


# False positives in breast cancer screening with one-view breast tomosynthesis: An analysis of findings leading to recall, work-up and biopsy rates in the Malmö Breast Tomosynthesis Screening Trial

Kristina Lång<sup>1</sup>  · Matilda Nergården<sup>1</sup> · Ingvar Andersson<sup>1</sup> · Aldana Rosso<sup>2</sup> ·  
Sophia Zackrisson<sup>1</sup>

Received: 29 October 2015 / Revised: 23 January 2016 / Accepted: 1 February 2016 / Published online: 4 March 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

## Abstract

**Objectives** To analyse false positives (FPs) in breast cancer screening with tomosynthesis (BT) vs. mammography (DM). **Methods** The Malmö Breast Tomosynthesis Screening Trial (MBTST) is a prospective population-based study comparing one-view BT to DM in screening. This study is based on the first half of the MBTST population (n = 7,500). Differences in FP recall rate, findings leading to recall, work-up and biopsy rate between cases recalled on BT alone, DM alone and BT+DM were analysed.

**Results** The FP recall rate was 1.7 % for BT alone (n = 131), 0.9 % for DM alone (n = 69) and 1.1 % for BT+DM (n = 81). The FP recall rate for BT alone was halved after the initial phase of the trial, stabilising at 1.5 %. BT doubled the recall of stellate distortions compared to DM (n = 64 vs. n = 33). There were fewer fibroadenomas and cysts, and the biopsy rate was slightly lower for FP recalled on BT alone compared to DM alone (15.3 % vs. 27.6 %: p = 0.037 and 33.8 % vs. 36.2 %; p = 0.641, respectively).

**Conclusions** FPs increased with BT screening mainly due to the recall of stellate distortions. The FP recall rate was still well within the European guidelines and showed evidence of a learning curve. Characterisation of rounded lesions was improved with BT.

## Key Points

- Tomosynthesis screening gave a higher false-positive recall rate than mammography
- There was a decline in the false-positive recall rate for tomosynthesis
- The recall due to stellate distortions simulating malignancy was doubled with tomosynthesis
- Tomosynthesis found more radial and postoperative scar tissue than mammography
- Tomosynthesis is better at characterising rounded lesions

**Keywords** Digital mammography · Breast tomosynthesis · Screening · Diagnostic imaging · Breast cancer

## Abbreviations

AGD	Average glandular dose
BT	Breast tomosynthesis
CC	Craniocaudal
DM	Digital mammography
FPs	False positives
MBTST	Malmö Breast Tomosynthesis Screening Trial
MLO	Mediolateral oblique
NOS	Not otherwise specified

## Introduction

Digital mammography (DM) is the standard modality in breast cancer screening today, but the sensitivity has been shown to be suboptimal especially in women with dense breasts [1–3]. In recent years, the use of breast tomosynthesis (BT) in screening has been investigated in several large prospective population-based trials, showing a substantial increase in the

✉ Kristina Lång  
kristina.lang@med.lu.se

<sup>1</sup> Department of Medical Radiology, Translational Medicine Malmö, Lund University, Inga Marie Nilssons gata 49, SE-20502 Malmö, Sweden

<sup>2</sup> Epidemiology and Register Centre South, Skåne University Hospital, Klinikgatan 22, SE-221 85 Lund, Sweden

cancer detection rate, that is by 30–40 %, when used in addition to DM [4–7] or as a stand-alone modality [8] and is, therefore, a candidate for becoming the next generation breast cancer screening modality. However, data from two of these prospective screening trials show an increase in the number of false positives (FPs) when using BT in screening [4, 6, 8] and might, therefore, be an adverse effect that needs to be analysed before a largescale implementation of the technique.

Women having an FP result when participating in breast cancer screening can experience both short- and long-term psychological distress [9, 10]. As a consequence, these women may be less likely to participate in subsequent screening rounds [9]. A high participation rate is an important screening performance measure in community-based screening programmes, since low participation rates are associated with low cost-effectiveness [11]. Furthermore, it has been shown that women with an FP screening mammogram have an increased risk of breast cancer [9, 12] and it is, therefore, important that these women are not lost to the screening programme. The estimated cumulative risk of an FP screening result in women aged 50–69 years undergoing ten biennial screening tests in Europe varies from 8 % to 21 % [13]. There are several factors influencing the risk of an FP result, such as a family history of breast cancer, oestrogen use, breast density and time between screenings [14, 15].

BT is a tomographic technique where the X-ray tube moves at an angle over the breast acquiring multiple low-dose projections that are reconstructed into a tomosynthesis image volume, in which each slice typically represents a 1-mm thin cross-section of the breast [16]. This approach reduces the detrimental effect of overlapping breast tissue on diagnostic performance in DM [2, 17], resulting in higher sensitivity [18] and improved lesion characterisation [19, 20], especially of spiculated tumours [21, 22]. As a consequence, BT will also reveal benign lesions that are concealed in DM and sometimes enhance normal parenchymal components, with the possibility of increasing the FP rate.

Several large retrospective-screening studies in the USA have shown the importance of adding two-view BT to two-view DM in a so-called *combination mode* in order to lower the recall rate. In these studies, the recall rate has been reported to vary between 8.7 % and 16.2 % for DM compared to 5.4 % and 13.6 % for the combination mode [23–29]. In Europe, where the recall rates are typically lower (below 5 %) [30], interim results from two large prospective population-based screening trials: the Oslo Tomosynthesis Screening Trial and the Malmö Breast Tomosynthesis Screening Trial (MBTST), have shown a statistically significant increase in the recall rate with independent double-reading of DM compared to the combination mode (2.9 % vs. 3.7 %) [6] or compared to one-view BT as a stand-alone modality (2.6 % vs. 3.8 %) [8]. The substantial increase in cancer detection rate contributed to increasing the recall rate, but there was also a slight

increase in the FP recall rate [6, 8]. Yet, another large prospective population-based screening trial, the Screening with Tomosynthesis OR Mammography (STORM) trial, designed to study the effect of sequential reading of two-view DM and the combination mode, gave an overall FP recall rate of 5.5 %, but the FP recall rate for the combination mode was lower compared to DM (3.5 % vs. 4.4 %). Nevertheless, the combination mode contributed to an addition of 73 FP cases (FP recall rate 1.0 %) that were negative on DM alone [7].

The aim of the current study was to characterise FP cases in breast cancer screening with one-view BT vs. DM with data from the MBTST, in terms of FP recall rate after arbitration, the findings leading to recall, the results of the work-up and biopsy rates. By characterising these cases we might improve our understanding of the causes leading to a FP result and the consequences for clinical practice in order to reduce a potentially negative effect of BT in breast cancer screening.

## Methods and materials

### The Malmö Breast Tomosynthesis Screening Trial

The MBTST is a prospective single-institution one-arm population-based study designed to compare the efficacy of one-view BT (mediolateral oblique view (MLO)) as a stand-alone breast cancer screening modality with two-view DM (MLO+ craniocaudal view (CC)) in women aged 40–74 years eligible for the screening programme in the City of Malmö ([www.clinicaltrials.gov](http://www.clinicaltrials.gov); NCT01091545). The study was approved by the Regional Ethical Review Board at Lund University (Dnr 2009/770) and the local Radiation Safety Board at Skåne University Hospital in Malmö. The MBTST is described in more detail elsewhere [8]. In short, participating women underwent a two-view DM as well as a one-view BT examination (Mammomat Inspiration, Siemens AG, Erlangen, Germany). The DM and BT images were subjected to independent blinded double reading and scoring in two independent reading arms, where findings were rated on a 5-point scale: (1) normal, (2) benign findings, (3) non-specific finding with low probability of malignancy, (4) findings suspicious of malignancy, and (5) findings highly suspicious of malignancy.

The two reading arms comprised three reading steps each. Each step was scored before proceeding to the following step. In the BT reading arm, BT alone was scored first (step 1); then with the addition of the DM CC view (step 2); and finally with the addition of prior DM if available (step 3). In the DM reading arm, DM alone was scored first (step 1) followed by the addition of prior DM (step 2). Breast density was classified according to BI-RADS (4th Edition) in the DM reading arm at reading step 3 [31]. If in either or both of the reading arms a case was given a score of 3 or higher by one of the two

readers, it was referred for arbitration. At the arbitration meeting, at least two readers re-evaluated the images and decided whether to recall the woman for further work-up, irrespective of the scores in the other reading arm. Thus, women could be recalled based on findings only on the BT reading arm, the DM reading arm or both reading arms (here called *BT alone*, *DM alone* and *BT+DM recall group*). Women reporting symptoms at the screen examination, e.g. palpable lump, could be recalled in spite of negative imaging findings. As reported previously [8] the cancer detection rate for BT was 8.9/1,000 screens and 6.3/1,000 screens DM. The recall rate after arbitration was 3.8 % for BT and 2.6 % for DM. The PPV was 24 % for both BT and DM. The MBTST is planned to include 15,000 women with a complete set of BT and DM images. This study is based on data from the first half of the MBTST population (n=7,500 women) who participated in the study between January 2010 and December 2012.

### Image acquisition

Two-view DM was immediately followed by one-view BT (Mammomat Inspiration, Siemens AG, Erlangen, Germany). BT images were acquired using the same beam quality and anode/filter combination (W/Rh) as DM. The automatic exposure control was set to yield an average glandular dose (AGD) of 1.2 mGy per DM image and 1.6 mGy for BT, for a standard breast of 53 mm consisting of 50 % glandular tissue and 50 % fatty tissue. Hence, the absorbed dose in a one-view BT was approximately 70 % of the absorbed dose in a two-view DM. The BT examination consisted of 25 projection images acquired over an angular range of 45°. These images were reconstructed into 1-mm slices using a generalised filtered back-projection reconstruction algorithm [32]. The BT examination was performed with reduced compression force of the breast compared to the previously acquired DM examination, with the goal of a 50 % reduction [33].

### False-positive (FP) cases

An FP case was defined as a recalled woman who was considered disease-free after work-up and at least a 3-year follow-up and through record linkage with the South Swedish Cancer Registry. Parameters collected from the MBTST were *FP recall rate after arbitration*, including *FP recall rate over time* and *population characteristics* (i.e. age and breast density) and *biopsy rates*. The FP cases were also analysed retrospectively by an expert panel consisting of three breast radiologists (mean 17 years' experience, range 1–42 years) and one medical student to assess the radiographic *finding leading to recall* and *outcome of the work-up*. Radiographic findings leading to recall were rendered through the primary description carried out by the radiologists at the arbitration meeting or by the

radiologist performing the initial work-up. If there was no distinct description of the finding, the expert panel categorised the finding by consensus. The following categories were used: stellate distortion, rounded lesion, indistinct density, calcifications, architectural distortion and symptoms. In the evaluation of the radiographic findings in the BT+DM recall group, the appearance on BT was chosen if there was a discrepancy between the modalities.

### Statistical analyses

Descriptive statistics (numbers and percentages) were used to analyse and present the data. A Chi<sup>2</sup> test was used to analyse differences in the proportions of findings leading to recall between the DM-alone and BT-alone recall group. Fisher's exact test was used to analyse the outcomes of the work-up and biopsy rate since there were few observations. Although there were three recall groups, we found the comparison of the DM-alone and BT-alone recall groups most relevant from an imaging perspective, since almost one-third of all women recalled in the BT+DM group were recalled due to the reporting of symptoms.

## Results

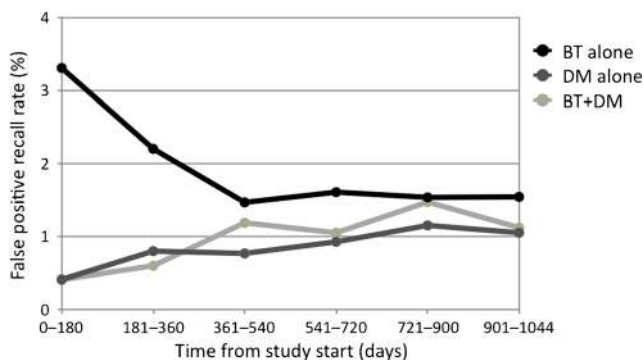
### FP recall rate after arbitration

Out of 7,500 screened women a total of 352 were recalled for work-up. Three women were excluded from the analysis, including one woman diagnosed with lymphoma and two women declining the work-up. Sixty-eight women were shown to have breast cancer and 281 were FPs. FP recall rate after arbitration for BT alone was 1.7 % (n=131), for DM alone 0.9 % (n=69) and for women recalled on both BT+DM 1.1 % (n=81) (Fig. 1). The majority of the cases were selected at reading step 1 (Table 1). The contribution of FP cases with the addition of prior DM was minor in the BT-alone and DM-alone recall groups (two cases per group). As expected, symptomatic women were mainly found in the BT+DM recall group.

The mean FP recall rate over time (1.5 years) for BT alone was 1.9 % (range 1.5–3.3), for DM alone 0.9 % (range 0.4–1.2) and for BT+DM 1.0 % (range 0.6–1.5). The FP recall rate for BT alone was halved during the first 1.5 year of the MBTST, stabilising on an FP recall rate of about 1.5 % (Fig. 1).

### Characteristics of the FP cases

The characteristics of the FP cases in the different recall groups are shown in Table 2.



**Fig. 1** False-positive recall rate over time. False-positive recall rate for breast tomosynthesis (BT) alone, digital mammography (DM) alone and for cases recalled on both BT+DM during the first half of the Malmö Breast Tomosynthesis Screening Trial

### Age and density

Women recalled on BT+DM were slightly younger and had denser breasts compared to the women recalled on DM and BT alone. The women in the BT-alone recall group had slightly fattier breasts, compared to the other recall groups.

### Finding leading to recall

Overall, the finding of an area of stellate distortion was the major cause of an FP in both modalities. There was a higher proportion of stellate distortions leading to a recall on BT alone compared to DM alone ( $n=53$ , 40.5 % (95 % CI 32.1–49.4) vs.  $n=22$ , 31.9 % (22.6–42.8);  $p=0.234$ ,  $\chi^2(1)=1.418$ ). In total, BT led to a doubling of the recall of stellate distortions compared to DM ( $n=64$  vs.  $n=33$ ). Furthermore, there were slightly fewer rounded lesions in the BT-alone group compared to DM alone ( $n=32$ , 24.4 % (16.9–33.9) vs.  $n=18$ , 26.1 % (16.1–39.3);  $p=0.797$ ,  $\chi^2(1)=0.066$ ). Women presenting with symptoms was the

**Table 1** Reading steps. Number of positive scores (rated 3 or higher) in the recall groups that resulted in a false-positive case in the two independent reading arms: Reading arm BT: one-view breast tomosynthesis (BT) alone (step 1); the addition of one-view digital mammography (DM) craniocaudal view (step 2); comparison with prior two-view DM, if available (step 3). Reading arm DM: two-view DM (step 1); and comparison with prior DM (step 2)

	BT alone (n)	DM alone (n)	BT+DM (n)	
			BT	DM
Symptoms	4	3	29	29
Step 1	122	64	45	51
Step 2	3	2	6	1
Step 3	2	n/a	1	n/a
Total	131	69	81	81

main reason for an FP in the BT+DM recall group ( $n=29$ , 35.8 %).

### Outcome of the work-up

The most frequent outcome for all FP cases was tissue that was considered free of abnormality, i.e. typically normal glandular tissue. This was also true for the majority of the additional FP cases attributed to screening with BT ( $n=74$ , 56.5 %) (Fig. 2, Case 1).

The work-up of all BT-alone cases resulted in the finding of more radial scars ( $n=5$ ) (Fig. 3, Case 2), postoperative scar tissue ( $n=8$ ) and benign lesions not otherwise specified (NOS) ( $n=5$ ), compared to the DM-alone group ( $n=0$ ,  $n=1$  and  $n=1$ , respectively). On the other hand, there were significantly fewer rounded lesions (fibroadenomas and cysts) in the BT-alone group compared to the DM-alone group ( $n=20$ , 15.3 % (95 % CI 10.9–21.0) vs.  $n=19$ , 27.6 % (18.2–39.3);  $p=0.037$ ). The work-up of women recalled on both BT+DM resulted mostly in the finding of benign cysts ( $n=20$ , 24.7 %).

In most cases, the assessment stellate distortions in all three recall groups ( $n=86$ ) led to the finding of normal breast tissue (Fig. 4). Stellate distortions recalled on BT alone contributed to the finding of normal tissue ( $n=43$ ), radial scars ( $n=5$ ), postoperative scar tissue ( $n=4$ ) and one cyst (Fig. 4). The work-up of radiographic findings with a rounded appearance recalled on BT alone resulted in cysts ( $n=10$ ), fibroadenomas ( $n=5$ ), normal tissue ( $n=8$ ), benign NOS ( $n=4$ ), lymph node ( $n=2$ ), atheroma ( $n=2$ ) and postoperative scar tissue ( $n=1$ ) (Fig. 4).

### Biopsy rate

The work-up of FP cases recalled on both BT+DM needed most biopsies (Fig. 5). The assessment of BT-alone cases had a slightly lower total biopsy rate compared to the DM-alone recall group ( $n=43$ , 32.8 % (95 % CI 24.9–41.6) vs.  $n=25$ , 36.2 % (25.0–48.7);  $p=0.641$ ). This was due to a lower fine needle aspiration rate ( $n=37$ , 28.2 % (20.7–36.8) vs.  $n=23$ , 33.3 % (22.5–45.7);  $p=0.517$ ), but the core-needle biopsy rate was slightly higher in the BT-alone group compared to DM alone ( $n=6$ , 4.6 % (1.7–9.7) vs.  $n=2$ , 2.9 % (0.4–10.1);  $p=0.717$ ).

### Discussion

The result of this study indicates that breast cancer screening with BT will lead to an increase in the recall of stellate distortions, of which the majority will show



**Table 2** False positives. Characteristics of the false-positive cases in the different recall groups: breast tomosynthesis (BT) alone, digital mammography (DM) alone and women recalled on both BT+DM

	BT alone	DM alone	BT+DM
Mean age, years (range)	53 (40–74)	55 (40–75)	51 (40–72)
Density			
Fatty (BI-RADS 1 + 2)	53 (40.5)	26 (37.7)	27 (33.3)
Dense (BI-RADS 3 + 4)	78 (59.5)	43 (62.3)	54 (67)
Total	131 (100)	69 (100)	81 (100)
Finding leading to recall			
Stellate distortion	53 (40.5)	22 (31.9)	11 (13.6)
Rounded lesion	32 (24.4)	18 (26.1)	24 (29.6)
Indistinct density	28 (21.4)	20 (29.0)	8 (9.9)
Calcifications	12 (9.9)	5 (7.2)	9 (11.1)
Architectural distortion	2 (1.7)	1 (1.4)	0 (0)
Symptoms	4 (2.5)	3 (4.3)	29 (35.8)
Total	131 (100)	69 (100)	81 (100)
Outcome of the work-up			
Normal breast tissue	74 (56.5)	39 (56.5)	21 (25.9)
Benign cyst	15 (11.5)	13 (18.8)	20 (24.7)
Benign calcifications	12 (9.2)	7 (10.1)	10 (12.3)
Postoperative scar tissue	8 (6.1)	1 (1.4)	1 (1.2)
Other <sup>a</sup>	6 (6.9)	2 (2.9)	8 (9.9)
Fibroadenoma	5 (3.8)	6 (8.7)	10 (12.3)
Benign lesion (NOS) <sup>b</sup>	5 (3.8)	1 (1.4)	3 (3.7)
Radial scar	5 (3.8)	0 (0)	4 (4.9)
Symptomatic women with negative work-up	1 (0.8)	0 (0)	4 (4.9)
Total	131 (100)	69 (100)	81 (100)

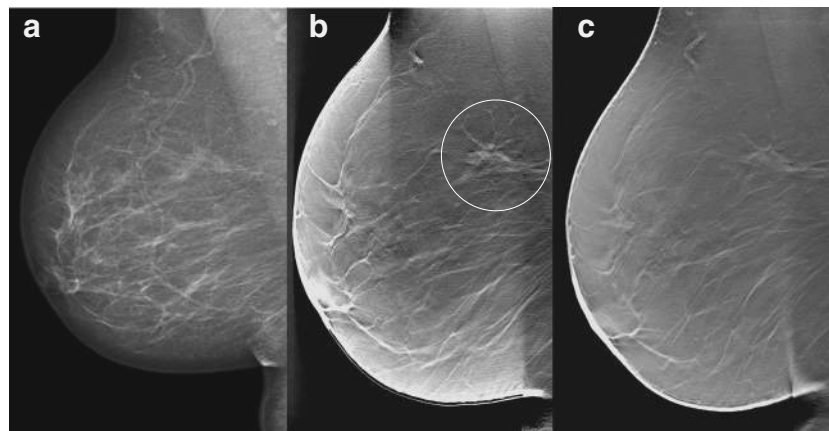
Data are n (%) or range when stated

<sup>a</sup> Various benign findings ( $\leq 2$  findings per lesion type), e.g., lymph node, atheroma, lipoma, oil cyst, hamartoma, papilloma

<sup>b</sup> Benign lesion not otherwise specified (NOS) was the finding of a small lesion with benign features that was stable on follow-up, but where needle-biopsies could not be performed

no evidence of abnormality after assessment and follow-up, but will also result in a higher frequency of radial

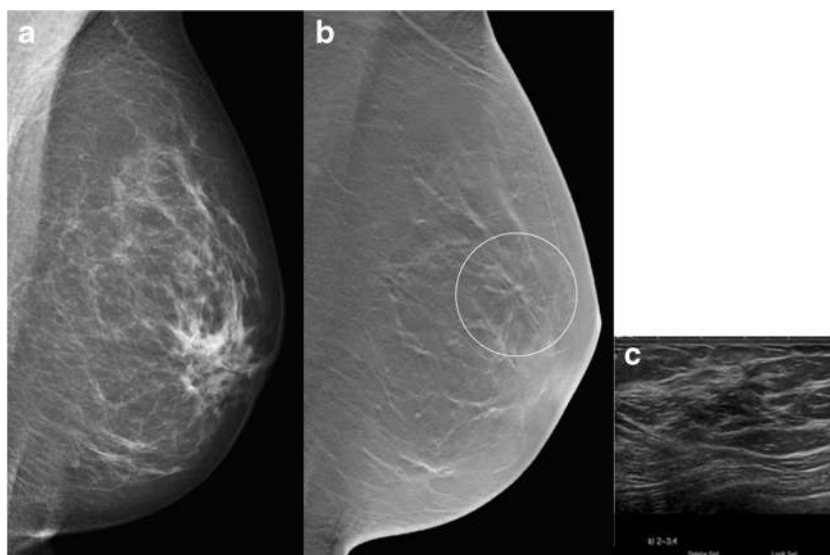
scars and postoperative scar tissue. On the other hand, BT was found to be better at characterising rounded



**Fig. 2** False positive case recalled on breast tomosynthesis alone. **Case 1.** A 60-year-old asymptomatic woman was considered to have a negative screening mammogram at double-reading (a), but was recalled due to the finding of a stellate distortion on breast tomosynthesis (mediolateral

oblique view) (b). However, at work-up there was no discernible lesion at ultrasonography or magnetic resonance imaging. The finding was stable at 1-year follow-up (c) and was considered to be ordinary fibroglandular tissue

**Fig. 3** False positive case recalled on breast tomosynthesis alone. **Case 2.** A 54-year-old asymptomatic woman with negative screening mammography (a) was recalled based on the finding of a small area of stellate distortion visible only on breast tomosynthesis (b). At work-up, ultrasound (c) showed a subtle stellate distortion without a distinguishable nucleus and fine needle aspiration showed no evidence of malignancy. It was considered to most likely represent a radial scar



lesions, reducing the assessment of benign cysts and fibroadenomas compared to cases recalled on DM alone.

The drop in the FP recall rate for BT alone during the first 1.5 years of the trial implies that the specificity can be improved with increased experience. The FP recall rate stabilised at 1.5 %. Assuming that this is the likely level in a routine screening in our group the difference against DM would be small. Furthermore, if the readers had had access to prior BT examinations, a further reduction in the FP recall rate might have been achieved [34, 35]. Also, since BT is a more sensitive method, the use of BT in the MBTST population should be regarded as a prevalence screening round, with a higher

recall and cancer detection rate compared to incidence screening. Had the same population been screened with BT for a second round, the recall rate and cancer detection rate would most likely be lower. Nevertheless, the observed FP recall rate for BT in this study was low and in accordance with the European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis [36], and is probably outweighed by the benefits of a significant increased cancer detection rate [8]. Hence, using only one-view BT in the MBTST did not seem to compromise the diagnostic performance since the results are comparable with the population-based screening trials that used a combination of two-view BT and two-view DM.

**Fig. 4** Work-up of false-positive cases. The result (number) of the work-up of women recalled due to a finding of an area of stellate distortion or with a rounded radiographic appearance for the different recall groups: Breast tomosynthesis (BT) alone, digital mammography (DM) alone and women recalled on both BT+DM

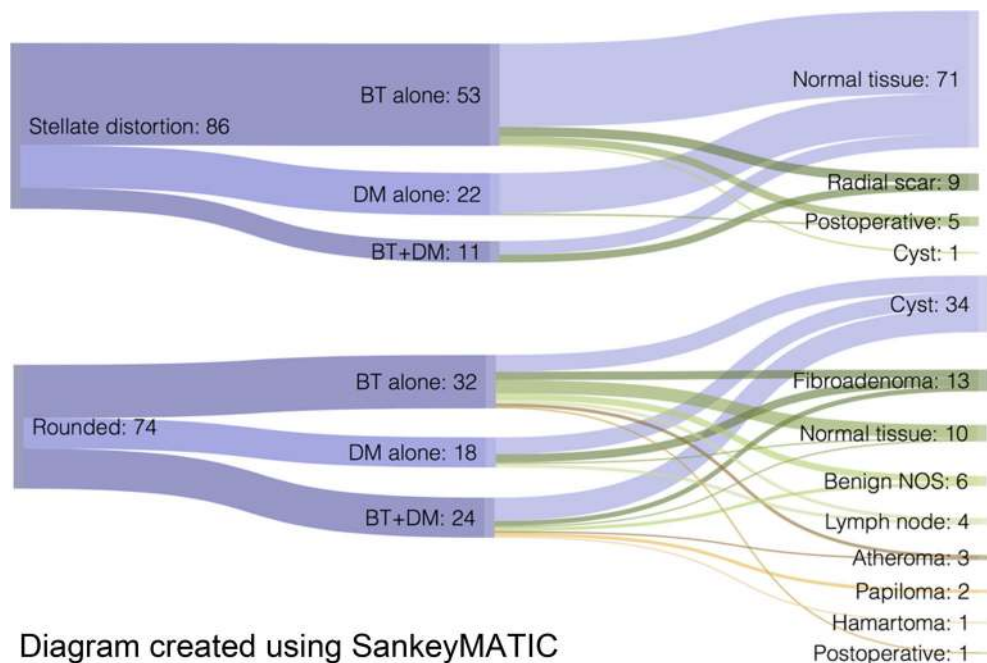
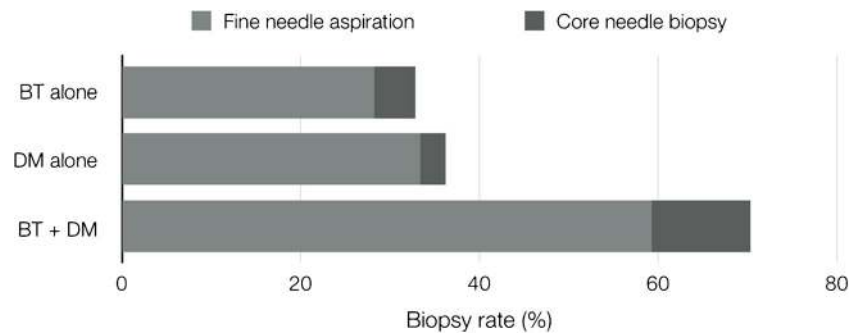


Diagram created using SankeyMATIC

**Fig. 5** Biopsy rate. The biopsy rate (total number of biopsies per number of recalled women) for the work-up of false-positive cases in the different recall groups: Breast tomosynthesis (BT) alone, digital mammography (DM) alone and women recalled on both BT+DM



However, it should be borne in mind, that a significant factor when comparing FP rates between different studies is the cut-off-level between different radiographic abnormalities. Furthermore, in addition to a high accuracy a mass screening modality should be fast, easy to read and reasonably inexpensive. One-view BT has the potential to meet these criteria. Further follow-up of the MBTST will show whether this holds true.

The observed increase of radial scars and postoperative scar tissue is attributed to the fact that BT is especially sensitive to stellate lesions, including both benign and malignant lesions, as observed in the additional cancers detected in the MBTST [8], as well as in the Oslo Tomosynthesis Screening Trial [4, 6]. The higher sensitivity of BT could also have contributed to the increased detection of benign lesions NOS – lesions that, due to their subtlety, were not discernible on ultrasonography, and hence not accessible to needle biopsy, but stable at follow-up. Hypothetically, this type of assessment could lead to more distress for the woman. Longer follow-up of the FP cases recalled on BT alone might add important knowledge, since it could answer the question of whether some of these findings actually represent a very early sign of a developing malignancy [9, 12]. At the breast clinic where the study was performed, there was no access to BT-guided biopsy, which could also explain the lower biopsy rate in the BT-alone group. Previous studies have shown increased performance by using vacuum-assisted biopsy with the aid of BT compared to prone stereotactic biopsy [37, 38]. Although it has been shown that most women with an FP DM do not undergo an invasive assessment [13], there is no doubt that access to this technology will be useful if BT is to be used in screening, especially to assess subtle lesions not visible on conventional mammography and ultrasound.

Lourenco et al. [26] showed that the implementation of the combination mode in screening gave an overall reduction in the recall rate, mainly due to fewer recalls of abnormalities presented as focal asymmetries. There were too few observations of asymmetries in this study to draw any similar conclusions. This discrepancy could be explained by the choice of nomenclature in the retrospective assessment of the radiographic findings and possibly by differences in the study populations.

This study did not show any major differences in age and breast density in the FP cases recalled on BT alone versus DM

alone. In a previous study performed by this group, the same data set was analysed with another statistical approach in order to obtain a model to predict the total number of FP when screening with BT [15]. The study showed that the FP fraction for both screening modalities, BT and DM, increased with breast density.

A clinical implication of this study is that, regardless of whether BT will increase or reduce the FP recall rate in breast cancer screening (depending on baseline recall rates), there will most likely be a shift in the type of FP cases that the radiologist needs to assess. Some of these cases – only visible on BT – will also be difficult to assess without access to BT-guided biopsies, magnetic resonance imaging or short-term follow-up. Before a large-scale implementation of BT in screening takes place, further analyses of the cost-benefit is needed. The cost of FP has been estimated to be almost one-third of the cost of a DM screening programme [39]. This warrants further studies of what type of examinations and investigations are needed to assess the FP cases generated with BT screening.

We have chosen to present the FP recall rate after arbitration, since it reflects the actual impact on clinical practice. However, the pre-arbitration FP recall rate could also add valuable information, since the arbitration meetings for the separate reading arms were not fully blinded, but this is beyond the scope for this study. A limitation of this study is the lack of prior BT examinations for comparison, as discussed above. Other limitations related to the design of the MBTST are discussed in detail elsewhere [8].

In conclusion, in the first half of this population-based screening trial with one-view BT the number of FPs increased mainly due to the recall of stellate distortions simulating malignancy. On the other hand, the characterisation of rounded lesions was improved with BT compared to DM, reducing the need to assess cysts and fibroadenomas. With increased experience the FP recall rate can be reduced.

**Acknowledgments** Preliminary results of this study were presented orally at a scientific session and as a poster at ECR 2015. The scientific guarantor of this publication is Sophia Zackrisson. The authors of this manuscript declare relationships with the following companies: Siemens AG (Erlangen, Germany) sponsored the study by providing the breast tomosynthesis equipment. KL, IA and SZ, have received speakers' fees

and travel grants from Siemens. This study has received funding from the Swedish Research Council, Swedish Cancer Society, Skane county council's research and development foundation, Lund University Medical Faculty, Unilabs, Sweden, and Siemens. One of the authors (AR) has significant statistical expertise. Institutional Review Board approval was obtained. Written informed consent was obtained from all subjects (patients) in this study. Some study subjects or cohorts have been previously reported in: Lång K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S (2015) Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmö Breast Tomosynthesis Screening Trial, a population-based study. *Eur Radiol*. doi:10.1007/s00330-015-3803-3. Methodology: prospective, diagnostic or prognostic study, performed at one institution.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Carney PA, Miglioretti DL, Yankaskas BC et al (2003) Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med* 138:168–175
- Laming D, Warren R (2000) Improving the detection of cancer in the screening of mammograms. *J Med Screen* 7:24–30
- Kolb TM, Lichy J, Newhouse JH (2002) Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology* 225:165–175
- Skaane P, Bandos AI, Gullien R et al (2013) Comparison of Digital Mammography Alone and Digital Mammography Plus Tomosynthesis in a Population-based Screening Program. *Radiology*. doi:10.1148/radiol.12121373
- Ciatto S, Houssami N, Bernardi D et al (2013) Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol* 14:583–589
- Skaane P, Bandos AI, Gullien R et al (2013) 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* 23:2061–2071
- Houssami N, Macaskill P, Bernardi D et al (2014) Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading—evidence to guide future screening strategies. *Eur J Cancer* 50:1799–1807
- Lång K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S (2015) Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmö Breast Tomosynthesis Screening Trial, a population-based study. *Eur Radiol*. doi:10.1007/s00330-015-3803-3
- Bond M, Pavey T, Welch K et al (2013) Systematic review of the psychological consequences of false-positive screening mammograms. *Health Technol Assess*. doi:10.3310/hta17130
- Bolejko A, Hagell P, Wann-Hansson C, Zackrisson S (2015) Prevalence, Long-term development, and predictors of psychosocial consequences of false-positive mammography among women attending population-based screening. *Cancer Epidemiol Biomark Prev* 24:1388–1397
- Bonfill X, Marzo M, Pladevall M, Marti J, Empanaza JI (2001) Strategies for increasing women participation in community breast cancer screening. *Cochrane Database Syst Rev*. doi:10.1002/14651858.cd002943:Cd002943
- von Euler-Chelpin M, Kuchiki M, Vejborg I (2014) Increased risk of breast cancer in women with false-positive test: the role of misclassification. *Cancer Epidemiol* 38:619–622
- Hofvind S, Ponti A, Patnick J et al (2012) False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. *J Med Screen* 19:57–66
- Christiansen CL, Wang F, Barton MB et al (2000) Predicting the cumulative risk of false-positive mammograms. *J Natl Cancer Inst* 92:1657–1666
- Rosso A, Lång K, Petersson IF, Zackrisson S (2015) Factors affecting recall rate and false positive fraction in breast cancer screening with breast tomosynthesis – A statistical approach. *Breast*. doi:10.1016/j.breast.2015.08.007
- Baker JA, Lo JY (2011) Breast tomosynthesis: state-of-the-art and review of the literature. *Acad Radiol* 18:1298–1310
- Bochud FO, Valley JF, Verdun FR, Hessler C, Schnyder P (1999) Estimation of the noisy component of anatomical backgrounds. *Med Phys* 26:1365–1370
- Houssami N, Skaane P (2013) Overview of the evidence on digital breast tomosynthesis in breast cancer detection. *Breast*. doi:10.1016/j.breast.2013.01.017
- Förnvik D, Zackrisson S, Ljungberg O et al (2010) Breast tomosynthesis: Accuracy of tumor measurement compared with digital mammography and ultrasonography. *Acta Radiol* 51:240–247
- Andersson I, Ikeda DM, Zackrisson S et al (2008) Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and BIRADS classification in a population of cancers with subtle mammographic findings. *Eur Radiol* 18:2817–2825
- Lång K, Andersson I, Zackrisson S (2014) Breast cancer detection in digital breast tomosynthesis and digital mammography—a side-by-side review of discrepant cases. *Br J Radiol* 87:20140080
- Skaane P, Gullien R, Bjørndal H et al (2012) Digital breast tomosynthesis (DBT): initial experience in a clinical setting. *Acta Radiol* 53:524–529
- Friedewald SM, Rafferty EA, Rose SL et al (2014) Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 311:2499–2507
- Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE (2014) Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *AJR Am J Roentgenol* 203:687–693
- Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE (2013) Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology* 269:694–700
- Lourenco AP, Barry-Brooks M, Baird GL, Tuttle A, Mainiero MB (2015) Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. *Radiology* 274:337–342
- Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R Jr (2013) Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR Am J Roentgenol* 200:1401–1408
- Durand MA, Haas BM, Yao X et al (2015) Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology* 274:85–92



29. McCarthy AM, Kontos D, Synnestvedt M et al (2014) Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *J Natl Cancer Inst.* doi: [10.1093/jnci/dju316](https://doi.org/10.1093/jnci/dju316)
30. Kemp Jacobsen K, O'Meara ES, Key D et al (2015) Comparing sensitivity and specificity of screening mammography in the United States and Denmark. *Int J Cancer.* doi: [10.1002/ijc.29593](https://doi.org/10.1002/ijc.29593)
31. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA et al (2013) ACR BI-RADS® Atlas, breast imaging reporting and data system. American College of Radiology, Reston, VA
32. Mertelmeier T (2006) Optimizing filtered backprojection reconstruction for a breast tomosynthesis prototype device. *SPIE*, pp 6142 61420F-61421-61412
33. Förnvik D, Andersson I, Svahn T, Timberg P, Zackrisson S, Tingberg A (2010) The effect of reduced breast compression in breast tomosynthesis: human observer study using clinical cases. *Radiat Prot Dosims* 139:118–123
34. Thurfjell MG, Vitak B, Azavedo E, Svane G, Thurfjell E (2000) Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. *Acta Radiol* 41: 52–56
35. Roelofs AA, Karssemeijer N, Wedekind N et al (2007) Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology* 242:70–77
36. van Engen R BH, Bouwman R, Dance D, Heid P, Lazzari B, Marshall N, Schopphoven S, Strudley C, Thijssen M, Young K (2013) Protocol for the quality control of the physical and technical aspects of digital breast tomosynthesis system. European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis. EUREF
37. Viala J, Gignier P, Perret B et al (2013) Stereotactic vacuum-assisted biopsies on a digital breast 3D-tomosynthesis system. *Breast J* 19:4–9
38. Schradang S, Distelmaier M, Dirrichs T et al (2015) Digital breast tomosynthesis-guided vacuum-assisted breast biopsy: initial experiences and comparison with prone stereotactic vacuum-assisted biopsy. *Radiology* 274:654–662
39. Lidbrink E, Elfving J, Frisell J, Jonsson E (1996) Neglected aspects of false positive findings of mammography in breast cancer screening: analysis of false positive cases from the Stockholm trial. *BMJ* 312:273–276