



ARTICLE OPEN



Epidemiology

Differential detection by breast density for digital breast tomosynthesis versus digital mammography population screening: a systematic review and meta-analysis

Tong Li^{1,2}, Nehmat Houssami^{1,2}, Naomi Noguchi², Aileen Zeng^{1,2} and M. Luke Marinovich^{1,2,3}

© The Author(s) 2022

BACKGROUND: We examined whether digital breast tomosynthesis (DBT) detects differentially in high- or low-density screens. **METHODS:** We searched six databases (2009–2020) for studies comparing DBT and digital mammography (DM), and reporting cancer detection rate (CDR) and/or recall rate by breast density. Meta-analysis was performed to pool incremental CDR and recall rate for DBT (versus DM) for high- and low-density (dichotomised based on BI-RADS) and within-study differences in incremental estimates between high- and low-density. Screening settings (European/US) were compared. **RESULTS:** Pooled within-study difference in incremental CDR for high- versus low-density was 1.0/1000 screens (95% CI: 0.3, 1.6; $p = 0.003$). Estimates were not significantly different in US (0.6/1000; 95% CI: 0.0, 1.3; $p = 0.05$) and European (1.9/1000; 95% CI: 0.3, 3.5; $p = 0.02$) settings (p for subgroup difference = 0.15). For incremental recall rate, within-study differences between density subgroups differed by setting ($p < 0.001$). Pooled incremental recall was less in high- versus low-density screens (−0.9%; 95% CI: −1.4%, −0.4%; $p < 0.001$) in US screening, and greater (0.8%; 95% CI: 0.3%, 1.3%; $p = 0.001$) in European screening. **CONCLUSIONS:** DBT has differential incremental cancer detection and recall by breast density. Although incremental CDR is greater in high-density, a substantial proportion of additional cancers is likely to be detected in low-density screens. Our findings may assist screening programmes considering DBT for density-tailored screening.

British Journal of Cancer (2022) 127:116–125; <https://doi.org/10.1038/s41416-022-01790-x>

BACKGROUND

Digital breast tomosynthesis (DBT) provides reconstructed, quasi-three-dimensional mammographic images of the breast, and has been proposed to improve cancer detection in screening through better visualisation of lesions that may be obscured by dense and/or overlapping breast tissue on conventional (two-dimensional) digital mammography (DM) [1]. In addition, by minimising cancer-mimicking artefacts associated with overlapping breast tissue, DBT may reduce high baseline rates of recall to further assessment [2]. Multiple studies have compared DBT and DM in breast cancer screening, including six published systematic reviews [3–8]. All of these reviews reported that detection measures favoured DBT (compared to DM) for breast cancer screening; however, none reported screening detection measures by high and low breast density. High mammographic density (having heterogeneously or extremely dense breasts [9]) is associated with an increased risk of breast cancer [10], including interval breast cancer [11]. Evidence on whether screening performance measures for DBT compared to DM differ by breast density is of interest to population breast cancer screening programmes, and could inform potential adoption of DBT screening in whole or subgroups of the population.

We address this critical knowledge gap by specifically examining cancer detection and recall rates for DBT versus DM by breast density, substantially extending our previous meta-analytic methods [4] to generate new evidence relevant to breast cancer screening practice. We therefore performed a systematic review and meta-analysis to determine if DBT screening is detecting differentially from DM screening in women with dense (high density) or non-dense breasts (low density).

METHODS

Search strategy

This systematic review and meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [12]. As our previous work [4] (PROSPERO, CRD42016038998) conducted a systematic literature search from 2009 to July 2017, we updated the search strategy to perform a literature search from 1 January 2017 to 23 November 2020 in six Ovid electronic databases (EMBASE, PREMEDLINE, ACP Journal Club, Cochrane Controlled Trials Register, Cochrane Database of Systematic Reviews, and Database of Abstracts of Reviews of Effectiveness). The full search strategy is detailed in Supplementary Method 1.

¹The Daffodil Centre, The University of Sydney, a joint venture with Cancer Council New South Wales, Sydney, NSW, Australia. ²School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia. ³Curtin School of Population Health, Curtin University, Bentley, WA, Australia. ✉email: t.li@sydney.edu.au

Received: 25 September 2021 Revised: 27 February 2022 Accepted: 8 March 2022

Published online: 28 March 2022

Eligibility criteria

Studies were eligible when they included asymptomatic women who attended population-based breast cancer screening programmes; compared DBT with DM; reported cancer detection and/or recall by breast density using American College of Radiology Breast Imaging Reporting and Database System (BI-RADS) [9] (any edition); and were reported in English. Detailed inclusion and exclusion criteria are available in Supplementary Method 2. Studies using either a paired design (i.e. all participants underwent DBT and DM, allowing within-participant comparison) or unpaired design (i.e. comparison of separate groups that underwent DBT, with or without DM, versus DM alone) were both eligible for inclusion. To ensure that density classification was consistent for the purpose of pooling estimates by density strata, we did not include studies using an automated density.

Study selection

Titles and abstracts were screened by one author (TL) to determine whether studies met the eligibility criteria for full-text assessment and a sample of 25% was screened independently by another author (MLM) as a quality assurance process. The full-text assessment was conducted by one author (TL) with consultation from a second author (MLM) if required.

Data extraction

Data extraction was performed by one author (TL), with another independent extraction by one of two other authors (NN and AZ). Any disagreements were resolved by discussion and consensus, or with arbitration by a third author (MLM) when needed.

The following data were extracted into an Excel spreadsheet using predefined cells: first author, publication date, country, study design, screening interval, years of participant enrolment, DBT views, DBT modality, DBT screening reading strategy, participants' age (median or mean), and the number of participants and outcomes (cancers detected, recalls) per modality in each density category. Breast density information was extracted according to the BI-RADS density classification of a-d [9] (or 1–4 [13]) when available, and the combined categories of low density (BI-RADS a + b/1 + 2) and high density (BI-RADS c + d/3 + 4) when studies did not report the full BI-RADS classification. Because there were more studies reporting by combined (and less reporting by four) categories of density, we used the binary low- and high-density classification to standardise these data and allow statistical pooling across studies. This approach avoided excluding a substantial number of otherwise eligible studies.

Quality assessment criteria

Quality assessment of all eligible studies was performed by one author (TL) in consultation with two other authors (MLM and NH) when required, using appraisal criteria adapted from QUADAS-2 [14]. Each study was assessed for risk of bias under four domains covering patient selection, index test, reference standard, and flow and timing. The first three domains were also assessed in terms of concerns regarding applicability.

Statistical analysis

Study characteristics were summarised descriptively using median values and ranges. For both DBT and DM, estimates of cancer detection rate (CDR; per 1000 screens) and recall rate (percent) were calculated for low- and high-density strata within each study, and exact 95% confidence intervals (CIs) were computed. Summary estimates of CDR and recall rates for DM (baseline) and DBT were derived and compared between screening settings using PROC GLIMMIX with random effects for study in SAS 9.4 (SAS Institute, Cary, NC, US). Incremental estimates (risk differences), calculated as the study-level differences between modalities (DBT minus DM) in CDR and recall rate, were pooled separately for low- and high-density strata using the inverse variance method with random effects (DerSimonian and Laird method as implemented in RevMan 5.4.1, The Cochrane Collaboration, 2020 [15]). Standard errors of the risk 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. These estimates were then input into RevMan for meta-analysis. Chi-squared tests of differences between the separate pooled estimates for density strata were not performed due to inappropriate standard errors (arising from the same studies contributing to both density strata) and the potential for bias [16, 17].

For the main analyses comparing density strata, we used PROC GENMOD (with the REPEATED statement for paired studies) to model the interaction between modality (DBT versus DM) and breast density (high versus low) for each study. Interaction terms (corresponding to the within-study difference between density strata in incremental CDR and recall rate) and their standard errors were input into RevMan and pooled using the inverse variance method with random effects [16, 17].

Analyses were stratified by screening setting (European versus US studies) based on a priori evidence of a difference in CDR and recall rate [4]. Differences between screening setting subgroups were assessed using the Chi-squared test. Sensitivity analyses were undertaken to include only studies that reported both CDR and recall rate to investigate the effect on pooled estimates. Heterogeneity was assessed using the I^2 statistic with values >50% representing substantial or considerable heterogeneity [15].

Pooled estimates of incremental CDR and recall rate, and the within-study differences between density strata, were incorporated in an epidemiological model simulating plausible scenarios in population screening practice. Simplified decision trees (Supplementary Method 3) were used to apply conditional probabilities to a hypothetical screening population of 10,000 women where the screening setting and proportion of the population with low density were varied. Estimates of the proportion of the population with low density were derived from the median and range of study-specific values reported by European and US studies. For each screening setting and density subgroup, predictions of the number of additional cancers detected and additional women recalled by DBT per 10,000 screens were calculated by multiplying the total number in the population, the proportion of low (or high) density, and the relevant pooled incremental estimate derived from the meta-analysis. A detailed description of this modelled prediction can be found in Supplementary Method 3.

All tests of statistical significance were two-sided. The level chosen for statistical significance was 0.05.

RESULTS

Study characteristics

An initial 565 studies were identified for title and abstract screening, of which 13 studies were eligible for inclusion (Supplementary Fig. 1) in our data synthesis and pooling [18–30], enrolling 1,238,735 participants/examinations (522,846 with DBT and 715,889 with DM) between 2010 and 2017. The study-level median age was 57.0 (range 54.5–59.0) years for DBT and 57.6 (range 53.8–58.6) years for DM cohorts, and the median proportion of women with low breast density was 63% (range 37%–83%) [18–30]. As summarised in Table 1, eight studies were based on US populations (using single reading) and five studies were Europe-based (using double-reading). European studies predominantly used a biennial screening interval, while US studies were assumed to employ mostly annual screening. Compared to US studies, European studies had significantly higher pooled CDR and lower pooled recall rate with DM (baseline) (Supplementary Tables 1 and 2; $p < 0.05$ for all comparisons between screening settings). These studies covered three modes of DBT, including DBT + DM ($n = 11$), DBT + SM (synthesised mammography) ($n = 1$), and DBT + DM + SM ($n = 1$), hereby referred to as DBT.

Risk of bias and applicability

Supplementary Fig. 2a–c provide a summary of the risk of bias and applicability concerns. Five of the 13 studies were rated as low risk of bias and low concerns of applicability [18–21, 27]. The main reason for a high or unclear risk of bias and/or applicability concerns was in the domain of patient selection due to reported [24, 26, 29, 30] or possible [23, 25, 28] selection bias from women with specific characteristics such as dense breasts being differentially referred to DBT or DM [24, 30], self-selection or referral [24, 28], equipment availability [26, 29], and/or hybrid settings of concurrent DBT and DM [23, 25]. Other reasons were related to flow and timing under risk of bias (e.g. examinations using non-standard density terms were excluded [29]) and index test under applicability concerns (e.g. one-view DBT [22]).

Table 1. Characteristics of the included studies.

Study ^a	Region	Design	Screening interval	Enrolment years	DBT mode	DBT reading strategy	DBT mean/median age	DM mean/median age	Breast density analysis level	Reported metrics
Bernardi et al. [18]	Europe	Prospective	Biennial	2013–2015	DBT + DM	Double ^e	58	58	Examinations	CDR and RR
Caumo et al. [19]	Europe	Prospective	Biennial	2014–2016	DBT + SM	Double ^e	59	58	Participants	CDR only
Ciatto et al. [20]	Europe	Prospective	Biennial	2011–2012	DBT + DM	Double ^e	58	58	Examinations	CDR and RR
Romero-Martín et al. [21]	Europe	Prospective	Biennial	2015–2016	DBT + DM + SM ^d	Double ^e	57.6	57.6	Examinations	CDR only
Zackrisson et al. [22]	Europe	Prospective	Biennial ^b	2010–2015	DBT + DM	Double ^f	57	57	Participants	CDR and RR
Alsheik et al. [23]	US	Retrospective	Annual ^c	2015–2017	DBT + DM	Single	57.8	58.6	Examinations	CDR and RR
Conant et al. [24]	US	Retrospective	Annual ^c	2011–2014	DBT + DM	Single	NR	NR	Examinations	CDR and RR
Friedewald et al. [25]	US	Retrospective	Annual ^c	2010–2012	DBT + DM	Single	56.2	57	Examinations	CDR and RR
Haas et al. [26]	US	Retrospective	Annual ^c	2011–2012	DBT + DM	Single	55.8	57.5	Participants	RR only
McCarthy et al. [27]	US	Retrospective	Annual ^c	2010–2013	DBT + DM	Single	56.7	56.9	Participants	CDR and RR
Rose et al. [28]	US	Retrospective	Annual ^c	2010–2012	DBT + DM	Single	54.5	53.8	Examinations	CDR and RR
Sharpe et al. [29]	US	Retrospective	Annual ^c	2011–2014	DBT + DM	Single	55.7	57.6	Examinations	RR only
Starikov et al. [30]	US	Retrospective	Annual ^c	2013–2013	DBT + DM	Single	NR	NR	Examinations	CDR and RR

DBT digital breast tomosynthesis, DM digital mammography, SM synthesised mammography, CDR cancer detection rate, RR recall rate, NR not reported.

^aAll studies used two-view DBT except for Zackrisson et al. [22] which used one-view (mediolateral oblique) DBT.

^bBiennial for women aged 55–74 years and 18 months for women aged 40–54 years.

^cNot reported but the US studies are likely to be predominantly annual screening.

^dDouble reading of DBT + DM + SM and DBT + SM.

^eRecall at either read.

^fWith consensus meeting.

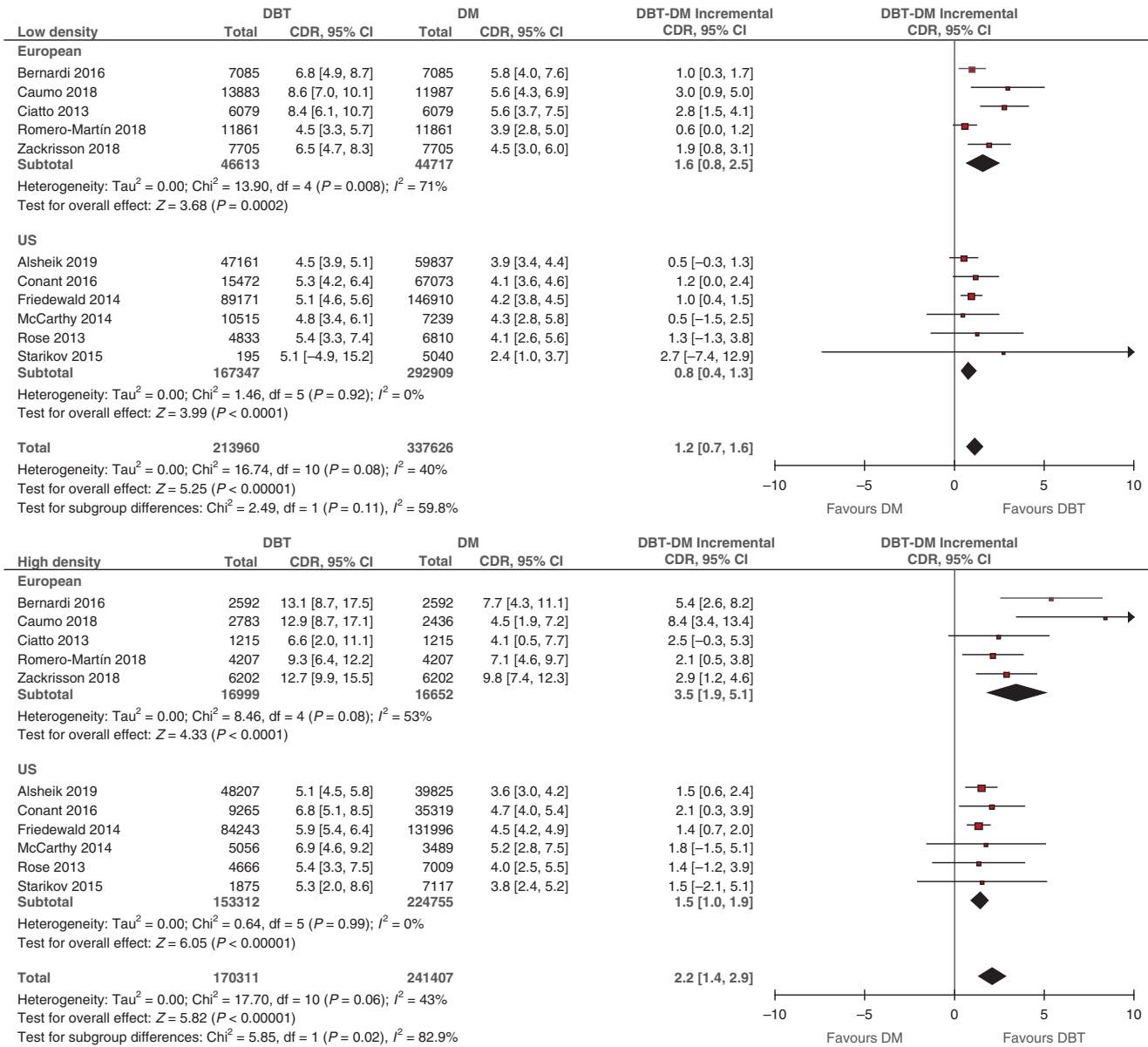


Fig. 1 Difference in cancer detection rate (incremental CDR) between DBT and DM stratified by breast density. Breast density was classified as low (BI-RADS a + b) and high (BI-RADS c + d) (see Data extraction). Squares with horizontal lines represent individual study estimates and 95% CIs. Diamonds represent pooled estimates of incremental CDR for DBT over DM and 95% CIs. Additional data were supplied by study authors for Alsheik et al. [23]. CI confidence interval, df degrees of freedom.

Incremental cancer detection rate (CDR) for DBT by breast density

Eleven studies reported on CDR, covering 384,271 participants/examinations undergoing DBT (213,960 in low and 170,311 in high density) and/or 579,033 participants/examinations undergoing DM (337,626 in low and 241,407 in high density). Supplementary Table 1 details study-specific number of screens, detected cancers and CDR for DBT and DM, stratified by breast density and screening setting. Pooled estimates of incremental CDR for DBT over DM, stratified by density and screening setting, are displayed in Fig. 1. DBT was associated with a statistically significant increase in CDR in both density strata. For European studies, the pooled incremental CDR in low density was 1.6 per 1000 screens (95% CI: 0.8, 2.5; $p < 0.001$; $I^2 = 71%$) and the pooled incremental CDR in high density was 3.5 per 1000 screens (95% CI: 1.9, 5.1; $p < 0.001$; $I^2 = 53%$). For US studies, the pooled incremental CDR in low

density was 0.8 per 1000 screens (95% CI: 0.4, 1.3; $p < 0.001$; $I^2 = 0%$) and the pooled incremental CDR in high density was 1.5 per 1000 screens (95% CI: 1.0, 1.9; $p < 0.001$; $I^2 = 0%$).

Within-study differences between density strata were pooled in Fig. 2. For all studies combined, the pooled difference in incremental CDR for high versus low density was 1.0 per 1000 screens (95% CI: 0.3, 1.6; $p = 0.003$; $I^2 = 10%$), indicating that the increase in CDR associated with DBT was statistically significantly greater in high compared with low density. When stratified by screening setting, a larger difference between density strata in incremental CDR was found for European studies (1.9 per 1000 screens; 95% CI: 0.3, 3.5; $p = 0.02$; $I^2 = 46%$) compared to US studies (0.6 per 1000 screens; 95% CI: 0.0, 1.3; $p = 0.05$; $I^2 = 0%$), but this subgroup difference was not statistically significant ($p = 0.15$).

In sensitivity analyses where only studies reporting both CDR and recall rate were included ($n = 9$), pooled estimates of

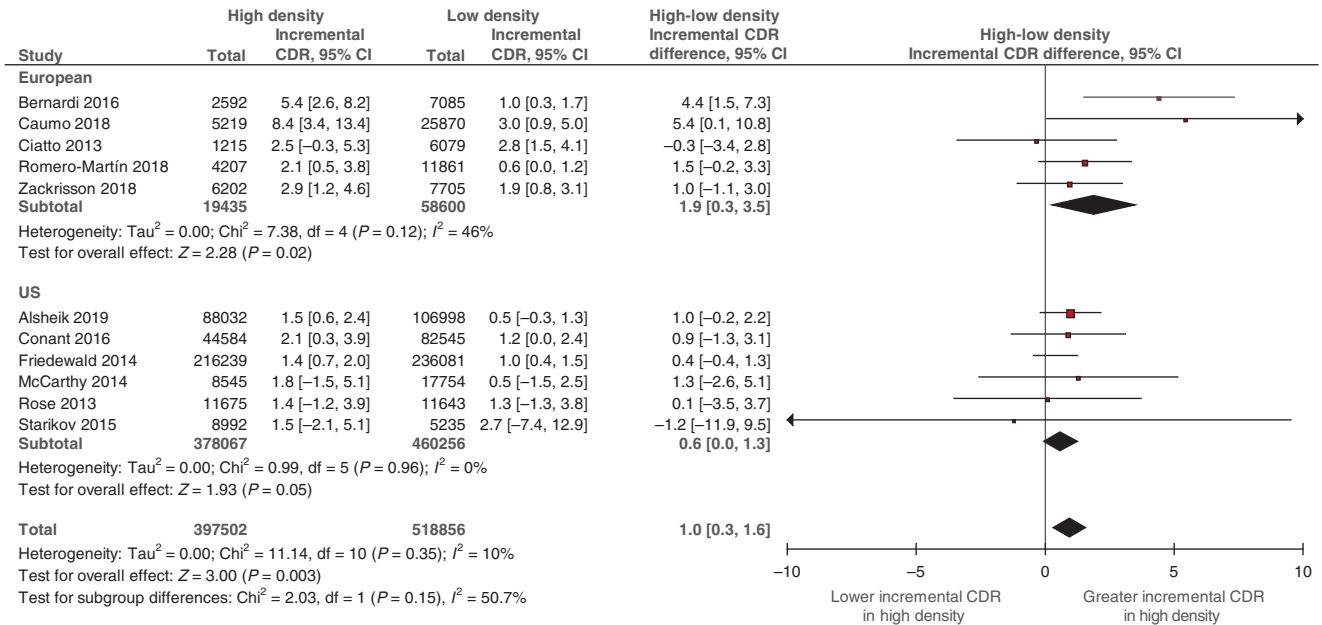


Fig. 2 Difference between high- and low-density subgroups in DBT's incremental CDR. Breast density was classified as low (BI-RADS a + b) and high (BI-RADS c + d) (see Data extraction). Squares with horizontal lines represent individual study estimates and 95% CIs. Diamonds represent pooled estimates in incremental CDR for high versus low density and 95% CIs. Additional data were supplied by study authors for Alsheik et al. [23]. CI confidence interval, df degrees of freedom.

incremental CDR (Supplementary Fig. 3) and the within-study difference between density strata (Supplementary Fig. 4) did not change substantially.

Incremental recall rate for DBT by breast density

Eleven studies reported recall data, including 490,112 participants/examinations undergoing DBT (264,675 in low and 225,437 and in high density) and/or 685,398 participants/examinations undergoing DM (396,597 in low and 288,801 in high density). Supplementary Table 2 details study-specific data for screen numbers, recalled cases and recall rates for DBT and DM, stratified by breast density and screening setting. Pooled estimates of incremental recall rate for DBT over DM, stratified by density and screening setting, are displayed in Fig. 3. For European studies, the pooled incremental recall rate was 0.2% (95% CI: -0.6%, 1.1%; $p = 0.60$; $I^2 = 93\%$) in low density and 1.0% (95% CI: -0.1%, 2.1%; $p = 0.07$; $I^2 = 84\%$) in high density. For US studies, DBT was associated with a statistically significant absolute decrease in recall rate of -1.8% (95% CI: -2.4%, -1.2%; $p < 0.001$; $I^2 = 91\%$) in low density and -3.5% (95% CI: -4.5%, -2.6%; $p < 0.001$; $I^2 = 94\%$) in high-density strata.

The pooling of within-study differences between density strata is displayed in Fig. 4. Stratification by screening setting showed a statistically significant difference between US and European studies subgroups ($p < 0.001$). In European screening studies the incremental recall rate was statistically significantly greater in high versus low density (0.8%; 95% CI: 0.3%, 1.3%; $p = 0.001$; $I^2 = 9\%$). In contrast, for US screening studies, the incremental recall rate was statistically significantly less in high versus low density (-0.9%; 95% CI: -1.4%, -0.4%; $p < 0.001$; $I^2 = 61\%$).

In sensitivity analyses where only studies reporting both CDR and recall rate were included ($n = 9$), pooled estimates of incremental recall rate (Supplementary Fig. 5) and the within-study difference between density strata (Supplementary Fig. 6) did not change substantially.

Modelled predictions of additional cancers detected and women recalled by DBT in population screening

Pooled estimates of incremental CDR and recall rate (Figs. 1 and 3), and the differences between density strata (Figs. 2 and 4), were

applied to different scenarios defined by screening setting (US or European) and the percentage of low breast density in the screening population ('median', 'maximum' and 'minimum' percentage based on density distributions in the included studies; see Supplementary Method 3). Across all scenarios, the predicted total number of additional cancers detected by DBT ranged from 9 to 25 per 10,000 screens (Table 2). The ratio of the numbers of additional cancers detected in high versus low density depended on the percentage of screens with low density. Despite evidence of greater incremental CDR in high density (Fig. 2), the number of additional cancers detected by DBT in women with low density exceeded the number in high density for the 'maximum' percentage of low-density screens. The reverse was apparent for the 'minimum' percentage estimates. These patterns were observed regardless of the screening setting.

The estimated number of additional women recalled by DBT ranged from -237 to 56 per 10,000 screens (Table 2). For European screening, DBT was associated with a relatively small increase in the number of recalls. At the 'maximum' percentage of low-density screens, the number of additional recalls was equal in the high- and low-density groups, but at lower percentages, the number of recalls was greater in high compared with low density. For US screening, the ratio of additional recalls in high versus low density reflected the pattern observed for CDR. At the 'maximum' percentage of low-density screens, the reduction in the number of women recalled was greater in the low-density than in the high-density group (and vice versa at the 'minimum' percentage).

DISCUSSION

The adoption of DBT in place of DM for population breast cancer screening has progressed rapidly, particularly in the US, whereas elsewhere there is conditional approval or restricted use of DBT in screening programmes [31]. Some population-based screening programmes do not currently endorse using DBT instead of DM but encourage its evaluation in prospective trials [32, 33]. Mammographic breast density, a long-established independent risk factor for breast cancer, has gained increased attention since the introduction of breast density legislation in the US [34, 35],

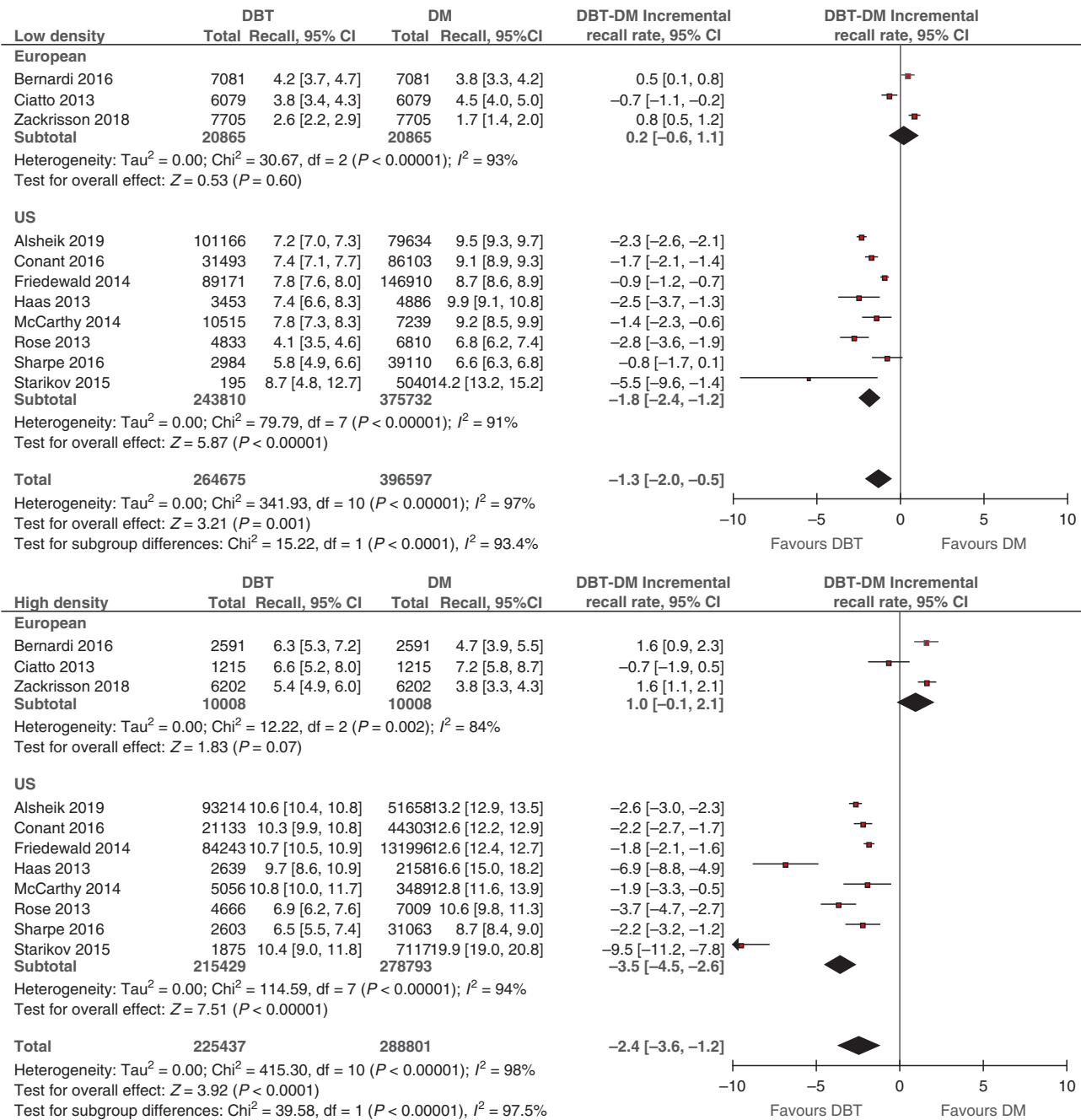


Fig. 3 Difference in recall rate (incremental recall rate) between DBT and DM stratified by breast density. Breast density was classified as low (BI-RADS a + b) and high (BI-RADS c + d) (see Data extraction). Squares with horizontal lines represent individual study estimates and 95% CIs. Diamonds represent pooled estimates of incremental recall rate for DBT over DM and 95% CIs. Additional data were supplied by study authors for Alsheik et al. [23] and Zackrisson et al. [22]. CI confidence interval, df degrees of freedom.

and there is a suggestion that DBT may be more effective for screening women with dense breasts [4, 36]. In this systematic review, we focused on estimating changes in cancer detection and recall associated with screening by DBT versus DM according to breast density. Our meta-analysis provides evidence that DBT increases cancer detection in both low- and high-density screening examinations regardless of the screening setting. Importantly, we also show that DBT has differential incremental detection (versus DM) by breast density, meaning that the increase in CDR is greater in high (versus low) density screens. Conversely, both the incremental recall rate for DBT and the differential incremental recall by density varied by screening setting.

Our estimates provide new synthesised evidence on the performance of DBT, noting that other systematic reviews [3–8] have not investigated the differential performance of DBT by density. One other review reported detection for DBT versus DM solely in screens classified as dense [37]. Our work showed that DBT detected more cancers than DM in both low- and high-density screens, and that DBT substantially improved CDR in high-density compared to low-density screens (pooled difference in incremental CDR 1.0 per 1000 screens). This improvement was more evident in studies undertaken in Europe (1.9 per 1000 screens) than in US studies (0.6 per 1000 screens). Although the difference between screening settings was not statistically

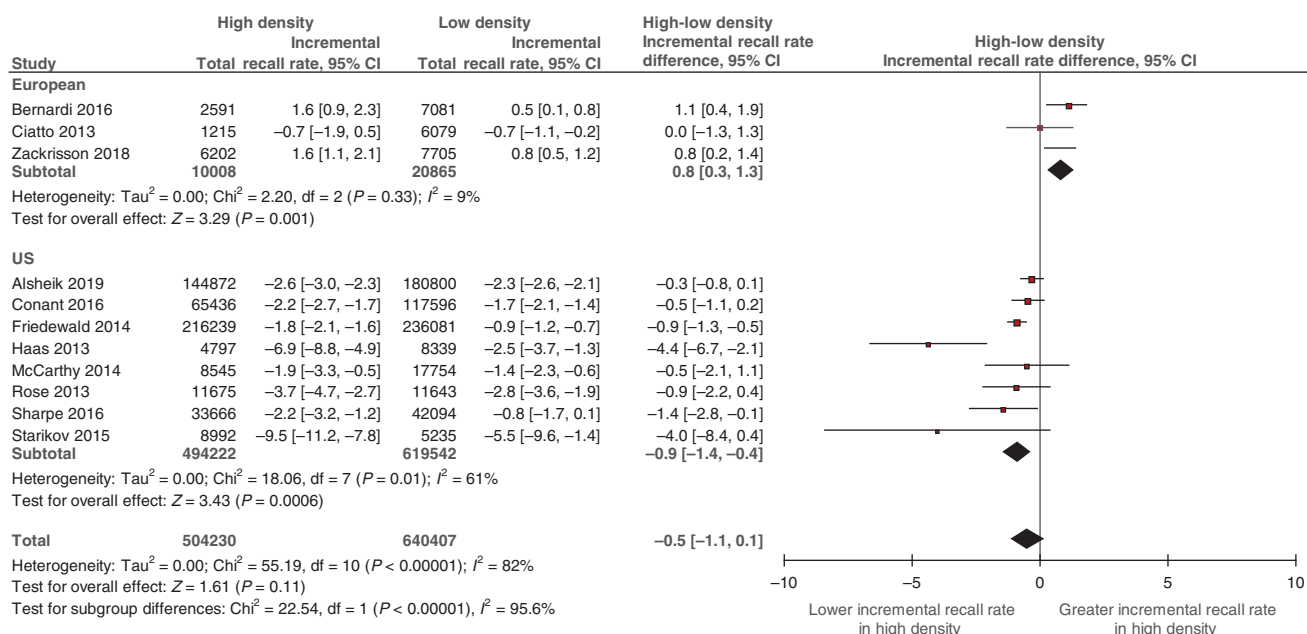


Fig. 4 Difference between high- and low-density subgroups in DBT's incremental recall rate. Breast density was classified as low (BI-RADS a + b) and high (BI-RADS c + d) (see Data extraction). Squares with horizontal lines represent individual study estimates and 95% CIs. Diamonds represent pooled estimates in incremental recall rate for high versus low and 95% CIs. Additional data were supplied by study authors for Alsheik et al. [23] and Zackrisson et al. [22]. CI confidence interval, df degrees of freedom.

significant, pooling within-study interactions is likely to have low power to detect such subgroup differences [16]. A greater contribution by DBT to cancer detection in European screening practice is likely to reflect a longer time interval between screens, however other differences between European and US screening practices (e.g. double versus single-reading) may also contribute to this difference.

The pooled difference in the incremental recall rate between low- and high-density screens differed between the screening settings. For European screening studies, a greater *increase* in DBT's incremental recall was observed in high compared with low-density screens (pooled absolute difference in recall rate 0.8%) with little heterogeneity (I² = 9%). In contrast, for US screening studies, there was a greater *decrease* in recall for DBT in high- than in low-density screens (pooled absolute difference in recall rate -0.9%). Although there was substantial heterogeneity in the magnitude of this estimate (I² = 61%), all US screening studies were consistent in the direction of the difference (Fig. 4). The opposing directions of the estimates from European and US studies are likely because the 'baseline' recall rates for DM in US screening studies were larger than those reported in European screening studies (Supplementary Table 2). Our results suggest that DBT has a beneficial effect in reducing recalls in women with dense breasts in US screening practice but may lead to increased recall in high-density screens in European screening programmes.

Our estimates of DBT's differential incremental detection and recall (versus DM) by breast density are relevant to screening programmes worldwide contemplating whether DBT should be used for population screening, if such decisions were to be based on conventional screening measures. The data provided in Table 2, for example, showing the additional detection (or effect on recall) if DBT replaced DM screening, according to the observed percentage of breast density and screening setting, can inform plans for trials or implementation studies. A screening programme targeting women aged 50-years-old and above with a large proportion of participants with low-density breasts (as would be

expected in many European programmes, and Australia's programme [38]) would improve CDR overall through more detection in both low and high-density screens. In that setting, limiting DBT to those with high density would not achieve optimal outcomes from DBT screening. In contrast, if a European screening programme comprised a large proportion of participants with high breast density, much of the incremental CDR would be achieved by offering DBT to women with dense breasts. Our results may also be relevant to planning new research in risk-tailored screening [39].

There are limitations to this work that should be considered when using our findings. The included studies reported initial detection measures and lacked data on long-term health outcomes from DBT screening. It is therefore unknown whether DBT's incremental detection will lead to incremental screening benefits by reducing breast cancer mortality. Also, most of the data reported on prevalent (initial) DBT screening, even though repeat breast screening represents the majority of screens in screening programmes. Therefore, it is possible our results may be less generalisable to repeat (incident) DBT screening. Another limitation is that we included studies that assessed breast density using BI-RADS density classification, but excluded studies using automated assessment for consistency in meta-analysis. Given that automated density measures have only been recently introduced into practice [40–43], automated density should be assessed in future meta-analyses as the evidence develops. These issues reflect the still-evolving evidence base for DBT, a limitation inherent in evaluations of new health technologies that aim to inform implementation before practice becomes established and therefore more challenging to modify [44].

In addition, we have used 'US screening' and 'European screening' to classify studies, but this classification is only broadly indicative of screening practice—we acknowledge that varying practices exist in an inter-screen time interval and screen-reading strategy. For example, US studies may not have performed all screening annually, and other factors that differ between US and European studies, such as single versus double-reading and the

Table 2. Modelled predictions of the number of additional cancers detected and cases recalled by DBT in a cohort of 10,000 screens.

Screening setting	Additional cancer detection in a cohort of 10,000 screens						All women	
	Women with low breast density			Women with high breast density				
Percentage of screens with low density ^a	Number of women	DBT-DM incremental CDR (per 1000 screens) ^b	Predicted number of additional cancers detected by DBT ^c	Number of women	DBT-DM incremental CDR (per 1000 screens) ^d	Predicted number of additional cancers detected by DBT ^c	Total predicted number of additional cancers detected by DBT per 10,000 screens ^e	
European	Median (74%)	7400	1.6	12	2600	9	21	
	Maximum (83%)	8300	1.6	13	1700	6	19	
	Minimum (55%)	5500	1.6	9	4500	16	25	
US	Median (55%)	5500	0.8	4	4500	6	10	
	Maximum (68%)	6800	0.8	5	3200	4	9	
	Minimum (37%)	3700	0.8	3	6300	9	12	
Screening setting	Additional recalls in a cohort of 10,000 screens						All women	
	Women with low breast density			Women with high breast density				
Percentage of screens with low density ^a	Number of women	DBT-DM incremental recall rate (%) ^b	Predicted number of additional cases recalled by DBT ^c	Number of women	DBT-DM incremental recall rate (%) ^d	Predicted number of additional cases recalled by DBT ^c	Total predicted number of additional cases recalled by DBT per 10,000 screens ^e	
European	Median (74%)	7400	0.2	15	2600	1.0	26	41
	Maximum (83%)	8300	0.2	17	1700	1.0	17	34
	Minimum (55%)	5500	0.2	11	4500	1.0	45	56
US	Median (55%)	5500	-1.8	-99	4500	-2.7	-121	-220
	Maximum (68%)	6800	-1.8	-122	3200	-2.7	-86	-208
	Minimum (37%)	3700	-1.8	-67	6300	-2.7	-170	-237

^aRefers to median, maximum and minimum study-level percentage of women with low breast density derived from European studies (N = 5) and US studies (N = 8) (Supplementary Method 3).

^bPooled estimates for low-density screens by screening setting subgroup from Fig. 1 (CDR) or Fig. 3 (recall).

^cNumber of women in density subgroup multiplied by incremental CDR or recall rate.

^dPooled incremental estimate for high versus low density from Fig. 2 (CDR) or Fig. 4 (recall) summed with a pooled estimate for low-density screens (Supplementary Method 3).

^eSum of additional cancers detected or cases recalled in both low- and high-density subgroups.

generally high recall rates in US studies, may account for some of the observed differences in incremental CDR and recall rates.

Internationally, the majority of population breast cancer screening programmes use DM, but many are contemplating the potential role of DBT screening. This is occurring in an evolving population screening landscape that includes deliberation regarding density notification, and risk-tailored breast screening. Our meta-analysis provides timely comparative estimates for DBT and DM screening showing that DBT has differential incremental cancer detection and recall by breast density. Therefore, our synthesised evidence may assist screening policy, planning of research and individual screening recommendations.

DATA AVAILABILITY

All data generated or analysed during this study are included in this published article and its Supplementary Information files.

REFERENCES

- Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. *Breast*. 2013;22:101–8.
- Li T, Marinovich ML, Houssami N. Digital breast tomosynthesis (3D mammography) for breast cancer screening and for assessment of screen-recalled findings: review of the evidence. *Expert Rev Anticancer Ther*. 2018;18:785–91.
- Alabousi M, Wadera A, Kashif Al-Ghita M, Kashef Al-Ghetara R, Salameh J-P, Pozdnyakov A, et al. Performance of digital breast tomosynthesis, synthetic mammography, and digital mammography in breast cancer screening: a systematic review and meta-analysis. *J Natl Cancer Inst*. 2021;113:680–90.
- 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;110:942–9.
- Yun SJ, Ryu C-W, Rhee SJ, Ryu JK, Oh JY. 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;164:557–69.
- Hodgson R, Heywang-Köbrunner SH, Harvey SC, Edwards M, Shaikh J, Arber M, et al. Systematic review of 3D mammography for breast cancer screening. *Breast*. 2016;27:52–61.
- Zeng B, Yu K, Gao L, Zeng X, Zhou Q. Breast cancer screening using synthesized two-dimensional mammography: a systematic review and meta-analysis. *Breast*. 2021. <https://doi.org/10.1016/j.breast.2021.07.016>
- Alabousi M, Zha N, Salameh J-P, Samoilov L, Sharifabadi AD, Pozdnyakov A, et al. Digital breast tomosynthesis for breast cancer detection: a diagnostic test accuracy systematic review and meta-analysis. *Eur Radiol*. 2020;30:2058–71.
- Sickles E, D'Orsi C, Bassett L. ACR BI-RADS® mammography. In: D'Orsi, C, Sickles, E, Mendelson, E, Morris, E, editors. *ACR BI-RADS® atlas, breast imaging reporting and data system*. Reston, VA: American College of Radiology; 2013.
- Huo CW, Chew GL, Britt KL, Ingman WV, Henderson MA, Hopper JL, et al. Mammographic density—a review on the current understanding of its association with breast cancer. *Breast Cancer Res Treat*. 2014;144:479–502.
- Houssami N, Hunter K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *NPJ Breast Cancer*. 2017;3:12–3.
- Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339:b2535.
- D'Orsi CJ, Bassett LW, Berg WA, Feig SA, Jackson VP, Kopans DB. BI-RADS: mammography. In: D'Orsi CJ, Mendelson EB, Ikeda DM, et al., editors. *Breast imaging reporting and data system: ACR BI-RADS – breast imaging atlas*. Reston, VA: American College of Radiology; 2003.
- Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–36.
- Deeks JH, Higgins JPT, Altman DG. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane handbook for systematic reviews of interventions version 6.2*. London: Cochrane; 2021.
- Fisher DJ, Copas AJ, Tierney JF, Parmar MKB. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. *J Clin Epidemiol*. 2011;64:949–67.
- Thompson SG, Higgins JPT. Treating individuals 4: Can meta-analysis help target interventions at individuals most likely to benefit? *Lancet*. 2005;365:341–6.
- Bernardi D, Macaskill P, Pellegrini M, Valentini M, Fantò C, Ostillo L, 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;17:1105–13.
- Caumo F, Zorzi M, Brunelli S, Romanucci G, Rella R, Cugola L, et al. Digital breast tomosynthesis with synthesized two-dimensional images versus full-field digital mammography for population screening: outcomes from the Verona screening program. *Radiology*. 2018;287:37–46.
- Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol*. 2013;14:583–9.
- Romero Martín S, Raya Povedano JL, Cara García M, Santos Romero AL, Pedrosa Garriguet M, Álvarez Benito M. Prospective study aiming to compare 2D mammography and tomosynthesis + synthesized mammography in terms of cancer detection and recall. From double reading of 2D mammography to single reading of tomosynthesis. *Eur Radiol*. 2018;28:2484–91.
- Zackrisson S, Lång K, Rosso A, Johnson K, Dustler M, Förnvik D, et al. One-view breast tomosynthesis versus two-view mammography in the Malmö Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. *Lancet Oncol*. 2018;19:1493–503.
- Alsheik NH, Dabbous F, Pohlman SK, Troeger KM, Glikich RE, Donadio GM, et al. Comparison of resource utilization and clinical outcomes following screening with digital breast tomosynthesis versus digital mammography: findings from a learning health system. *Acad Radiol*. 2019;26:597–605.
- Conant EF, Beaber EF, Sprague BL, Herschorn SD, Weaver DL, Onega T, 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;156:109–16.
- Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *J Am Med Assoc*. 2014;311:2499–507.
- Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology*. 2013;269:694–700.
- McCarthy AM, Kontos D, Synnestvedt M, Kay See TAN, Heitjan DF, Schnell M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *J Natl Cancer Inst*. 2014;106:dju316.
- Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton JR. Implementation of breast tomosynthesis in a routine screening practice: an observational study. *Am J Roentgenol*. 2013;200:1401–8.
- Sharpe JRE, Venkataraman S, Phillips J, Dialani V, Fein-Zachary VJ, Prakash S, 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;280:981–981.
- Starikov A, Drotman M, Hentel K, Katzen J, Min RJ, Arleo EK. 2D mammography, digital breast tomosynthesis, and ultrasound: which should be used for the different breast densities in breast cancer screening? *Clin Imaging*. 2015;40:68–71.
- European Commission Initiative on Breast Cancer. Tomosynthesis plus digital mammography vs. digital mammography alone in screening. <https://healthcare-quality.jrc.ec.europa.eu/european-breast-cancer-guidelines/screening-tests/DBT-plus-DM-or-DM>. Accessed 1 Aug 2021.
- BreastScreen Australia. Position statement on the use of tomosynthesis within BreastScreen Australia Services. Department of Health. 2014.
- Houssami N, Lockie D, Clemson M, Pridmore V, Taylor D, Marr G, et al. Pilot trial of digital breast tomosynthesis (3D mammography) for population-based screening in BreastScreen Victoria. *Med J Aust*. 2019;211:357–62.
- Kyanko KA, Hoag J, Busch SH, Aminawung JA, Xu X, Richman IB, et al. Dense breast notification laws, education, and women's awareness and knowledge of breast density: a nationally representative survey. *J Gen Intern Med*. 2020;35:1940–5.
- Huang S, Houssami N, Brennan M, Nickel B. The impact of mandatory mammographic breast density notification on supplemental screening practice in the United States: a systematic review. *Breast Cancer Res Treat*. 2021;187:11–30.
- Østerås BH, Martinsen ACT, Gullien R, Skaane P. Digital mammography versus breast tomosynthesis: impact of breast density on diagnostic performance in population-based screening. *Radiology*. 2019;293:60–8.
- Phi XA, Tagliafico A, Houssami N, Greuter MJW, de Bock GH. Digital breast tomosynthesis for breast cancer screening and diagnosis in women with dense breasts – a systematic review and meta-analysis. *BMC Cancer*. 2018;18:380.
- Noguchi N, Marinovich ML, Wylie EJ, Lund HG, Houssami N. Screening outcomes by risk factor and age: evidence from BreastScreen WA for discussions of risk-stratified population screening. *Med J Aust*. 2021. <https://doi.org/10.5694/mja2.51216>

39. Esserman LJ. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer*. 2017;3:34–7.
40. Aase HS, Holen AS, Pedersen K, Houssami N, Haldorsen IS, Sebuodegard S, et al. A randomized controlled trial of digital breast tomosynthesis versus digital mammography in population-based screening in Bergen: interim analysis of performance indicators from the To-Be trial. *Eur Radiol*. 2019;29:1175–86.
41. Hofvind S, Moshina N, Holen AS, Danielsen AS, Lee CI, Houssami N, et al. Interval and subsequent round breast cancer in a randomized controlled trial comparing digital breast tomosynthesis and digital mammography screening. *Radiology*. 2021;300:66–76.
42. Moshina N, Aase HS, Danielsen AS, Haldorsen IS, Lee CI, Zackrisson S, et al. Comparing screening outcomes for digital breast tomosynthesis and digital mammography by automated breast density in a randomized controlled trial: results from the to-be trial. *Radiology*. 2020;297:522–31.
43. Osteras BH, Martinsen ACT, Gullien R, Skaane P. Digital mammography versus breast tomosynthesis: impact of breast density on diagnostic performance in population-based screening. *Radiology*. 2019;293:60–8.
44. Buxton, M. Problems in the economic appraisal of new health technology: the evaluation of heart transplants in the UK. *Economic appraisal of health technology in the European Community*. Oxford: Oxford Medical Publications; 1987. pp. 103–18.

AUTHOR CONTRIBUTIONS

NH and MLM conceived the study. MLM did the literature search. TL identified and assessed eligible studies, and MLM made final decisions about the inclusion of articles. TL, AZ and NN extracted data from included studies. TL and MLM undertook quality assessment. MLM designed the data analyses, and TL and MLM conducted the analyses. TL, NH and MLM interpreted findings. TL drafted the initial manuscript, and MLM and NH edited the subsequent versions. All authors read and approved the final version of the manuscript.

FUNDING

Open Access funding enabled and organized by CAUL and its Member Institutions.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All analyses were based on previously published studies; thus, no ethical approval and patient consent are required.

CONSENT FOR PUBLICATION

Not applicable.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41416-022-01790-x>.

Correspondence and requests for materials should be addressed to Tong Li.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022