Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study

Stefano Ciatto†, Nehmat Houssami, Daniela Bernardi, Francesca Caumo, Marco Pellegrini, Silvia Brunelli, Paola Tuttobene, Paola Bricolo, Carmine Fantò, Marvi Valentini, Stefania Montemezzi, Petra Macaskill

Background Digital breast tomosynthesis with 3D images might overcome some of the limitations of conventional 2D mammography for detection of breast cancer. We investigated the effect of integrated 2D and 3D mammography in population breast-cancer screening.

Methods Screening with Tomosynthesis OR standard Mammography (STORM) was a prospective comparative study. We recruited asymptomatic women aged 48 years or older who attended population-based breast-cancer screening through the Trento and Verona screening services (Italy) from August, 2011, to June, 2012. We did screen-reading in two sequential phases—2D only and integrated 2D and 3D mammography—yielding paired data for each screen. Standard double-reading by breast radiologists determined whether to recall the participant based on positive mammography at either screen read. Outcomes were measured from final assessment or excision histology. Primary outcome measures were the number of detected cancers, the number of detected cancers per 1000 screens, the number and proportion of false positive recalls, and incremental cancer detection attributable to integrated 2D and 3D mammography. We compared paired binary data with McNemar's test.

Findings 7292 women were screened (median age 58 years [IQR 54–63]). We detected 59 breast cancers (including 52 invasive cancers) in 57 women. Both 2D and integrated 2D and 3D screening detected 39 cancers. We detected 20 cancers with integrated 2D and 3D only versus none with 2D screening only (p<0·0001). Cancer detection rates were 5·3 cancers per 1000 screens (95% CI 3·8–7·3) for 2D only, and 8·1 cancers per 1000 screens (6·2–10·4) for integrated 2D and 3D screening. The incremental cancer detection rate attributable to integrated 2D and 3D mammography was 2·7 cancers per 1000 screens (1·7–4·2). 395 screens (5·5%; 95% CI 5·0–6·0) resulted in false positive recalls: 181 at both screen reads, and 141 with 2D only versus 73 with integrated 2D and 3D screening (p<0·0001). We estimated that conditional recall (positive integrated 2D and 3D mammography as a condition to recall) could have reduced false positive recalls by 17·2% (95% CI 13·6–21·3) without missing any of the cancers detected in the study population.

Interpretation Integrated 2D and 3D mammography improves breast-cancer detection and has the potential to reduce false positive recalls. Randomised controlled trials are needed to compare integrated 2D and 3D mammography with 2D mammography for breast cancer screening.

Funding National Breast Cancer Foundation, Australia; National Health and Medical Research Council, Australia; Hologic, USA; Technologic, Italy.

Lancet Oncol 2013; 14: 583–89

Published Online
April 25, 2013
http://dx.doi.org/10.1016/S1470-2045(13)70134-7

UO Senologia Clinica e Screening Mammografico, Department of Diagnostics, Azienda Provinciale Servizi Sanitari, Trento, Italy (S Ciatto PhD, D Bernardi MD, M Pellegrini MD, P Tuttobene MD, C Fantò MD, M Valentini MD); **Centro di Prevenzione Senologica, Marzana, Verona, Italy** (S Ciatto, F Caumo MD, S Brunelli MD, P Bricolo MD, S Montemezzi MD); **and Screening and Test Evaluation Program, School of Public Health, Sydney Medical School, University of Sydney, Sydney, NSW, Australia** (Prof N Houssami MBBS/PhD, Prof P Macaskill PhD)

†Died May, 2012

Correspondence to:
Prof Nehmat Houssami, School of Public Health, Sydney Medical School, University of Sydney, Sydney 2006, NSW, Australia
nehmath@med.usyd.edu.au

Introduction

Although controversial, mammography screening is the only population-level early detection strategy that has been shown to reduce breast-cancer mortality in randomised trials.[1,2] Irrespective of which side of the mammography screening debate one supports,[1–3] efforts should be made to investigate methods that enhance the quality of (and hence potential benefit from) mammography screening. A limitation of standard 2D mammography is the superimposition of breast tissue or parenchymal density, which can obscure cancers or make normal structures appear suspicious. This shortcoming reduces the sensitivity of mammography and increases false-positive screening. Digital breast tomosynthesis with 3D images might help to overcome these limitations. Several reviews[4,5] have described the development of breast tomosynthesis technology, in which several low-dose radiographs are used to reconstruct a pseudo-3D image of the breast.[4–6]

Initial clinical studies of 3D mammography,[6–10] though based on small or selected series, suggest that addition of 3D to 2D mammography could improve cancer detection and reduce the number of false positives. However, previous assessments of breast tomosynthesis might have been constrained by selection biases that distorted the potential effect of 3D mammography; thus, screening trials of integrated 2D and 3D mammography are needed.[6]

We report the results of a large prospective study (Screening with Tomosynthesis OR standard Mammography [STORM]) of 3D digital mammography. We investigated the effect of screen-reading using both standard 2D and 3D imaging with tomosynthesis compared with screening with standard 2D digital mammography only for population breast-cancer screening.

Methods

Study design and participants

STORM is a prospective population-screening study that compares mammography screen-reading in two sequential phases (figure)—2D only versus integrated 2D and 3D mammography with tomosynthesis—yielding paired results for each screening examination. Women aged 48 years or older who attended population-based screening through the Trento and Verona screening services, Italy, from August, 2011, to June, 2012, were invited to be screened with integrated 2D and 3D mammography. Participants in routine screening mammography (once every 2 years) were asymptomatic women at standard (population) risk for breast cancer. The study was granted institutional ethics approval at each centre, and participants gave written informed consent. Women who opted not to participate in the study received standard 2D mammography. Digital mammography has been used in the Trento breast-screening programme since 2005, and in the Verona programme since 2007; each service monitors outcomes and quality indicators as dictated by European standards, and both have published data for screening performance.[11,12]

Procedures

All participants had digital mammography using a Selenia Dimensions Unit with integrated 2D and 3D mammography done in the COMBO mode (Hologic, Bedford, MA, USA): this setting takes 2D and 3D images at the same screening examination with a single breast position and compression. Each 2D and 3D image consisted of a bilateral two-view (mediolateral oblique and craniocaudal) mammogram. Screening mammograms were interpreted sequentially by radiologists, first on the basis of standard 2D mammography alone, and then by the same radiologist (on the same day) on the basis of integrated 2D and 3D mammography (figure). Thus, integrated 2D and 3D mammography screening refers to non-independent screen reading based on joint interpretation of 2D and 3D images, and does not refer to analytical combinations. Radiologists had to record whether or not to recall the participant at each screen-reading phase before progressing to the next phase of the sequence. For each screen, data were also collected for breast density (at the 2D screen-read), and the side and quadrant for any recalled abnormality (at each screen-read). All eight radiologists were breast radiologists with a mean of 8 years (range 3–13 years) experience in mammography screening, and had received basic training in integrated 2D and 3D mammography. Several of the radiologists had also used 2D and 3D mammography for patients recalled after positive conventional mammography screening as part of previous studies of tomosynthesis.[8,13]

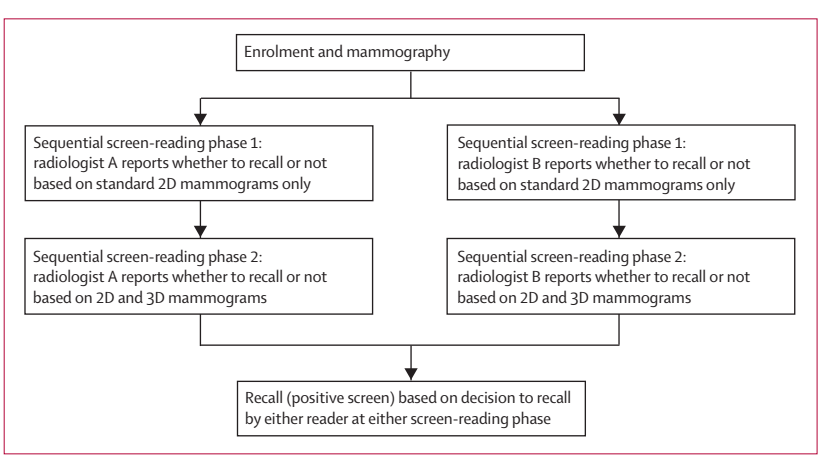

Figure: **Study design**

Mammograms were interpreted in two independent screen-reads done in parallel, as practised in most population breast-screening programmes in Europe. A screen was considered positive and the woman recalled for further investigations if either screen-reader recorded a positive result at either 2D or integrated 2D and 3D screening (figure). When previous screening mammograms were available, these were shown to the radiologist at the time of screen-reading, as is standard practice. For assessment of breast density, we used Breast Imaging Reporting and Data System (BI-RADS)[14] classification, with participants allocated to one of two groups (1–2 [low density] or 3–4 [high density]). Disagreement between readers about breast density was resolved by assessment by a third reader.

Our primary outcomes were the number of cancers detected, the number of cancers detected per 1000 screens, the number and percentage of false positive recalls, and the incremental cancer detection rate attributable to integrated 2D and 3D mammography screening. We compared the number of cancers that were detected only at 2D mammography screen-reading and those that were detected only at 2D and 3D mammography screen-reading; we also did this analysis for false positive recalls. To explore the potential effect of integrated 2D and 3D screening on false-positive recalls, we also estimated how many false-positive recalls would have resulted from using a hypothetical conditional false-positive recall approach; ie—positive integrated 2D and 3D mammography as a condition of recall (screening recalled at 2D mammography only would not be recalled). Pre-planned secondary analyses were comparison of outcome measures by age group and breast density.

Outcomes were assessed by excision histology for participants who had surgery, or the complete assessment outcome (including investigative imaging with or without histology from core needle biopsy) for all recalled participants. Because our study focuses on the difference in detection by the two screening methods, some cancers might have been missed by both 2D and integrated 2D and 3D mammography; this possibility could be assessed at future follow-up to identify interval cancers. However, this outcome is not assessed in the present study and

does not affect estimates of our primary outcomes—ie, comparative true or false positive detection for 2D-only versus integrated 2D and 3D mammography.

Statistical analysis

The sample size was chosen to provide 80% power to detect a difference of 20% in cancer detection, assuming a detection probability of 80% for integrated 2D and 3D screening mammography and 60% for 2D only screening, with a two-sided significance threshold of 5%. Based on the method of Lachenbruch[15] for estimating sample size for studies that use McNemar's test for paired binary data, a minimum of 40 cancers were needed. Because most screens in the participating centres were incident (repeat) screening (75%–80%), we used an underlying breast-cancer prevalence of 0·5% to estimate that roughly 7500–8000 screens would be needed to identify 40 cancers in the study population.

We calculated the Wilson CI for the false-positive recall ratio for integrated 2D and 3D screening with conditional recall compared with 2D only screening.[16] All of the other analyses were done with SAS/STAT (version 9.2), using exact methods to compute 95 CIs and p-values.

Role of the funding source

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author (NH) had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

7292 participants with a median age of 58 years (IQR 54–63, range 48–71) were screened between Aug 12, 2011, and June 29, 2012. Roughly 5% of invited women declined integrated 2D and 3D screening and received standard 2D mammography. We present data for 7294 screens because two participants had bilateral cancer (detected with different screen-reading techniques for one participant). We detected 59 breast cancers in 57 participants (52 invasive cancers and seven ductal carcinoma in-situ). Of the invasive cancers, most were invasive ductal (n=37); others were invasive special types (n=7), invasive lobular (n=4), and mixed invasive types (n=4). Table 1 shows the characteristics of the cancers. Mean tumour size (for the invasive cancers with known exact size) was 13·7 mm (SD 5·8) for cancers detected with both 2D alone and integrated 2D and 3D screening (n=29), and 13·5 mm (SD 6·7) for cancers detected only with integrated 2D and 3D screening (n=13).

Of the 59 cancers, 39 were detected at both 2D and integrated 2D and 3D screening (table 2). 20 cancers were detected with only integrated 2D and 3D screening compared with none detected with only 2D screening (p<0·0001; table 2). 395 screens were false positive (5·5%, 95% CI 5·0–6·0); 181 occurred at both screen-readings,

	Cancers detected at both 2D only and integrated 2D and 3D screening (n=39)	Cancers detected only at integrated 2D and 3D screening (n=20)
pT category		
pTis (in-situ cancer)	4 (10%)	3 (15%)
pT1a (≤5 mm)	3 (8%)	0 (0%)
pT1b (>5 mm but ≤10 mm)	10 (26%)	8 (40%)
pT1c (>10 mm but ≤20 mm)	20 (51%)	8 (40%)
pT2 (>20 mm but ≤50 mm)	2 (5%)	1 (5%)
Node status		
Negative for metastases	24 (62%)	11 (55%)
Positive for metastases	9 (23%)	3 (15%)
Micrometastases or isolated tumour cells	2 (5%)	1 (5%)
No node surgery or data not available	4 (10%)	5 (25%)

Data are n (%). pT=pathological tumour size.

Table 1: **Breast-cancer characteristics**

	Breast cancer			No breast cancer (including false recalls)		
	Integrated 2D and 3D positive	Integrated 2D and 3D negative	Total	Integrated 2D and 3D positive	Integrated 2D and 3D negative	Total
2D positive	39	0	39	181	141	322
2D negative	20	0*	20	73	6840	6913
Total	59	0	59	254	6981	7235
p value	<0·0001			<0·0001		

p values are exact for McNemar's test for paired binary data. *Does not include follow-up data on interval cancers.

Table 2: **Outcomes of screening**

	Number of cancers	Cancer detection rate (cancers per 1000 screens; 95% CI)	p value	Incremental cancer detection rate attributed to integrated 2D and 3D screening (95% CI)
Overall (7294 screens)				
2D mammography	39	5·3 (3·8–7·3)	..	..
Integrated 2D and 3D mammography	59	8·1 (6·2–10·4)	<0·0001	2·7 (1·7–4·2)
Age <60 years (4044 screens)				
2D mammography	20	4·9 (3·0–7·6)	..	..
Integrated 2D and 3D mammography	27	6·7 (4·4–9·7)	0·016	1·7 (0·7–3·6)
Age ≥60 years (3250 screens)				
2D mammography	19	5·8 (3·5–9·1)	..	..
Integrated 2D and 3D mammography	32	9·8 (6·7–13·9)	<0·0001	4·0 (2·1–6·8)
Breast density 1–2 (6079 screens)				
2D mammography	34	5·6 (3·9–7·8)	..	..
Integrated 2D and 3D mammography	51	8·4 (6·3–11·0)	<0·0001	2·8 (1·6–4·5)
Breast density 3–4 (1215 screens)				
2D mammography	5	4·1 (3·1–9·6)	..	..
Integrated 2D and 3D mammography	8	6·6 (4·1–18·6)	0·25	2·5 (0·5–7·2)

Table 3: **Breast-cancer detection rates, and incremental detection from integrated 2D and 3D screening mammography**

	Age <60 years			Age ≥60 years		
	Integrated 2D and 3D positive	Integrated 2D and 3D negative	Total	Integrated 2D and 3D positive	Integrated 2D and 3D negative	Total
Breast cancer						
2D positive	20	0	20	19	0	19
2D negative	7	0*	7	13	0*	13
Total	27	0	27	32	0	32
p value	0·016			<0·0001		
No breast cancer (including false recalls)						
2D positive	129	89	218	52	52	104
2D negative	41	3758	3799	32	3082	3114
Total	170	3847	4017	84	3134	3218
p value	<0·0001			0·038		

p values are exact for McNemar's test for paired binary data. *Does not include follow-up data on interval cancers, this does not affect the comparative detection data.

Table 4: **Outcomes of screening, stratified by age**

	Breast density 1–2 (low density)			Breast density 3–4 (high density)		
	Integrated 2D and 3D positive	Integrated 2D and 3D negative	Total	Integrated 2D and 3D positive	Integrated 2D and 3D negative	Total
Breast cancer						
2D positive	34	0	34	5	0	5
2D negative	17	0*	17	3	0*	3
Total	51	0	51	8	0	8
p value	<0·0001			0·25		
No breast cancer (including false recalls)						
2D positive	130	109	239	51	32	83
2D negative	52	5737	5789	21	1103	1124
Total	182	5846	6028	72	1135	1207
p value	<0·0001			0·17		

p values are exact for McNemar's test for paired binary data. *Does not include follow-up data on interval cancers, this does not affect the comparative detection data.

Table 5: **Outcomes of screening, stratified by breast density**

and 141 occurred at 2D screening only compared with 73 at integrated 2D and 3D screening (p<0·0001; table 2). These differences were still significant in sensitivity analyses that excluded the two participants with bilateral cancer (data not shown).

5·3 cancers per 1000 screens (95% CI 3·8–7·3; table 3) were detected with 2D mammography only versus 8·1 cancers per 1000 screens (95% CI 6·2–10·4) with integrated 2D and 3D mammography (p<0·0001). The incremental cancer detection rate attributable to integrated 2D and 3D screening was 2·7 cancers per 1000 screens (95% CI 1·7–4·2), which is 33·9% (95% CI 22·1–47·4) of the cancers detected in the study population. In a sensitivity analysis that excluded the two participants with bilateral cancer the estimated incremental cancer detection rate attributable to integrated 2D and 3D screening was 2·6 cancers per 1000 screens (95% CI 1·4–3·8). The stratified results show that integrated 2D and 3D mammography was associated with an incrementally increased cancer detection rate in both age-groups and density categories (tables 3–5). A minority (16·7%) of breasts were of high density (category 3–4) reducing the power of statistical comparisons in this subgroup (table 5). The incremental cancer detection rate was much the same in low density versus high density groups (2·8 per 1000 *vs* 2·5 per 1000; p=0·84; table 3).

Overall recall—any recall resulting in true or false positive screens—was 6·2% (95% CI 5·7–6·8), and the false-positive rate for the 7235 screens of participants who did not have breast cancer was 5·5% (5·0–6·0). Table 6 shows the contribution to false-positive recalls from 2D mammography only, integrated 2D and 3D mammography only, and both, and the estimated number of false positives if positive integrated 2D and 3D mammography was a condition for recall (positive 2D only not recalled). Overall, more of the false-positive rate was driven by 2D mammography only than by integrated 2D and 3D, although almost half of the false-positive rate was a result of false positives recalled at both screen-reading phases (table 6). The findings were much the same when stratified by age and breast density (table 6). Had a conditional recall rule been applied, we estimate that the false-positive rate would have been 3·5% (95% CI 3·1–4·0%; table 6) and could have potentially prevented 68 of the 395 false positives (a reduction of 17·2%; 95% CI 13·6–21·3). The ratio between the number of false positives with integrated 2D and 3D screening with conditional recall (n=254) versus 2D only screening (n=322) was 0·79 (95% CI 0·71–0·87).

Discussion

Our study showed that integrated 2D and 3D mammography screening significantly increases detection of breast cancer compared with conventional mammography screening. There was consistent evidence of an incremental improvement in detection from integrated 2D and 3D mammography across age-group and breast density strata, although the analysis by breast density was limited by low number of women with breasts of high density.

One should note that we investigated comparative cancer detection, and not absolute screening sensitivity. By integrating 2D and 3D mammography using the study screen-reading protocol, 1% of false-positive recalls resulted from 2D and 3D screen-reading only (table 6). However, significantly more false positives resulted from 2D only mammography compared with integrated 2D and 3D mammography, both overall and in the stratified analyses. Application of a conditional recall rule would have resulted in a false-positive rate of 3·5% instead of the actual false-positive rate of 5·5%. The estimated false positive recall ratio of 0·79 for integrated 2D and 3D screening with conditional recall compared with 2D only screening suggests that integrated 2D and 3D screening

could reduce false recalls by roughly a fifth. Had such a condition been adopted, none of the cancers detected in the study would have been missed because no cancers were detected by 2D mammography only, although this result might be because our design allowed an independent read for 2D only mammography whereas the integrated 2D and 3D read was an interpretation of a combination of 2D and 3D imaging. We do not recommend that such a conditional recall rule be used in breast-cancer screening until our findings are replicated in other mammography screening studies—STORM involved double-reading by experienced breast radiologists, and our results might not apply to other screening settings. Using a test set of 130 mammograms, Wallis and colleagues[7] report that adding tomosynthesis to 2D mammography increased the accuracy of inexperienced readers (but not of experienced readers), therefore having experienced radiologists in STORM could have underestimated the effect of integrated 2D and 3D screen-reading.

No other population screening trials of integrated 2D and 3D mammography have reported final results (panel); however, an interim analysis of the Oslo trial[17]—a large population screening study—has shown that integrated 2D and 3D mammography substantially increases detection of breast cancer. The Oslo study investigators screened women with both 2D and 3D mammography, but randomised reading strategies (with *vs* without 3D mammograms) and adjusted for the different screen-readers,[17] whereas we used sequential screen-reading to keep the same reader for each examination. Our estimates for comparative cancer detection and for cancer detection rates are consistent with those of the interim analysis of the Oslo study.[17] The applied recall methods differed between the Oslo study (which used an arbitration meeting to decide recall) and the STORM study (we recalled based on a decision by either screen-reader), yet both studies show that 3D mammography reduces false-positive recalls when added to standard mammography.

An editorial in *The Lancet*[18] might indeed signal the closing of a chapter of debate about the benefits and harms of screening. We hope that our work might be the beginning of a new chapter for mammography screening: our findings should encourage new assessments of screening using 2D and 3D mammography and should factor several issues related to our study. First, we compared standard 2D mammography with integrated 2D and 3D mammography—the 3D mammograms were not interpreted independently of the 2D mammograms—therefore 3D mammography only (without the 2D images) might not provide the same results. Our experience with breast tomosynthesis—and a review[6] of 3D mammography—underscore the importance of 2D images in integrated 2D and 3D screen-reading. The 2D images form the basis of the radiologist's ability to integrate the information from 3D images with that from 2D images. Second, although most screening in STORM was incident screening, the substantial increase in cancer detection rate with integrated 2D and 3D mammography results from the enhanced sensitivity of integrated 2D and 3D screening and is probably also a result of a prevalence effect (ie, the effect of a first screening round with integrated 2D and 3D mammography). We did not assess the effect of repeat (incident) screening with integrated 2D and 3D mammography on cancer detection—it might provide a smaller effect on cancer detection rates than what we report. Third, STORM was not designed to measure biological differences between the cancers detected at integrated 2D and 3D screening compared with those detected at both screen-reading phases. Descriptive analyses suggest that, generally, breast cancers detected only at integrated 2D and 3D screening had similar features (eg, histology, pathological tumour size, node status) as those detected at both screen-reading phases. Thus, some of the cancers detected only at 2D and

	n	% (95% CI)
Overall (7235 screens*)		
Recalled at either 2D or integrated 2D and 3D mammography	395	5·5% (5·0–6·0)
Recalled at both 2D and integrated 2D and 3D mammography	181	2·5% (2·2–2·9)
Recalled at 2D mammography only	141	2·0% (1·6–2·3)
Recalled at integrated 2D and 3D mammography only	73	1·0% (0·8–1·3)
Conditional false positive recalls†	254	3·5% (3·1–4·0)
Age <60 years (4017 screens)		
Recalled at either 2D or integrated 2D and 3D mammography	259	6·5% (5·7–7·3)
Recalled at both 2D and integrated 2D and 3D mammography	129	3·2% (2·7–3·8)
Recalled at 2D mammography only	89	2·2% (1·8–2·7)
Recalled at integrated 2D and 3D mammography only	41	1·0% (0·7–1·4)
Conditional false positive recalls†	170	4·2% (3·6–4·9)
Age ≥60 years (3218 screens)		
Recalled at either 2D or integrated 2D and 3D mammography	136	4·2% (3·6–5·0)
Recalled at both 2D and integrated 2D and 3D mammography	52	1·6% (1·2–2·1)
Recalled at 2D mammography only	52	1·6% (1·2–2·1)
Recalled at integrated 2D and 3D mammography only	32	1·0% (0·7–1·4)
Conditional false positive recalls†	84	2·6% (2·1–3·2)
Breast density 1–2 (6028 screens)		
Recalled at either 2D or integrated 2D and 3D mammography	291	4·8% (4·3–5·4)
Recalled at both 2D and integrated 2D and 3D mammography	130	2·2% (1·8–2·6)
Recalled at 2D mammography only	109	1·8% (1·5–2·2)
Recalled at integrated 2D and 3D mammography only	52	0·9% (0·6–1·1)
Conditional false positive recalls†	182	3·0% (2·6–3·5)
Breast density 3–4 (1207 screens)		
Recalled at either 2D or integrated 2D and 3D mammography	104	8·6% (7·1–10·3)
Recalled at both 2D and integrated 2D and 3D mammography	51	4·2% (3·2–5·5)
Recalled at 2D mammography only	32	2·7% (1·8–3·7)
Recalled at integrated 2D and 3D mammography only	21	1·7% (1·1–2·7)
Conditional false positive recalls†	72	6·0% (4·7–7·5)

*Did not have breast cancer. †False-positive recalls using positive integrated 2D and 3D mammography as a condition to recall (ie—positive 2D mammography only would not be recalled).

Table 6: **False-positive recalls for mammography screening**

Panel: Research in context

Systematic review

NH searched the English language published work up to October, 2012, to assess the evidence of the accuracy of digital breast tomosynthesis (3D mammography), using the search strategy and eligibility criteria detailed in a systematic review.[6] The search consisted of a Medline search (exploded "breast neoplasm", combined with "tomosyn$" in title) and contact with experts. We did not identify any randomised controlled trials of digital breast tomosynthesis, or any population screening trials of digital breast tomosynthesis that had reported final results.[6] Studies of digital breast tomosynthesis were generally small test-set (observer) studies with a high proportion of patients with cancer or were based on selected clinical series with various limitations, however these studies provided evidence that addition of digital breast tomosynthesis to standard mammography increases the accuracy of interpretation.[6] An interim analysis of the Oslo trial—a large screening trial in which women had both standard mammography and digital breast tomosynthesis with randomly assigned screen-reading strategies—reported that the addition of digital breast tomosynthesis to digital mammography significantly increased detection of breast cancer and reduced false positive recalls.[17]

Interpretation

Our work is, to the best of our knowledge, the first population breast-cancer screening trial of integrated 2D and 3D mammography screening to report its final results. Our findings—that integrated 2D and 3D screening significantly increased breast cancer detection (relative to standard 2D mammography) and has the potential to reduce false-positive recalls—accord with the interim analysis of the Oslo trial.[17] These two screening trials provide evidence that integrated 2D and 3D mammography screening could be used as part of population screening, provided that screening outcome measures are carefully monitored. To assess whether the measured increase in breast cancer detection with integrated 2D and 3D mammography is likely to translate into improved screening efficacy, a randomised controlled trial is needed to compare integrated 2D and 3D mammography with standard 2D mammography with interval cancer rates as the endpoint.

3D screening might represent early detection (and would be expected to receive screening benefit) whereas some might represent over-detection and a harm from screening, as for conventional screening mammography.[1,19] The absence of consensus about over-diagnosis in breast-cancer screening should not detract from the importance of our study findings to applied screening research and to screening practice; however, our trial was not done to assess the extent to which integrated 2D and 3D mammography might contribute to over-diagnosis.

The average dose of glandular radiation from the many low-dose projections taken during a single acquisition of 3D mammography is roughly the same as that from 2D mammography.[6,20–22] Using integrated 2D and 3D entails both a 2D and 3D acquisition in one breast compression, which roughly doubles the radiation dose to the breast. Therefore, integrated 2D and 3D mammography for population screening might only be justifiable if improved outcomes were not defined solely in terms of improved detection. For example, it would be valuable to show that the increased detection with integrated 2D and 3D screening leads to reduced interval cancer rates at follow-up. A limitation of our study might be that data for interval cancers were not available; however, because of the paired design we used, future evaluation of interval cancer rates from our study will only apply to breast cancers that were not identified using 2D only or integrated 2D and 3D screening. We know of two patients from our study who have developed interval cancers (follow-up range 8–16 months). We did not get this information from cancer registries and follow-up was very short, so these data should be interpreted very cautiously, especially because interval cancers would be expected to occur in the second year of the standard 2 year interval between screening rounds. Studies of interval cancer rates after integrated 2D and 3D mammography would need to be randomised controlled trials and have a very large sample size. Additionally, the development of reconstructed 2D images from a 3D mammogram[23] provides a timely solution to concerns about radiation by providing both the 2D and 3D images from tomosynthesis, eliminating the need for two acquisitions.

We have shown that integrated 2D and 3D mammography in population breast-cancer screening increases detection of breast cancer and can reduce false-positive recalls depending on the recall strategy. Our results do not warrant an immediate change to breast-screening practice, instead, they show the urgent need for randomised controlled trials of integrated 2D and 3D versus 2D mammography, and for further translational research in breast tomosynthesis. We envisage that future screening trials investigating this issue will include measures of breast cancer detection, and will be designed to assess interval cancer rates as a surrogate endpoint for screening efficacy.

Contributors

SC had the idea for and designed the study, and collected and interpreted data. NH advised on study concepts and methods, analysed and interpreted data, searched the published work, and wrote and revised the report. DB and FC were lead radiologists, recruited participants, collected data, and commented on the draft report. MP, SB, PT, PB, PT, CF, and MV did the screen-reading, collected data, and reviewed the draft report. SM collected data and reviewed the draft report. PM planned the statistical analysis, analysed and interpreted data, and wrote and revised the report.

Conflicts of interest

SC, DB, FC, MP, SB, PT, PB, CF, MV, and SM received assistance from Hologic (Hologic USA; Technologic Italy) in the form of tomosynthesis technology and technical support for the duration of the study, and travel support to attend collaborators' meetings. NH receives research support from a National Breast Cancer Foundation (NBCF Australia) Practitioner Fellowship, and has received travel support from Hologic to attend a collaborators' meeting. PM receives research support through Australia's National Health and Medical Research Council programme grant 633003 to the Screening & Test Evaluation Program.

References

1 Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012; **380:** 1778–86.

2 Glasziou P, Houssami N. The evidence base for breast cancer screening. *Prev Med* 2011; **53:** 100–102.

3 Autier P, Esserman LJ, Flowers CI, Houssami N. Breast cancer screening: the questions answered. *Nat Rev Clin Oncol* 2012; **9:** 599–605.

4 Baker JA, Lo JY. Breast tomosynthesis: state-of-the-art and review of the literature. *Acad Radiol* 2011; **18:** 1298–310.

5 Helvie MA. Digital mammography imaging: breast tomosynthesis and advanced applications. *Radiol Clin North Am* 2010; **48:** 917–29.

6 Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. *Breast* 2013; **22:** 101–08.

7 Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution X-ray imaging observer study. *Radiology* 2012; **262:** 788–96.

8 Bernardi D, Ciatto S, Pellegrini M, et al. Prospective study of breast tomosynthesis as a triage to assessment in screening. *Breast Cancer Res Treat* 2012; **133:** 267–71.

9 Michell MJ, Iqbal A, Wasan RK, et al. A comparison of the accuracy of film-screen mammography, full-field digital mammography, and digital breast tomosynthesis. *Clin Radiol* 2012; **67:** 976–81.

10 Skaane P, Gullien R, Bjorndal H, et al. Digital breast tomosynthesis (DBT): initial experience in a clinical setting. *Acta Radiol* 2012; **53:** 524–29.

11 Pellegrini M, Bernardi D, Di MS, et al. Analysis of proportional incidence and review of interval cancer cases observed within the mammography screening programme in Trento province, Italy. *Radiol Med* 2011; **116:** 1217–25.

12 Caumo F, Vecchiato F, Pellegrini M, Vettorazzi M, Ciatto S, Montemezzi S. Analysis of interval cancers observed in an Italian mammography screening programme (2000–2006). *Radiol Med* 2009; **114:** 907–14.

13 Bernardi D, Ciatto S, Pellegrini M, et al. Application of breast tomosynthesis in screening: incremental effect on mammography acquisition and reading time. *Br J Radiol* 2012; **85:** e1174–78.

14 American College of Radiology. ACR BI-RADS: breast imaging reporting and data system, Breast Imaging Atlas. Reston: American College of Radiology, 2003.

15 Lachenbruch PA. On the sample size for studies based on McNemar's test. *Stat Med* 1992; **11:** 1521–25.

16 Bonett DG, Price RM. Confidence intervals for a ratio of binomial proportions based on paired data. *Stat Med* 2006; **25:** 3039–47.

17 Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology* 2013; published online Jan 3. http://dx.doi.org/10.1148/radiol.12121373.

18 The Lancet. The breast cancer screening debate: closing a chapter? *Lancet* 2012; **380:** 1714.

19 Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. *Lancet Oncol* 2007; **8:** 1129–38.

20 Tagliafico A, Astengo D, Cavagnetto F, et al. One-to-one comparison between digital spot compression view and digital breast tomosynthesis. *Eur Radiol* 2012; **22:** 539–44.

21 Tingberg A, Fornvik D, Mattsson S, Svahn T, Timberg P, Zackrisson S. Breast cancer screening with tomosynthesis—initial experiences. *Radiat Prot Dosimetry* 2011; **147:** 180–83.

22 Feng SS, Sechopoulos I. Clinical digital breast tomosynthesis system: dosimetric characterization. *Radiology* 2012; **263:** 35–42.

23 Gur D, Zuley ML, Anello MI, et al. Dose reduction in digital breast tomosynthesis (DBT) screening using synthetically reconstructed projection images: an observer performance study. *Acad Radiol* 2012; **19:** 166–71.