BJR https://doi.org/10.1259/bjr.20230081

Received: 25 January 2023 | Revised: 07 April 2023 | Accepted: 18 April 2023

Cite this article as:
Houssami N, Lockie D, Giles M, Noguchi N, Marr G, Marinovich ML. Two-year follow-up of participants in the BreastScreen Victoria pilot trial of tomosynthesis versus mammography: breast density-stratified screening outcomes. *Br J Radiol* (2023) 10.1259/bjr.20230081.

FULL PAPER

Two-year follow-up of participants in the BreastScreen Victoria pilot trial of tomosynthesis versus mammography: breast density-stratified screening outcomes

[1,2]NEHMAT HOUSSAMI, MBBS, PhD, [3]DARREN LOCKIE, MBBS, [3]MICHELLE GILES, BAppSc, [2]NAOMI NOGUCHI, PhD, [4]GEORGINA MARR, GradDip (EpiBiostats) and [1,2]M LUKE MARINOVICH, PhD

[1]The Daffodil Centre, the University of Sydney, a joint venture with Cancer Council NSW, Sydney, Australia
[2]Sydney School of Public Health, Faculty of Medicine & Health, University of Sydney, Sydney, Australia
[3]Maroondah BreastScreen, Eastern Health, Victoria, Australia
[4]BreastScreen Victoria, Victoria, Australia

Address correspondence to: Prof Nehmat Houssami
E-mail: *nehmat.houssami@sydney.edu.au*

Objective: This follow-up study of BreastScreen Victoria's pilot trial of digital breast tomosynthesis aimed to report interval cancer rates, screening sensitivity, and density-stratified outcomes for tomosynthesis *vs* mammography screening.
Methods: Prospective pilot trial [ACTRN-12617000947303] in Maroondah BreastScreen recruited females ≥ 40 years presenting for screening (August 2017–November 2018) to DBT; concurrent screening participants who received mammography formed a comparison group. Follow-up of 24 months from screen date was used to ascertain interval cancers; automated breast density was measured.
Results: There were 48 screen-detected and 9 interval cancers amongst 4908 tomosynthesis screens, and 34 screen-detected and 16 interval cancers amongst 5153 mammography screens. Interval cancer rate was 1.8/1000 (95%CI 0.8–3.5) for tomosynthesis *vs* 3.1/1000 (95%CI 1.8–5.0) for mammography (p = 0.20). Sensitivity of tomosynthesis (86.0%; 95% CI 74.2–93.7) was significantly higher than mammography (68.0%; 95% CI 53.3–80.5), p = 0.03. Cancer detection rate (CDR) of 9.8/1000 (95%CI 7.2–12.9) for tomosynthesis was higher than that of 6.6/1000 (95%CI 4.6–9.2) for mammography (p = 0.08); density-stratified analyses showed CDR was significantly higher for tomosynthesis than mammography (10.6/1000 *vs* 3.5/1000, p = 0.03) in high-density screens. Recall rate for tomosynthesis was significantly higher than for mammography (4.2% *vs* 3.0%, $p < 0.001$), and this increase in recall for tomosynthesis was evident only in high-density screens (5.6% *vs* 2.9%, $p < 0.001$).
Conclusion: Although interval cancer rates did not significantly differ between screened groups, sensitivity was significantly higher for tomosynthesis than mammography screening.
Advances in knowledge: In a program-embedded pilot trial, both increased cancer detection and recall rates from tomosynthesis were predominantly observed in high-density screens.

INTRODUCTION

In the context of growing adoption and evaluation of digital breast tomosynthesis (DBT),[1] a pilot trial was undertaken in Australia's BreastScreen program to examine detection metrics and the feasibility of DBT screening.[2] Obtaining evidence from local screening programs was relevant given that initial studies of DBT, both prospective and retrospective, showed heterogeneity in detection measures across screening settings.[1] The BreastScreen Victoria DBT trial was a *prospective* population-based trial that compared DBT and mammography screens from concurrent cohorts presenting for breast screening to Maroondah BreastScreen. It showed that DBT increased cancer detection (DBT 9.8/1000 screens, DM 6.6/1000 screens), as well as recall rates (DBT 4.2%, DM 3.0%), and prolonged screen-reading time, compared to mammography.[2] Details of the trial and its initial outcomes have been previously reported.[2]

From an international perspective, although there is increasing use of DBT imaging, many organised screening programs use digital mammography for screening but

support further investigation of DBT screening or allow conditional use of DBT.[3] In more recent years, three randomised trials[4–6] reported from European programs have shown increased cancer detection rate for DBT relative to mammography for initial screening outcomes, and another RCT from BreastScreen Norway showed a reduction in recall but no increased cancer detection rate from DBT.[7] However, there is limited and conflicting evidence on differences in interval cancer rate and screening sensitivity between DBT and mammography at follow-up of screened women.[6,8,9]

Hence, in the present study, we *update* the BreastScreen Victoria pilot trial at 2-year follow-up of screening participants to estimate interval cancer rates and screening sensitivity, for DBT and mammography. We also extend our work through automated breast density measurement, to provide knowledge on density-stratified screening outcomes for DBT *vs* mammography population-based screening.

METHODS

Trial design and population

The setting and methods of the trial have been detailed in our earlier report,[2] so will be described briefly for the present study. The BreastScreen DBT pilot was a prospective non-randomised trial [registered ACTRN-12617000947303] conducted in Maroondah BreastScreen (Eastern Health), which recruited females ≥ 40 years presenting for routine screening (August 2017–November 2018).[2] The trial was designed to estimate detection measures for 5000 DBT screens and to determine the feasibility of DBT population screening. Screening participants were informed that they could receive DBT or standard digital mammography screening. Therefore, a concurrent group of females who received digital mammography screening formed a comparison group from the same population. Ethics approval for the trial was granted by Eastern Health HREC (LR36/2017).

The trial was implemented with an 'opt-out' option, where females who declined participation (opted-out of having DBT) had mammography screening. BreastScreen services provide written information about screening and require written consent from participants, therefore trial-specific information and consent forms were integrated into existing BreastScreen information processes. During the pilot trial, Maroondah BreastScreen had two screening rooms, one equipped with a DBT-capable mammography unit, and the other with a standard digital mammography unit. Therefore, except for those who opted-out of having DBT, receipt of DBT screening was driven by the next available screening room when the woman was called into a mammography room for acquisition of the screen. Online-only Figure 1 shows a flow-diagram of the study.

Screening and screen-reading

DBT screening consisted of tomosynthesis acquisitions from which synthesized 2D-images were reconstructed [Selenia® Dimensions Unit, C-View™ 2D-software]. Digital mammography was performed using one of two units (Hologic Selenia® Dimensions Unit, or Siemens Mammomat Inspiration). Mediolateral oblique and craniocaudal views of each breast were obtained for DBT acquisitions and for mammography. All other aspects of screening, screen-reading and assessment, and follow-up to ascertain outcomes, were based on BreastScreen Australia's National Accreditation Standards.[10] Double-reading practice (two independent readings per screen) was used with arbitration by a third read for disagreement. Seven radiologists performed screen-reads in the timeframe of the trial, therefore the same radiologists who interpreted DBT also read standard mammograms.

For the *updated* report, we included follow-up to ascertain interval cancers to estimate interval cancer rates, and measured breast density to report outcomes stratified by density, to provide evidence requested by Australia's Medical Services Advisory Committee.

Interval breast cancers

Interval cancers were ascertained by BreastScreen Victoria through established data linkage with the Victorian Cancer Registry; this is routinely performed via the program's quality assurance processes to monitor interval cancer rates as a key performance indicator of cancer screening.[11] Follow-up was set at 24 months from the date of screening given that the program provides biennial screening. Date of diagnosis and tumour characteristics were obtained for interval cancers.

Breast density

Mammographic breast density is *not* routinely measured in the BreastScreen Australia program, therefore density was not measured in the DBT pilot trial. However, we retrospectively measured breast density to enable density-stratified reporting of outcomes at follow-up. Automated density measurement was assessed using Densitas® breast density software (densitas® densityAI™) which is comprised of two distinct deep learning algorithms, one that computes quantitative breast density percentage ranging from 0 to 100%, and another that computes a qualitative breast density classification that aligns with the BI-RADS density categories (A–D). For analytic purposes, we classified categories A–B as 'low density' and C–D as 'high density'.

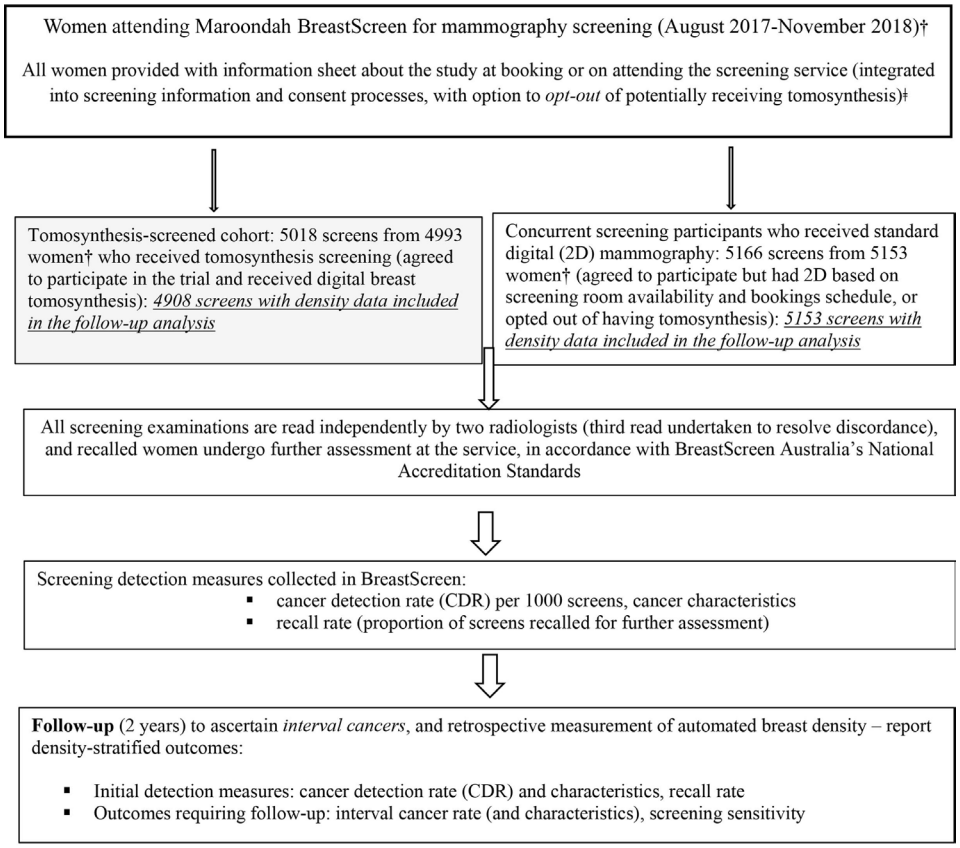
Outcomes

Primary outcomes for the trial were the number of detected cancers and cancer detection rate (CDR) per 1000 screens, and the number and percentage of recalls. Outcomes were ascertained based on excision histology, or the completed assessment inclusive of work-up imaging (and histology from needle biopsy) in recalled participants, or otherwise by linking with the cancer registry as described for interval cancer ascertainment. Secondary outcomes have been reported in the initial trial report,[2] including cancer characteristics, and the assessment of recalled abnormalities has also been described.[12] Descriptive data for breast density categories, and for interval cancer characteristics, were included in the updated trial.

Statistical methods

Age and breast density characteristics of screening participants were described and compared by the type of screen received using the χ^2 test for proportions, and independent samples *t*-test for

Figure 1. Pilot trial of digital breast tomosynthesis* population-based screening in BreastScreen Victoria



* Digital breast tomosynthesis acquisition from which synthesized 2D-images are also reconstructed.

†From 18 August 2017 to 8 November 2018, a total of 10,146 women presented to Maroondah BreastScreen, and received 10,184 screening examinations shown above by the screening modality received (38 women had annual screening hence had 2 screening episodes within the trial's timeframe, accounting for the small difference between number of women and number of screening examinations)

continuous data. For each screening modality, the CDR, interval cancer rate, proportion of recalls, sensitivity (proportion of all cancers detected at screening), specificity (proportion of screens without cancer correctly excluded) and positive-predictive value (PPV, proportion of recalls with cancer) were computed, and estimates were stratified by breast density (low *vs* high) and age groups (<60 years *vs* ≥60 years). Exact (Clopper-Pearson) 95% confidence intervals (CIs) for proportions were computed. Differences in estimates between screened groups were calculated with Miettinen-Nurminen 95% CI, and the estimates were compared with χ^2 or Fisher's exact tests, as appropriate. Interval cancer characteristics (pathological tumour size, nodal status, tumour grade, and density categories) were tabulated descriptively. Analyses were undertaken in SAS 9.4.

RESULTS

From the initial 10,184 screening examinations (10,146 women) performed at Maroondah BreastScreen as part of the trial, from 18 August 2017 to 8 November 2018, this follow-up study included 10,061 screens—123 (1%) screening exams missing density data due to technical reasons were excluded. Because the characteristics of screening participants have been previously reported, here we describe only age and breast density characteristics (Table 1) showing significant differences in mean age and in the distribution of age and density categories between screened groups. Table 1 also shows that these differences in age descriptive data between screened groups were evident for low- and high-density screening examinations.

Table 2 summarises detection measures by screening modality inclusive of follow-up for interval breast cancers. Amongst 4908 (from initial 5018) tomosynthesis screens, there were 48 screen-detected cancers and 9 interval cancers at 2-year follow-up. Amongst 5153 (from initial 5166) mammography screens, there were 34 screen-detected cancers and 16 interval cancers at follow-up. Interval cancer rate was 1.8/1000 (95%CI 0.8, 3.5) for tomosynthesis *vs* 3.1/1000 (95%CI 1.8, 5.0) for mammography screens ($p = 0.20$). Interval cancer rates are also shown for each screening modality stratified by density and by age-group (Table 2).

CDRs overall and by age-groups have been previously reported; some of these results are shown in Table 2 for transparency and to ensure a cohesive follow-up report. A CDR of 9.8/1000 (95%CI

Table 1. Age and density by screening modality in the BreastScreen Victoria tomosynthesis screening trial

	Digital mammography (2D)		Digital breast tomosynthesis (3D)		*p*-value
Age descriptive data	**N screens**	**Mean (SD) years, or %**	**N screens**	**Mean (SD) years, or %**	
All screens[a]					
Mean	5153	62.3 (8.1)	4908	58.1 (8.5)	<0.001
Age <60 years	1985	38.5%	2943	60.0%	<0.001
Age ≥60 years	3168	61.5%	1965	40.0%	
Age, high density screens					
Mean	1415	59.7 (8.1)	1986	55.6 (8.0)	<0.001
Age <60 years	704	49.7%	1421	71.5%	<0.001
Age ≥60 years	711	50.3%	565	28.5%	
Age, low density screens					
Mean	3738	63.2 (7.9)	2922	59.7 (8.3)	<0.001
Age <60 years	1281	34.3%	1522	52.1%	<0.001
Age ≥60 years	2457	65.7%	1400	47.9%	
Density descriptive data	N	%	N	%	<0.001
Category A	1718	33.3%	705	14.4%	
Category B	2020	39.2%	2217	45.2%	
Category C	1196	23.2%	1473	30.0%	
Category D	219	4.3%	513	10.5%	

2D, two-dimensional; 3D, three-dimensional.
[a]Total number of screens shown differs from the numbers reported in our initial publication due to exclusion of 123 (from 10,184) screens in whom density could not be measured.

7.2, 12.9) for tomosynthesis was higher than that of 6.6/1000 (95%CI 4.6, 9.2) for mammography ($p = 0.08$); the higher CDR for tomosynthesis was statistically significant in females aged 60 years and older ($p = 0.03$). Density-stratified analyses showed that CDR was significantly higher for tomosynthesis (10.6/1000; 95% CI 6.6, 16.1) than mammography (3.5/1000; 95% CI 1.1, 8.2) in high density screens ($p = 0.03$).

Screening sensitivity for tomosynthesis (86.0%; 95% CI 74.2, 93.7) was significantly higher than that for mammography (68.0%; 95% CI 53.3, 80.5), $p = 0.03$. Although a higher sensitivity was shown for tomosynthesis in both density strata, a statistically significant increase in sensitivity was only evident in our data for high-density screens (Table 2), reflecting the pattern observed also for CDR. Although screening sensitivity was higher for tomosynthesis than mammography across both age-groups, estimates did not significantly differ between screening modalities in age-stratified analyses (Table 2). A sensitivity analysis that reclassified (from true-positive to false-negative) one interval cancer, recalled at tomosynthesis screening but with a false-negative assessment, did not substantially alter the above-reported findings; in this reanalysis, screening sensitivity for tomosynthesis slightly decreased to 84.2% ($p = 0.05$ compared to mammography, all screens). Further, in this sensitivity analysis (reanalysis) our estimates for sensitivity and PPV did not substantially change for all screens nor in stratified analyses, and statistical associations (or lack of) did not change.

Recall rate for tomosynthesis screening was significantly higher than for mammography (4.2% *vs* 3.0%, $p < 0.001$), and this was also evident in age-stratified analyses. However, density-stratified estimates (Table 2) show that the significant increase in recall for tomosynthesis was evident only in high-density screens (5.6% *vs* 2.9%, $p < 0.001$). Specificity was significantly lower for tomosynthesis than for mammography (96.7% *vs* 97.6%, $p = 0.006$), and this lower specificity was statistically significant only in high-density screens (95.4%% *vs* 97.4%, $p = 0.002$). PPV for recall (cancer yield from recall to assessment) was generally slightly higher for tomosynthesis in all analyses shown in Table 2.

The characteristics of screen-detected cancers have been described in the initial trial report and are included as online-only Supplementary Material 1; we now report the number of interval cancers by tumour characteristics and density categories for each screened group in Table 3 (noting the small number of cases does not support further analysis of these data).

DISCUSSION

In *BreastScreen* Australia, as in many national population-based screening programs, digital mammography is the recommended screening modality, although evaluation of tomosynthesis is encouraged to inform future screening practice. BreastScreen Victoria's pilot trial was designed to provide estimates for tomosynthesis detection metrics, and to assess its feasibility, in the Australian population screening context—the initial findings

Table 2. Screening performance measures by screening modality for all screens, and stratified by breast density and age-group, in the BreastScreen Victoria tomosynthesis screening trial

	Digital mammography (2D)		Digital breast tomosynthesis (3D)		Difference, 3D-2D	*p* value
	N	% or per 1000 (95% CI)	N	% or per 1000 (95% CI)	% or per 1000 (95% CI)	
All screens	5153*	–	4908*	–	–	–
Screen-detected cancers	34	6.6 per 1000 (4.6, 9.2)	48	9.8 per 1000 (7.2, 12.9)	3.2 per 1000 (−0.3, 6.9)	0.08
Recalls	155ł	3.0% (2.6, 3.5)	208	4.2% (3.7, 4.8)	1.2% (0.5, 2.0)	<0.001
Interval cancers	16	3.1 per 1000 (1.8, 5.0)	9	1.8 per 1000 (0.8, 3.5)	−1.3 per 1000 (−3.4, 0.7)	0.20
Sensitivity	–	68.0% (53.3, 80.5)	–	86.0%** (74.2, 93.7)	18.0% (2.1, 33.9)	0.03
Specificity	–	97.6% (97.2, 98.0)	–	96.7% (96.2, 97.2)	−0.9% (−1.6,−0.3)	0.006
PPV	–	21.9% (15.7, 29.3)	–	23.6%** (18.0, 29.9)	1.6% (−7.3, 10.2)	0.72
High density	1415	–	1986	–	–	–
Screen-detected cancers	5	3.5 per 1000 (1.1, 8.2)	21	10.6 per 1000 (6.6, 16.1)	7.0 per 1000 (1.3, 13.0)	0.03***
Recalls	41	2.9% (2.1, 3.9)	111	5.6% (4.6, 6.7)	2.7% (1.3, 4.0)	<0.001
Interval cancers	8	5.7 per 1000 (2.4, 11.1)	5	2.5 per 1000 (0.8, 5.9)	−3.1 per 1000 (−8.8, 1.1)	0.17***
Sensitivity	–	38.5% (13.9, 68.4)	–	80.8% (60.7, 93.5)	42.3% (10.2, 67.7)	0.01***
Specificity	–	97.4% (96.5, 98.2)	–	95.4% (94.4, 96.3)	−2.0% (−3.3,−0.7)	0.002
PPV	–	12.2% (4.1, 26.2)	–	18.9% (12.1, 27.5)	6.7% (−8.0, 17.9)	0.33
Low density	3738	–	2922	–	–	–
Screen-detected cancers	29	7.8 per 1000 (5.2, 11.1)	27	9.2 per 1000 (6.1, 13.4)	1.5 per 1000 (−2.9, 6.3)	0.51
Recalls	114 ł	3.1% (2.5, 3.7)	97	3.3% (2.7, 4.0)	0.3% (−0.6, 1.1)	0.53
Interval cancers	8	2.1 per 1000 (0.9, 4.2)	4	1.4 per 1000 (0.4, 3.5)	−0.8 per 1000 (−3.0, 1.6)	0.57***
Sensitivity	–	78.4% (61.8, 90.2)	–	90.3%** (74.3, 98.0)	11.9% (−6.5, 29.5)	0.11***
Specificity	–	97.7% (97.2, 98.2)	–	97.6% (97.0, 98.1)	−0.1% (−0.9, 0.6)	0.81
PPV	–	25.4% (17.7, 34.5)	–	28.9%** (20.1, 38.9)	3.4% (−8.5, 15.6)	0.58
Age <60 years	1985	–	2943	–	–	–
Screen-detected cancers	7	3.5 per 1000 (1.4, 7.3)	18	6.1 per 1000 (3.6, 9.6)	2.6 per 1000 (−1.7, 6.6)	0.21
Recalls	63	3.2% (2.5, 4.0)	127	4.3% (3.6, 5.1)	1.1% (0.1, 2.2)	0.04
Interval cancers	5	2.5 per 1000 (0.8, 5.9)	3	1.0 per 1000 (0.2, 3.0)	−1.5 per 1000 (−5.0, 0.9)	0.28***
Sensitivity	–	58.3% (27.7, 84.8)	–	85.7% (63.7, 96.9)	27.4% (−3.4, 57.3)	0.11***

(Continued)

Table 2. (Continued)

	Digital mammography (2D)		Digital breast tomosynthesis (3D)		Difference, 3D-2D	*p* value
Specificity	–	97.2% (96.3, 97.9)	–	96.3% (95.5, 96.9)	−0.9% (−1.9, 0.1)	0.09
PPV	–	11.1% (4.6, 21.6)	–	14.2% (8.6, 21.5)	3.1% (−8.2, 12.4)	0.56
Age ≥ 60 years	3168	–	1965	–	–	–
Screen-detected cancers	27	8.5 per 1000 (5.6, 12.4)	30	15.3 per 1000 (10.3, 21.7)	6.7 per 1000 (0.8, 13.7)	0.03
Recalls	92 ǂ	2.9% (2.3, 3.5)	81	4.1% (3.3, 5.1)	1.2% (0.2, 2.3)	0.02
Interval cancers	11	3.5 per 1000 (1.7, 6.2)	6	3.1 per 1000 (1.1, 6.6)	−0.4 per 1000 (−3.6, 3.5)	0.80
Sensitivity	–	71.1% (54.1, 84.6)	–	86.1%** (70.5, 95.3)	15.1% (−4.0, 33.5)	0.12
Specificity	–	97.9% (97.4, 98.4)	–	97.4% (96.6, 98.1)	−0.5% (−1.4, 0.3)	0.23
PPV	–	29.3% (20.3, 39.8)	–	38.3%** (27.7, 49.7)	8.9% (−5.2, 22.9)	0.21

2D, two-dimensional; 3D, three-dimensional; PPV, positive-predictive value.
* Total number of screens shown differs from the numbers reported in our initial publication due to exclusion of screens missing density data in the present analysis.
ǂTechnical recalls (*n* = 2) excluded.
** One recalled case with no cancer found at assessment but with subsequent interval cancer diagnosis was classified as a true-positive screen: when this was reclassified as false negative, sensitivity for tomosynthesis was re-estimated as 84.2% for all screens—in this reanalysis, sensitivity, PPV, and statistical associations did not substantially alter in all screens or in stratified analyses.
*** Fisher exact test.

from the trial have been described.[2] The purpose of the present study is to complete follow-up of the trial's participants to estimate interval cancer rates, and to integrate density-stratified outcomes. Although there were fewer interval cancers in the tomosynthesis-screened group than in the mammography-screened group, interval cancer rates did not significantly differ between screened groups (Table 2)—it should be noted, however, that this was a secondary outcome for this pilot trial which was not powered to compare interval cancer rates but to provide estimates to inform future evaluation. Nonetheless, these data add to the limited evidence on this outcome from tomosynthesis screening trials[8,13] and are generally similar to interval cancer rate estimates from the Norwegian randomised trial.[8]

Our findings add to the growing body of evidence that tomosynthesis has higher CDR than mammography screening at prevalent screening, and further highlights the variability of findings on the effect of tomosynthesis on recall rates.[1,14,15] Having reported that tomosynthesis had higher CDR than mammography (based on the overall screened cohorts),[2] we found that tomosynthesis had higher CDR than mammography across age-groups although this was only statistically significant in females ≥ 60 years. We now additionally show that most of the increased cancer detection was in high-density screens, where a significant difference in CDR was found for tomosynthesis compared to mammography screens ($p = 0.03$). Aligning with this finding, screening sensitivity at 2-year follow-up of participants (allowing inclusion of interval cancers) for tomosynthesis was significantly higher than that for mammography in comparison of the overall cohorts ($p = 0.03$) and was also significantly higher for tomosynthesis in high-density screens ($p = 0.01$). One of the largest trials (TOSYMA) embedded in the German screening program also found that DBT significantly increased cancer detection particularly in dense breasts.[16]

This study's focus on comparative cancer detection and interval cancer rates and screening sensitivity at follow-up provides evidence on key outcomes for population screening programs; we acknowledge that these initial screening outcomes may not equate with screening benefit because increased cancer screen-detection may represent (or contribute to) overdiagnosis.[13,17] Our earlier report of the trial described the characteristics of the screen-detected cancers (by imaging modality) highlighting generally similar distributions for most prognostic features, however, there were relatively more Grade 1 (and fewer Grade 3) invasive cancers amongst tomosynthesis-detected cancers than those detected in the mammography group.[2] Given generally small number of cancers (in the context of a pilot study) inferences cannot be made from our findings on cancer characteristics, although this could suggest potential overdiagnosis or additional lead time from tomosynthesis.

Our findings for recall to assessment had shown that tomosynthesis has a higher recall rate than mammography (based on the overall screened cohorts),[2] and in our updated work the density-stratified estimates (Table 2) indicate that this significant increase in recall for tomosynthesis was evident only in high-density screens ($p < 0.001$). Commensurate with the higher

Table 3. Interval cancer characteristics by screening modality (numbers shown)

	Digital mammography (2D)	Digital breast tomosynthesis (3D)
Number of interval cancers	16	9
Pathological tumour size (pT) category[a]		
pTis (ductal carcinoma *in situ*)	0	2
pT1a (≤5 mm)	1	0
pT1b (>5 mm-≤10 mm)	5	2
pT1c (>10 mm-≤20 mm)	4	2
pT2 (>20 mm-≤50 mm)	5	3
Nodal status[a]		
Negative for metastases	12	4
Positive for metastases	3	3
Tumour grade[a]		
Grade 1	4	1
Grade 2	4	4
Grade 3	7	3
Density category[a]		
A	3	0
B	5	4
C	6	5
D	2	0

2D, two-dimensional; 3D, three-dimensional.
[a]Numbers do not sum to corresponding totals because of missing data.

recall rates for tomosynthesis, specificity was significantly lower for tomosynthesis than for mammography overall (p = 0.006), and this was statistically significant only in high-density screens (p = 0.002). Placing these findings into context, recall is known to be higher in dense breasts and this trial was the Australian program's first experience with tomosynthesis screening—it is possible that the recall could be lower (and specificity improved) with further experience and with availability of prior tomosynthesis images to compare with. In our descriptive study of the mammographic lesions recalled to assessment, we showed that there was a higher proportion of lesions depicted as calcifications for tomosynthesis (32.4%) than mammography (21.3%), and a lower proportion of asymmetric densities for tomosynthesis (3.2%) than mammography (15.7%) recalls in a lesion-based analysis.[12] On further exploration of these data, we observed the same pattern when the analysis focused on false-positive recall: the types of imaging lesions amongst false-positives were generally similar for both modalities except for calcifications (tomosynthesis 37.3%, mammography 21.3%) and asymmetric densities (tomosynthesis 4.1%, mammography 19.1%) as shown in Supplementary Material 2.

Limitations of this study include the non-randomised design, leading to imbalances in age and density group distributions—hence, we stratified our findings by age and density groups but acknowledge the possibility of residual confounding. Mean age of the participants in the tomosynthesis group was lower than in the mammography group (Table 1); a possible contributor to this age imbalance between groups is that amongst the females who opted out of receiving DBT (5% of screening participants) there were more females older than 60 years (mean age 61.4 years)[2] who preferred to have 'standard' mammography. Measurement of density was undertaken retrospectively, with around 1% of screens not recording a density measure for technical reasons; density is not routinely measured in BreastScreen Australia, so density information is unlikely to have caused preferential selection to DBT. Given that the trial was a pilot study, the sample size was smaller than other international prospective trials but was appropriate for estimating DBT detection measures in line with the trial's primary aims. The pilot trial was not powered to support stratified analyses, therefore when interpreting our findings estimates of effect should be considered with the associated imprecision from smaller subgroup data (as occurs with stratification).

This trial also has a similar limitation to other trials of tomosynthesis to date in that longer-term outcomes have not been considered: the effect of tomosynthesis, relative to mammography, on breast cancer mortality and overdiagnosis requires much larger studies and sustained follow-up. Despite these limitations, BreastScreen Victoria's pilot trial is the only population-based study of tomosynthesis screening in Australia and provides relevant evidence for the program to determine further research directions in tomosynthesis screening.

CONCLUSIONS

At 2-year follow-up of screened participants in this population-based trial, interval cancer rates did not significantly differ between the tomosynthesis screened and mammography screened groups. Screening sensitivity at follow-up, inclusive of interval cancers, was significantly higher for tomosynthesis than for mammography in the overall cohorts and in high-density screens. In addition, the previously reported significant increase in recall for tomosynthesis was shown to be evident only in high-density screens. Therefore, the main effects of tomosynthesis screening in this pilot trial, both increased cancer detection rate and recall rate, occurred predominantly in high-density screens.

ACKNOWLEDGEMENTS

We thank the radiology, radiography, data and administrative staff at Maroondah BreastScreen, and also thank the information technology and data management staff at BreastScreen Victoria for helping us implement the trial. We thank the women who attended Maroondah BreastScreen and participated in the study.

FUNDING

This project was funded by MRFF Targeted Health System Research grant (No. 1199927). NH is funded via the National Breast Cancer Foundation (NBCF) Chair in Breast Cancer Prevention grant (EC-21-001) and a National Health and Medical Research Council Investigator (Leader) grant (No.1194410). MLM is funded by a National Breast Cancer Foundation (NBCF) Investigator Initiated Research Scheme grant (IIRS-20-011). Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australian University Librarians.

REFERENCES

1. 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**: 942–49. https://doi.org/10.1093/jnci/djy121
2. 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. https://doi.org/10.5694/mja2.50320
3. European Commission Initiative on Breast Cancer (ECIBC). Use of Digital Breast Tomosynthesis (DBT). 2020. Available from: https://healthcare-quality.jrc.ec.europa.eu/european-breast-cancer-guidelines/screening-tests)
4. Heindel W, Weigel S, Gerß J, Hense H-W, Sommer A, Krischke M, et al. Digital breast tomosynthesis plus synthesised mammography versus digital screening mammography for the detection of invasive breast cancer (TOSYMA): a multicentre, open-label, randomised, controlled, superiority trial. *Lancet Oncol* 2022; **23**: 601–11: . https://doi.org/10.1016/S1470-2045(22)00194-2
5. Pattacini P, Nitrosi A, Giorgi Rossi P, Iotti V, Ginocchi V, Ravaioli S, et al. Digital mammography versus digital mammography plus tomosynthesis for breast cancer screening: the reggio emilia tomosynthesis randomized trial. *Radiology* 2018; **288**: 375–85. https://doi.org/10.1148/radiol.2018172119
6. Armaroli P, Frigerio A, Correale L, Ponti A, Artuso F, Casella D, et al. A randomised controlled trial of digital breast tomosynthesis vs digital mammography as primary screening tests: screening results over subsequent episodes of the Proteus donna study. *Int J Cancer* 2022; **151**: 1778–90. https://doi.org/10.1002/ijc.34161
7. Hofvind S, Holen ÅS, Aase HS, Houssami N, Sebuødegård S, Moger TA, et al. Two-View digital breast tomosynthesis versus digital mammography in a population-based breast cancer screening programme (to-be): a randomised, controlled trial. *Lancet Oncol* 2019; **20**: 795–805: . https://doi.org/10.1016/S1470-2045(19)30161-5
8. Hofvind S, Moshina N, Holen ÅS, 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. https://doi.org/10.1148/radiol.2021203936
9. Johnson K, Lång K, Ikeda DM, Åkesson A, Andersson I, Zackrisson S. Interval breast cancer rates and tumor characteristics in the prospective population-based Malmö breast tomosynthesis screening trial. *Radiology* 2021; **299**: 559–67. https://doi.org/10.1148/radiol.2021204106
10. Australian Institute of Health and Welfare (AIHW). BreastScreen Australia Monitoring Report 2014-2015. 2017. Cancer Series no.6. Cat. no.CAN 105(available at. Canberra, AIHW. Available from: https://www.aihw.gov.au/reports-data)
11. Australian Institute of Health and Welfare (AIHW). BreastScreen Australia Monitoring Report 2020.Canberra, AIHW 2023. Cancer series no.129. Cat. no.CAN. Available from: athttps://www.aihw.gov.au/reports-data
12. Li T, Lockie D, Clemson M, Houssami N. Assessment of screen-recalled abnormalities for digital breast tomosynthesis versus digital mammography screening in the BreastScreen Maroondah trial . *J Med Imag Rad Onc* 2023; **67**: 242–51. https://doi.org/10.1111/1754-9485.13452
13. Houssami N, Hofvind S, Soerensen AL, Robledo KP, Hunter K, Bernardi D, et al. Interval breast cancer rates for digital breast tomosynthesis versus digital mammography population screening: an individual participant data meta-analysis. *EClinicalMedicine* 2021; **34**: : 100804. https://doi.org/10.1016/j.eclinm.2021.100804
14. 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. https://doi.org/10.1007/s00330-019-06549-2
15. 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. https://doi.org/10.1080/14737140.2018.1483243
16. Weigel S, Heindel W, Hense HW, et al. Breast density and breast cancer screening with digital breast tomosynthesis: a TOSYMA trial subanalysis. *Radiology* 2022.
17. Farber R, Houssami N, Barnes I, McGeechan K, Barratt A, Bell KJL. Considerations for evaluating the introduction of new cancer screening technology: use of interval cancers to assess potential benefits and harms. *IJERPH* 2022; **19**: 14647. https://doi.org/10.3390/ijerph192214647