



Research Paper

Interval breast cancer rates for digital breast tomosynthesis versus digital mammography population screening: An *individual participant data* meta-analysis

Nehmat Houssami^{a,*}, Solveig Hofvind^b, Anne L. Soerensen^c, Kristy P. Robledo^c, Kylie Hunter^c, Daniela Bernardi^d, Kristina Lång^e, Kristin Johnson^e, Camilla F. Aglen^b, Sophia Zackrisson^e

^a Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

^b Section of Breast Screening, Cancer Registry of Norway, Oslo, Norway

^c NHMRC Clinical Trials Centre, University of Sydney, Sydney, Australia

^d Breast Imaging and Screening Unit, Department of Radiology, Humanitas Research Hospital, Milan, Italy

^e Department of Translational Medicine, Diagnostic Radiology, Lund University, Skane University Hospital, Malmö, Sweden

ARTICLE INFO

Article History:

Received 19 January 2021

Revised 2 March 2021

Accepted 3 March 2021

Available online 20 March 2021

Keywords:

Breast cancer

Cancer screening

Digital breast tomosynthesis

Interval cancer

Mammography

Meta-analysis

ABSTRACT

Background: Digital breast tomosynthesis (DBT) improves breast cancer (BC) detection compared to mammography, however, it is unknown whether this reduces *interval cancer rate* (ICR) at follow-up.

Methods: Using *individual participant data* (IPD) from DBT screening studies (identified via periodic literature searches July 2016 to November 2019) we performed an IPD meta-analysis. We estimated ICR for DBT-screened participants and the difference in pooled ICR for DBT and mammography-only screening, and compared interval BC characteristics. Two-stage meta-analysis (study-specific estimation, pooled synthesis) of ICR included random-effects, adjusting for study and age, and was estimated in age and density subgroups. Comparative screening sensitivity was calculated using screen-detected and interval BC data.

Findings: Four prospective DBT studies, from European population-based programs, contributed IPD for 66,451 DBT-screened participants: age-adjusted pooled ICR was 13.17/10,000 (95%CI: 8.25–21.02). Pooled ICR was higher in the high-density (21.08/10,000; 95%CI: 6.71–66.27) than the low-density (8.63/10,000; 95%CI: 5.25–14.192) groups ($P = 0.03$) however estimates did not differ across age-groups ($P = 0.32$). Based on two studies that also provided data for 153,800 mammography screens (age-adjusted ICR 17.69/10,000; 95%CI: 13.22–23.66), DBT's pooled ICR was 16.83/10,000 (95%CI: 11.89–23.82). Comparative meta-analysis showed a non-significant difference in ICR (-0.44/10,000; 95%CI: -11.00–10.11) and non-significant difference in screening sensitivity (6.79%; 95%CI: -0.73–14.87%) between DBT and DM but a significant pooled *difference* in cancer detection rate of 33.49/10,000 (95%CI: 23.88–43.10). Distribution of interval BC prognostic characteristics did not differ between screening modalities except that those occurring in DBT-screened participants were significantly more likely to be negative for axillary-node metastases ($P = 0.005$).

Interpretation: Although heterogeneity in ICR estimates and few datasets limit recommendations, there was no difference between DBT and mammography in pooled ICR despite DBT increasing cancer detection.

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Introduction

Mammography population screening has been shown to reduce breast cancer (BC) mortality [1–3]; in some screening settings, mammography is being replaced by digital breast tomosynthesis (DBT). This pseudo-3D-mammography technology has been adopted in breast imaging practice as it reduces overlapping parenchyma (relative to mammography) which might reduce false-positive

interpretations and enhance visualisation and detection of BC [4–7]. Studies of DBT screening have shown that DBT, used in combination with acquired or synthesized 2D-images, detects an estimated additional 1.6 cancers/1000 screens compared to mammography [6]. However, it remains unknown whether DBT's incremental BC detection improves outcomes at follow-up, such as reducing the risk of an interval BC, beyond what can be achieved with mammography screening.

BC mortality as an endpoint requires very large studies with long follow-up, so conducting such trials may not be feasible for comparing the effectiveness of DBT and digital mammography (DM) [8]. We

* Corresponding author.

E-mail address: nehmat.houssami@sydney.edu.au (N. Houssami).

Research in Context

Evidence before this study

To address the knowledge gap about interval breast cancer rates for DBT versus mammography screening, and to plan an *individual participant data* (IPD) meta-analysis, we searched the literature in July 2016 to identify prospective studies that investigated DBT screening and reported data on cancer detection and (potentially) interval cancer data at follow-up; four non-randomised prospective studies met eligibility criteria for our proposed study. Since our initial literature search (using an updated search April 2020) two studies have reported that at follow-up of DBT screening participants (Houssami 2018; Bernardi 2020) there was no significant difference in interval cancer rates between DBT and mammography screening, and two randomised trials (Weigel 2018; Pattacini 2018) have indicated that interval cancer rates will be reported as part of trial outcomes.

Added value of this study

This is the only IPD meta-analysis of interval breast cancer rates in DBT-screened populations, showing that there was no difference in interval breast cancer rates between DBT and mammography screening, although DBT improved screening sensitivity by significantly increasing the rate of screen-detected cancers. Our comparative meta-analysis of interval cancer characteristics provides new prognostic insights showing that interval cancers in DBT-screened participants were more likely to be negative for axillary nodal metastases than those occurring in mammography-screened participants.

Implications of all the available evidence

Breast cancer population screening programs grappling with decisions regarding the incremental benefits and harms of transitioning to DBT, for whom to date there is little evidence on outcomes for DBT vs mammography screening beyond evidence on initial detection metrics, may find our results relevant to screening policy decisions. Our findings can be used to prioritise efforts towards large-scale trials of DBT, rather than prompt adoption, and ensuring that such evaluations consider a broad range of screening end-points including interval BC rates.

proposed using interval BC rate as a surrogate for screening effectiveness, and described sourcing data on this outcome from prospective studies in a published study plan [8]. Given that interval cancers are not detected at screening and progress to clinical presentation (often due to symptoms), we hypothesised that increased screen-detection of BC would lead to a reduction in interval cancers from DBT relative to DM screening which would be expected to have benefit [8]. This could also inform inferences on whether the additional BC detection from DBT is adding screening benefit, or adding to over-diagnosis (through detection of indolent tumours) or lead-time.

We therefore undertook a collaborative *individual participant data* (IPD) meta-analysis to assess the impact of DBT screening (versus DM) on interval BC rates in population screening. Because interval cancers are an important but *infrequent* outcome, pooling datasets gives increased precision in estimation of this endpoint. Using IPD also allows adjustments for differences between studies and study-groups and supports subgroup analyses. This paper reports the primary endpoint of the IPD meta-analysis, interval BC rates, based on prospective studies comparing DBT and DM screening.

Methods

An international collaborative meta-analysis using IPD was established in 2016 and detailed in a protocol outlining rationale, study eligibility criteria, literature search method, end-points, and requested data [8]. Study eligibility was based on the following pre-defined criteria: studies that investigated DBT (interpreted alone, or with acquired or synthetic 2D-images) compared to DM screening; used a prospective design for DBT screening; reported data on screen-detected *and* interval cancers (numbers or rates), or stated that interval cancer data would be obtained at follow-up [8]; and used predominantly *biennial* screening (aligning with practice in population-based screening programs) [8]. Because some prospective DBT trials used a paired design, comparing DBT and DM screen-reads *within-woman*, [6] interval BC rates from those studies would represent outcomes for participants screened with DBT (since all had DBT and none had DM-alone). For those studies, interval BC data were requested for a comparison cohort screened with DM only from the same screening program and population.

2.1 Literature searching and collaborative process

Literature searches (MEDLINE, EMBASE and EBM Reviews) were performed at several time-points from July 2016 to November 2019; Fig. 1 describes the literature search, the cumulative search results

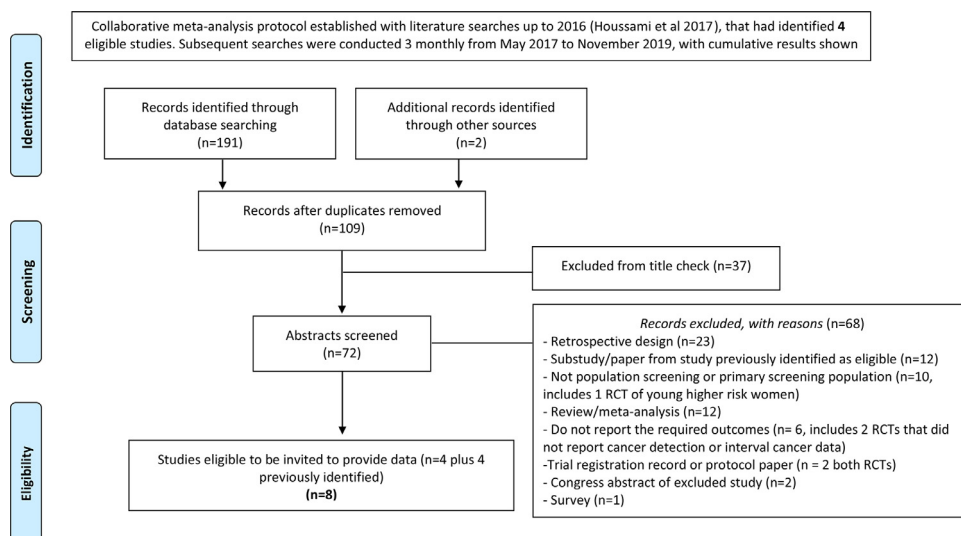


Fig. 1. Search strategy and PRISMA flow diagram.

and study identification process in a flow-diagram (including eligible studies [4,9–15] and reasons for excluding studies). Investigators of eligible studies were contacted via email and invited to collaborate (up to three contact attempts were made), and were informed of the study protocol and required data variables. Agreement to provide IPD was documented, and data were accepted for inclusion in the analysis up to June 2020; we anticipate updating the analyses if additional studies provide data in the future.

Ethics: The IPD meta-analysis involved a data-provision plan for which individual studies secured institutional ethics approval, and was also approved by the University of Sydney Ethics Committee for the proposed analyses of IPD sourced from existing studies (Project no. 2017/143) [8].

2.2 Data variables and endpoints

Requested data for screened participants included: age, screen date, mammographic density (as measured in each study), screening round, self-reported symptoms, whether recalled to assessment, outcome (negative, screen-detected or interval cancer), cancer characteristics (histology, size, node status), date of interval BC diagnosis. For each study, variables were collected in a standardised database, including standardised categorisation, to enable pooling of data. Study databases were checked against published reports – differences were discussed with study investigators to achieve resolution or identify the reason for discrepancies (including those imposed by data privacy requirements, such as provision of month/year only for date of birth by some trials).

The *primary endpoint* was an interval BC (invasive or *in-situ*), and screening sensitivity (inclusive of BC detection data). Interval BC classification was based on ascertainment of screened participants within 24 months following a negative screen or assessment. Secondary endpoints, comprising cancer detection rates and characteristics, screening recall and predictive value will be reported in future publications.

2.3 Study quality

Eligibility criteria for this IPD meta-analysis required that DBT studies be *prospective*, to minimise selection of participants to DBT screening. In addition, we describe the characteristics of each study to guide interpretation of quality in the context of population screening, specifically: methods of the prospective DBT study (whether randomisation used); design of comparison groups (whether concurrent, and whether from the same screening program); and the methods used to ascertain interval BCs (whether population cancer registry checked). This information is reported in study-specific summaries and flow-charts (Table 1; Online Appendix-1) supplemented by quality appraisal using the QUADAS-2 tool [16] (online Appendix-1).

2.4 Statistical methods

Study characteristics were summarised descriptively. The number of participants and screens, number of detected and interval cancers, and time to interval cancer diagnosis, were calculated. Screen characteristics were summarised within studies and study-groups (DBT and DM) using the median [interquartile range, IQR] for continuous variables and counts (percent) for categorical variables. Interval cancer characteristics were summarised by study-groups and compared using the Chi-squared test for categorical variables and Kruskal-Wallis test for continuous variables.

Two-stage IPD meta-analysis was performed: study-specific estimates were first calculated using logistic regression, and then synthesised using random-effects meta-analysis. Random-effects models allow for both within-study sampling variability as well as

heterogeneity between studies when calculating pooled estimates. If a study had multiple screens from the same participant, a random effect for participant was included within the logistic regression. All meta-analyses were adjusted, by age-group (<50, 50–59, 60+) or by age (continuous) using the median age as reference, and where data were provided also by density (reclassified into low, high). Rate ratios and risk difference with 95% confidence intervals were reported. Sensitivity analyses examined the effect of excluding the small number of participants self-reporting symptoms. *P*-values <0.05 were considered statistically significant. Analyses were performed using the metafor package [17] (version 2.4–0) for R (version 3.6.3) [18].

Results

We identified eight eligible studies of DBT screening [4,9–15] (Fig. 1), and from these four studies [4,10–12] contributed data according to the IPD meta-analysis protocol inclusive of interval BC data. Collectively these four studies, which have reported outcomes in seven papers, [4,10–12,19–21] contributed IPD for 66,451 DBT-screened participants; two of these studies [10,12] also provided data for 153,800 DM-only screens from 101,532 participants (Table 1). Three of the studies were prospective non-randomised trials, [4,10,11] one was a prospective cohort study [12]; all four studies investigated DBT in organised population-based screening programs. Study-specific characteristics, including methods and quality evaluation, and data contribution to the meta-analysis are summarised in Table 1, supplemented by study flow-charts and appraisal (online-only Appendix-1); none of the studies were randomised controlled trials (RCT). Median age of DBT-screened populations was similar across studies (Table 1), however it differed between DBT and DM groups for one study [10]. Screening participants in all studies received DBT acquisitions with either DM or synthesised 2D-images, hence for the purpose of interval BC analyses, these will be referred to as DBT-screened participants. However, some studies compared screen-reading (*within* participants) using DBT alone, DBT with DM, or DBT with synthesised 2D-images (Table 1).

3.1 Interval BC in DBT-screened participants

Across all studies 105 interval BCs were diagnosed in DBT-screened participants at a median 1.3 [IQR 0.9, 1.7] years from screening date (study-specific data shown in Fig. 2). Crude interval BC rate was 15.97/10,000 DBT screens (95%CI: 13.11–19.23), and the *age-adjusted* (referent 56 years) modelled estimate was 13.17/10,000 (95%CI: 8.25–21.02); Fig. 2 shows there was significant heterogeneity in estimates across studies. Fig. 3 displays study-specific and pooled estimates in age subgroups: although interval BC rate was lower in the 50–59 years age-group (10.08/10,000; 95%CI: 4.88–20.82) relative to younger and older age-groups, there was no evidence that estimates differed significantly by age-groups (*P* = 0.32).

Fig. 4 shows modelled estimates for density subgroups (three studies contributed data [4,10,11]), indicating significantly higher interval BC rate in the high-density (21.08/10,000; 95%CI: 6.71–66.27) than the low-density (8.63/10,000; 95%CI: 5.25–14.192) group (*P* = 0.03). The modelled interval BC rate ratio for density (adjusted for study and age) was 2.92 (95%CI: 1.48–5.84) for high versus low density (*P* = 0.002). Analysis by screen-round was limited to two studies [11,12] of which one was predominantly repeat screening (online-only Appendix-2: Fig. A).

3.2 Interval BC for DM screens & comparative interval BC rates for DBT versus DM

For the two larger studies (Table 1: MBTST, OVVV) [10,12] contributing DM-only screening data, there were 276 interval cancers at a median 1.2 [IQR 0.7, 1.6] years from screening. Crude interval BC rate was 17.58/10,000 DM screens (95%CI: 13.78–22.44) and the *age-*

Table 1Characteristics of *prospective* studies of DBT population screening contributing data to the IPD meta-analysis.

Study [publications reporting outcomes]	Design	Number of screens [N interval cancer analysis] [*]	Median age [IQR]	Double-reading and recall process	Comparison group [#] screened with digital mammography (DM) only			Interval cancer ascertainment
					Same time-frame? same program?	N screens [N interval cancer analysis] [*]	Median age [IQR]	
STORM-1 trial [Ciatto [4]; Houssami [19]]	Prospective trial comparing sequential screen-reading of DM with DBT+DM (paired data) in women \geq 48 years	7292 [7235]	58 [54, 63]	Independent double-read, recall if either reader recalls	NA	NA	NA	Check of regional cancer registry, and pathology and hospital databases; 2 year follow-up
MBTST (Malmo trial) [Zackrisson [10]; Johnson [20]]	Prospective trial comparing screen-reading using one-view DBT with 2-view DM (paired data) in women \geq 40 years	14,848 [14,711]	57.4 [48.9, 65.3]	Independent double-read, consensus meeting if score \geq 3 by either reader	Yes (slightly longer in 2015); yes	96,037 from 43,769 women [95,497]	52.3 [45.5, 61.4]	Population cancer registry linkage; 2 year follow-up (18 months in subgroup where this screening interval recommended)
STORM-2 trial [Bernardi [11]]	Prospective trial comparing in separate arms in women aged \geq 49 years: screen-reading of DBT [†] with DM (paired data), and screen-reading DBT +DM with DM (paired data); only one set of paired data used in analysis	9672 [9587]	58 [53, 63]	Two parallel independent double-reading arms, recall if any reader recalls	NA	NA	NA	Check of regional cancer registry, and pathology and hospital databases; 2 year follow-up
OVVV study [Hofvind [12]; Hovda [21]]	Prospective cohort study of DBT [†] screening vs concurrent (DM-only screened) cohort in women 50–69 years at screening	34,639 [34,315]	59 [54, 64]	Independent double-read, consensus meeting if discordant scores	Yes; yes	57,763 [57,408]	59 [54, 64]	Cancer Registry of Norway Population cancer registry linkage; 2 year follow-up

STORM = Screening with Tomosynthesis or Mammography; MBTST= Malmo Breast Tomosynthesis Screening Trial; OVVV = Oslo, Vestfold & Vestre Viken study; NA = not applicable (did not provide data).

^{*} Number of screens represents individual participants except for the DM comparison cohort which included repeat screens - number in squared brackets represents number used in the analysis of interval cancer rates using negative screens as the denominator.[#] Comparison groups from same population screening programs allowed comparison of interval cancer rates (OVVV compared services within Norwegian program).[†] Indicates study used DBT with synthetic 2D-images reconstructed from the DBT acquisitions.

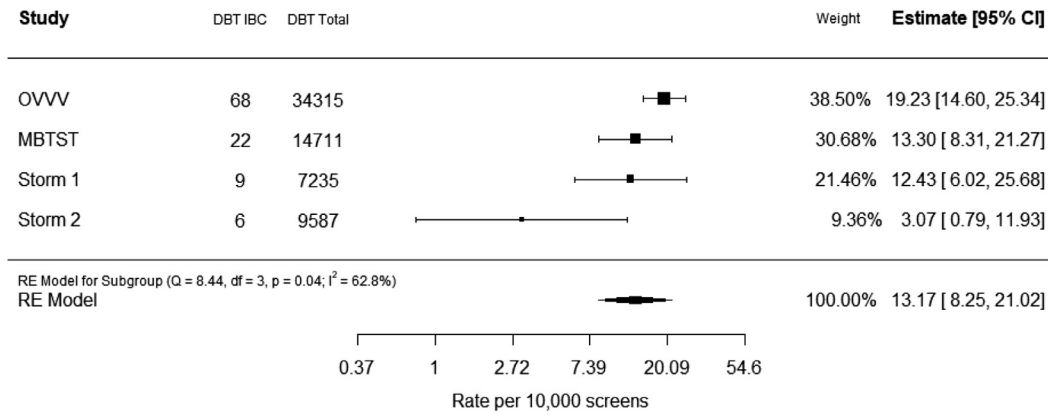


Fig. 2. Interval breast cancers (IBC) in DBT-screened participants: study-specific data and rates, and pooled estimate, adjusted for age per 10,000 screens (reference age 56 years).

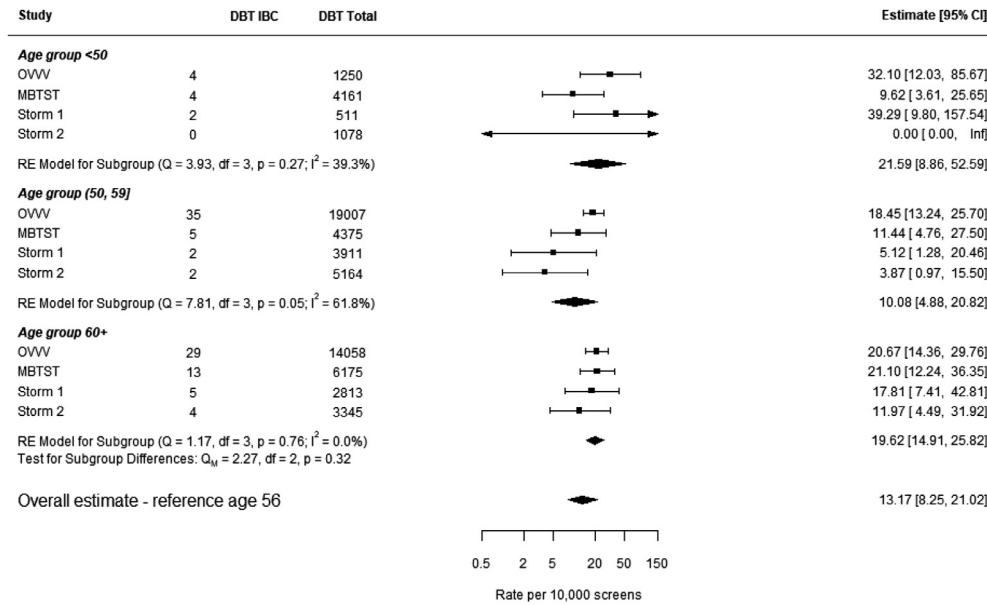


Fig. 3. Two-stage subgroup meta-analysis: Estimated interval breast cancer (IBC) rates by age-group in DBT-screened participants.

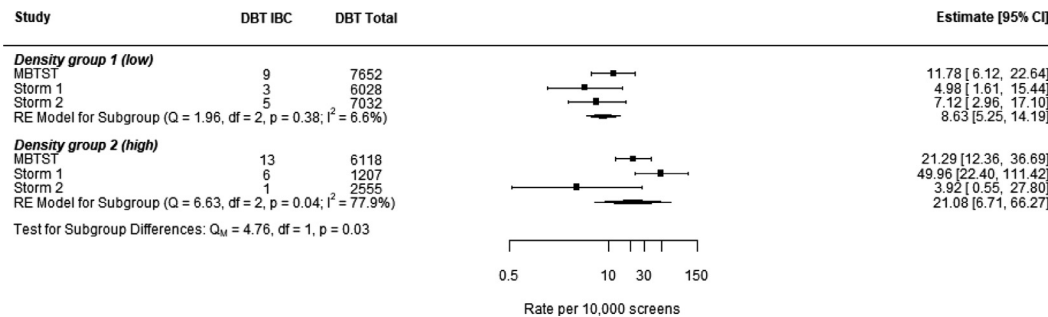


Fig. 4. Two-stage subgroup meta-analysis: Estimated interval breast cancer (IBC) rates by breast density* in DBT-screened participants. *Some studies reported density using four BI-RADS density categories (1–4, also referred to as A–D) and some reclassified these categories into two density groups, low (1–2) vs high (3–4) density, we have therefore used the two-category classification to enable pooled analysis.

adjusted (referent 56 years) modelled estimate was 17.69/10,000 (95%CI: 13.22–23.66). For these two studies, the age-adjusted modelled estimate of interval BC rate for DBT screening was 16.83/10,000 (95%CI: 11.89–23.82). Temporal distribution showed that a lower proportion of interval BCs was diagnosed in year one of the inter-screening interval for DBT than DM (31.1% for DBT and 40.2% for DM) as shown in online Appendix-2 (Fig. B).

Comparative meta-analyses (Figs. 5 and 6) show an unadjusted interval BC rate difference of 0.08/10,000 (95%CI: –8.94–9.11), and an age-

adjusted estimate of –0.44/10,000 (95%CI: –11.00–10.11) indicating a non-significant difference in interval BC rate between DBT and DM screening; Fig. 5 also shows significant heterogeneity in estimates between studies. Fig. 6 shows the rate ratio in subgroup meta-analysis for age, indicating there were no significant differences in interval BC rates between DBT and DM across age-groups (P = 0.86). Sensitivity analyses for these comparative meta-analyses, whereby the small number of screens with self-reported breast symptoms (<0.4% both groups) were excluded, did not alter results (data not shown).

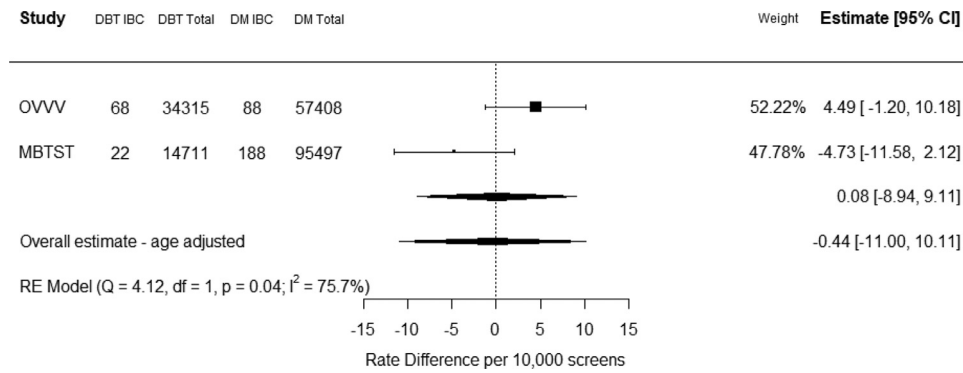


Fig. 5. Two-stage comparative meta-analysis of the difference in *interval breast cancer* (IBC) rates between DBT and DM screens (unadjusted and age-adjusted estimates).

Comparative meta-analysis was also stratified by age subgroups showing similar results (ie. there were no differences in interval breast cancer rates between modalities across age-groups).

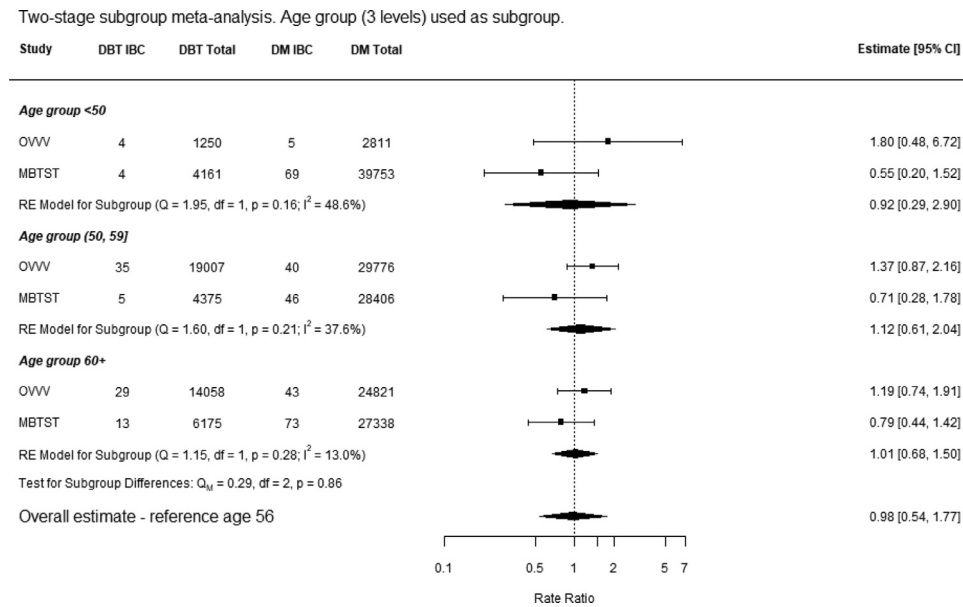


Fig. 6. Comparative meta-analysis: Interval breast cancer (IBC) rate ratios for DBT vs DM (referent), shown overall and by age subgroups.

3.3 Cancer detection rate (CDR) and comparative screening sensitivity

Amongst DBT-screened participants across all studies there were 603 screen-detected BCs, a pooled crude CDR of 91.71/10,000 DBT screens; online-only Appendix-2, Fig. C, shows study-specific detection data, and pooled estimates, indicating homogeneous cancer detection rates for DBT. For the two studies [10,12] that also provided data for DM-only screening, there were 895 screen-detected BCs yielding a pooled crude CDR of 58.81/10,000 DM screens (Fig. C, online-Appendix-2, shows study-specific detection data, and age-adjusted pooled estimate). There was a significant pooled *difference* in CDR of 33.49/10,000 (95%CI: 23.88–43.10) between DBT and DM. As also shown in Fig. D (online Appendix-2) this significant difference in CDR between DBT and DM screens was evident in comparative meta-analysis across all age-groups.

Meta-analysis comparing screening sensitivity adjusted for age (Fig. 7) estimated a sensitivity difference of 6.79% (95%CI: -0.73–14.87%) between DBT and DM, or a risk ratio of 1.07 (95%CI: 0.99–1.15). As also shown in study-specific estimates [10,12] (Fig. 7), the age-adjusted estimate for one of the studies (MBTST [10]) showed a significant sensitivity difference of 11.2% (95%CI: 3.02–20.01%) between DBT and DM. Sensitivity analyses for the above-reported meta-analyses, that excluded screens with self-reported symptoms (<0.4% both groups), did not alter results.

3.4 Interval cancer characteristics

Table 2 summarises the characteristics of the 366 interval BCs, comprising predominantly (94.3%) invasive cancers, from studies contributing data for DBT and DM screening [10,12]. The median tumour size of invasive interval cancers was 18.0 mm (IQR 13.0, 27.0): median size was 16.0 mm (IQR 12.5, 25.0) in the DBT group vs 19.5 mm (IQR 14.0, 27.0) in the DM group ($P = 0.47$). Table 2 describes the distribution of histological type, pathological tumour size (pT) category and grade, and lymph node status. Lower proportions of invasive lobular cancer, pT2 tumours, and grade 3 cancers were noted amongst interval cancers in DBT-screened relative to DM-screened groups, although distributions did not significantly differ between modalities (all $p > 0.13$). There was a significant difference in the distribution of axillary node metastases amongst interval cancers between DBT and DM screening ($P = 0.005$) – those in DBT-screened participants were more likely to be negative for metastases.

Discussion

Digital breast tomosynthesis (DBT) is being adopted for BC screening in some settings because it increases BC detection rates compared to DM; its effect on recall is variable in European studies however DBT reduces screening recall in studies conducted in the USA[6]. The

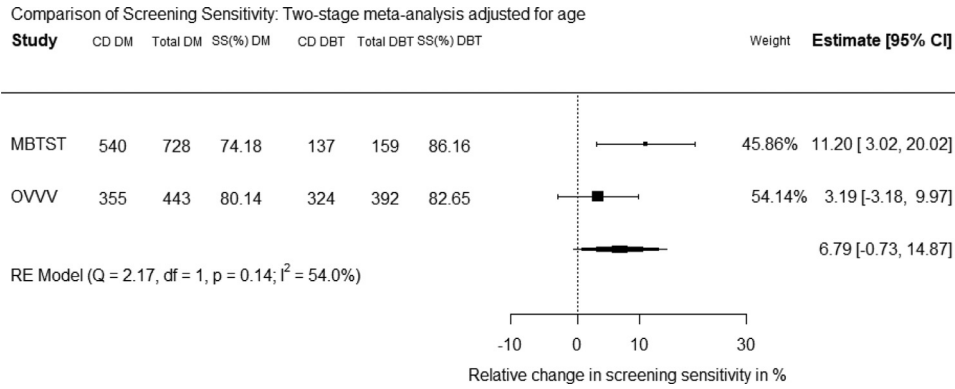


Fig. 7. Comparative screening sensitivity: pooled difference in sensitivity for DBT and DM screening adjusted for age.

Figure shows *age-adjusted* pooled difference in sensitivity (study-specific estimates also adjusted using median age); unadjusted pooled difference in sensitivity was 9.28% (95%CI: -2.74–22.77%).

Table 2

Characteristics of *interval breast cancers* occurring in participants screened with digital breast tomosynthesis (DBT) versus digital mammography (DM).

Variable		DBT (n = 90)	DM (n = 276)	Total (n = 366)	P value*
Study	MBTST	22 (24.4)	188 (68.1)	210 (57.4)	
	OVVV	68 (75.6)	88 (31.9)	156 (42.6)	
Histological type	Ductal carcinoma in-situ	7 (7.8)	14 (5.1)	21 (5.8)	0.35
	Invasive ductal cancer	71 (78.9)	206 (75.2)	277 (76.1)	
	Invasive lobular cancer	9 (10.0)	39 (14.2)	48 (13.2)	
	other invasive types	3 (3.3)	15 (5.5)	18 (5.0)	
	Not reported [†]	0	2	2	
Pathological tumour (pT) size	pTis	7 (8.3)	14 (5.3)	21 (6.1)	0.43
	pT1a-b (<=10 mm)	15 (17.9)	32 (12.2)	47 (13.5)	
	pT1c (>10 and <= 20 mm)	34 (40.5)	106 (40.3)	140 (40.3)	
	pT2 (>20 and <= 50 mm)	26 (31.0)	105 (39.9)	131 (37.8)	
	pT3 (>50 mm)	2 (2.4)	6 (2.3)	8 (2.3)	
	Not reported [†]	6	13	19	
Tumour grade	1	13 (16.5)	35 (14.5)	48 (15.0)	0.13
	2	44 (55.7)	109 (45.0)	153 (47.7)	
	3	22 (27.8)	98 (40.5)	120 (37.4)	
	Not reported [†]	11	34	45	
Axillary lymph nodes	Negative	61 (73.5)	141 (54.7)	202 (59.2)	0.005
	Positive (metastases)	20 (24.1)	92 (35.7)	112 (32.8)	
	Micro-metastases or isolated tumour cells	2 (2.4)	25 (9.7)	27 (7.9)	
	Not reported [†]	7	18	25	

MBTST= Malmö Breast Tomosynthesis Screening Trial; OVVV= Oslo, Vestfold & Vestre Viken study.

* P value for comparison of the distribution of interval cancer variables between DBT and DM screens.

[†] Not reported: includes missing data or not applicable (these were mostly not applicable in reference to ductal carcinoma in-situ cases); these data were not counted in distribution of proportions.

growing body of evidence on DBT's detection metrics is contrasted by sparse evidence on its effect on outcomes at follow-up of screened populations,^[15] yet such evidence is essential to inform BC screening policy decisions^[1]. In Europe, for example, DBT has only recently received conditional approval for BC screening in recommendations which point out evidence gaps regarding DBT's effect, including limited evidence on interval BC rates ^[22]. Interval BCs emerge after a negative screen and have similar prognostic features to clinically-diagnosed cancers ^[23]; they are therefore routinely monitored in many BC screening programs as an indication of screening sensitivity ^[23–25]. We report interval BC rate as the primary end-point, acknowledging this to be one surrogate measure for screening benefit, using IPD sourced from prospective DBT population screening studies ^[4,10–12]. Our pooled results show an age-adjusted interval BC rate of 13.17/10,000 based on 66,451 DBT-screened participants, although there was heterogeneity between studies in estimated interval BC rates. In particular, the STORM-2 ^[11] trial had a relatively low interval BC rate, possibly due to its design comprising parallel double-reading arms – although other studies had several screen-reading arms, STORM-2 was the only study with quadruple screen-reading using DBT. Pooled interval BC rates were significantly higher

in dense breasts; adjusting for study and age, the rate ratio was almost 3-times in dense (vs low-density) breasts, similar to findings for mammography screening.

For the two larger prospective studies ^[10,12] in this meta-analysis that also provided IPD for concurrent cohorts screened with DM alone (153,800 screens) from the same population programs, the *age-adjusted* pooled interval BC rate for DM was similar to that for DBT. Comparative meta-analysis showed no significant difference in interval BC rates (*rate difference* -0.44/10,000; 95%CI: -11.00–10.11) between DBT and DM screening. This finding was consistent across age-groups, so there was no evidence that DBT reduced interval BC rates in any age-group compared to DM. Considering other studies that have evaluated DBT population-based screening, and have reported on interval BCs at follow-up, both Skaane et al. ^[26] and Bernardi et al. ^[15] found that there was no significant difference in interval BC rates between DBT-screened groups and DM-screened historical cohorts. Therefore even if these additional studies had been included in the IPD meta-analysis, the findings of our comparative meta-analysis of interval BC rates are unlikely to have changed.

Meta-analysis comparing screening sensitivity showed weak evidence of improved sensitivity from DBT, although DBT significantly

increased cancer detection rate (and this effect was consistent between studies). Comparison of interval BC characteristics did not find any significant differences between screening modalities, with the exception of axillary-node metastases ($P = 0.005$) – a higher proportion of node-negative and lower proportion of node-positive interval BCs were shown in DBT-screened than DM-screened participants. This finding raises the possibility that DBT screening could reduce the rate of advanced BCs at repeat screening rounds. Temporal distribution of interval BCs, though limited by small numbers at each year, suggests that DBT may shift the time at which interval BC events are diagnosed, so radiological review may provide insights as to whether DBT alters the imaging pattern of interval BCs (for example, fewer false-negatives) [23].

Overall our results indicate that DBT screening, despite increasing BC detection, did not have a measurable impact on interval BC rates, and the increase in screening sensitivity (which was not statistically significant) was due to DBT increasing the number of screen-detected cancers. Our findings should be interpreted with caution despite the strengths of IPD meta-analysis, which enabled more precise analyses of the primary end-point than individual studies and supported statistical adjustments of interval BC rates, because there were limitations that hinder definitive conclusions from our work. The main limitation is that only four from eight eligible studies provided data (some investigators were unable or unwilling to contribute data), and only two studies provided data for concurrent DM screens, so pooled estimates have modest precision. These two studies forming the comparative meta-analysis were from Norway (OVVV) and Sweden (MBTST) [10,12] and were the largest of the four studies contributing data. Although all studies in our meta-analysis had ascertained interval BC events (Table 1), it is noteworthy that both studies contributing to comparative meta-analysis used population cancer registry linkage for this ascertainment. Since both studies were from Nordic nations with highly-organised population screening programs, this could potentially limit generalisability of our findings. Further, we focused on trials conducted in biennial screening programs, so our results may not apply to annual screening practice (where interval BC rates would be relatively lower) [23]. Another limitation is the heterogeneity in interval BC rate estimates, as discussed earlier; we used random-effects meta-analysis to allow for heterogeneity between studies, and explored its possible sources in subgroup analyses. Our findings suggest that there was relatively less heterogeneity in interval BC rates in the 60+ age-group and in the low-density group. Heterogeneity was also evident in comparative estimates based on the MBTST and OVVV studies [10,12]; these two studies used different DBT technology and screen-reading protocols (Table 1) which might account for the relatively divergent estimates of interval BC rates.

Our IPD meta-analysis did not focus on DBT's initial performance measures (screen-detection and recall rates) because this has been done in study-level meta-analyses from Marinovich et al., [6] Alabousi et al., [27] and Giampietro et al. [28] – however, most DBT screening studies do not report interval BC data so pooled interval BC rates have not been reported in these study-level analyses. We focused on interval BC rates to provide knowledge on an evidence gap in DBT screening and because interval BCs represent clinically progressive cancers [1,22]. We reported cancer detection rates showing that DBT significantly increased BC detection compared to DM (pooled CDR difference of 33.49/10,000). So the lack of effect from DBT (vs DM) on interval BC rates is discouraging, raising the possibility that DBT might increase over-diagnosis or extend lead-time without adding benefit above that from DM screening. Given the above-noted limitations of heterogeneity in estimates of interval BC rates and few datasets in this meta-analysis, we cannot make definitive inferences on the extent of over-diagnosis from DBT screening. Interval BC rate represents one proxy outcome for screening benefit (traditionally determined using mortality outcomes), and this relatively infrequent

outcome may require larger datasets particularly if there is subgroup heterogeneity in DBT's effect. Substantially larger IPD datasets that include more studies (including RCTs) would be needed to assess DBT's effect on interval BC rates in subgroups. We are aware of RCTs that will report other endpoints for screening benefit, such as a reduction in advanced BC rates, and which will report interval BC data potentially supporting larger analyses [29,30].

Using IPD meta-analysis of prospective DBT screening studies, a few studies with heterogeneous results on ICR could be included; our work did not find a significant difference in interval BC rates between DBT and DM screening. Screening sensitivity was not statistically significantly higher for DBT than DM screening, and there was evidence that interval BCs were more frequently node-negative in DBT-screened than DM-screened groups. The findings from our work highlight the need for much larger collaborative analyses of DBT screening to allow comparative evaluation of important but relatively infrequent screening outcomes such as interval BC rates, or advanced breast cancer rates. Our findings can inform future evaluation of DBT screening particularly for population screening programs contemplating potential transition to DBT or retention of standard mammography screening, suggesting that investment in large comparative studies may be a more appropriate strategy than prompt adoption of DBT.

Declaration of Competing Interest

Dr. Houssami reports grants from National Breast Cancer Foundation (NBCF Australia), grants from National Health & Medical Research Council (NHMRC, Australia), during the conduct of the study. Dr. Zackrisson reports grants from the Swedish Cancer Society, grants from Governmental funding (Region Skane), during the conduct of the study; in addition, Dr. Zackrisson has a patent approved, US (application no PCT/EP2014/057,372) issued. The other authors have not reported any conflicts of interest.

Acknowledgment

We thank Mr Sol Libesman for assistance with the study database.

Funding

Study was supported by a National Breast Cancer Foundation (NBCF Australia) *Breast Cancer Research Leadership Fellowship* (awarded to NH).

Data sharing statement

IPD will not be made available due to privacy regulations that affect IPD provisioning as part of this collaborative study; investigators may be contacted to discuss processes required for potential data access.

Authors' contributions

Conceptualisation: Houssami, Zackrisson, Hofvind; writing original draft: Houssami; writing (review and edit): all authors; literature search: Hunter, Houssami; data sourcing and collection: Bernardi, Lång, Hunter, Aglen, Johnson, Zackrisson, Hofvind; Project administration: Houssami, Hunter; Statistical analysis: Soerensen, Robledo; access to data: Soerensen, Houssami, Hunter; figures: Soerensen, Hunter; data interpretation: all authors.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.eclinm.2021.100804.

References

- [1] Lauby-Secretan B, Scoccianti C, Loomis D, et al. Breast-cancer screening –viewpoint of the IARC working group. *New Engl J Med* 2015;372:2353–8.
- [2] Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 U.S. preventive services task force recommendation. *Ann Intern Med* 2016;164:244–55.
- [3] Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 2013;108:2205–40.
- [4] 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;583–9.
- [5] Houssami N, Miglioretti DL. Digital breast tomosynthesis: a brave new world of mammography screening. *JAMA Oncol* 2016;2:725–7.
- [6] 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.
- [7] Lång K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S. 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;26:184–90.
- [8] 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):869–77. doi: 10.21037/tcr.2017.06.39.
- [9] 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;23:2061–71.
- [10] Zackrisson S, Lång K, Rosso A, et al. One-view breast tomosynthesis versus two-view mammography in the malmo breast tomosynthesis screening trial (MBTST): a prospective, population-based, diagnostic accuracy study. *Lancet Oncol* 2018;19:1493–503.
- [11] 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;17:1105–13.
- [12] Hofvind S, Hovda T, Holen AS, et al. Digital breast tomosynthesis and synthetic 2D mammography versus digital mammography: evaluation in a population-based screening program. *Radiology* 2018;287:787–94.
- [13] Caumo F, Zorzi M, Brunelli S, 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.
- [14] Romero Martin S, Povedano JLR, Garcia MC, Santos Romero AL, Garriguet MP, Alvarez 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.
- [15] Bernardi D, Gentilini MA, De Nisi M, et al. Effect of implementing digital breast tomosynthesis (DBT) instead of mammography on population screening outcomes including interval cancer rates: results of the Trento DBT pilot evaluation. *Breast* 2020;50:135–40.
- [16] Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- [17] Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw* 2021;10(36):1–48 A.D.
- [18] Core Team R. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017 <https://www.r-project.org/> 2017.
- [19] Houssami N, Bernardi D, Caumo F, et al. Interval breast cancers in the 'screening with tomosynthesis or standard mammography' (STORM) population-based trial. *Breast* 2018;38:150–3.
- [20] Johnson K, Lång K, Ikeda D, Åkesson A, Andersson I, Zackrisson S. Interval breast cancer rates and tumor characteristics in the prospective population-based malmö breast tomosynthesis screening trial (MBTST). *Radiology* 2021;204106. doi: 10.1148/radiol.202104106.
- [21] Hovda T, Holen AS, Lång K, et al. Interval and consecutive round breast cancer after digital breast tomosynthesis and synthetic 2D mammography versus standard 2D digital mammography in breastcreen norway. *Radiology* 2020;294:256–64.
- [22] European Commission Initiative on Breast Cancer (ECIBC). Use of Digital Breast Tomosynthesis (DBT). Website (<https://healthcare-quality.jrc.ec.europa.eu/european-breast-cancer-guidelines/screening-tests/DBT-or-DM/>); accessed December 2020.
- [23] Houssami N, Hunter K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *NPJ Breast Cancer* 2017;3:12.
- [24] Tornberg S, Kemetti L, Ascunce N, et al. A pooled analysis of interval cancer rates in six European Countries. *Eur J Cancer Prev* 2010;19:87–93.
- [25] Nederend J, Duijm LE, Louwman MW, 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;50:31–9.
- [26] Skaane P, Sebuodegard S, Bandos AI, et al. Performance of breast cancer screening using digital breast tomosynthesis: results from the prospective population-based Oslo tomosynthesis screening trial. *Breast Cancer Res Treat* 2018;169:489–96.
- [27] Alabousi M, Zha N, Salameh JP, et al. Digital breast tomosynthesis for breast cancer detection: a diagnostic test accuracy systematic review and meta-analysis. *Eur Radiol* 2020;30:2058–71.
- [28] Giampietro RR, Cabral MVG, Lima SAM, Weber SAT, Dos SN-N, Dos Santos Nunes-Nogueira V. Accuracy and effectiveness of mammography versus mammography and tomosynthesis for population-based breast cancer screening: a systematic review and meta-analysis. *Sci Rep* 2020;10:7991.
- [29] Weigel S, Gerss J, Hense HW, et al. Digital breast tomosynthesis plus synthesised images versus standard full-field digital mammography in population-based screening (TOSYMA): protocol of a randomised controlled trial. *BMJ Open* 2018;8:e020475.
- [30] Pattacini P, Nitrosi A, Giorgi RP, et al. Digital mammography versus digital mammography plus tomosynthesis for breast cancer screening: the reggio emilia tomosynthesis randomized trial. *Radiology* 2018;288:375–85.