OXFORD

Impact of Full-Field Digital Mammography Versus Film-Screen Mammography in Population Screening: A Meta-Analysis

Rachel Farber, MPH (iD), * Nehmat Houssami, PhD (iD), Sally Wortley, PhD, Gemma Jacklyn, PhD (iD), Michael L. Marinovich, PhD (iD), Kevin McGeechan, PhD, Alexandra Barratt, PhD, Katy Bell, PhD (iD)

Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, New South Wales, Australia

*Correspondence to: Rachel Farber, MPH, Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, New South Wales, Australia (e-mail: rachel.farber@sydney.edu.au).

Abstract

Background: Breast screening programs replaced film mammography with digital mammography, and the effects of this practice shift in population screening on health outcomes can be measured through examination of cancer detection and interval cancer rates. **Methods:** A systematic review and random effects meta-analysis were undertaken. Seven databases were searched for publications that compared film with digital mammography within the same population of asymptomatic women and reported cancer detection and/or interval cancer rates. **Results:** The analysis included 24 studies with 16 583 743 screening examinations (10 968 843 film and 5 614 900 digital). The pooled difference in the cancer detection rate showed an increase of 0.51 per 1000 screens (95% confidence interval [CI] = 0.19 to 0.83), greater relative increase for ductal carcinoma in situ (25.2%, 95% CI = 17.4% to 33.5%) than invasive (4%, 95% CI = −3% to 13%), and a recall rate increase of 6.95 (95% CI = 3.47 to 10.42) per 1000 screens after the transition from film to digital mammography. Seven studies (80.8% of screens) reported interval cancers: the pooled difference showed no change in the interval cancer rate with −0.02 per 1000 screens (95% CI = −0.06 to 0.03). Restricting analysis to studies at low risk of bias resulted in findings consistent with the overall pooled results for all outcomes. **Conclusions:** The increase in cancer detection following the practice shift to digital mammography did not translate into a reduction in the interval cancer rate. Recall rates were increased. These results suggest the transition from film to digital mammography did not result in health benefits for screened women. This analysis reinforces the need to carefully evaluate effects of future changes in technology, such as tomosynthesis, to ensure new technology leads to improved health outcomes and beyond technical gains.

Population mammography-screening programs aim to prevent women dying from breast cancer through earlier detection and treatment (1). Improvements in mammographic technology hold the promise of increasing the benefit of screening by increasing the detection of clinically important breast cancers. However, determining whether such theoretical benefits are realized or not is challenging outside the context of randomized controlled trials with long follow-up to assess for reductions in morbidity and mortality. Such trials are impractical to conduct each time an incremental technological change is made.

Population-based observational studies, in contrast, may enable rapid, real-world evaluation of how changes in screening technologies affect health outcomes (2). A change in technology that increases cancer detection rates (CDR) may reduce subsequent interval cancer rates if the additional cancers detected would have otherwise presented clinically at a more advanced

stage (3, 4). This pattern suggests more benefit from screening due to reduced underdiagnosis. In contrast, an increase in cancer detection rate without a decrease in interval cancer rate would suggest more overdetection and therefore less net benefit of screening. Thus, considering how much interval cancer rates decrease, as well as how much cancer detection rates increase, after a change in technology provides a rapid and efficient assessment of the likely effectiveness of a change in screening technology (5).

A new screening technology should be at minimum as efficient and ideally have an improved effect on health outcomes as the existing technology. A change in a screening program should aim to further reduce morbidity or harm without unduly increasing the diagnosis of cancers that would otherwise never have caused morbidity or mortality (6). The introduction of a new technology needs to prudently determine the effect that

the change in modality will have on detecting biologically relevant cancers (7).

Most breast screening programs worldwide have replaced screen film mammography with full field digital mammography because of technical and practical advantages. These advantages include enabling images to be stored and transmitted electronically. Although these advantages have driven widespread replacement of film mammography with digital mammography, the absence of long-term follow-up studies assessing the impact of the change on breast cancer morbidity and mortality means that the impact on health outcomes is unclear. A systematic consideration of effects of the switch to digital mammography on cancer detection rate and interval cancer rate provides a means to address this evidence gap and is the aim of this systematic review. Furthermore, our analysis provides a blueprint for more rapid and efficient assessment of future technological innovation, such as practice change to 3-dimensional or breast tomosynthesis.

Before the rollout of digital mammography, early studies did not show any overall difference in cancer detection rate between the 2 types of mammography technologies (5, 8). However, they did show increased detection in women who had dense breasts, were younger than 50 years, or were pre- or perimenopausal (9). Following the transition to digital mammography, studies from several real-world screening programs have reported conflicting and heterogeneous results on the effect of this change on cancer detection rate and interval cancer rate (10–13). Furthermore, some studies with a higher cancer detection rate during digital mammography found the increase is mostly for ductal carcinoma in situ (DCIS) (13–15). In addition to changes in the cancer detection rate, there have been inconsistent findings on recall rates (14–16).

This systematic review and meta-analysis (17) aimed to summarize, to our knowledge for the first time, all the available evidence on cancer detection rates, interval rates, and recall rates for digital mammography compared with film mammography for population breast cancer screening. In doing so, the review provides important new information for current breast cancer screening practice and demonstrates an approach to the acquisition of evidence to inform decisions on the adoption of future mammography technologies (18).

Methods

Search Strategy and Selection Criteria

We conducted a systematic review and meta-analysis of the medical literature, comparing digital to film mammography for screening asymptomatic women who are at average risk of breast cancer within the same population. We searched MEDLINE, PREMEDLINE, PubMed, Embase, National Health Service Economic Evaluation Database, Database of Abstracts of Reviews of Effects, American College of Physicians Journal Club, Cochrane Central Register of Controlled Trials, and Cochrane Database of Systematic Reviews and followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for reporting (19, 20). Search terms were broad to encompass all studies comparing film and digital mammography using key words "digital*", "film*", and "mammogra*" or the Medical Subject Headings term "Mammography". The search did not have any date, study type, or language restrictions.

Using prespecified inclusion criteria (supplementary material), decisions regarding inclusion and exclusion of studies

were made independently by 2 researchers (R.F. and S.W.); any disagreement or uncertainty was discussed and resolved with a third researcher (K.B.). Titles and abstracts of all studies identified from the databases searched were reviewed together with a review of articles that cited, or were citations of, the included articles. Articles were excluded at the title and abstract review stage (R.F. and M.L.M.) if the study did not look at both film and digital mammography, were investigating women at high risk of breast cancer, or were a review, editorial, or commentary. Articles were excluded at the full-text review stage (R.F. and S.W.) if they did not provide counts for screen-detected breast cancers or if cancer diagnosis was not verified with histopathology. Risk of bias was measured by 2 authors using the ROBINS-I tool for all included studies (R.F. and G.J.). The ROBINS-I tool is designed to compare the level of bias that would occur in a theoretical, high-quality, randomized trial with nonrandomized interventional cohorts or observational studies (20). The PROSPERO systematic review protocol is available online and summarized in Supplementary Figure 1 available at: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD4201707060119 (21).

Statistical Analysis

Two researchers independently extracted data from included studies using a standardized template (R.F. and S.W.). Where 2 or more studies had cohorts with populations that overlapped by 20% or more, we chose the one with highest number of screenings, longest time, and/or were most recent. Comparisons between the 2 types of mammography were limited to within-study to limit confounding from differing background rates and practices between different populations. Studies that measured computed radiography in addition to digital mammography were included, but the data relating to computed radiography were not extracted.

The primary outcomes were differences in cancer detection rate and interval cancer rate. We calculated point estimates and 95% confidence intervals (CI) for each study to ensure consistent calculation of risk differences between the 2 types of mammography. Screen-detected breast cancers were defined as breast cancers diagnosed as a result of a positive screening result, and cancer detection rate was calculated as the number of women diagnosed with screen-detected breast cancer (including DCIS) per 1000 screens. Interval cancers were defined as cancers diagnosed after a negative screening result and before the subsequent scheduled screening episode; interval cancer rate was calculated as the number of interval cancers (including DCIS) per 1000 screened women. Secondary outcomes included recall rates and false positive rates. Recall rates were defined as the number of women with a positive screening examination per 1000 screened women, and false positive recall rates were calculated as the recall rate minus cancer detection rate.

We used random effects meta-analysis utilizing the DerSimonian-Laird method to pool risk differences and relative risks for each outcome for digital vs film mammography. We assessed heterogeneity between studies using χ^2 and I^2 statistics and visual inspection of forest plots (METAN in Stata/IC 15.1). We conducted subgroup analyses by screening frequency, age, breast density (Breast Imaging Reporting and Data System I + II vs III+ IV), and initial vs subsequent screening round, where outcomes were reported for these variables. We used generalized linear mixed models with a logit link using the Wald test to check for interactions between the type of mammography used
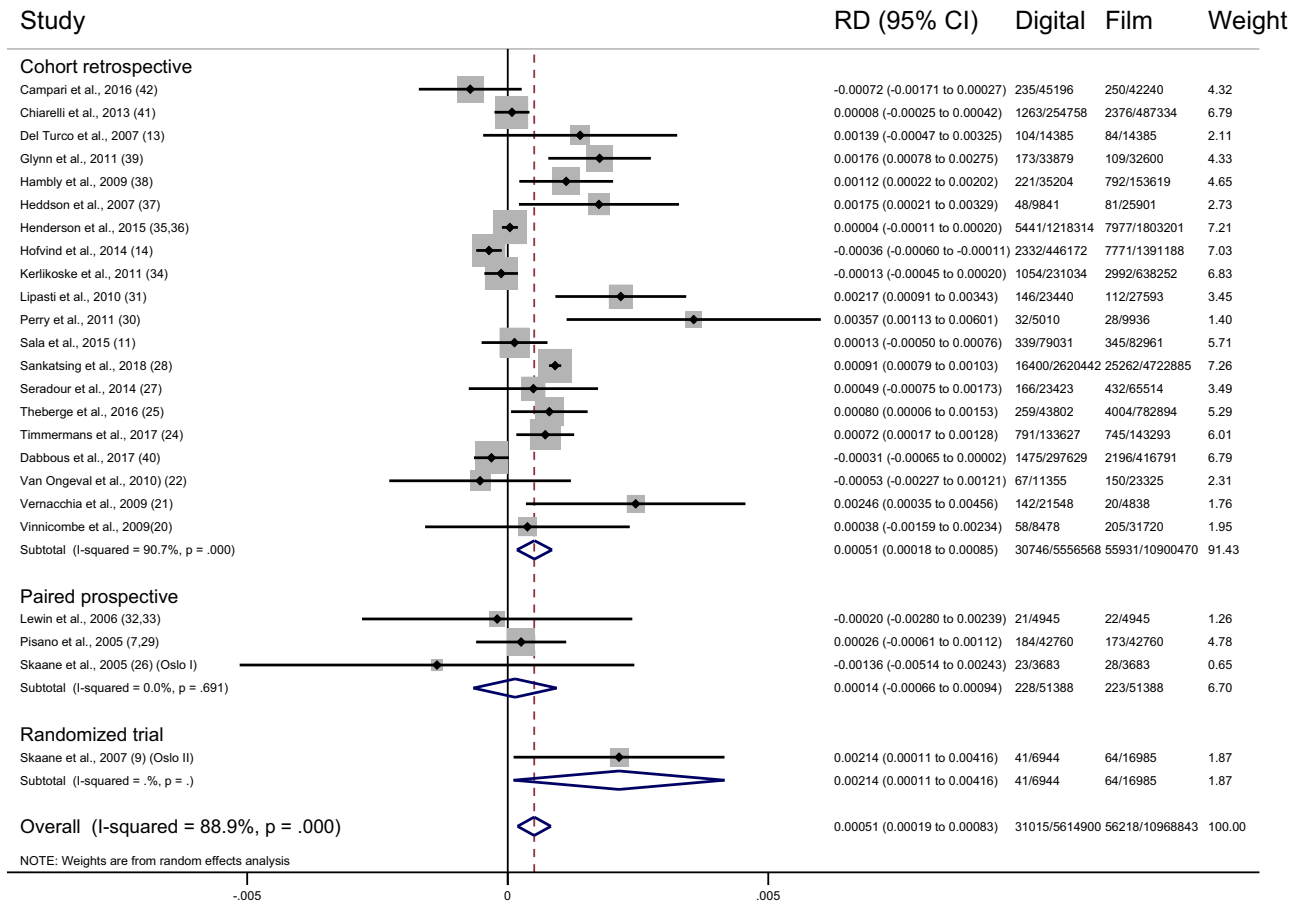
Figure 1. Forest plot of cancer detection rates by study type. CI = confidence interval; RD = risk difference.

and each subgroup to account for each study contributing measurements for each modality (PROC GLIMMIX in SAS 9.4). Because not all studies reported all outcomes, besides cancer detection, we conducted sensitivity analyses of the pooled cancer detection rate differences for the subsets of studies that reported each of these different outcomes or subgroups (DCIS/invasive, recalls, interval cancers, screening round, age, density, and by risk of bias).

Results

The database search retrieved 1030 publications, and after screening of titles and abstracts, 177 full-text articles were screened. Twenty-nine articles met the eligibility criteria and reported on 24 individual studies (Supplementary Figure 1, available online) (9, 11–13, 15, 16, 22–44). The included studies, from 12 different countries (Supplementary Table 1, available online), had a total of 16 583 743 screening examinations. Of the 10 968 843 film-screens, 56 218 breast cancers were detected, and 31 015 cancers were detected of the 5 614 900 digital screens (Table 1). In the 18 studies measuring DCIS as an outcome, of the 87 233 cancers detected, 14 706 were DCIS and 63 726 were invasive cancers. Three of the studies used a paired design, 1 was a randomized trial, and the remaining were retrospective observational studies.

Based on the ROBINS-I tool, the risk of bias ranged from moderate to critical; study quality is summarized in Supplementary Table 2 (available online). Studies were generally assessed to be at moderate to serious risk of bias, primarily due to the potential for confounders to change over time and affect the background cancer rates in the populations under study and apparent differences in outcomes between screening modalities. Studies that received a low rating for risk of bias (ie, least likely to be affected by bias) had either a paired or randomized design. The studies that received a moderate rating, measured exposures concurrently for the majority of the study period and/or measured all important potential confounders, and found the unadjusted and adjusted rates compared in the original study did not greatly differ; all fell within the 95% confidence intervals. Studies with a serious risk of bias compared some important potential confounders but either had long periods of nonconcurrence for use of the different screening modalities, with greater potential for the confounders to change over time, or found the confounders measured to vary between exposures. The studies found to have critical risk of bias did not measure enough potential confounders to be able to determine the effect confounding may have had on the results. All studies that measured interval cancer rates were found to be of either moderate or serious risk of bias. Using the ROBIN-I tool, only prospective studies were found to be of low risk of bias.

Table 1. Summary of results[a]

Outcome	No. of studies	Total No. of screenings	Total No. of screenings: film	Total No. of screenings: digital	Total No. of detected	Total No. of detected: film	Total No. of detected: digital	Risk difference per 1000 screens (95% CI)	Relative risk (95% CI)
Screen-detected cancer rate	24	16 583 743	10 968 843	5 614 900	87 233	56 218	31 015	0.51 (0.19 to 0.83)	1.10 (1.04 to 1.17)
Screen-detected DCIS	18	14 841 193	9 838 833	5 002 360	14 706	8925	5781	0.22 (0.15 to 0.30)	1.25 (1.17 to 1.33)
Screen-detected invasive	18	14 841 193	9 838 833	5 002 360	63 726	41 691	22 035	0.19 (−0.14 to 0.51)	1.04 (0.97 to 1.13)
Recalls	21	11 978 522	8 111 599	3 867 923	381 290	261 652	119 638	6.95 (3.47 to 10.42)	1.12 (1.02 to 1.23)
False positives	21	11 978 522	8 110 599	3 867 923	311 254	216 278	94 976	6.33 (3.24 to 9.42)	1.10 (0.99 to 1.23)
Interval cancers	7	13 048 505	8 571 434	4 477 071	23 353	15 479	7874	−0.02 (−0.06 to 0.03)	1.00 (0.96 to 1.03)
Interval DCIS	4	5 877 887	3 748 603	1 770 654	538	358	181	0.01 (−0.01 to 0.02)	1.08 (0.89 to 1.30)
Interval invasive	4	5 877 887	3 748 603	1 770 654	6373	4554	1819	−0.03 (−0.1 to 0.17)	1.01 (0.92 to 1.12)
Screen-detected cancer rate in studies with interval cancers	7	13 407 135	8 647 847	4 759 288	71 147	44 540	26 607	0.33 (−0.16 to 0.82)	1.06 (0.97 to 1.17)
Round	11	11 499 199	8 112 651	3 390 231	58 136	40 381	17 755	0.37 (0.02 to 0.71)	1.07 (1.00 to 1.14)
Initial	11	1 747 513	1 290 864	460 332	10 447	7558	2889	0.61 (−0.04 to 1.27)	1.11 (0.99 to 1.25)
Subsequent	8	9 751 686	6 821 787	2 929 899	47 689	32 823	14 866	0.18 (−0.26 to 0.61)	1.03 (0.95 to 1.13)
Age, y	16	7 555 361	5 048 405	2 534 770	37 199	25 318	11 881	0.26 (−0.03 to 0.56)	1.06 (0.99 to 1.12)
<50	7	1 159 221	681 959	483 464	3257	1846	1411	0.28 (−0.26 to 0.83)	1.12 (0.92 to 1.36)
50–69	16	5 757 548	3 972 558	1 806 602	29 341	20 635	8706	0.34 (−0.06 to 0.74)	1.07 (0.99 to 1.16)
70+	4	638 592	393 888	244 704	4601	2837	1764	0.03 (−0.4 to 0.45)	1.00 (0.94 to 1.06)
Density	4	3 262 666	1 889 853	1 372 813	14 670	8511	6159	−0.02 (−0.38 to 0.35)	1.01 (0.93 to 1.09)
BI-RAD I+II	4	1 747 192	1 024 643	722 549	7542	4392	3150	−0.06 (−0.05 to 0.42)	0.99 (0.89 to 1.09)
BI-RAD III+IV	4	1 515 474	865 210	650 264	7128	4119	3009	0.21 (−0.51 to 0.92)	1.08 (0.93 to 1.26)

[a] BI-RAD = Breast Imaging Reporting and Data System; CI = confidence interval; DCIS = ductal carcinoma in situ.

Table 2. Number of cancers detected, CDRs, and difference in CDRs

| Study | | Digital | | | Film | | | |
Author	Country	Screenings	Cancers[a]	CDR/1000	Screenings	Cancers[a]	CDR/1000	Difference in CDR/1000
Campari et al., 2016 (43)	Italy	45 196	235	5.2	42 240	250	5.92	−0.72
Chiarelli et al., 2013 (42) and Prummel et al., 2016 (10)	Canada	254 758	1263	4.96	487 334	2376	4.88	0.08
Dabbous et al., 2017 (41)	United States	297 629	1475	4.96	416 791	2196	5.27	−0.31
Del Turco et al., 2007 (13)	Italy	14 385	104	7.23	14 385	84	5.84	1.39
Glynn et al., 2011 (40)	United States	33 879	173	5.11	32 600	109	3.34	1.76
Hambly et al., 2009 (39)	Ireland	35 204	221	6.28	153 619	792	5.16	1.12
Heddson et al., 2007 (38)	Sweden	9841	48	4.88	25 901	81	3.13	1.75
Henderson et al., 2015 (36, 37)	United States	1 218 314	5441	4.47	1 803 201	7977	4.42	0.04
Hofvind et al., 2014 (14)	Norway	446 172	2332	5.23	1 391 188	7771	5.59	−0.36
Kerlikoske et al., 2011 (35)	United States	231 034	1054	4.56	638 252	2992	4.69	−0.13
Lewin et al., 2006 (33, 34)	United States	4945	21	4.25	4945	22	4.45	−0.2
Lipasti et al., 2010 (32)	Finland	23 440	146	6.23	27 593	112	4.06	2.17
Perry et al., 2011 (31)	UK	5010	32	6.39	9936	28	2.82	3.57
Pisano et al., 2005 (7, 30)	United States	42 760	184	4.3	42 760	173	4.05	0.26
Sala et al., 2015 (11)	Spain	79 031	339	4.29	82 961	345	4.16	0.13
Sankatsing et al., 2018 (29)	Netherlands	2 620 442	16 400	6.26	4 722 885	25 262	5.35	0.91
Seradour et al., 2014 (28)	France	23 423	166	7.09	65 514	432	6.59	0.49
Skaane et al., 2005 (27) (Oslo I)	Norway	3683	23	6.24	3683	28	7.6	−1.36
Skaane et al., 2007 (9) (Oslo II)	Norway	6944	41	5.9	16 985	64	3.77	2.14
Theberge et al., 2016 (26)	Canada	43 802	259	5.91	782 894	4004	5.11	0.80
Timmermans et al., 2017 (25)	Belgium	133 627	791	5.92	143 293	745	5.2	0.72
Van Ongeval et al., 2010) (23)	Belgium	11 355	67	5.9	23 325	150	6.43	−0.53
Vernacchia et al., 2009 (22)	United States	21 548	142	6.59	4838	20	4.13	2.46
Vinnicombe et al., 2009 (20)	UK	8478	58	6.84	31 720	205	6.46	0.38

[a] Cancer rates include DCIS and invasive. CDR = cancer detection rate; DCIS = ductal carcinoma in situ.

Overall, the pooled difference in cancer detection rate showed an increase of 0.51 per 1000 screens (95% CI = 0.19 to 0.83, I^2 = 88.9%; 24 studies) after the transition from film to digital mammography. This represents a relative increase of 10% (95% CI = 4% to 17%, I^2 = 88.0%; 24 studies) (Figure 1). In the studies reporting DCIS, there was an increase in cancer detection rate for DCIS of 0.22 (95% CI = 0.15 to 0.30, I^2 = 41.4%; 18 studies) per 1000 and 0.19 (95% CI = −0.14 to 0.51, I^2 = 89.7%; 18 studies) per 1000 screens for invasive breast cancer. The estimated change in the cancer detection rate following transition to digital was more evident for DCIS than invasive cancer, with a relative increase in cancer detection rate of 25% (95% CI = 17% to 33%, I^2 = 33.1%; 18 studies) for DCIS and 4% (95% CI = −3% to 13%, I^2 = 88.7%; 18 studies) for invasive cancers (Table 1) (9, 11–13, 15, 16, 22, 24–27, 29–31, 34, 35, 37–40, 43, 44).

The difference in the cancer detection rate in the studies with low risk of bias was 0.14 (95% CI = −0.66 to 0.94, I^2 = 0.0%; 3 studies). The differences in the cancer detection rate for the studies of moderate, serious, and critical risk of bias were 0.35 (95% CI = 0.01 to 0.68, I^2 = 65.1%; 9 studies), 0.35 (95% CI = −0.29 to 1.00, I^2 = 93.0%; 9 studies), and 1.27 (95% CI = −0.24 to 2.78, I^2 = 90.5%; 4 studies), respectively (Figure 4).

Even though only 7 of the 24 included studies reported both cancer detection and interval cancer rates, these studies make up 80.8% of the total screens (8 647 847 film and 459 288 digital). In these 7 studies, there were 71 147 screen-detected cancers (44 540 film and 26 607 digital) and 23 353 interval cancers (15 479 film and 7874 digital). These studies show a pooled increase in cancer detection rate of 0.33 per 1000 screens (95% CI = −0.16 to 0.82, I^2 = 95.7%; 7 studies) after the transition from film to digital mammography and no change in interval cancer rates of −0.02 per 1000 screens (95% CI = −0.06 to 0.03, I^2 = 0.0%; 7 studies) (Figure 2; Table 3) (11–13, 15, 26, 30, 38, 42, 43). There appeared to be no pattern in interval cancer rates between studies according to their risk of bias.

Of the studies reporting data for recalls, the pooled difference in recalls showed an increase of 6.95 (95% CI = 3.47 to 10.42, I^2 = 99.1%; 21 studies) per 1000 screens after the transition to digital mammography, due mostly to an increase in false-positive results (6.33/1000, 95% CI = 3.24 to 9.42, I^2 = 99.1%; 21 studies) (Figure 3; Table 4) (9, 11–13, 15, 16, 22–35, 37, 39–41, 43, 44). There appeared to be no pattern in recall rates between studies according to their risk of bias.

There was a large amount of heterogeneity in cancer detection rate between studies (I^2 = 88.9%; P < .001), and we explored possible reasons for this in the subgroup analysis. There was no evidence that screening round, age group, or breast density modified the effect of the transition on cancer detection rate. Fourteen studies from Europe and Canada had a screening interval of 2 years. In these studies, the increase in cancer detection rate was 0.61 (95% CI = 0.16 to 1.06, I^2 = 89.3%; 14 studies) compared with 0.20 (95% CI = −0.23 to 0.63, I^2 = 81.6%; 5 studies) in the studies from the United States where a screening interval of 1 year was the usual practice.

There was no strong evidence of an interaction between type of screening technology and screening round, density, or age in mixed models of cancer detection rate. The transition to digital mammography was associated with 0.53 (95% CI = −0.36
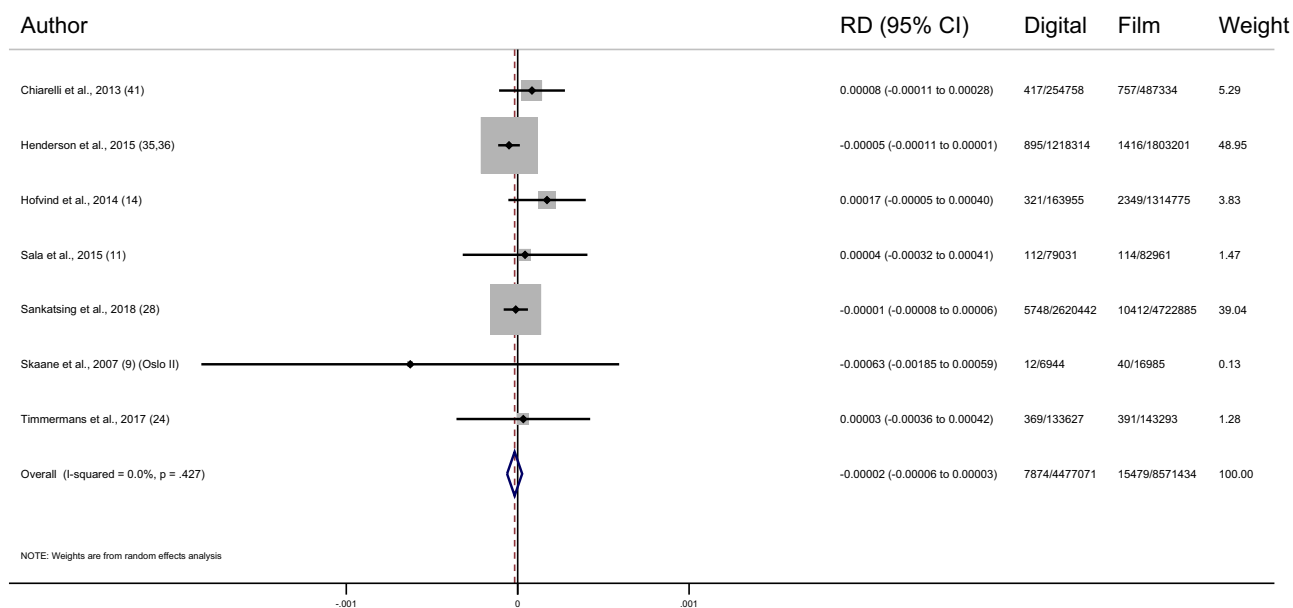
Figure 2. Forest plot of interval cancer rates. CI = confidence interval; RD = risk difference.

Table 3. Number of interval cancers, interval cancer rates, and difference in interval cancer rates

Study		Digital			Film			Difference in
Author	Country	Screenings	Interval cancers[a]	Interval cancers/1000	Screenings	Interval cancers[a]	Interval cancers/1000	interval cancers/1000
Chiarelli et al., 2013 (43) and Prummel et al., 2016 (12)	Canada	254 758	417	1.64	487 334	757	1.55	0.08
Henderson et al., 2015 (37, 38)	United States	1 218 314	895	0.73	1 803 201	1416	0.79	−0.05
Hofvind et al., 2014 (15)	Norway	163 955	321	1.96	1 314 775	2349	1.79	0.17
Sala et al., 2015 (13)	Spain	79 031	112	1.42	82 961	114	1.37	0.04
Sankatsing et al., 2018 (30)	Netherlands	2 620 442	5748	2.19	4 722 885	10 412	2.2	−0.01
Skaane et al., 2007 (11) (Oslo II)	Norway	6944	12	1.73	16 985	40	2.36	−0.63
Timmermans et al., 2017 (26)	Belgium	133 627	369	2.76	143 293	391	2.73	0.03

[a] Cancer rates include ductal carcinoma in situ and invasive.

to 1.42) and 0.18 (95% CI = −0.26 to 0.61) per 1000 screens higher cancer detection rate for digital than film for initial and subsequent screening rounds, respectively (I^2 = 89.6%; 8 studies, $P_{\text{interaction}}$ = .61). When the additional 3 studies that included only data for the initial screening round were added, the cancer detection rate for initial rounds was 0.61 (95% CI = −0.04 to 1.27, I^2 = 69.9%; 11 studies) per 1000 screens higher for digital than film (11, 13, 15, 16, 24, 25, 30, 38, 40, 44). The incremental cancer detection rates were 0.28 (95% CI = −0.26 to 0.83, I^2 = 61.2%; 6 studies), 0.34 (95% CI = −0.06 to 0.74, I^2 = 79.8%; 16 studies), and 0.03 (95% CI = −0.40 to 0.45, I^2 = 0.0%; 4 studies) per 1000 screens in women aged younger than 50 years, 50 to 69 years, and older than 70 years, respectively ($P_{\text{interaction}}$ = .02) (9, 11, 13, 15, 16, 22, 24, 26, 27, 29, 32, 33, 38, 40, 42, 44). Only 4 studies stratified screen-detected cancers by breast density. The incremental cancer detection rates were −0.06 (95% CI = −0.053 to 0.42, I^2 = 52.4%; 4 studies) and 0.21 (95% CI = −0.51 to 0.92, I^2 = 76.4%; 4 studies) per 1000 screens in women with less dense and more dense breasts, respectively ($P_{\text{interaction}}$ = .12) (9, 29, 38, 42). Sensitivity analyses of the pooled cancer detection rate

differences for the subsets of studies included for each outcome and subgroup are provided in Figure 4. Each subgroup of studies was directionally consistent with the overall results for the pooled difference of cancer detection rate.

Discussion

The systematic review of the observed changes on screening outcomes associated with the transition from film to digital mammography combined data from 16 583 743 screens to provide robust pooled estimates for cancer detection and recall rates. Overall, there was a higher screen-detected cancer rate (0.51/1000, 95% CI = 0.19 to 0.83) following transition from film to digital mammography screening. The higher cancer detection rate was largely attributable to greater detection of DCIS, with a smaller difference in invasive cancer detection. Only a proportion of DCIS will become invasive cancer; however, it is not currently possible to accurately predict which DCIS will progress. Therefore, all DCIS lesions are surgically treated as routine care
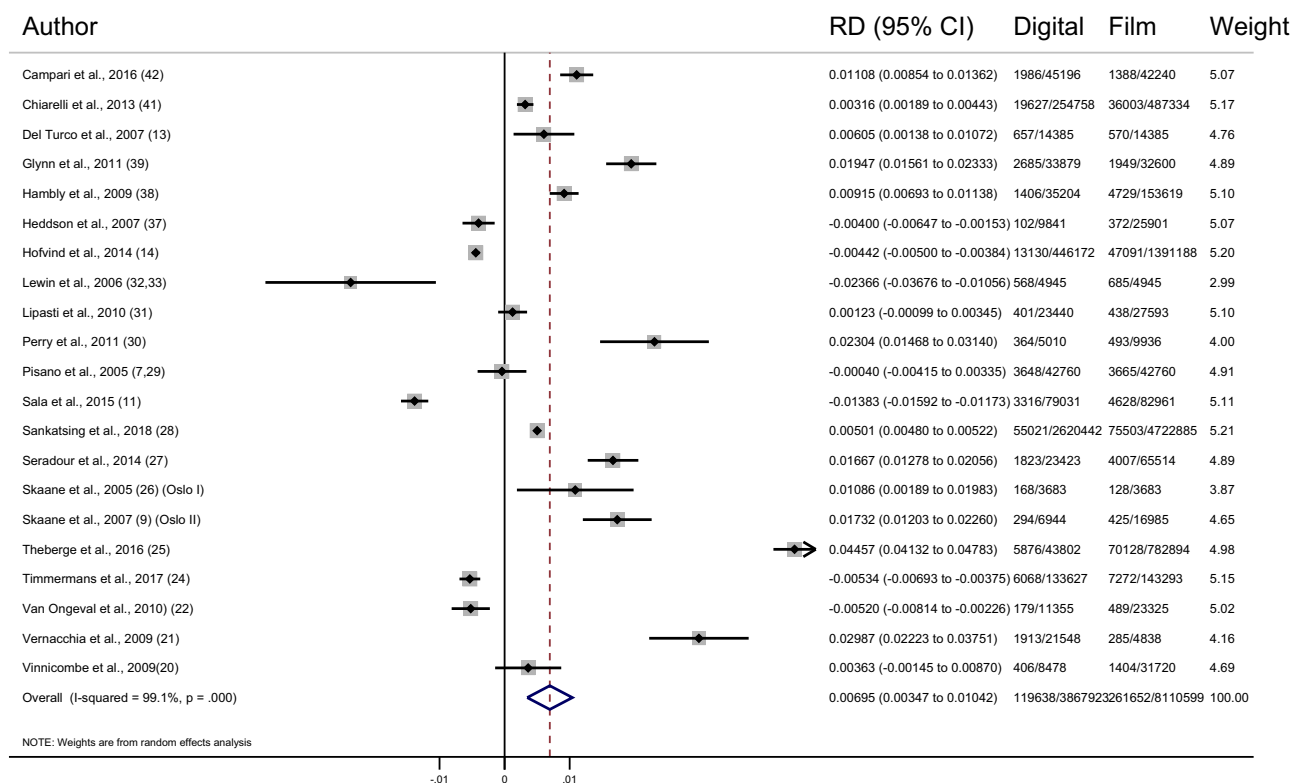
REVIEW

Figure 3. Forest plot of recall rates. CI = confidence interval; RD = risk difference.

(45). Whether this additional detection (and treatment) of DCIS following transition to digital translates into benefits for health outcomes is unclear (2, 46, 47).

Among the 7 studies that reported interval cancers, the difference in cancer detection rate was smaller but consistent with that estimated for all studies combined. However, there was no change in interval cancer rate, with a lower confidence interval limit that excludes a large effect (−0.02/ 1000, 95% CI = −0.06 to 0.03). This suggests that the (modest) extra detection may not be delivering additional benefit, although it is possible that more screening rounds are needed for a lower interval rate to become apparent. Interval cancers represent a measure of the limits of the effectiveness of a screening program in detecting clinically important breast cancers in the population. Some of the interval cancers will be false negatives that were missed at screening, and others are cancers that have developed since the last screening (12, 13, 48).

The pooled estimates show that following the transition to digital mammography there were higher recall rates (6.95/1000, 95% CI = 3.47 to 10.42), most of which were false positives. If 10 000 women are screened with digital mammography rather than film, there would be approximately 70 additional women recalled, 7 of whom would have the additional cases of cancer detected and 63 would have additional false positives. More false-positive mammograms may result in more women experiencing increased short-term, sometimes severe, anxiety (2, 46, 47). In the Netherlands, the increased recall was only temporary; as familiarity with the technology improved, the recall rates stabilized over time (25). Under this notion, the Spanish study excluded the first screening round with digital, and the Swedish study and one of the Canadian studies excluded results

from the first 6 months after the changeover (13, 27, 39). These exclusions from recall data negate the fact that real women were recalled during that time. Further, although recall rates may stabilize in some instances, in many cases they may not have a chance to reach a steady state before a new technology is introduced. Breast cancer diagnoses were histologically verified in all studies, minimizing the chance that outcomes were misclassified. Although there is potential for misestimation of false positive rates due to variation in the further imaging and/ or procedures that are performed when a woman is recalled for assessment and biopsy is not done, this is unlikely to affect our comparative analyses because the review was limited to within-population comparisons.

The Digital Mammographic Imaging Screening Trial reported that screening with digital mammography detected more cancers than film mammography in women who have dense breasts, who are younger than 50 years, or who are premenopausal or perimenopausal (9). The Digital Mammographic Imaging Screening Trial did not provide the interval cancer rate and therefore was not able to contribute evidence to our assessment of the impact of digital mammography on net health benefit or harm from the change in technology. The small differences in pooled incremental cancer detection rate by age and by density, in the few studies that stratified by these variables, had point estimates suggesting higher incremental detection for women aged younger than 70 years and for women with dense breasts. The wide confidence intervals for these reflect the small number of studies that provided data on these subgroups.

Although a priori we allowed for up to 20% overlap in study populations, there was no overlap between the included

Table 4. Number of recalls, recall rates, and difference in recall rates

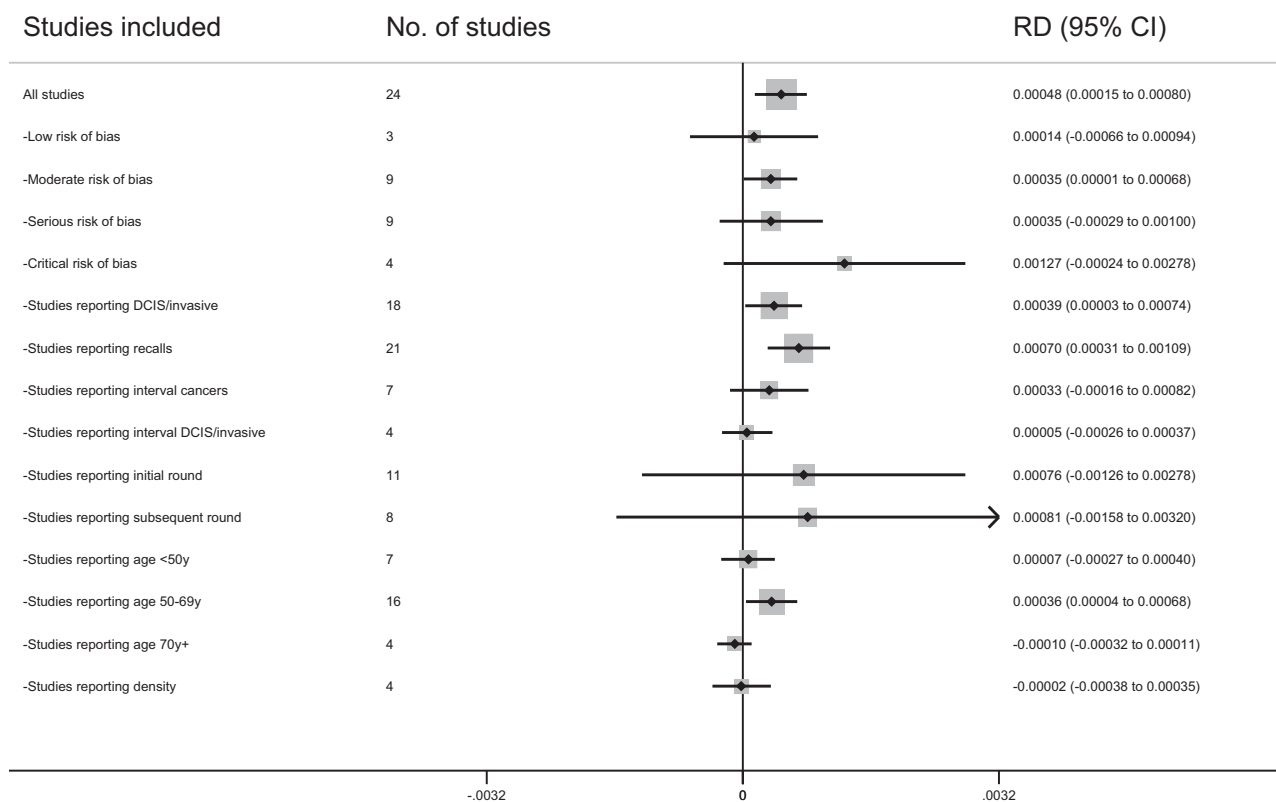| Study Author	Country	Digital No. screening	Digital No. recalls	Digital Recall/1000	Film No. screening	Film No. recalls	Film Recall/1000	Difference in recalls/1000
Campari et al., 2016 (44)	Italy	45 196	1986	43.94	42 240	1388	32.86	11.08
Chiarelli et al., 2013 (43) and Prummel et al., 2016 (12)	Canada	254 758	19 627	77.04	487 334	36 003	73.88	3.16
Del Turco et al., 2007 (16)	Italy	14 385	657	45.67	14 385	570	39.62	6.05
Glynn et al., 2011 (41)	United States	33 879	2685	79.25	32 600	1949	59.79	19.47
Hambly et al., 2009 (40)	Ireland	35 204	1406	39.94	153 619	4792	31.19	8.74
Heddson et al., 2007 (39)	Sweden	9841	102	10.36	25 901	372	14.36	-4
Hofvind et al., 2014 (15)	Norway	446 172	13 130	29.43	1 391 188	47 091	33.85	-4.42
Lewin et al., 2006 (34, 35)	United States	4945	568	114.86	4945	685	138.52	-23.66
Lipasti et al., 2010 (33)	Finland	23 440	401	17.11	27 593	438	15.87	1.23
Perry et al., 2011 (32)	UK	5010	364	72.65	9936	493	49.62	23.04
Pisano et al., 2005 (9, 31)	United States	42 760	3592	84	42 760	3592	84	0
Sala et al., 2015 (13)	Spain	79 031	3316	41.96	82 961	4628	55.79	-13.83
Sankatsing et al., 2018 (30)	The Netherlands	2 620 442	55 021	21	4 722 885	75 503	15.99	5.01
Seradour et al., 2014 (29)	France	23 423	1823	77.83	65 514	4007	61.16	16.67
Skaane et al., 2005 (28) (Oslo I)	Norway	3683	168	45.61	3683	128	34.75	10.86
Skaane et al., 2007 (11) (Oslo II)	Norway	6944	294	42.34	16 985	425	25.02	17.32
Theberge et al., 2016 (27)	Canada	43 802	5876	134.15	782 894	70 128	89.58	44.57
Timmermans et al., 2017 (26)	Belgium	133 627	6068	45.41	143 293	7272	50.75	-5.34
Van Ongeval et al., 2010 (24)	Belgium	11 355	179	15.76	23 325	489	20.96	-5.2
Vernacchia et al., 2009 (23)	United States	21 548	1913	88.78	4838	285	58.91	29.87
Vinnicombe et al., 2009 (22)	UK	8478	406	47.89	31 720	1405	44.29	3.59

Studies included	No. of studies		RD (95% CI)
All studies	24		0.00048 (0.00015 to 0.00080)
-Low risk of bias	3		0.00014 (-0.00066 to 0.00094)
-Moderate risk of bias	9		0.00035 (0.00001 to 0.00068)
-Serious risk of bias	9		0.00035 (-0.00029 to 0.00100)
-Critical risk of bias	4		0.00127 (-0.00024 to 0.00278)
-Studies reporting DCIS/invasive	18		0.00039 (0.00003 to 0.00074)
-Studies reporting recalls	21		0.00070 (0.00031 to 0.00109)
-Studies reporting interval cancers	7		0.00033 (-0.00016 to 0.00082)
-Studies reporting interval DCIS/invasive	4		0.00005 (-0.00026 to 0.00037)
-Studies reporting initial round	11		0.00076 (-0.00126 to 0.00278)
-Studies reporting subsequent round	8		0.00081 (-0.00158 to 0.00320)
-Studies reporting age <50y	7		0.00007 (-0.00027 to 0.00040)
-Studies reporting age 50-69y	16		0.00036 (0.00004 to 0.00068)
-Studies reporting age 70y+	4		-0.00010 (-0.00032 to 0.00011)
-Studies reporting density	4		-0.00002 (-0.00038 to 0.00035)

-.0032   0   .0032

Figure 4. Sensitivity analysis of the pooled cancer detection rate differences for the subsets of studies. CI = confidence interval; DCIS = ductal carcinoma in situ; RD = risk difference.

studies. The included study from the Netherlands covered the entirety of the country, rendering unnecessary the other 8 potential studies. Additionally, 3 studies were excluded from Norway, 1 from Ireland, and 2 from Spain due to overlapping cohorts with other included studies.

There was substantial heterogeneity in the included studies, which was explored in the risk of bias assessment through identifying potential confounding due to changes in underlying breast cancer rates over time. To limit confounding of background rates, only studies that compared digital with film within the same study population were included; however, each of the studies varied in country as well as population size. Studies that capture longer time periods are more robust; however, there is a trade-off with the potential for confounders (eg, reproductive status, density, age distribution, postmenopause obesity, alcohol consumption, hormone replacement therapy use, and socioeconomic status) to change over time. The ROBINS-I risk of bias tool incorporates the confounders measured by each study to determine the extent of confounding that may have occurred. For the studies that calculated adjusted rates, these were compared with the unadjusted rates used in the meta-analysis and were not found to differ (Supplementary Table 2.1, available online). Studies that utilized both types of mammography concurrently were less prone to bias from differences in rates of potential confounders over time. Fourteen of the included studies used the 2 types of mammography concurrently for at least 50% of their study period (8, 9, 11, 12, 15, 16, 22, 26–30, 32, 36, 40, 43).

Because the potential confounders measured by each study were inconsistent, we only compared unadjusted rates, which

is an important limitation of our meta-analysis. Specifically, our estimates assume that there were no changes in confounding within the screening population over the study period. This assumption may be avoided by analyzing longitudinal data over time to account for changes in confounding (49), but this approach requires individual patient data. Additional sources of heterogeneity are from differences in screening program eligibility and design that could affect outcomes. These include age range, length of screening interval, 1 vs 2 views, number of readers, training and time with each technology, and the mammography units used.

Robust evaluation of health outcomes of screening programs or changes in screening policy have traditionally used long-term follow-up studies, and ideally randomized controlled trials. However, such studies, although providing high-quality evidence, take many years to complete, thus delaying the availability of essential information to evaluate the effectiveness of screening policy. However, the potential for selection bias and confounding bias means that evidence from observational screening studies should be interpreted with caution. These biases may explain the larger difference in the cancer detection rate found in the included observational studies compared with the trials. Nevertheless, it is crucial to measure the impact an intervention has had after the rollout on a population-wide level, and comparison of cancer detection rates and interval cancer rates enables a rigorous yet timely evaluation (3).

If the introduction of a new technology results in an increase in cancer detection rate, then one should expect a reduction in the interval cancer rate, indicating improved early detection

rather than increased overdetection (3, 50). In fact, an increase in cancer detection rate may not be necessary to decrease interval cancer rates if the new technology is better at detecting clinically important cancers. The small increase in cancer detection rate following the transition to DM may not be sufficient to translate into a decrease in interval cancers. Any improvement in the sensitivity of a screening program needs to be evaluated with respect to the associated increase in false positives and with appropriate economic assessment (48).

In summary, this review based on population breast-screening studies shows that following the transition from film to digital mammography, there was a modest but statistically significant increase in cancer detection rate, a statistically significant increase in recall and false-positive screens, and no effect on interval cancer rates. The increase in cancer detection rate noted with the transition from film to digital mammography was largely attributable to more detection of DCIS, with little difference in invasive cancer detection. Although the transition from film to digital may have been for technological reasons and for efficiencies in service screening, it seems unlikely to have translated into a beneficial effect for screening participants. At a time when new mammography and other imaging technologies are proposed for adoption in population screening, it is critical to carefully consider the effect this could have on benefits and harms and to ensure that new technologies are not adding substantial harm with little benefit.

Funding

Notes

Role of the funder: The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Disclosures: The authors have no conflicts of interest.

Role of the authors: RF and MM screened title and abstracts. RF and SW completed the full text read and data extraction. RF completed the data synthesis with support from KB and KM. RF and GJ conducted the risk of bias assessment. RF wrote the manuscript with input from KB, AB, NH, and KM.

References

1. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M; The Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer*. 2013; 108(11):2205–2240.
2. Nelson HD, Pappas M, Cantor A, Griffin J, Daeges M, Humphrey L. Harms of breast cancer screening: systematic review to update the 2009 U.S. Preventive Services Task Force Recommendation. *Ann Intern Med*. 2016;164(4): 256–267.
3. Irwig L, Houssami N, Armstrong B, Glasziou P. Evaluating new screening tests for breast cancer. *BMJ*. 2006;332(7543):678–679.
4. Bell KJL, Bossuyt P, Glasziou P, Irwig L. Assessment of changes to screening programmes: why randomisation is important. *BMJ*. 2015;350(mar30 14): h1566–h1566.
5. Skaane P, Young K, Skjennald A. Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading - Oslo I Study. *Radiology*. 2003;229(3):877–884.
6. Esserman LJ, Shieh Y, Rutgers EJT, et al. Impact of mammographic screening on the detection of good and poor prognosis breast cancers. *Breast Cancer Res Treat*. 2011;130(3):725–734.
7. Houssami N. Overdiagnosis of breast cancer in population screening: does it make breast screening worthless? *Cancer Biol Med*. 2017;14(1):1–8.
8. Lewin JM, D'Orsi CJ, Hendrick RE, et al. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. *Am J Roentgenol*. 2002;179(3):671–677.
9. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med*. 2005; 353(17):1773–1783.
10. Weber RJ, van Bommel RM, Louwman MW, et al. Characteristics and prognosis of interval cancers after biennial screen-film or full-field digital screening mammography. *Breast Cancer Res Treat*. 2016;158(3):471–483.
11. Skaane P, Hofvind S, Skjennald A. Randomized trial of screen-film versus full-field digital mammography with soft-copy reading in population-based screening program: follow-up and final results of Oslo II study. *Radiology*. 2007;244(3):708–717.
12. Prummel MV, Muradali D, Shumak R, et al. Digital compared with screen-film mammography: measures of diagnostic accuracy among women screened in the Ontario Breast Screening Program. *Radiology*. 2016;278(2):365–373.
13. Sala M, Domingo L, Macià F, Comas M, Burón A, Castells X. Does digital mammography suppose an advance in early diagnosis? Trends in performance indicators 6 years after digitalization. *Eur Radiol*. 2015;25(3):850–859.
14. Nederend J, Duijm LEM, Louwman MWJ, et al. Impact of the transition from screen-film to digital screening mammography on interval cancer characteristics and treatment - a population based study from the Netherlands. *Eur J Cancer*. 2014;50(1):31–39.
15. Hofvind S, Skaane P, Elmore JG, Sebuødegård S, Hoff SR, Lee CI. Mammographic performance in a population-based screening program: before, during, and after the transition from screen-film to full-field digital mammography. *Radiology*. 2014;272(1):52–62.
16. Del Turco MR, Mantellini P, Ciatto S, et al. Full-field digital versus screen-film mammography: comparative accuracy in concurrent screening cohorts. *Am J Roentgenol*. 2007;189(4):860–866.
17. de Gelder R, Fracheboud J, Heijnsdijk EAM, et al. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. *Prev Med*. 2011;53(3):134–140.
18. Marinovich ML, Hunter KE, Macaskill P, Houssami N. Breast cancer screening using tomosynthesis or mammography: a meta-analysis of cancer detection and recall. *J Natl Cancer Inst*. 2018;110(9):942–949.
19. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62(10): e1–e34.
20. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ (Online)*. 2016;355:i4919.
21. Farber R, Bell K, McGeechan K, Houssami K, Marinovich M, Barratt A. Impact of the practice shift from plain film mammography to digital mammography: systematic review and meta-analysis. *PROSPERO* 2017:CRD42017070601. http://www.crd.york.ac.uk/PROSPERO/display_record.php? ID=CRD42017070601.
22. Vinnicombe S, Pinto Pereira SM, McCormack VA, Shiel S, Perry N, Dos Santos Silva IM. Full-field digital versus screen-film mammography: comparison within the UK breast screening program and systematic review of published data. *Radiology*. 2009;251(2):347–358.
23. Vernacchia FS, Pena ZG. Digital mammography: its impact on recall rates and cancer detection rates in a small community-based radiology practice. *Am J Roentgenol*. 2009;193(2):582–585.
24. Van Ongeval C, Van Steen A, Vande Putte G, et al. Does digital mammography in a decentralized breast cancer screening program lead to screening performance parameters comparable with film-screen mammography? *Eur Radiol*. 2010;20(10):2307–2314.
25. van Luijt PA, Fracheboud J, Heijnsdijk EA, den Heeten GJ, de Koning HJ, National Evaluation Team for Breast Cancer Screening in Netherlands Study Group. Nation-wide data on screening performance during the transition to digital mammography: observations in 6 million screens. *Eur J Cancer*. 2013; 49(16):3517–3525.
26. Timmermans L, Bleyen L, Bacher K, et al. Screen-detected versus interval cancers: effect of imaging modality and breast density in the Flemish Breast Cancer Screening Programme. *Eur Radiol*. 2017;27(9):3810–3810.
27. Theberge I, Vandal N, Langlois A, Pelletier E, Brisson J. Detection rate, recall rate, and positive predictive value of digital compared to screen-film

REVIEW

mammography in the Quebec population-based breast cancer screening program. *Can Assoc Radiol J*. 2016;67(4):330–338.

28. Skaane P, Skjennald A, Young K, et al. Follow-up and final results of the Oslo I Study comparing screen-film mammography and full-field digital mammography with soft-copy reading. *Acta Radiol* 2005;46(7):679–689.

29. Seradour B, Heid P, Esteve J. Comparison of direct digital mammography, computed radiography, and film-screen in the French National Breast Cancer Screening Program. *Am J Roentgenol*. 2014;202(1):229–236.

30. Sankatsing VDV, Fracheboud J, de Munck L, et al.; National Evaluation Team for Breast cancer screening, NETB. Detection and interval cancer rates during the transition from screen-film to digital mammography in population-based screening. *BMC Cancer*. 2018;18(1):256.

31. Pisano ED, Hendrick RE, Yaffe MJ, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. *Radiology*. 2008;246(2):376–383.

32. Perry NM, Patani N, Milner SE, et al. The impact of digital mammography on screening a young cohort of women for breast cancer in an urban specialist breast unit. *Eur Radiol*. 2011;21(4):676–682.

33. Lipasti S, Anttila A, Pamilo M. Mammographic findings of women recalled for diagnostic work-up in digital versus screen-film mammography in a population-based screening program. *Acta Radiol*. 2010;51(5):491–497.

34. Lewin JM, Hendrick RE, D'Orsi CJ, et al. Comparison of full-field digital mammography with screen-film mammography for cancer detection: results of 4,945 paired examinations. *Radiology*. 2001;218(3):873–880.

35. Lewin J. Clinical trials in full-field digital mammography. *Semin Breast Dis*. 2006;9(3):87–91.

36. Kerlikowske K, Hubbard RA, Miglioretti DL, et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. *Ann Intern Med*. 2011;155(8):493–502.

37. Henderson LM, Miglioretti DL, Kerlikowske K, Wernli KJ, Sprague BL, Lehman CD. Breast cancer characteristics associated with digital versus film-screen mammography for screen-detected and interval cancers. *Am J Roentgenol*. 2015;205(3):676–684.

38. Henderson LM, Benefield T, Marsh MW, et al. The influence of mammographic technologists on radiologists' ability to interpret screening mammograms in community practice. *Acad Radiol*. 2015;22(3):278–289.

39. Heddson B, Ronnow K, Olsson M, Miller D. Digital versus screen-film mammography: a retrospective comparison in a population-based screening program. *Eur J Radiol*. 2007;64(3):419–425.

40. Hambly NM, McNicholas MM, Phelan N, Hargaden GC, O'Doherty A, Flanagan FL. Comparison of digital mammography and screen-film mammography in breast cancer screening: a review in the Irish breast screening program. *Am J Roentgenol*. 2009;193(4):1010–1018.

41. Glynn CG, Farria DM, Monsees BS, Salcman JT, Wiele KN, Hildebolt CF. Effect of transition to digital mammography on clinical outcomes. *Radiology*. 2011; 260(3):664–670.

42. Dabbous F, Dolecek TA, Friedewald SM, et al. Performance characteristics of digital vs film screen mammography in community practice. *Breast J*. 2018; 24(3):369–372.

43. Chiarelli AM, Edwards SA, Prummel MV, et al. Digital compared with screen-film mammography: performance measures in concurrent cohorts within an organized breast screening program. *Radiology*. 2013;268(3):684–693.

44. Campari C, Giorgi Rossi P, Mori CA, et al. Impact of the introduction of digital mammography in an organized screening program on the recall and detection rate. *J Digit Imaging*. 2016;29(2):235–242.

45. Ozanne EM, Shieh Y, Barnes J, Bouzan C, Hwang ES, Esserman LJ. Characterizing the impact of 25 years of DCIS treatment. *Breast Cancer Res Treat*. 2011;129(1):165–173.

46. Tosteson ANA, Fryback DG, Hammond CS, et al. Consequences of false-positive screening mammograms. *JAMA Intern Med*. 2014;174(6):954–961.

47. Lauby-Secretan B, Scoccianti C, Loomis D, et al. Breast-cancer screening—viewpoint of the IARC Working Group. *N Engl J Med*. 2015;372(24):2353–2358.

48. Houssami N, Irwig L, Ciatto S. Radiological surveillance of interval breast cancers in screening programmes. *Lancet Oncol*. 2006;7(3):259–265.

49. Bluekens AMJ, Karssemeijer N, Beijerinck D, et al. Consequences of digital mammography in population-based breast cancer screening: initial changes and long-term impact on referral rates. *Eur Radiol*. 2010;20(9):2067–2073.

50. Houssami N, Lång K, Hofvind S, et al. Effectiveness of digital breast tomosynthesis (3D-mammography) in population breast cancer screening: a protocol for a collaborative individual participant data (IPD) meta-analysis. *Transl Cancer Res*. 2017;6(4):869–877.

REVIEW