OXFORD

JNCI J Natl Cancer Inst (2018) 110(9): djy121

doi: 10.1093/jnci/djy121
First published online August 9, 2018
Review

REVIEW

Breast Cancer Screening Using Tomosynthesis or Mammography: A Meta-analysis of Cancer Detection and Recall

M. Luke Marinovich, Kylie E. Hunter, Petra Macaskill, Nehmat Houssami

See the Notes section for the full list of authors' affiliations.

Correspondence to: Michael Luke Marinovich, MPH, PhD, Sydney School of Public Health, Faculty of Medicine and Health, A27, Edward Ford Building, The University of Sydney NSW 2006, Australia (e-mail: luke.marinovich@sydney.edu.au).

Abstract

Background: Tomosynthesis approximates a 3D mammogram of the breast, reducing parenchymal overlap that masks cancers or creates false "lesions" on 2D mammography, and potentially enabling more accurate detection of breast cancer. We compared breast cancer screening detection and recall in asymptomatic women for tomosynthesis vs 2D mammography.
Methods: A systematic review and random effects meta-analysis were undertaken. Electronic databases (2009–July 2017) were searched for studies comparing tomosynthesis and 2D mammography in asymptomatic women who attended population breast cancer screening and reporting cancer detection rate (CDR) and recall rate. All statistical tests were two-sided.
Results: Seventeen studies (1 009 790 participants) were included from 413 citations. The pooled incremental CDR for tomosynthesis was 1.6 cancers per 1000 screens (95% confidence interval [CI] = 1.1 to 2.0, $P < .001$, $I^2 = 36.9\%$). Incremental CDR was statistically significantly higher for European/Scandinavian studies, all using a "paired" design where women had both tests (2.4 per 1000 screens, 95% CI = 1.9 to 2.9, $P < .001$, $I^2 = 0.0\%$) compared with US ("unpaired") studies (1.1 per 1000 screens, 95% CI = 0.8 to 1.5, $P < .001$, $I^2 = 0.0\%$; $P < .001$ between strata). The recall rate for tomosynthesis was statistically significantly lower than for 2D mammography (pooled absolute reduction = –2.2%, 95% CI = –3.0 to –1.4, $P < .001$, $I^2 = 98.2\%$). Stratified analyses showed a decrease in US studies (pooled difference in recall rate = –2.9%, 95% CI = –3.5 to –2.4, $P < .001$, $I^2 = 92.9\%$) but not European/Scandinavian studies (0.5% increase in recall, 95% CI = –0.1 to 1.2, $P = .12$, $I^2 = 93.5\%$; $P < .001$ between strata). Results were similar in sensitivity analyses excluding studies with overlapping cohorts.
Conclusions: Tomosynthesis improves CDR and reduces recall; however, effects are dependent on screening setting, with greater improvement in CDR in European/Scandinavian studies (biennial screening) and reduction in recall in US studies with high baseline recall.

Population breast cancer screening has been implemented in most developed health care systems based on evidence from randomized trials that mammography screening confers breast cancer mortality reduction (1,2). Observational studies have provided complementary evidence on the benefits and also the harms accrued in real-world screening (1,3). Technological advances in image acquisition provided the impetus for transition from film screen to digital mammography, but in more recent years, digital breast tomosynthesis (quasi-3D mammograhy) has been translated into screening practice and touted as a mammography technology that addresses the limitations of conventional (2D) mammography. Through acquisition of multiple low-dose x-rays of the breast at different angles, and reconstruction of these projection images into thin slices that can be viewed sequentially by scrolling in an image stack or as a cine loop, tomosynthesis approximates a 3D mammogram of the breast. This imaging approach reduces the breast parenchymal overlap inherent in conventional

Received: March 23, 2018; **Revised:** May 17, 2018; **Accepted:** June 15, 2018

2D mammography (which masks some cancers and creates false "lesions"); the reduction in overlapping tissue appears to translate into more accurate detection using tomosynthesis (4).

Evidence on the detection capability of tomosynthesis for population breast cancer screening has grown rapidly over recent years and has been the subject of descriptive reviews and commentaries (4–7) discussing the merits and pitfalls of tomosynthesis. Published systematic reviews on tomosynthesis screening to date have had a narrow scope (8), have not included most of the currently available studies (9,10), and have not considered jointly the outcomes of cancer detection and recall (10), the latter representing a frequent harm of screening given that the majority of recalled women are not diagnosed with breast cancer. In this work, we report a systematic review and meta-analysis of the evidence on tomosynthesis for population breast cancer screening. Our aims are to summarize all the available evidence on cancer detection and recall for tomosynthesis (3D) vs 2D mammography screening and to assess heterogeneity in the evidence.

Methods

Identification of Studies

A systematic search of the biomedical literature up to July 2017 was undertaken to identify studies comparing tomosynthesis and 2D mammography in the breast cancer screening setting. The EMBASE, PREMEDLINE, Database of Abstracts of Reviews of Effects (DARE), Heath Technology Assessment (HTA), NHS Economic Evaluation Database (NHSEED), ACP Journal Club, and Cochrane databases were searched via Ovid. Search terms were selected to link tomosynthesis with breast cancer and screening. Key words and medical subject headings included "breast cancer," "tomosynthesis," "DBT," "3D mammography," and "screening." The search was limited to studies published in 2009 or later, to align with the early clinical application of the technology. The full search strategy is available in the Supplementary Methods (available online). Reference lists were also searched, and content experts consulted, to identify additional studies.

Review of Studies and Eligibility Criteria

All abstracts were screened for eligibility by one author (LM), and a sample of 25% was assessed independently by another author (NH) to ensure consistent application of the eligibility criteria. Eligible studies were required to have compared tomosynthesis and 2D mammography in asymptomatic women attending population breast cancer screening. Studies using either a paired design (ie, all participants underwent 2D mammography and tomosynthesis, allowing within-participant comparison) or unpaired design (ie, comparison of separate groups that underwent tomosynthesis, with or without 2D mammography, vs 2D mammography alone) were eligible for inclusion. Hereafter, we use the labels "paired" and "unpaired" to differentiate these study subgroups. Studies were required to report measures of both cancer detection (eg, cancer detection rate [CDR]) and recall (eg, recall rate, false recall rate). Studies enrolling symptomatic or high-risk women or conducted in nonscreening settings (eg, assessment, diagnosis, staging) were ineligible. Study inclusion and exclusion criteria are fully described in the Supplementary Methods (available online).

Potentially eligible citations were reviewed in full by one author to determine eligibility (LM), in consultation with a second author (NH) as required. The screening and inclusion process is summarized in Supplementary Figure 1 (available online). In accordance with methodology recommended by the Cochrane Collaboration (11), the potential for publication bias was not formally assessed as the methods developed for randomized controlled trials (ie, tests of funnel plot asymmetry) are not appropriate for studies of screening and diagnostic tests (12).

Data Extraction

Data relating to cancer detection and recall, study design, patient characteristics, and technical details of tomosynthesis were extracted independently by two authors (LM and KH). Breast density was dichotomized as either "low density" (Breast Image Reporting and Data System [BIRADS] density 1 or 2, equivalent to BIRADS A or B) or "high density" (BIRADS density 3 or 4, equivalent to BIRADS C or D) (13). Nonbreast cancers were excluded from cancer detection outcomes. Cancer detection and recall rates for participant subgroups (age, breast density, cancer characteristics) were not extracted due to the infrequent and inconsistent presentation of those data between studies. Quality appraisal was undertaken using the Revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) checklist (14), which we modified for application to studies in the screening setting. Disagreements were resolved by discussion and consensus, with arbitration by a third author (NH) when required.

Statistical Analysis

Study characteristics were summarized descriptively using median values and their associated ranges and interquartile ranges (IQRs). Estimates of CDR per 1000 screens and recall rate were calculated for each study, and exact 95% confidence intervals (CIs) were computed. For each subgroup of studies defined by modality and study design (paired vs unpaired), a summary estimate for both CDR and recall outcomes was computed using a logistic regression model with random effects for study (PROC GLIMMIX in SAS 9.4, SAS Institute, Cary, NC). For the main analysis, the differences between tests in CDR and recall rate within studies (stratified by study design) were pooled as risk differences using the inverse variance method with random effects for study (DerSimoneon and Laird method, as implemented in Revman 5.3) (15). Standard errors of the differences were calculated based on differences in two independent proportions for unpaired study designs. For paired study designs, PROC GENMOD in SAS was used to take account of the pairing of results within an individual when computing the standard error of the difference in proportions. Estimates were then input into Revman for meta-analysis. Log odds ratios (ORs) and their standard errors were also calculated and were pooled in sensitivity analyses for comparison with risk differences.

For unpaired studies, forest plots were ordered by decreasing difference between tomosynthesis and 2D mammography cohorts in the proportion of women with high breast density (BIRADS 3 and 4), and the plots were visually inspected for evidence of a relationship between that variable and CDR and recall outcomes. Where plots were suggestive of an association, mixed models with random effects for study were used to test for linear trend (PROC MIXED in SAS).

All tests of statistical significance were two-sided; the level chosen for statistical significance was 5%.

Results

Eligible Studies and Study Characteristics

A total of 413 citations were identified. Seventeen studies were eligible for inclusion in our meta-analysis (16–32), reporting data on 1 009 790 participants undergoing tomosynthesis and/or 2D mammography, as shown in Table 1. Studies enrolled participants between 2010 and 2014 (median midpoint of recruitment: 2011) and included a median of 23 355 participants. Four studies used a paired design, comparing the addition of tomosynthesis to 2D mammography with 2D mammography alone; all those studies were prospective trials and were conducted in Europe/Scandinavia. The remaining 13 studies used retrospective, unpaired designs and were conducted in the United States. In all unpaired studies, participants in the tomosynthesis cohort also underwent 2D mammography. Therefore, although for brevity we refer simply to "tomosynthesis," our analyses reflect the combination of tomosynthesis and 2D mammography compared with 2D mammography alone. Characteristics of included studies are summarized in Table 1.

Study Quality and Risk of Bias

Study quality is summarized in Supplementary Figures 2 and 3 (available online). Studies were generally assessed to be at low risk of bias; however, there was the possibility that some unpaired studies may be affected by selection bias from hybrid environments where tomosynthesis and mammography screening were available concurrently (21,22,30) or explicit referral of patients with high breast density to tomosynthesis (32).

Cancer Detection

Study-specific data for overall cancer detection, stratified by study design (paired vs unpaired), are described in Table 2 and Supplementary Table 1 (available online); corresponding pooled estimates for incremental CDR are presented in Figure 1. The pooled incremental CDR for tomosynthesis from all studies (n = 17) was 1.6 cancers per 1000 screens (95% CI = 1.1 to 2.0, $P < .001$, $I^2 = 36.9\%$). When the analysis was stratified by study design, a statistically significantly higher incremental CDR was found for paired studies (2.4 per 1000 screens, 95% CI = 1.9 to 2.9, $P < .001$, $I^2 = 0.0\%$) compared with unpaired studies (1.1 per 1000 screens, 95% CI = 0.8 to 1.5, $P < .001$, $I^2 = 0.0\%$; $P < .001$ for difference between strata). These findings were consistent when results were expressed as odds ratios (odds of cancer detection by tomosynthesis vs 2D mammography) and when an outcome of invasive cancer detection was used (Supplementary Figures 4 and 5, available online). Pooled estimates and *P* values did not change substantially in sensitivity analyses where studies with overlapping patient cohorts were excluded (Supplementary Table 2, available online) (21,23,24,27,29).

Due to the theoretical advantage of tomosynthesis in visualizing cancer in dense breasts, forest plots of incremental CDR for unpaired studies were ordered by the difference between cohorts in the proportion of women with high breast density (BIRADS density 3/4) (Supplementary Table 3, available online). Visual inspection of the plot suggested a tendency for higher incremental CDR for studies with a greater proportion of women with high breast density in the tomosynthesis arm; however, this trend was not statistically significant ($P = .21$; outcome expressed as odds ratio, $P = .12$).

Recall

Table 3 and Supplementary Table 1 (available online) present study-specific recall data stratified by study design, and pooled estimates for incremental recall rate are presented in Figure 2. The overall pooled estimate showed a statistically significant absolute decrease in recall rate for tomosynthesis compared with 2D mammography (–2.2%, 95% CI = –3.0 to –1.4, $P < .001$, $I^2 = 98.2\%$). Stratified analyses showed that decrease to be attributable to unpaired studies (pooled difference in recall rate = –2.9%, 95% CI = –3.5 to –2.4, $P < .001$, $I^2 = 92.9\%$), with no statistically significant difference in recall rates observed for paired studies (0.5% increase in recall, 95% CI = –0.1 to 1.2, $P = .12$, $I^2 = 93.5\%$; $P < .001$ for difference between strata). Similar results were observed when results were expressed as odds ratios and for the outcome of false recall (Supplementary Figures 6 and 7, available online), as well as in sensitivity analyses that excluded studies with overlapping cohorts (Supplementary Table 2, available online).

Visual inspection of the forest plot of differences in recall rates for unpaired studies suggested greater reductions in recall when a higher proportion of women with dense breasts was included in the tomosynthesis cohort relative to the 2D mammography cohort, and this linear trend was statistically significant ($P = .03$); however, no such association was observed when recall outcomes were expressed as odds ratios ($P = .40$) (Supplementary Figure 6, available online). In a sensitivity analysis excluding one potentially influential study (32) that had both the highest 2D mammography recall rate (17.5%) and the largest imbalance between cohorts in the proportions of high breast density (32.0%; due to women with dense breasts being preferentially referred to tomosynthesis), there was no statistically significant association between the difference in recall rate and the difference between cohorts in relation to breast density ($P = .42$).

Discussion

Digital breast tomosynthesis has been introduced into the breast cancer screening environment relatively recently, in the absence of long-term screening efficacy data, based on a rapidly emerging body of evidence on its detection capability relative to 2D mammography screening. In this meta-analysis, we synthesise and evaluate the quality of the evidence on breast tomosynthesis for population screening, providing up-to-date pooled estimates for cancer detection and recall in comparison with 2D mammography and elucidating sources of heterogeneity in published studies. Our work clearly indicates that tomosynthesis improves CDR (incremental CDR: 1.6 cancers per 1000 screens); however, contrary to a previous meta-analysis of fewer studies that found no differences by study design (10), we found that the improvement in cancer detection was more evident in the "paired" (prospective, European/Scandinavian) trials (2.4 per 1000 screens) than the "unpaired" (retrospective, US) studies (1.1 per 1000 screens). This difference is unlikely to be a direct effect of design per se (noting that none of these studies were RCTs) but is likely explained by the screening context in which the studies were conducted: the prospective (paired) studies were embedded in European biennial screening

Table 1. Summary of study, patient, and testing characteristics of included studies

Variable	No. who provided data			Study-level estimates		
	No. of studies	No. of patients	% of patients	Median	IQR	Range
No.	17	1 009 790	–	23 355	14 588–59 617	1048–454 850
Tomosynthesis	17	350 810	34.7	9499	6100–15 571	524–173 663
2D mammography	17	658 980	65.3	12 157	9364–32 076	524–281 187
Study region (design*)						
United States (retrospective)	13	935 606	92.7	25 498	14 227–62 637	1048–454 850
Europe/Scandinavia (prospective)	4	74 184	7.3	17 177	14 794–22 298	14 588–25 242
Recruitment midpoint, y	17	1 009 790	–	2011	2011–2012	2010–2014
Age, mean (or median), y						
Tomosynthesis	13	826 498	–	56.2	55.7–58.0	54.5–59.6
2D mammography	13	826 498	–	57.5	56.6–58.0	53.8–59.5
Breast density (BIRADS 3/4), %						
Tomosynthesis	17	151 024	43.1†	46.6	38.0–55.8	16.7–90.6
2D mammography	17	258 067	39.2†	42.0	31.4–54.3	16.7–63.3
No. of views (tomosynthesis)						
1	1	15 000	1.5	15 000	–	–
2	16	994 790	98.5	24 299	14 408–61 127	1048–454 850
Screen reading practice						
Single-read	13	935 606	92.7	25 498	14 227–62 637	1048–454 850
Double-read	4‡	74 184	7.3	17 177	14 794–22 298	14 588–25 242

*All included studies were nonrandomized. There were no randomized controlled trials.
IQR = interquartile range.
†Percentage of total number of patients within each testing group.
‡Two studies used arbitration for disagreement between readers, and two studies recalled based on recall by either reader.

Table 2. Number of cancers detected, CDRs, and incremental CDRs for tomosynthesis vs 2D mammography (stratified by study design)

Study*	Tomosynthesis			2D mammography			Incremental CDR/1000† (95% CI)
	No.	Cancers, No.	CDR/1000 (95% CI)	No.	Cancers, No.	CDR/1000 (95% CI)	
Paired studies‡							
Bernardi et al. 2016 (17)	9677	82	8.5 (6.6 to 10.3)	9677	61	6.3 (4.7 to 7.9)	2.2 (1.2 to 3.1)
Ciatto et al. 2013 (18)	7294	59	8.1 (6.0 to 10.1)	7294	39	5.3 (3.7 to 7.0)	2.7 (1.5 to 3.9)
Lang et al. 2016 (25)	7500	67	8.9 (6.8 to 11.1)	7500	47	6.3 (4.5 to 8.1)	2.7 (1.4 to 3.9)
Skaane et al. 2013 (31)	12 621	119	9.4 (7.7 to 11.1)	12 621	90	7.1 (5.7 to 8.6)	2.3 (1.4 to 3.2)
Summary estimate§			8.8 (7.4 to 10.5)			6.4 (5.2 to 7.9)	2.4 (1.9 to 2.9)
Unpaired studies							
Starikov et al. 2015 (32)	2070	11	5.3 (2.7 to 9.5)	12 157	39	3.2 (2.3 to 4.4)	2.1 (–1.2 to 5.4)
Powell et al. 2017 (28)	2304	18	7.8 (4.6 to 12.3)	10 477	54	5.2 (3.9 to 6.7)	2.7 (–1.2 to 6.5)
Haas et al. 2013 (24)	6100	35	5.7 (4.0 to 8.0)	7058	37	5.2 (3.7 to 7.2)	0.5 (–2.0 to 3.0)
Durand et al. 2015 (21)	8591	51	5.9 (4.4 to 7.8)	9364	54	5.8 (4.43 to 7.5)	0.2 (–2.1 to 2.4)
Conant et al. 2016 (19)	25 268	149	5.9 (5.0 to 6.9)	113 061	499	4.4 (4.0 to 4.8)	1.5 (0.5 to 2.5)
Destounis et al. 2014 (20)	524	3	5.7 (1.2 to 16.6)	524	2	3.8 (0.5 to 13.7)	1.9 (–6.4 to 10.3)
Greenberg et al. 2014 (23)	20 943	130	6.2 (5.2 to 7.4)	38 674	188	4.9 (4.2 to 5.6)	1.3 (0.1 to 2.6)
Friedewald et al. 2014 (22)	173 663	950	5.5 (5.1 to 5.8)	281 187	1207	4.3 (4.1 to 4.5)	1.2 (0.8 to 1.6)
Lourenco et al. 2015 (26)	12 921	60	4.6 (3.5 to 6.0)	12 577	68	5.4 (4.2 to 6.8)	–0.8 (–2.5 to 1.0)
Sharpe et al. 2016 (30)	5703	31	5.4 (3.7 to 7.7)	80 149	280	3.5 (3.1 to 3.9)	1.9 (0.0 to 3.9)
McCarthy et al. 2014 (27)	15 571	83	5.3 (4.2 to 6.6)	10 728	49	4.6 (3.4 to 6.0)	0.8 (–1.0 to 2.5)
Rose et al. 2013 (29)	9499	51	5.4 (4.0 to 7.1)	13 856	56	4.0 (3.1 to 5.2)	1.3 (–0.5 to 3.1)
Aujero et al. 2017 (16)	30 561	194	6.3 (5.5 to 7.3)	32 076	169	5.3 (4.5 to 6.1)	1.1 (–0.1 to 2.3)
Summary estimate§			5.7 (5.3 to 6.0)			4.5 (4.1 to 5.0)	1.1 (0.8 to 1.5)

*Unpaired studies are presented in decreasing order of difference in percentage of high breast density between cohorts (Supplementary Table 3, available online). For paired studies, where all participants underwent tomosynthesis and 2D mammography, this difference is zero; these studies are ordered alphabetically.
CDR = cancer detection rate; CI = confidence interval.
†Incremental refers to additional cancers detected when comparing modalities in the same women in the paired studies and the difference between groups in CDR for the unpaired studies.
‡Concordant and discordant cell counts for paired studies are available in Supplementary Table 1 (available online).
§Summary CDRs pooled the estimates for each test separately, by study type. Summary incremental CDRs pooled the difference in estimates between tests (Figure 1).

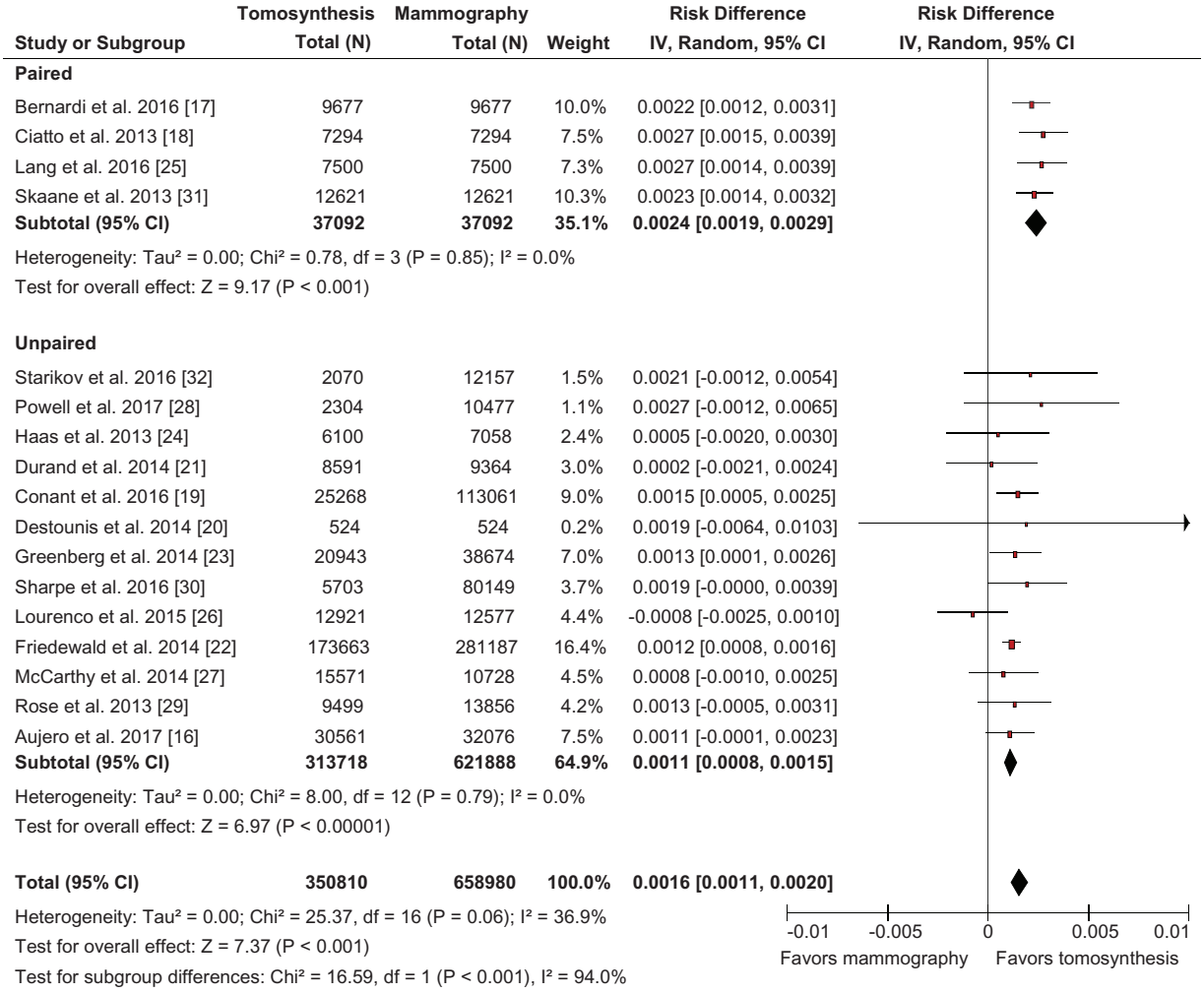

Figure 1. Differences in cancer detection rate (CDR) for tomosynthesis vs 2D mammography stratified by study design. **Squares with horizontal lines** represent individual study estimates and 95% confidence intervals (CIs). **Diamonds** represent pooled estimates and 95% CIs. Tests of overall effect are based on the Z test. Tests of heterogeneity and subgroup differences are based on the χ^2 test. All tests of significance were two-sided. CI = confidence interval; df = degrees of freedom; IV = inverse variance.

programs, whereas the retrospective (unpaired) studies were undertaken in the US screening context, which is predominantly annual screening. Therefore, the CDR at "baseline" (ie, at 2D mammography) differs in these different settings; the pooled baseline CDR in the European/Scandinavian studies was 6.4 per 1000 screens, compared with 4.5 per 1000 in the US studies. Consequently, the gain in cancer detection from tomosynthesis also differs accordingly; the higher incremental CDR attributed to tomosynthesis from the European/Scandinavian studies reflects that, given the longer time interval between screenings, there is greater gain in CDR from tomosynthesis screening.

Previous meta-analyses have either not addressed the impact of tomosynthesis on recall (a frequent harm of screening) (10) or have presented only a qualitative synthesis of relatively few studies (9). A key finding from this analysis is that tomosynthesis screening had a substantial effect in reducing recall rates in the US studies, yielding a pooled absolute reduction of 2.9% (and similarly false-positive recall, which constitutes most of the recall) (Supplementary Figure 7A, available online), where recall rates at 2D mammography were in the range of 7.5% to 17.5% (pooled baseline recall: 11.3%) (Table 3). This absolute reduction in recall is clinically relevant as it reduces the burden of unnecessary testing of women and reduces screening program costs due to false-positive screens. In contrast, tomosynthesis screening had little effect on recall in the European/Scandinavian trials, which reported modest recall rates at 2D mammography (relative to the US studies), in the range of 2.6% to 4.9% (pooled baseline recall: 3.5%). Our interpretation of these data is that tomosynthesis screening has a beneficial effect on reducing recall in screening settings where recall rates at standard 2D mammography are relatively high, but a limited effect on recall in screening services with relatively low recall rates.

Interpretation of our pooled CDR should consider that studies contributing to this analysis were evaluating first (prevalent) screening rounds using tomosynthesis technology. Hence a "prevalence effect" may exist that exaggerates the incremental CDR attributed to tomosynthesis; studies evaluating repeat (incident) screening rounds (of the same women) with tomosynthesis will be important to elucidate whether improved CDR will persist at repeat tomosynthesis screening (33). This would also shed light on whether repeat screening with tomosynthesis

Table 3. Number of recalls, recall rates, and absolute differences in recall rates for tomosynthesis vs 2D mammography (stratified by study design)

Study*	Tomosynthesis			2D mammography			Recall rate difference (95% CI), %
	No.	Recalls, No.	Recall rate % (95% CI)	No.	Recalls, No.	Recall rate (95% CI), %	
Paired studies†							
Bernardi et al. 2016 (17)	9677	463	4.8 (4.4 to 5.2)	9677	389	4.0 (3.6 to 4.4)	0.8 (0.4 to 1.1)
Ciatto et al. 2013 (18)	7294	313	4.3 (3.8 to 4.8)	7294	361	4.9 (4.5 to 5.5)	–0.7 (–1.1 to –0.3)
Lang et al. 2016 (25)	7500	282	3.8 (3.3 to 4.2)	7500	197	2.6 (2.3 to 3.0)	1.1 (0.7 to 1.5)
Skaane et al. 2013 (31)	12 621	463	3.7 (3.3 to 4.0)	12 621	375	2.9 (2.6 to 3.2)	0.8 (0.5 to 1.0)
Summary estimate‡			4.1 (3.3 to 5.0)			3.5 (2.2 to 5.6)	0.5 (–0.1 to 1.2)
Unpaired studies							
Starikov et al. 2015 (32)	2070	212	10.2 (9.0 to 11.6)	12 157	2128	17.5 (16.8 to 18.2)	–7.3 (–8.7 to –5.8)
Powell et al. 2017 (28)	2304	319	13.8 (12.5 to 15.3)	10 477	1694	16.2 (15.5 to 16.9)	–2.3 (–3.9 to –0.7)
Haas et al. 2013 (24)	6100	513	8.4 (7.7 to 9.1)	7058	847	12.0 (11.3 to 12.8)	–3.6 (–4.6 to –2.6)
Durand et al. 2015 (21)	8591	671	7.8 (7.3 to 8.4)	9364	1154	12.3 (11.7 to 13.0)	–4.5 (–5.4 to –3.6)
Conant et al. 2016 (19)	55 998	4856	8.7 (8.4 to 8.9)	142 883	14 884	10.4 (10.3 to 10.6)	–1.7 (–2.0 to –1.5)
Destounis et al. 2014 (20)	524	22	4.2 (2.7 to 6.3)	524	60	11.5 (8.9 to 14.5)	–7.3 (–10.5 to –4.0)
Greenberg et al. 2014 (23)	20 943	2845	13.6 (13.1 to 14.1)	38 674	6247	16.2 (15.8 to 16.5)	–2.6 (–3.2 to –2.0)
Friedewald et al. 2014 (22)	173 663	15 541	8.9 (8.8 to 9.1)	281 187	29 726	10.6 (10.5 to 10.7)	–1.6 (–1.8 to –1.5)
Lourenco et al. 2015 (26)	12 921	827	6.4 (6.0 to 6.8)	12 577	1175	9.3 (8.8 to 9.9)	–2.9 (–3.6 to –2.3)
Sharpe et al. 2016 (30)	5587	341	6.1 (5.5 to 6.8)	70 173	5270	7.5 (7.3 to 7.7)	–1.4 (–2.0 to –0.7)
McCarthy et al. 2014 (27)	15 571	1366	8.8 (8.3 to 9.2)	10 728	1112	10.4 (9.8 to 11.0)	–1.6 (–2.3 to –0.9)
Rose et al. 2013 (29)	9499	518	5.5 (5.0 to 5.9)	13 856	1208	8.7 (8.3 to 9.2)	–3.3 (–3.9 to –2.6)
Aujero et al. 2017 (16)	30 561	1785	5.8 (5.6 to 6.1)	32 076	2799	8.7 (8.4 to 9.0)	–2.9 (–3.3 to –2.5)
Summary estimate‡			8.0 (6.5 to 9.8)			11.3 (9.6 to 13.3)	–2.9 (–3.5 to –2.4)

*Unpaired studies are presented in decreasing order of difference in percentage of high breast density between cohorts (Supplementary Table 3, available online). For paired studies, where all participants underwent tomosynthesis and 2D mammography, this difference is zero; these studies are ordered alphabetically.
CI = confidence interval.
†Concordant and discordant cell counts for paired studies are available in Supplementary Table 1 (available online).
‡Summary recall rates pooled the estimates for each test separately, by study type. Summary differences in recall rates pooled the difference in estimates between tests (Figure 2).

will continue to impact recall rates, with only one study to date suggesting that it may have a sustained effect in reducing recall (34).

Although this meta-analysis provides convincing evidence that tomosynthesis screening improves screen detection measures (compared with 2D mammography), given its relatively recent introduction into practice, there are no long-term efficacy data for tomosynthesis. Little is known on whether the improved CDR estimated for tomosynthesis screening will have additional benefit (ie, whether it will further reduce breast cancer mortality) or whether it will add to overdiagnosis, compared with 2D mammography screening. These evidence gaps relating to tomosynthesis screening will persist in the foreseeable future; hence screening policy and recommendations are likely to look at evidence from detection measures (such as those reported in our meta-analysis) and may also consider surrogate measures. It has been suggested that a potential surrogate for incremental screening benefit when transitioning to a new screening technology is to measure the effect on interval cancer rates; this could also be used as an indicator that tomosynthesis is "early-detecting" cancers that would have otherwise progressed clinically. However, at present, there are very few (or incomplete) data on whether tomosynthesis impacts interval cancer rates at follow-up of screened cohorts (33,34). Randomized trials of tomosynthesis screening have been initiated in several countries; however, these are still in progress, so results may not be available in the near future. Research efforts are being directed toward addressing the outcome of interval cancer rates through individual participant data meta-analysis (35), but reporting is not planned until 2020 due to the necessity for follow-up data and assembling comparison cohorts for paired studies (by nature of their design, these studies are unable to assess the impact of tomosynthesis on interval cancer rates without comparison with a cohort screened by 2D mammograhy alone).

The pooling of study-level data in our analysis allowed not only for precise estimates of incremental CDR and recall, but also for the exploration of differences in those estimates by study design and setting. However, the study-level nature of the analysis also carries certain limitations. Importantly, some studies used overlapping cohorts; that is, a subset of women included in one analysis may also have been included in cohorts analyzed in a later study (or studies). This has the potential to artefactually increase the precision of pooled estimates by reducing variability across studies. Where possible, we identified studies with overlapping cohorts and conducted sensitivity analyses excluding the smaller data set(s). The similarity between findings of the main and sensitivity analyses suggests that the impact of any residual overlap is likely to be minimal.

Although our systematic quality appraisal indicated that included studies were generally at low risk of bias, there are potential sources of bias for unpaired and paired studies that should be noted. For unpaired studies, there is the potential for the cohorts to be "unbalanced" due to their nonrandomized design; that is, for the distribution of participant characteristics to differ between groups screened with different tests. This is particularly so when tomosynthesis and 2D mammography are assessed in a hybrid environment (ie, both tests available concurrently), where patients with specific characteristics may be preferentially referred to a particular test (21–30). In one such

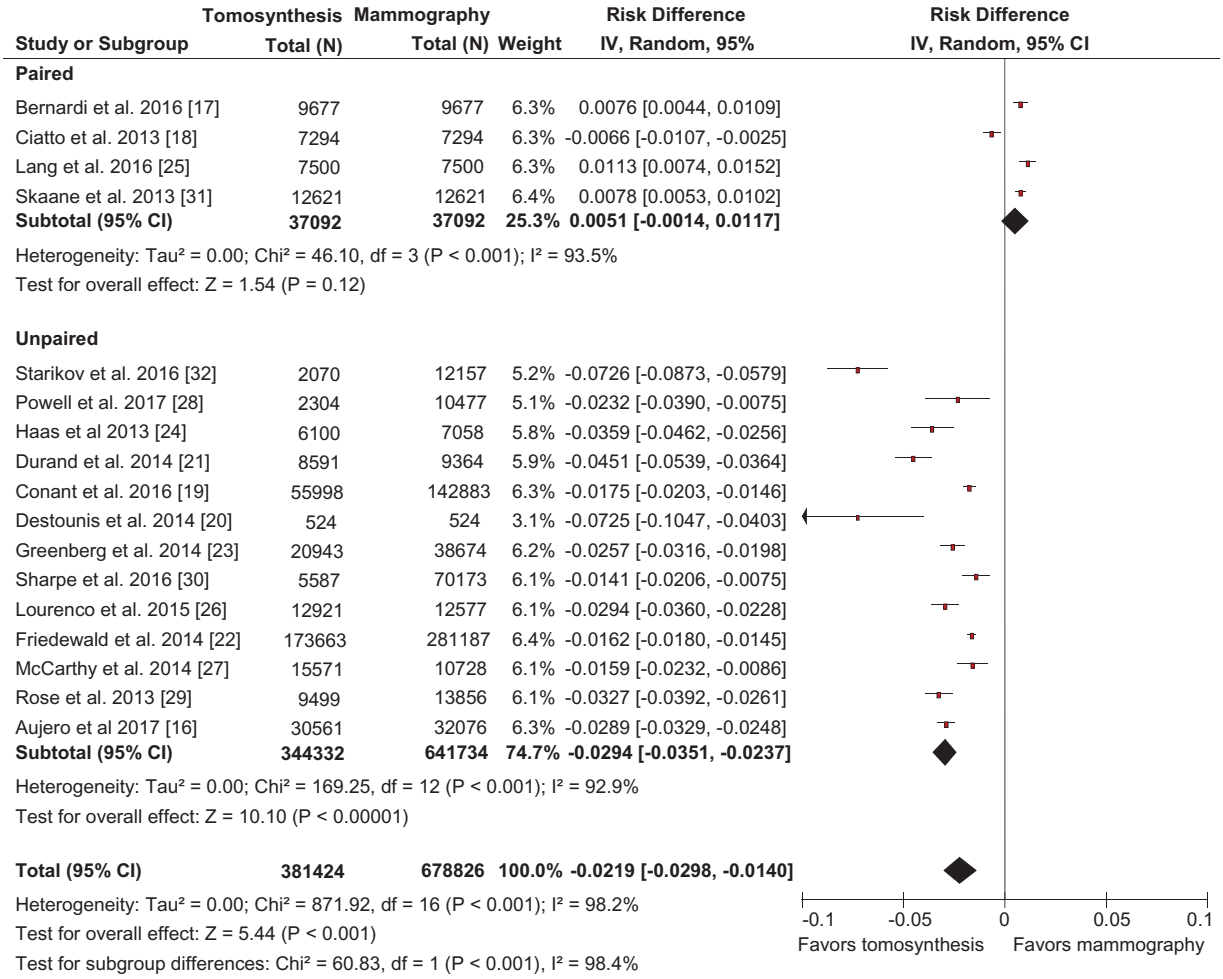

Figure 2. Differences in recall rate for tomosynthesis vs 2D mammography stratified by study design. **Squares with horizontal lines** represent individual study estimates and 95% confidence intervals (CIs). **Diamonds** represent pooled estimates and 95% CIs. Tests of overall effect are based on the Z test. Tests of heterogeneity and subgroup differences are based on the χ^2 test. All tests of significance were two-sided. CI = confidence interval; df = degrees of freedom; IV = inverse variance.

study, it was noted that women with high breast density were offered tomosynthesis when that test was available (32). We explored the potential impact of such an imbalance for breast density, given the theoretical advantages of tomosynthesis over 2D mammography in detecting cancer in dense breasts. Plots of the outcomes of incremental CDR (Figure 1) and recall rate (Figure 2) against the difference in high breast density proportion between cohorts suggested a weak relationship, whereby the observed increases in CDR and reductions in recall due to tomosynthesis were greater when high breast density was more prevalent in the tomosynthesis group. This suggests that tomosynthesis may have greater effectiveness in women with dense breasts. However, statistical analyses did not provide evidence of linear association with either incremental CDR or recall. A limitation of study-level meta-regression is relatively low power to detect such a relationship (15). Furthermore, the infrequent and inconsistent reporting of outcomes by breast density (as well as other subgroups, such as those defined by age or cancer characteristics) limits the ability of study-level meta-analysis to investigate subgroup effects. Reporting of outcomes stratified by breast density within studies would allow for such effects to be explored more comprehensively in future meta-analyses.

Imbalance between cohorts is not a concern in paired studies (despite the absence of random assignment) due to test performance being compared within each participant. However, the sequential reading of 2D mammography and tomosynthesis used by several of the paired studies has the potential to overestimate cancer detection, particularly when an "either-positive" recall rule is used, relative to the integrated interpretation of both tests common in screening practice. Of the paired studies in our analysis, two used an "either-positive" rule (17,18) and two used arbitration of the "either-positive" result that incorporated imaging and clinical information as the basis for recall (25,31). Estimates of incremental CDR were comparable between those studies, suggesting that an "either-positive" recall rule in paired (European/Scandinavian) studies is unlikely to account for the observed difference compared with unpaired (US) studies. However, it remains possible that the sequential reading performed in these studies may have resulted in a bias toward tomosynthesis.

Our meta-analysis provides up-to-date pooled estimates of effect for breast cancer screening using tomosynthesis compared with 2D mammography to inform screening research and practice. At present, nonrandomized studies provide evidence that tomosynthesis improves initial screen detection measures (cancer detection and/or recall rates); however, the heterogeneity in the evidence shown for different study designs/strata highlights that the effect of tomosynthesis may vary according

to screening setting. New research in tomosynthesis screening should focus on evaluating both the sustained effect (of repeat screening with tomosynthesis) and the intermediate to long-term outcomes, including effect on interval cancer rates.

Funding

This work was supported by a Cancer Institute NSW (CINSW) Early Career Fellowship, Grant ID No. 14/ECF/1–06 (MLM), and by a National Breast Cancer Foundation (NBCF Australia) Breast Cancer Research Leadership Fellowship (NH).

Notes

Affiliations of authors: Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia (MLM, KEH, PM, NH); NHMRC Clinical Trials Centre, Camperdown, NSW, Australia (KEH).

The funders had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

The authors have no conflicts of interest.

References

1. 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.
2. Marmot M, Altman DG, Cameron DA, et al. The benefits and harms of breast cancer screening: An independent review. *Lancet*. 2012;380(9855):1778–1786.
3. 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(4):256–267.
4. Houssami N, Miglioretti DL. Digital breast tomosynthesis: A brave new world of mammography screening. *JAMA Oncol*. 2016;2(6):725–727.
5. Gilbert FJ, Tucker L, Young KC. Digital breast tomosynthesis (DBT): A review of the evidence for use as a screening tool. *Clin Radiol*. 2016;71(2):141–150.
6. Conant EF. Clinical implementation of digital breast tomosynthesis. *Radiol Clin North Am*. 2014;52(3):499–518.
7. Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. *Breast*. 2013;22(2):101–108.
8. Houssami N, Turner RM. Rapid review: Estimates of incremental breast cancer detection from tomosynthesis (3D mammography) screening in women with dense breasts. *Breast*. 2016;30(01):141–145.
9. Hodgson R, Heywang-Kobrunner SH, Harvey SC, et al. Systematic review of 3D mammography for breast cancer screening. *Breast*. 2016;27(June):52–61.
10. Yun SJ, Ryu CW, Rhee SJ, et al. Benefit of adding digital breast tomosynthesis to digital mammography for breast cancer screening focused on cancer characteristics: A meta-analysis. *Breast Cancer Res Treat*. 2017;164(3):557–569.
11. Macaskill P, Gatsonis C, Deeks JJ, et al. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PMM, Gatsonis C, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 1.0. The Cochrane Collaboration; 2010. Available from: http://srdta.cochrane.org/.
12. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882–893.
13. D'Orsi CJ, Sickles EA, Mendelson EB, et al. *ACR BI-RADS Atlas, Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology; 2013.
14. Whiting PF, Rutjes AWS, Westwood ME, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011; 155(8):529–536.
15. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org.
16. Aujero MP, Gavenonis SC, Benjamin R, et al. Clinical performance of synthesized two-dimensional mammography combined with tomosynthesis in a large screening population. *Radiology*. 2017;283(1):70–76.
17. Bernardi D, Macaskill P, Pellegrini M, et al. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): A population-based prospective study. *Lancet Oncol*. 2016;17(8):1105–1113.
18. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): A prospective comparison study. *Lancet Oncol*. 2013;14(7):583–589.
19. Conant EF, Beaber EF, Sprague BL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: A cohort study within the PROSPR Consortium. *Breast Cancer Res Treat*. 2016;156(1):109–116.
20. Destounis S, Arieno A, Morgan R. Initial experience with combination digital breast tomosynthesis plus full field digital mammography or full field digital mammography alone in the screening environment. *J Clin Imaging Sci*. 2014; 4(1):9.
21. Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology*. 2015;274(1):85–92.
22. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *J Am Med Assoc*. 2014;311(24):2499–2507.
23. Greenberg JS, Javitt MC, Katzen J, et al. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *Am J Roentgenol*. 2014;203(3): 687–693.
24. Haas BM, Kalra V, Geisel J, et al. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology*. 2013;269(3):694–700.
25. Lang K, Andersson I, Rosso A, et al. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: Results from the Malmo Breast Tomosynthesis Screening Trial, a population-based study. *Eur Radiol*. 2016;26(1):184–190.
26. Lourenco AP, Barry-Brooks M, Baird GL, et al. Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. *Radiology*. 2015;274(2):337–342.
27. McCarthy AM, Kontos D, Synnestvedt M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *J Natl Cancer Inst*. 2014;106(11):dju316.
28. Powell JL, Hawley JR, Lipari AM, et al. Impact of the addition of digital breast tomosynthesis (DBT) to standard 2D digital screening mammography on the rates of patient recall, cancer detection, and recommendations for short-term follow-up. *Acad Radiol*. 2017;24(3):302–307.
29. Rose SL, Tidwell AL, Bujnoch LJ, et al. Implementation of breast tomosynthesis in a routine screening practice: An observational study. *Am J Roentgenol*. 2013;200(6):1401–1408.
30. Sharpe RE, Venkataraman S, Phillips J, et al. Increased cancer detection rate and variations in the recall rate resulting from implementation of 3D digital breast tomosynthesis into a population-based screening program. *Radiology*. 2016;278(3):698–706.
31. Skaane P, Bandos AI, Gullien R, et al. Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. *Eur Radiol*. 2013;23(8):2061–2071.
32. Starikov A, Drotman M, Hentel K, et al. 2D mammography, digital breast tomosynthesis, and ultrasound: Which should be used for the different breast densities in breast cancer screening? *Clin Imaging*. 2015;40(1):68–71.
33. Houssami N, Macaskill P, Bernardi D, et al. Breast screening using 2D mammography or integrating digital breast tomosynthesis (3D mammography) for single-reading or double-reading—evidence to guide future screening strategies. *Eur J Cancer*. 2014;50(10):1799–1807.
34. McDonald ES, Oustimov A, Weinstein SP, et al. Effectiveness of digital breast tomosynthesis compared with digital mammography: Outcomes analysis from 3 years of breast cancer screening. *JAMA Oncol*. 2016;2(6):737–743.
35. 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.