI 0.48–0.80; $I^2 = 0\%$),[23,24,27,45] which is equivalent to 65 fewer cases of advanced BC *(low certainty)*. In women aged 50–74, mammography screening may reduce the risk of advanced

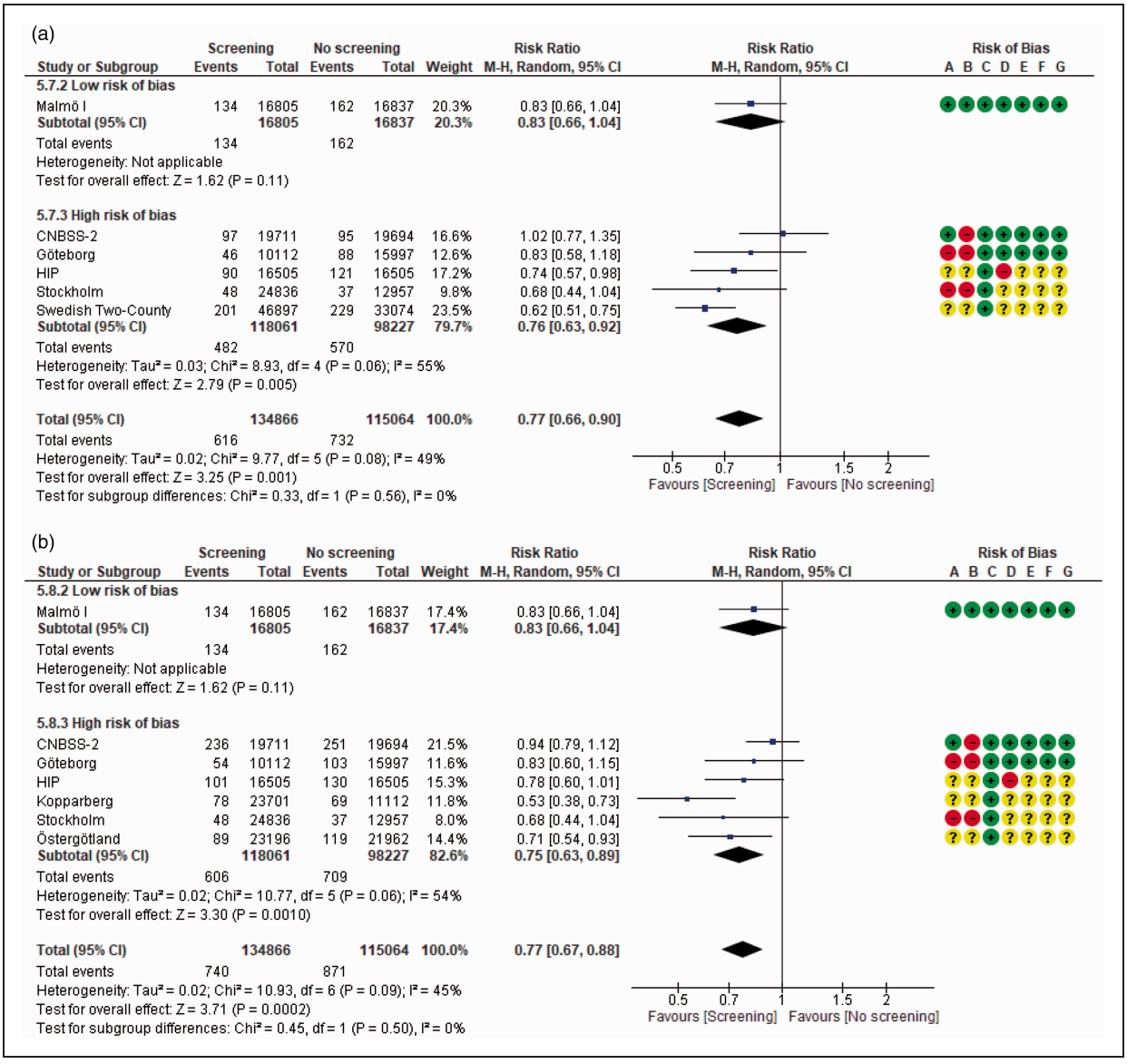

Figure 3. Effect on breast cancer mortality of mammography screening (women aged 50–69): (a) short case accrual, mean follow-up across studies 17.6 years; (b) longest case accrual, mean follow-up across studies 15.5 years. Risk of bias legend: (A) random sequence generation, (B) allocation concealment, (C) blinding of participants and personnel, (D) blinding of outcome assessment, (E) incomplete outcome data, (F) selective reporting and (G) other bias.

disease (RR 0.63, 95%CI 0.45–0.89),[45] equivalent to 63 fewer cases (95%CI 94 fewer to 19 fewer) (Table 2) *(low certainty)* (Supplemental material 3: Figures S2 and S7).

Overdiagnosis

We identified three trials, the CNBSS-1, the CNBSS-2 and a subgroup of women aged 55–69 from the Malmo-I trial (women aged 45–54 received screening at the end of the study). In women aged 40–49, the estimates of overdiagnosis were 12.4% (95%CI 9.9–14.9) from a population perspective and 22.7% (95%CI: 18.4–27.0) from an individual perspective *(moderate certainty)*.[28,29] In women aged between 50 and 69, we estimated a pooled overdiagnosis of 10.1% (95%CI: 8.6–11.6; $I^2 = 0\%$) from a population perspective and 17.3% (95%CI: 14.7–20.0; $I^2 = 10\%$) from an individual perspective (Table 3) *(moderate certainty)*.[27,29,40]

Rate of mastectomies

Across all age groups, women invited to screening may undergo more mastectomies (RR 1.20, 95%CI 1.11–1.30; $I^2 = 0\%$, 180 more in absolute terms) *(low certainty)* (Supplemental material 3: Figure S3).[18,31,39,46]

Provision of chemotherapy

Across all age groups, the evidence was uncertain with an RR of 0.86 (95%CI 0.53–1.40; I^2 71%)[18,39,46] (Table 3) *(very low certainty)* (Supplemental material 3: Figure S4).

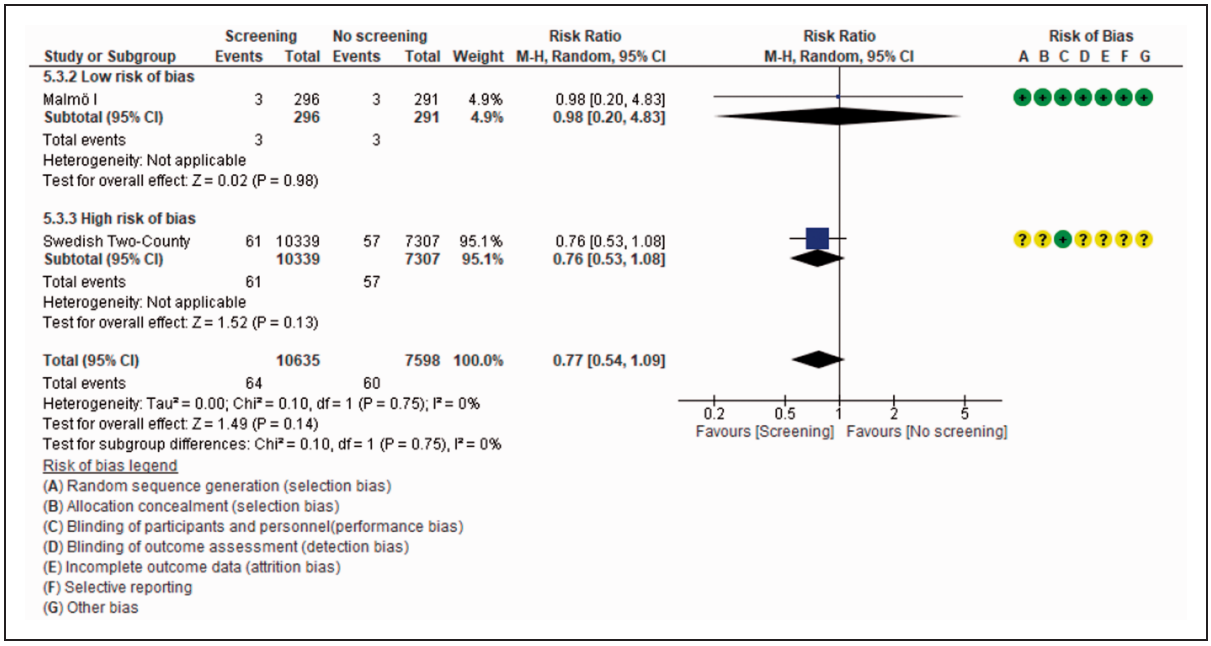

Figure 4. Effect on breast cancer mortality of mammography screening (women 70 years of age or older). Malmö I and Swedish two-county reported short case accrual estimate, follow-up across studies 9.5 years.

Psychological effects

Uncertain evidence showed that mammographic screening may not produce anxiety in women given a clear result after a mammogram.[53] However, those requiring further investigations may experience significant anxiety, in the short and long term, depending on the extent of the additional exams (Table 3) *(very low certainty)*.[53]

False-positive-related psychological distress

One review, including 17 studies, suggested an increase in the scores of disease-specific BC measures of psychological distress with false-positive results, being largest for anxiety about BC (r = 0.22; 95%CI: 0.18–0.27) and smallest for fear (r = 0.08; 95%CI: 0.03–0.14) *(low certainty)*.[55] In contrast, when using non-specific measures, the only suggested effect was a higher risk of generalised anxiety (r = 0.03; 95%CI: 0.00–0.07) *(low certainty)*.[55]

Another review included four studies evaluating psychological impact;[52] false-positive mammograms may be associated with negative psychological consequences when assessed using disease-specific measures (e.g. BC anxiety). Additionally, the risk of negative effects may be greater if a biopsy is required (RR 2.07; 95%CI 1.22–3.52) than if only further mammography is needed (RR 1.28; 95%CI 0.82–2.00) *(low certainty)*.[52]

False-positive-related procedures

One systematic review included four primary studies and an analysis of performance parameters from 20 screening programmes *(low certainty)*.[54] One Norwegian study reported a cumulative risk of undergoing fine needle aspiration cytology, core needle biopsy (CNB) and having a surgical intervention with a benign outcome of 3.9%, 1.5% and 0.9%, respectively.[56] The largest study, from a Spanish screening programme, reported an estimated cumulative risk of 1.8% for undergoing an invasive procedure with a benign outcome.[57]

RoB and certainty of the evidence

Our main concerns for BC mortality were: (1) the use of suboptimal random allocation methods, such as the date of birth to allocate women to each study arm in the Stockholm and Gothenburg trials;[36,42] (2) in the CNBSS trials, participants underwent clinical breast examination before randomisation, and this information was available to the personnel in charge of allocation.[27,28] However, no single trial drove the overall results in the subgroup analysis of low versus high RoB (Tables 2 and 3). This was corroborated in a post-hoc sensitivity analysis, where we explored the impact of excluding the CNBSS trials, which resulted in a non-meaningful change in the point estimates for the outcome breast cancer mortality (i.e. for women under 50, RR 0.84; 95%CI 0.73–0.98, for short case accrual) (Supplemental material 3: Figures S10 and S11).

For overdiagnosis, in the CNBSS trials, there was potential bias due to screening after the trial ended, as screening programmes were subsequently implemented under different province jurisdictions with an increased likelihood of attendance of women who had been screened during the trials.[29,40] Also, the overdiagnosis results were simple cumulative incidence differences with no adjustment for lead time and, therefore, might be overestimated.

Other relevant limitations were related to indirectness, due to difference in quality control of screening and improvement of care over the last two decades, but given

Table 2. Summary of available evidence on desirable effects of breast cancer mammography screening by age groups.

Quality assessment					Summary of findings		
Age group	No of studies (population)	Design	Certainty assessment	Certainty of evidence	Relative effect (95%CI)	Basal risk[a]	Absolute reduction
Breast cancer mortality (short case accrual)							
<50	8 (348,478)	Randomised trials	Downgraded due to imprecision (serious)	⊕⊕⊕◯[b,c,d] MODERATE	RR 0.88 (0.76–1.02)	0.4%	48 fewer per 100,000 (from 84 fewer to 8 more)
						0.7%	84 fewer per 100,000 (from 168 fewer to 14 more)
50–69	6 (249,930)	Randomised trials	No relevant considerations were found to justify downgrading the certainty of the evidence.	⊕⊕⊕⊕[b,c,e] HIGH	RR 0.77 (0.66–0.90)	0.6%	138 fewer per 100,000 (from 204 fewer to 60 fewer)
						1.0%	230 fewer per 100,000 (from 340 fewer to 100 fewer)
						2.1%	483 fewer per 100,000 (from 714 fewer to 210 fewer)
70 to 74	2 (18,233)	randomised trials	No relevant considerations were found to justify downgrading the certainty of the evidence.	⊕⊕⊕⊕[b,c,f] HIGH	RR 0.77 (0.54–1.09)	0.9%	207 fewer per 100,000 (from 81 more to 414 fewer)
						3.0%	690 fewer per 100,000 (from 1380 fewer to 270 more)
Breast cancer mortality (longest case accrual available)							
<50	8 (348,478)	Randomised trials	Downgraded due to imprecision (serious)	⊕⊕⊕◯[b,c,d] MODERATE	RR 0.92 (0.83–1.02)	0.5%	38 fewer per 100,000 (from 82 fewer to 10 more)
50–69	6 (249,930)	Randomised trials	No relevant considerations were found to justify downgrading the certainty of the evidence.	⊕⊕⊕⊕[b,c,e] HIGH	RR 0.77 (0.67–0.88)	0.8%	175 fewer per 1,00,000 (from 251 fewer to 91 fewer)
Other cause mortality							
<50	6 (290,600)	Randomised trials	Downgraded due to inconsistency (serious), indirectness (serious) and imprecision (serious).	⊕◯◯◯ VERY LOW	RR 1.04 (0.95–1.15)	2.5%	100 more per 100,000 (from 125 fewer to 375 more)
50–69	2 (119,376)	Randomised trials	Downgraded due to indirectness (serious) and imprecision (serious).	⊕⊕◯◯ LOW	RR 0.99 (0.95–1.04)	6.6%	66 fewer per 100,000 (from 330 fewer to 264 more)
70–74	1 (17,646)	Randomised trials	Downgraded due to indirectness (serious) and imprecision (serious).	⊕⊕◯◯ LOW	RR 1.01 (0.91–1.10)	27.0%	270 more per 100,000 (from 2430 fewer to 2700 more)
Breast cancer stage IIA or higher							
<50	5 (300,307)	Randomised trials	Downgraded due to risk of bias (serious), indirectness (serious) and imprecision (serious).	⊕◯◯◯[d,g,h] VERY LOW	RR 0.88 (0.78–0.99)	0.4%	46 fewer per 100,000 (from 84 fewer to 4 fewer)
50–69	4 (196,141)	Randomised trials	Downgraded due to risk of bias (serious), inconsistency	⊕◯◯◯[c,d,i,k] VERY LOW	RR 0.80 (0.64–1.00)	0.7%	140 fewer per 100,000 (from 252 fewer to 0 fewer)

(continued)

Table 2. Continued.

Quality assessment					Summary of findings		
Age group	No of studies (population)	Design	Certainty assessment	Certainty of evidence	Relative effect (95%CI)	Basal risk[a]	Absolute reduction
			(serious) indirectness (serious) and imprecision (serious).				
50–74	1 (97,617)	Randomised trials	Downgraded due to indirectness (very serious).	⨁⨁◯◯[b,j] LOW	RR 0.64 (0.55–0.73)	1.1%	385 fewer per 100,000 (from 482 fewer to 289 fewer)
Breast cancer stage – stage III+ or tumour size ⩾ 40 mm[k]							
<50	4 (274,212)	Randomised trials	Downgraded due to indirectness (serious) and imprecision (serious).	⨁⨁◯◯[b,c,d] LOW	RR 0.98 (0.74–1.29)	0.1%	2 fewer per 100,000 (from 23 fewer to 26 more)
50–69	3 (170,032)	Randomised trials	Downgraded due to risk of bias (serious) and indirectness (serious).	⨁⨁◯◯[c,i] LOW	RR 0.62 (0.48–0.80)	0.2%	65 fewer per 100,000 (from 88 fewer to 34 fewer)
50–74	1 (97,617)	Randomised trials[c]	Downgraded due to indirectness (very serious).	⨁⨁◯◯[b,j] LOW	RR 0.63 (0.45–0.89)	0.2%	63 fewer per 100,000 (from 94 fewer to 19 fewer)

[a]The GDG considered that baseline risks higher than 0.6% should be considered to evaluate absolute effects of breast cancer mortality (Breast Cancer Screening, IARC Handbook of Cancer Prevention Volume 15).
[b]Some studies used random allocation methods that would not be currently accepted. One study had a non-blinded assessment of 'cause of death'. The GDG felt that the CNBSS-2 possibly had issues with achieving prognostic balance. The GDG felt that lack of allocation concealment in this set of studies did not lead to high risk of bias. Given that lack of single trials driving the overall results and similarity in effect sizes (the test for subgroup differences – low vs. high risk of bias trials – was non-significant) and overlapping confidence intervals (CIs), the risk of bias was rated as 'not serious'.
[c]Trials were conducted more than 20 years ago. Currently, women have higher adherence to breast cancer screening, and the quality control of screening and the care of breast cancer have improved. A large non-randomised study (Hellquist B 2011)[79] showed a reduced risk for breast cancer deaths in women aged 40–49 years invited to screening, compared with women not invited (RR = 0.74; 95%CI, 0.66–0.83), which is consistent with the results seen in the RCTs. The GDG did not downgrade for indirectness for breast cancer mortality but considered it serious for other outcomes.
[d]The 95% CI limits crosses the decision threshold (as the CI is wide, a different clinical decision regarding the intervention may be taken depending on whether the lower or the higher limit is considered).
[e]Despite concerns about indirectness from the trials, including the fact that the population age range of 40–74 is broader than the age range in this question, after considering evidence from contemporary non-randomised studies (Broeders et al.[3]) the GDG decided not to downgrade the quality of evidence for indirectness.
[f]For the mortality-related outcomes, the GDG decided not to downgrade for imprecision because the relative effect is consistent with those in other age groups and that lends support that the estimate of the effect is close to what is reported here. This decision is also reinforced by the fact that, if the indirect evidence from the 50–69 age stratum were considered here, the certainty of the evidence for this outcome would also have been rated as 'moderate', as a result of downgrading that evidence from 'high' to 'moderate' by one level for indirectness and using it here.
[g]Some studies were sub-optimally randomised and had non-blinded assessment of stage of disease; when analysis was restricted to low risk of bias trials, the risk estimate was non-significant.
[h]Indirectness same as for women aged 50–69.
[i]Non-blinded assessment of breast cancer stage is a serious concern. GDG members decided to downgrade to 'serious' for risk of bias.
[j]Unexplained inconsistency with statistical heterogeneity ($I^2 = 70\%$, $p = 0.02$). While one study shows clear benefit, in three studies, the 95%CI does not exclude important benefit or harm.
[k]Analysis includes women aged 40–74 years, but only about 13% of women were ⩾ 70 years.
[l]In the group of older than 70 years, it only included tumour size ⩾50 mm.

Table 3. Summary of available evidence on undesirable effects of breast cancer mammography screening by age groups.

Quality assessment					Summary of findings		
Age group	No of studies (population)	Design	Comments	Certainty of evidence	Relative effect (95%CI)	Basal risk	Absolute reduction
Overdiagnosis (population perspective)							
<50	1 (50,430)	Randomised trials	Downgraded due to indirectness (serious).	⨁⨁⨁◯[a] MODERATE	—	12.4% (95% CI 9.9–14.9%)	
50–69	2 (64,117)	Randomised trials	Downgraded due to indirectness (serious).	⨁⨁⨁◯ MODERATE	—	10.1% (95% CI 8.6–11.6%)	
Overdiagnosis (individual perspective)							
<50 y	1 (50,430)	Randomised trials	Downgraded due to indirectness due to indirectness (serious).	⨁⨁⨁◯[a] MODERATE	—	22.7% (95% CI 18.4–27.0%)	
50–69	2 (64,117)	Randomised trials	Downgraded due to indirectness (serious).	⨁⨁⨁◯ MODERATE	—	17.3% (95%CI 14.7–20.0%)	
Rate of mastectomies							
All groups	5 (249,550)	Randomised trials	Downgraded due to indirectness (very serious).	⨁⨁◯◯[a,b,c] LOW	**RR 1.20** (1.11–1.30)	0.9%	**180 more per 100,000** (from 99 more to 270 more)
Provision of chemotherapy							
All groups	2 (99,454)	Randomised trials	Downgraded due to inconsistency (serious), indirectness (very serious) and imprecision (serious).	⨁◯◯◯[d,e,f] VERY LOW	**RR 0.86** (0.53 to 1.40)	0.4%	**56 fewer per 100,000** (from 188 fewer to 160 more)
Quality of life (inferred from psychological effects)							
All groups	1 SR including 54 primary studies.	Observational studies	No relevant considerations were found to justify downgrading the certainty of the evidence	⨁◯◯◯[g] VERY LOW	—	It does not appear to create anxiety in women who are given a clear result after a mammogram and subsequently placed on routine recall. Mixed results about anxiety in women recalled for further testing: several studies reported transient or long term (from 6 months to 1 year after recall) anxiety, while other studies reported no differences in anxiety levels.	
False-positive-related adverse effects (psychological distress)							
All groups	2 SR including 24 primary studies.	Observational studies	No relevant considerations were found to justify	⨁⨁◯◯ LOW	—	Psychological distress in women that needed further mammography: RR = 1.28, 95%CI 0.82-2.00; For	

(continued)

Table 3. Continued.

Quality assessment					Summary of findings		
Age group	No of studies (population)	Design	Comments	Certainty of evidence	Relative effect (95%CI)	Basal risk	Absolute reduction
			downgrading the certainty of the evidence				women placed in early recall: RR = 1.82, 95%CI 1.22-2.72; For women that needed a fine needle puncture aspiration: RR = 1.80, 95% CI 1.17-2.77; For women that needed a biopsy: RR = 2.07, 95%CI 1.22-3.52. No differences in generic measures of general anxiety and depression at six weeks after assessment and three months after screening
False-positive-related adverse effects (biopsies and surgeries)							
All groups	One SR including four primary studies and a cross sectional data analysis.	Observational studies	No relevant considerations were found to justify downgrading the certainty of the evidence	⨁⨁◯◯[h] LOW	—		Overall 2.9% pooled cumulative risk of an invasive procedure with benign outcome (range 1.8% to 6.3% based on 2 studies) and 0.9% risk of undergoing surgical intervention with benign outcome (based on 1 study) (Hofvind 2012).[54] Cross-sectional data from the EUNICE Project showed that 2.2% and 1.1% of all screening examinations resulted in needle biopsy among women without breast cancer (initial and subsequent screens).[54]

[a]Trials were conducted more than 20 years ago. Currently, women have higher adherence to breast cancer screening and the quality control of screening and the care of breast cancer have improved. A large non-randomised study (Hellquist B 2011) showed a reduced risk for breast cancer deaths in women aged 40–49 years invited to screening, compared with women not invited (RR = 0.74; 95%CI, 0.66–0.83) which is consistent with the results seen in the RCTs. The GDG did not downgrade for indirectness for breast cancer mortality but considered it serious for other outcomes.
[b]Some studies used methods that would not be accepted for random allocation today. One study had non-blinded assessment of 'cause of death'. The GDG felt that the CNBSS-1 possibly had issues with achieving prognostic balance. The GDG felt that lack of allocation concealment in this set of studies did not lead to high risk of bias. Given the lack of single trials driving the overall results and similarity in effect sizes (the test for subgroup differences – low vs high risk of bias trials – was non-significant) and overlapping confidence intervals (CIs), the risk of bias was rated as 'not serious'.
[c]Population include women aged 40–74. Therefore, a much broader age range than the age group studied here. Observational studies do not confirm these results; instead, they provide opposite results.
[d]95% CI probably crosses the clinical decision threshold (as the CI is wide, a different clinical decision regarding the intervention may be taken depending on whether the lower or the higher limit is considered).
[e]Unexplained inconsistency with statistical heterogeneity ($I^2 = 71\%$, $P = 0.06$).
[f]Chemotherapy protocols and indications have significantly changed (e.g. node status was not determined in earlier studies).
[g]Unexplained inconsistency for variability in anxiety in the group of women recalled for further testing.
[h]Studies included women aged 50–69. Estimates for the 45–49 age stratum are likely to be higher.

the consistency with more recent observational studies (see Discussion section), the GDG did not consider downgrading the evidence for this reason.

Discussion

Main findings

Our review shows that there is high certainty evidence that mammography screening reduces the risk of BC mortality in women between the ages of 50 and 69, with the number of deaths averted ranging from 138 fewer to 483 fewer per 100,000 women invited to screening, depending on the baseline risk assumed (from 0.6 to 2.1%). For other age groups, the evidence is not conclusive. Consistently, women invited to screening across all age groups showed a lower risk of advanced stages of BC.

There is moderate certainty that screening is associated with an increase in undesirable effects. Especially important was overdiagnosis, regardless of the calculation method, which was larger from an individual perspective in the younger age groups compared to older groups.[7] Mammography screening did not appear to produce significant negative psychological effects as long as the results were clearly communicated, while false-positive results, especially when further assessment is required, increased the number of invasive procedures and psychological distress.

Our results in the context of previous research

Consistent with our analysis, observational studies suggest a reduction of BC mortality after screening implementation. A systematic review of time trend studies estimated a BC mortality reduction from 1 to 9% per year and from 28 to 36% in studies comparing post- and pre-screening periods.[3] A pooled analysis of seven incidence-based mortality (IBM) studies from European countries showed a mortality reduction of 25% among invited women and 38% among those actually screened.[3] Another review which classified studies according to the quality of the methods used to estimate the expected mortality in absence of screening found an IBM risk reduction of 26% in women invited for screening from studies with robust approaches.[58]

Our results suggest that screened women are diagnosed with less advanced disease. Observational evidence shows earlier BC staging at diagnosis in women who had received mammography screening. One Canadian registry-based study showed that screening attendees were more likely to have in-situ disease alone, and in those with invasive cancer, a lower proportion of grade III histology.[59] Furthermore, two studies using the SEER-Medicare database showed that extending mammography screening to elderly women decreased advanced stage at diagnosis.[60,61]

The interpretation of BC stage results from RCTs is precluded by stage migration bias due to the introduction of sentinel lymph node dissection[62] and by modifications in coding and classification practices.[63] Consequently, ecological studies have yielded conflicting results; for example, the incidence of BC stages II–IV has been reported to remain unchanged since the introduction of screening in the Netherlands.[64] To overcome these limitations, a systematic review suggested to use the primary tumour size as the most direct link to radiological detection;[65] their findings from observational studies suggested a reduction in BC advanced stages after the introduction of screening.[65]

We observed an increased risk of mastectomies, which has not been consistently described in population studies. One Canadian study found that mastectomies were less frequently performed in screening attendees,[59] while women diagnosed in a New Zealand screening programme were more likely to undergo conservative surgery;[66] similar results were observed in women aged 40–49 from the US.[67] One Norwegian study reported that mammography screening was associated with an increase in mastectomy rates, which later declined, likely explained by changes in recommended surgical approaches.[68] It is noteworthy that a recent systematic review found that adherence to guideline recommendations on breast-conserving surgery is highly variable (35–95%).[69] Thus, the increase in mastectomies among RCTs may partly be due to lead time bias, the progress in BC care, or variation in clinical practice.

Our estimates on chemotherapy are limited by changes in clinical practices, but recent observational studies suggest similar results. One study using Italian population cancer registries, from 2009 to 2013, observed that the neo-adjuvant therapy indication was lower in provinces where a screening programme had been present for many years.[70] Another study that identified women aged 40–79 with incident BC from the British Columbia Cancer Registry (Canada) found that chemotherapy use was lower among regular screening participants after adjustment for age.[71]

Our overdiagnosis estimates were in the range described in the literature for European screening programmes, which have been roughly estimated to range from 0 to 54% using unadjusted data and from 1 to 10% after adjustment for BC risk and lead time.[72] However, a proportion of the excess of incidence from the CNBSS trials occurred years after screening ceased in the intervention arm[24] and should not be considered as overdiagnosis. Overall, the certainty of the evidence of overdiagnosis was moderate, due to potential RoB, as women in the control group of the CNBSS trial might have received opportunistic or programmatic mammography screening at the end of the trial period.

There is no consensus about the method to estimate overdiagnosis. Most common approaches assess the difference in cancer incidence in the presence and absence of screening or make inferences about the lead time of BC.[73] One study observed that a long follow-up time is needed to account for lead time, as the excess of cumulative BC incidence will fall below 10% after a follow-up of 25 years in a simulated population.[74] Another study,

applying a micro-simulation model to the Netherlands population, found that estimations made in earlier phases of the screening programme may overestimate overdiagnosis by a factor of 4, underlining the relevance of allowing an appropriate follow-up time to obtain reliable estimates.[75]

There are discrepancies among previous systematic reviews on the assessment of RoB.[4,76] In particular, a Cochrane review considered that only the CNBSS, Malmo and UK Age trials were at low RoB, showing a non-significant effect of mammography screening on BC mortality from those trials.[4] We considered the CNBSS trial as high RoB due to allocation of women by using open lists, and the inclusion of a clinical examination before randomisation which could have led to differential assignment;[27,28] thus, only two RCTs were at low RoB with similar BC mortality effects when compared to the remaining studies. Consequently, in the 50–69 years strata, we did not downgrade our certainty for RoB due to the similarity in effect estimates across studies.

Some authors have proposed all-cause mortality as a better estimate of screening impact. This measure would be less prone to ascertainment bias of the cause of death, an issue described in the Swedish trials with higher all-cause mortality in the control group. However, authors of the Swedish trials[4,77] reported no significant increased rate of death from other causes after appropriate adjustments for age distribution and lead time bias were implemented.[18,47] Moreover, all-cause mortality would be an inefficient measure given the unfeasibly large sample size required to detect differences between groups.[47] We complemented our estimation of mortality impact with the results for other-cause mortality which suggested no difference between women invited or not to screening.

Balancing potential benefits and overdiagnosis, we estimate for 100,000 women invited to screening from age 50 to 69, at least 138 BC deaths would be avoided and 3240 BC would be diagnosed (2.7 per 1000 annual rate-$\times$ 20 years $\times$ 0.6 mammography adherence) of which 550 (17%, individual perspective) could be overdiagnosed. Thus, for each BC death avoided, approximately four overdiagnosed cases will be managed. This estimate is in the range of previous systematic assessments of screening.[78] However, the potential bias in the overdiagnosis estimates means that this figure remains tentative.

Limitations and strengths

Our systematic review has some limitations, as no RCTs have sufficient statistical power to assess the benefit of screening on BC mortality according to age subgroups. Additionally, we included only English language articles; however, the risk of selection bias is probably small because we screened previous systematic reviews, and the GDG includes several international experts, making the possibility of missing studies unlikely. Although our original search was conducted up to April 2016, we conducted a new search in June 2018, and after looking at the results, the GDG decided not to update the systematic review.

Our review also has strengths: we used rigorous methods including the GRADE approach to rate the certainty of the evidence and included the longest follow-up data available from the RCTs and systematic reviews of observational studies. In contrast to previous systematic reviews, the consideration of contextual evidence allowed us to rate the certainty of evidence for BC mortality as high for women aged 50–69 and 70–74 and moderate for women aged 45–49. We also provided results stratified by age groups of interest for women, clinicians and policymakers.

Conclusions

Our findings have different implications depending on the stakeholder group. Guideline panellists (and policy makers) are more likely to formulate strong recommendations in women in the 50–69 age group than in other groups. In women under 50 or over 69, where the balance is less clear, conditional recommendations are more likely. Moreover, panels may specify further subgroups among women below 50, where baseline risk changes rapidly and recommendations could vary between the 40–44 and 45–49 age groups, as in the ECIBC guidelines[13] (https://ecibc.jrc.ec.europa.eu/recommendations/). Although informed decision-making should be recommended in all age groups, this will be especially important in these age groups where the balance is less clear.

A number of research priorities were identified with input from the GDG experts, which included: assessing the impact of different screening intervals; the identification of risk factors to stratify women who should start screening earlier (or at shorter examination intervals, such as women with dense breast tissue); better assessment of the magnitude of overdiagnosis with an emphasis on methods to estimate the actual impact across age groups; and the use of new technologies for screening (i.e. tomosynthesis).

Availability of data and materials

All data sources used during this study are described in this published article and its additional information files. The datasets analysed are available from the corresponding author on reasonable request.

Acknowledgements

The authors would like to sincerely thank all members of the Guidelines Development Group of the European Commission Initiative on Breast Cancer for their participation in the discussions generated by this systematic review which led to the different recommendations they developed in the European Guidelines on Breast Cancer Screening and diagnosis (https://healthcare-quality.jrc.ec.europa.eu/european-breast-cancer-guidelines).

Authors' contributions

Carlos Canelo-Aybar, Diogenes Seraphim Ferreira, Monica Ballesteros, Margarita Posso, Nadia Montero and Ivan Solá, Pablo Alonso-Coello were responsible for conducting the systematic review. Monica Ballesteros, Margarita Posso, Nadia Montero and Ivan Solá conducted the search and data extraction. Paolo Giorgi Rossi, Stephen Duffy, Markus Follmann, Zuleika Saz-Parkinson, and Axel Gräwingholt contributed to the definition of the research protocol and provided comments to the preliminary results of the systematic review. Carlos Canelo-Aybar

and Pablo Alonso-Coello drafted the first version of the article. All authors contributed to the interpretation and reporting of the results and provided comments on subsequent versions of the article. All authors read and approved the final manuscript prior submission.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Zuleika Saz-Parkinson and Donata Lerda are current employees of the Joint Research Centre, European Commission. Carlos Canelo-Aybar, Diogenes Seraphim Ferreira, Monica Ballesteros, Margarita Posso, Nadia Montero, Ivan Solà and Pablo Alonso-Coello are employees of the Iberoamerican Cochrane Collaboration. Paolo Giorgi Rossi, Stephen Duffy, Markus Follmann, and Axel Gräwingholt are members of the ECIBC Guidelines Development Group. Diogenes Seraphim Ferreira was supported by the fellowship MTF 2015–02 from the European Respiratory Society during the conduct of the study.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The systematic review was carried out by Iberoamerican Cochrane Collaboration under the Framework contract 443094 for procurement of services between the European Commission's Joint Research Centre and Asociación Colaboración Cochrane Iberoamericana.

ORCID iDs

Zuleika Saz-Parkinson https://orcid.org/0000-0002-3518-5899
Paolo G Rossi https://orcid.org/0000-0001-9703-2460
Stephen W Duffy https://orcid.org/0000-0003-4901-7922
Axel Gräwingholt https://orcid.org/0000-0003-4778-8731

Supplemental material

Supplemental material for this article is available online.

References

1. Ferlay JE, Lam F, Colombet M, et al. *Global cancer observatory: cancer today*. Lyon, France: International Agency for Research on Cancer, https://gco.iarc.fr/today (2018, accessed 25 July 2019).
2. ECIS. European Cancer Information System, https://ecis.jrc.ec.europa.eu (2018, accessed 20 July 2019).
3. Broeders M, Moss S, Nystrom L, EUROSCREEN Working Group, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen* 2012; 19 (Suppl 1): 14–25.
4. Gotzsche PC and Jorgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2013; CD001877.
5. Shapiro S. Screening: assessment of current studies. *Cancer* 1994; 74: 231–238.
6. Hubbard RA, Kerlikowske K, Flowers CI, et al. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. *Ann Intern Med* 2011; 155: 481–492.
7. Marmot MG, Altman DG, Cameron DA, et al. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 2013; 108: 2205–2240.
8. *WHO position paper on mammography screening*. Geneva: World health Organisation, 2014.
9. Tonelli M, Gorber SC, Joffres M, et al. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. *Canad Med Assoc J* 2011; 183: 1991–2001.
10. Oeffinger KC, Fontham ET, Etzioni R, American Cancer Society, et al. Breast cancer screening for women at average risk: 2015 guideline update from the American cancer society. *JAMA* 2015; 314: 1599–1614.
11. Schunemann HJ, Lerda D, Dimitrova N, for the European Commission Initiative on Breast Cancer Contributor Group, et al. Methods for development of the European commission initiative on breast cancer guidelines: recommendations in the era of guideline transparency. *Ann Intern Med* 2019; 171: 273.
12. (ECIBC) ECIOBC. Recommendations from European Breast Guidelines, https://ecibc.jrc.ec.europa.eu/recommendations/list/3 (2018, accessed 29 March 2018).
13. Schunemann HL, Quinn C, Follmann M, for the European Commission Initiative on Breast Cancer (ECIBC) Contributor Group, et al. Breast cancer screening and diagnosis: a synopsis of the European breast guidelines. *Ann Intern Med* 2020; 172: 46.
14. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011; 64: 395–400.
15. Higgins JP, Altman DG, Gotzsche PC, Cochrane Statistical Methods Group, et al. The cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; 343: d5928.
16. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009; 62: 1013–1020.
17. Nystrom L, Larsson LG, Rutqvist LE, et al. Determination of cause of death among breast cancer cases in the Swedish randomized mammography screening trials. A comparison between official statistics and validation by an endpoint committee. *Acta Oncol* 1995; 34: 145–152.
18. Nyström L, Andersson I, Bjurstam N, et al. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* 2002; 359: 909–919.
19. Guyatt GH, Oxman AD, Schunemann HJ, et al. GRADE guidelines: a new series of articles in the journal of clinical epidemiology. *J Clin Epidemiol* 2011; 64: 380–382.
20. Guyatt GH, Oxman AD, Vist GE, GRADE Working Group, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336: 924–926.
21. Roberts MM, Alexander FE, Anderson TJ, et al. The Edinburgh randomised trial of screening for breast cancer: description of method. *Br J Cancer* 1984; 50: 1–6.
22. Alexander FE, Anderson TJ, Brown HK, et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet* 1999; 353: 1903–1908.
23. Chu KC, Smart CR and Tarone RE. Analysis of breast cancer mortality and stage distribution by age for the health insurance plan clinical trial. *J Natl Cancer Inst* 1988; 80: 1125–1132.
24. Habbema JD, van Oortmarssen GJ, van Putten DJ, et al. Age-specific reduction in breast cancer mortality by screening: an analysis of the results of the health insurance plan of greater New York study. *J Natl Cancer Inst* 1986; 77: 317–320.
25. Shapiro S, Strax P, Venet L, et al. Proceedings: changes in 5-year breast cancer mortality in a breast cancer screening program. *Proc Natl Cancer Conf* 1972; 7: 663–678.
26. Shapiro S. Periodic screening for breast cancer: the HIP randomized controlled trial. Health insurance plan. *J Natl Cancer Inst Monogr* 1997; 22: 27–30.
27. Miller AB, To T, Baines CJ, et al. Canadian national breast screening study-2: 13-year results of a randomized trial in women aged 50-59 years. *J Natl Cancer Inst* 2000; 92: 1490–1499.
28. Miller AB, To T, Baines CJ, et al. The Canadian national breast screening study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med* 2002; 137: 305–312.
29. Miller AB, Wall C, Baines CJ, et al. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian national breast screening study: randomised screening trial. *BMJ* 2014; 348: g366.
30. Miller AB, Baines CJ, To T, et al. Canadian national breast screening study. 1. Breast cancer detection and death rates among women aged 40 to 49 years. *Canad Med Assoc J* 1992; 147: 1459–1476.
31. Miller AB. The costs and benefits of breast cancer screening. *Am J Prev Med* 1993; 9: 175–180.
32. Miller AB, Baines CJ, To T, et al. Canadian national breast screening study. 2. Breast cancer detection and death rates among women aged 50 to 59 years. *Canad Med Assoc J* 1992; 147: 1477–1488.
33. Moss S. A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40. Trial steering group. *J Med Screen* 1999; 6: 144–148.
34. Moss S, Waller M, Anderson TJ, Trial Management Group, et al. Randomised controlled trial of mammographic screening in women from age 40: predicted mortality based on surrogate outcome measures. *Br J Cancer* 2005; 92: 955–960.
35. Moss SM, Cuckle H, Evans A, Trial Management Group, et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet* 2006; 368: 2053–2060.
36. Frisell J, Glas U, Hellstrom L, et al. Randomized mammographic screening for breast cancer in Stockholm. Design, first round results and comparisons. *Breast Cancer Res Treat* 1986; 8: 45–54.
37. Frisell J, Lidbrink E, Hellstrom L, et al. Followup after 11 years – update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Res Treat* 1997; 45: 263–270.
38. Frisell J. *Mammographic screening for breast cancer*. PhD Thesis. Stockholm: Södersjukhuset, 1989.

39. Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the malmo mammographic screening trial. *BMJ* 1988; 297: 943–948.
40. Zackrisson S, Andersson I, Janzon L, et al. Rate of over-diagnosis of breast cancer 15 years after end of malmo mammographic screening trial: follow-up study. *BMJ* 2006; 332: 689–692.
41. Andersson I and Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the malmo mammographic screening program. *J Natl Cancer Inst Monogr* 1997; 22: 63–67.
42. Bjurstam N, Björneld L, Duffy SW, et al. The Gothenburg breast screening trial: first results on mortality, incidence, and mode of detection for women ages 39-49 years at randomization. *Cancer* 1997; 80: 2091–2099.
43. Bjurstam N, Bjorneld L, Warwick J, et al. The Gothenburg breast screening trial. *Cancer* 2003; 97: 2387–2396.
44. Tabar L, Fagerberg G, Duffy SW, et al. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Community Health* 1989; 43: 107–114.
45. Tabar L, Fagerberg G, Chen HH, et al. Efficacy of breast cancer screening by age. New results from the Swedish Two-County trial. *Cancer* 1995; 75: 2507–2517.
46. Tabar L, Chen HH, Duffy SW, et al. Primary and adjuvant therapy, prognostic factors and survival in 1053 breast cancers diagnosed in a trial of mammography screening. *Jpn J Clin Oncol* 1999; 29: 608–616.
47. Tabar L, Duffy SW, Yen MF, et al. All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point. *J Med Screen* 2002; 9: 159–162.
48. Tabar L, Vitak B, Chen TH, et al. Swedish Two-County trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology* 2011; 260: 658–663.
49. Tabar L, Vitak B, Chen HH, et al. The Swedish Two-County trial twenty years later. Updated mortality results and new insights from long-term follow-up. *Radiol Clin N Am* 2000; 38: 625–651.
50. Moss SM, Wale C, Smith R, et al. 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. *Lancet Oncol* 2015; 16: 1123–1132.
51. Bjurstam NG, Bjorneld LM and Duffy SW. Updated results of the Gothenburg trial of mammographic screening. *Cancer* 2016; 122: 1832–1835.
52. Bond M, Pavey T, Welch K, et al. Psychological consequences of false-positive screening mammograms in the UK. *Evid Based Med* 2013; 18: 54–61.
53. Brett J, Bankhead C, Henderson B, et al. The psychological impact of mammographic screening. A systematic review. *Psychooncology* 2005; 14: 917–938.
54. Hofvind S, Ponti A, Patnick J, EUNICE Project and Euroscreen Working Groups, et al. False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. *J Med Screen* 2012; 19 (Suppl 1): 57–66.
55. Salz T, Richman AR and Brewer NT. Meta-analyses of the effect of false-positive mammograms on generic and specific psychosocial outcomes. *Psychooncology* 2010; 19: 1026–1034.
56. Hofvind S, Thoresen S and Tretli S. The cumulative risk of a false-positive recall in the Norwegian breast cancer screening program. *Cancer* 2004; 101: 1501–1507.
57. Salas D, Ibanez J, Roman R, CFPR (Cumulative False Positive Risk) group, et al. Effect of start age of breast cancer screening mammography on the risk of false-positive results. *Prev Med* 2011; 53: 76–81.
58. Njor S, Nystrom L, Moss S, Euroscreen Working Group, et al. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. *J Med Screen* 2012; 19 (Suppl 1): 33–41.
59. Olivotto IA, Mates D, Kan L, et al. Prognosis, treatment, and recurrence of breast cancer for women attending or not attending the screening mammography program of British Columbia. *Breast Cancer Res Treat* 1999; 54: 73–81.
60. Randolph WM, Goodwin JS, Mahnken JD, et al. Regular mammography use is associated with elimination of age-related disparities in size and stage of breast cancer at diagnosis. *Ann Intern Med* 2002; 137: 783–790.
61. McCarthy EP, Burns RB, Freund KM, et al. Mammography use, breast cancer stage at diagnosis, and survival among older women. *J Am Geriatr Soc* 2000; 48: 1226–1233.
62. Maaskant AJ, van de Poll-Franse LV, Voogd AC, et al. Stage migration due to introduction of the sentinel node procedure: a population-based study. *Breast Cancer Res Treat* 2009; 113: 173–179.
63. Larsen IK, Myklebust TA, Johannesen TB, et al. Stage-specific incidence and survival of breast cancer in Norway: the implications of changes in coding and classification practice. *Breast* 2018; 38: 107–113.
64. Autier P, Boniol M, Koechlin A, et al. Effectiveness of and overdiagnosis from mammography screening in the Netherlands: population based study. *BMJ* 2017; 359: j5224.
65. Broeders MJM, Allgood P, Duffy SW, et al. The impact of mammography screening programmes on incidence of advanced breast cancer in Europe: a literature review. *BMC Cancer* 2018; 18: 860.
66. Samnakay N, Tinning J, Ives A, et al. Rates for mastectomy are lower in women attending a breast-screening programme. *ANZ J Surg* 2005; 75: 936–939.
67. James TA, Wade JE and Sprague BL. The impact of mammographic screening on the surgical management of breast cancer. *J Surg Oncol* 2016; 113: 496–500.
68. Suhrke P, Mæhlen J, Schlichting E, et al. Effect of mammography screening on surgical treatment for breast cancer in Norway: comparative analysis of cancer registry data. *BMJ* 2011; 343: d4692.
69. Niño de Guzmán E, Song Y, Alonso-Coello P, et al. Healthcare providers' adherence to breast cancer guidelines in Europe: a systematic literature review. *Breast Cancer Res Treat* 2020; 181: 499–518.
70. Mangone L, Mancuso P, Tagliabue G, et al. Neoadjuvant therapy for breast cancer. *Tumori* 2019; 105: 488–493.
71. Coldman AJ, Phillips N and Speers C. A retrospective study of the effect of participation in screening mammography on the use of chemotherapy and breast conserving surgery. *Int J Cancer* 2007; 120: 2185–2190.
72. Puliti D, Duffy SW, Miccinesi G, EUROSCREEN Working Group, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen* 2012; 19 (Suppl 1): 42–56.
73. Nelson HD, Pappas M, Cantor A, et al. Harms of breast cancer screening: systematic review to update the 2009 U.S. Preventive services task force recommendation. *Ann Intern Med* 2016; 164: 256–267.
74. Duffy SW and Parmar D. Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time. *Breast Cancer Res* 2013; 15: R41.
75. de Gelder R, Heijnsdijk EA, van Ravesteyn NT, et al. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* 2011; 33: 111–121.
76. Nelson HD, Fu R, Cantor A, et al. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 U.S. preventive services task force recommendation. *Ann Intern Med* 2016; 164: 244–255.
77. Black WC, Haggstrom DA and Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst* 2002; 94: 167–173.
78. *The Research Council of Norway: research-based evaluation of the Norwegian breast cancer screening program*. Lysaker: Division for Society and Health, 2015.
79. Hellquist BN, Duffy SW, Abdsaleh S, et al. Effectiveness of population-based service screening with mammography for women ages 40 to 49 years: evaluation of the Swedish Mammography Screening in Young Women (SCRY) cohort. *Cancer* 2011; 117: 714–722.