

# Clinical implementation of synthesized mammography with digital breast tomosynthesis in a routine clinical practice

Phoebe E. Freer<sup>1</sup> · Joanna Riegert<sup>1</sup> · Laura Eisenmenger<sup>2</sup> · Dominik Ose<sup>3</sup> ·  
Nicole Winkler<sup>1</sup> · Matthew A. Stein<sup>1</sup> · Gregory J. Stoddard<sup>4</sup> · Rachel Hess<sup>3,5</sup>

Received: 4 May 2017 / Accepted: 29 July 2017  
© Springer Science+Business Media, LLC 2017

## Abstract

**Background** Most published studies evaluating digital breast tomosynthesis (DBT) included a separate 2-dimensional full-field digital mammogram (FFDM) for DBT screening protocols, increasing radiation from screening mammography. Synthesized mammography (SM) creates a 2-dimensional image from the DBT source data, and if used in place of FFDM, it reduces radiation of DBT screening. This study evaluated the implementation of SM + DBT in routine screening practice in terms of recall rates, cancer detection rates (CDR), % of minimal cancers, % of node-positive cancers, and positive predictive values (PPV).

**Materials and methods** A multivariate retrospective institutional analysis was performed on 31,979 women who obtained screening mammography (10/2013–12/2015) with cohorts divided by modality (SM + DBT, FFDM + DBT, and FFDM). We adjusted for comparison mammograms, age, breast density, and the interpreting radiologist. Recall type was analyzed for differences (focal asymmetry, asymmetry, masses, calcifications, architectural distortion).

**Results** SM + DBT significantly decreased the recall rate compared to FFDM (5.52 vs. 7.83%,  $p < 0.001$ ) with no differences in overall CDR ( $p = 0.66$ ), invasive and/or in situ CDR, or percentages of minimal and node-negative cancers. PPV1 significantly increased with SM + DBT relative to FFDM (9.1 vs. 6.2%,  $p = 0.02$ ). SM + DBT did not differ significantly in recall rate or overall CDR compared to FFDM + DBT. There were statistically significant differences in certain findings recalled by screening modality (e.g., focal asymmetries).

**Conclusions** SM + DBT reduces false positives compared to FFDM, while maintaining the CDR and other desirable audit outcome data. SM + DBT is more accurate than FFDM alone, and is a desirable alternative to FFDM + DBT, given the added benefit of radiation reduction.

**Keywords** Breast cancer · Screening · Mammography · Tomosynthesis · Cancer detection · Digital breast tomosynthesis

## Introduction

Digital breast tomosynthesis (DBT) decreases the recall rate for screening mammography compared to full-field digital mammography (FFDM) while simultaneously maintaining or increasing the cancer detection rate [1–7]. As such, the National Cancer Comprehensive Network (NCCN) guidelines for breast cancer screening state to “consider tomosynthesis” for all women 40 and older, regardless of risk [8]. When used for screening, an additional 2-dimensional (2D) mammography image is recommended to be obtained in addition to DBT [9–11]. Acquiring both DBT and a separate 2D FFDM exam nearly doubles the radiation dose to the patient over FFDM alone

✉ Phoebe E. Freer  
phoebe.freer@hsc.utah.edu

<sup>1</sup> Radiology and Imaging Sciences, University of Utah/  
Huntsman Cancer Institute, Salt Lake City, UT 84132, USA

<sup>2</sup> Radiology and Imaging Sciences, University of Utah,  
Salt Lake City, UT, USA

<sup>3</sup> Population Health Sciences, University of Utah,  
Salt Lake City, UT, USA

<sup>4</sup> Family and Preventive Medicine, University of Utah,  
Salt Lake City, UT, USA

<sup>5</sup> Internal Medicine, University of Utah, Salt Lake City, UT,  
USA

[12, 13]. While this “double-dose” exposure for FFDM + DBT falls below the United States Food and Drug Administration (USFDA) thresholds for screening mammography [13], minimizing the radiation dose is of paramount importance if DBT is to become the new standard of care for mammographic screening within the general population.

Recent technology has developed a “virtual” or synthesized mammogram (SM) that is reconstructed from the DBT source data, potentially eliminating the need to obtain a standard FFDM. SM + DBT technology was recently approved for clinical use by the FDA, and has been incorporated in some clinical practices [14]. However, data regarding the clinical implementation of SM + DBT remain limited. Initial prototypes of SM technology demonstrated inferior sensitivity to FFDM + DBT [15, 16]. Subsequent versions of SM technology demonstrated improvement with a few clinical and reader studies suggesting that SM + DBT can be used as an alternative to FFDM + DBT [16–19]. To date, only one study in the United States has published the results of using SM + DBT technology in clinical practice. This study did not control for multiple possible confounders (including breast density, presence of comparison mammograms, and effect of the interpreting radiologist), thus limiting the ability to understand the feasibility of implementation and the implications of switching from either a FFDM or a FFDM + DBT program to a SM + DBT screening program [20].

The purpose of our study is to analyze the basic clinically relevant audit outcome data including recall rates and cancer detection rates (CDR) between the SM + DBT technology in our screening program compared to historical cohort of FFDM + DBT technique, as well as compared to the FFDM alone during the same time periods, adjusting for multiple possible confounders.

## Materials and methods

### Study population and screening interpretations

This retrospective study used HIPAA standards and was approved by the institutional review board with a waiver for informed consent. We extracted data from institutional databases (MagView, Epic Radiant, and GE Centricity) including all women ( $n = 31,979$ , age 19–100) who presented for screening mammography to any screening facility of our academic medical center from 10/1/13 to 12/31/15. As part of our workflow, two facilities have capability of performing DBT and/or FFDM (Dimensions, Hologic, Inc., Bedford, MA), whereas three have only

FFDM capability (Selenia, Hologic, Inc., Bedford, MA). All women who presented to our sites equipped with DBT from October 2013 through December 2014 were presented with the option to receive combination FFDM + DBT, while being informed that this option involved higher radiation dose than FFDM alone ( $n = 1019$  women chose FFDM + DBT screening, Cohort 2). Secondary to concerns over increased radiation from FFDM + DBT, our institution implemented SM + DBT technology (C-View software, Hologic, Inc.) starting January 2015. All women who presented to a site with SM technology from 1/1/2015 to 12/31/15 were screened using SM + DBT ( $n = 9525$ , Cohort 1). Throughout the same time periods (10/1/13–12/31/15), all women who presented to facilities with FFDM only capability, and those 10/1/13–12/31/14 who opted not to receive FFDM + DBT, were screened with standard FFDM ( $n = 21,435$ , Cohort 3). All interpreting radiologists ( $n = 10$ ) met MQSA and FDA guidelines for FFDM and DBT interpretation and were breast imaging specialists ( $n = 8$  fellowship trained or equivalent; experience range 1–26 years). We evaluated patient demographics of each cohort with respect to age, mammographic breast density (according to BIRADS recommended categories (almost entirely fat, scattered areas of fibroglandular density, heterogeneously dense, and extremely dense) [21]), and the presence of prior comparison mammograms (Table 1).

### Study outcomes measures

As the primary outcome, we evaluated the recall rate (% recalled from screening for additional evaluation) and the cancer detection rate (CDR) (# of cancers detected/1000 screening exams) of the SM + DBT group (Cohort 1) compared with the historical FFDM + DBT group (Cohort 2) and with the historical and concurrent FFDM only group (Cohort 3). We analyzed results both based on actual rates, as well as controlling for the presence of prior mammograms, patient age, breast density and the effect of the individual radiologist. We initially divided Cohort 3 into two sub-cohorts with time periods to match Cohort 1 and Cohort 2, respectively; however, upon initial analysis, the recall rate for the two sub-cohorts were identical (8.8%) and the CDR for the two cohorts were similar (5.4/1000 and 5.6/1000); therefore, the two cohorts were included together for comparison. Additionally, we evaluated the finding type recalled using BIRADS 5<sup>th</sup> edition lexicon (asymmetry, focal asymmetry, mass, architectural distortion, and calcifications) [21]. We measured the basic clinically relevant audit screening outcomes including invasive CDR (# invasive cancers/1000 screening exams); in situ CDR (# of ductal carcinoma in situ/1000 screening exams),

**Table 1** Patient demographics

	Cohort 1	Cohort 2	Cohort 3	<i>p</i> value
Screen type	SM + DBT	FFDM + DBT	FFDM only	
Study period	1/1/15–12/31/15	10/1/13–12/31/14	10/1/13–12/31/15	
Number	9525	1019	21435	
Age (year)				<.001
<40	147 (1.5)	17 (1.7)	298 (1.4)	
40–49	2592 (27.2)	280 (27.5)	5576 (26.0)	
50–59	3047 (32.0)	349 (34.2)	6790 (31.7)	
60–69	2445 (25.7)	258 (25.3)	5456 (25.5)	
70–74	681 (7.1)	70 (6.9)	1680 (7.8)	
>74	613 (6.4)	45 (4.4)	1635 (7.6)	
Breast density				<.001
Almost entirely fatty	1175 (12.3)	76 (7.5)	3365 (15.7)	
Scattered areas of fibroglandular densities	4358 (45.8)	422 (41.4)	9768 (45.6)	
Heterogeneously dense	3309 (34.7)	428 (42.0)	7157 (33.4)	
Extremely dense	683 (7.2)	93 (9.1)	1145 (5.3)	
Prior comparison mammogram	8310 (87.2)	783 (76.8)	16281 (76.0)	<.001

Unless otherwise indicated, each expressed as # of patients (%)

percentage of cancers that are “minimal” (defined as all invasive cancers  $\leq 10$  mm in size, or DCIS of any size), percentage of node-negative cancers, PPV 1 (# of cancers/all exams recalled at screen), PPV 2 (# of cancers/# of biopsy recommendations at the diagnostic work-up after recall), and PPV 3 (# of cancers/biopsies actually performed).

### Statistical analysis

Patient characteristics were compared using a  $\chi^2$  test for categorical variables and a Kruskal–Wallis analysis of variance for ordered categorical variables. Although there were no clinically significant differences between patient age and breast density between the cohorts, the differences were statistically significant (Table 1). Therefore, all multivariable models adjusted for age, breast density, and the presence of comparison mammograms at the time of interpretation in addition to the effect of the individual radiologist on interpretation. We also analyzed the recall rate in a multivariable analysis comparing dense (heterogeneously dense and extremely dense breasts) to that of non-dense breasts (almost entirely fat and scattered areas of fibroglandular density categories).

The analysis was performed using a multivariable mixed effects Poisson regression model, yielding adjusted proportions and relative risk for the outcome variables for cohort 1 relative to cohorts 2 and 3. Radiologist was included as a random effects term, with patients nested within radiologist in a multilevel fashion. Age category

was included as a fixed effects factor, using indicator or dummy variables, for the age categories. Presence of comparison mammograms at the time of interpretation was included as a single fixed effects indicator variable. Adjusted percentages, which are percentages estimated while controlling for the variables included in the model, were computed using the method of average marginal effects [22], also known as marginal standardization [23].

Including radiologist as a random effect controlled for radiologist experience and interpretation preferences of images. Frequently, the estimate of effect size from a random effect term in a model is reported as an intra-class correlation coefficient. A more intuitive way to express it, which is consistent with how other factors in the model are expressed, is as a risk ratio. However, there are many risk ratios, since there is one for each radiologist compared to every other radiologist. To provide a single risk ratio, we report it as the median risk ratio (MRR) [24, 25]. It is interpreted as the average, or median, risk ratio for the outcome when comparing a radiologist who has a higher propensity to have the outcome to a radiologist with a lower propensity to have the outcome. For example, if the recall rate  $MRR = 1.31$ , radiologists who have a higher recall rate are 1.31 times more likely, on average, to recall a mammogram compared to radiologists who have a lower recall rate.

Statistical tests were performed using STATA Version 14.2 (StataCorp LP, College Station, TX) and the database was stored in Excel (Microsoft Office 2013). All *p* values represent a two-sided comparison.

## Results

### Cohort characteristics

Cohort 1 consists of all women from 1/1/2015 to 12/31/15 who were screened using SM + DBT ( $n = 9525$ ). Cohort 2 is all women from October 2013 through December 2014 who received combination FFDM + DBT ( $n = 1019$ ). Cohort 3 includes all women from 10/1/13 to 12/31/15, who were screened with standard FFDM ( $n = 21,435$ ). Although there were no clinically significant differences between patient age and breast density between the cohorts, the differences were statistically significant (Table 1).

### Recall rate

The overall unadjusted actual recall rate for SM + DBT was significantly lower than standard FFDM alone (5.78 vs. 8.68%); when controlling for the presence of prior mammograms, breast density, patient age, and the effect of the individual radiologist on interpretation, the adjusted recall rate is 5.52 versus 7.83% respectively [adjusted risk ratio (aRR) = 0.70, 95% confidence interval (CI) (0.64, 0.78),  $p < 0.001$ , MRR = 1.31]. There was also a decrease in the unadjusted actual recall rate of SM + DBT versus FFDM + DBT, (5.78 vs. 6.97%) with adjusted recall rates of 5.52 versus 6.39%, respectively, although this was not statistically significant [aRR = 0.86 95% CI (0.67, 1.11),  $p = 0.25$ ] (Table 2).

When broken down by density (Table 3), the improvement in adjusted recall rate for SM + DBT compared to FFDM was significant for both dense and non-dense breasts with a relatively greater decrease for non-dense breasts [aRR = 0.61 non-dense (95% CI (0.53, 0.71),  $p < 0.001$ ) versus aRR = 0.80 dense (95% CI (0.70, 0.91),  $p = 0.001$ ]. For the non-dense breasts, the adjusted likelihood of recall was also significantly decreased for SM + DBT relative to FFDM + DBT [aRR = 0.65 (95% CI (0.45, 0.94),  $p = 0.021$ ).

In analyzing the recall rate by finding type (Table 4), a patient was significantly less likely to be recalled using SM + DBT compared to FFDM for findings of focal asymmetries, asymmetries, and calcifications [aRR, 0.42 (95% CI (0.34, 0.53)),  $p < 0.005$ ; 0.28 (95% CI (0.21, 0.38)),  $p < 0.005$ ; and 0.78 (95% CI (0.62, 0.98)),  $p < 0.05$  respectively]. Conversely, a patient was significantly more likely to be recalled for the finding of a mass using SM + DBT relative to FFDM alone [aRR, 1.33 (95% CI (1.13, 1.57),  $p < 0.005$ )]. There were no significant differences in recall by finding type between the SM + DBT and the FFDM + DBT exams, although there was a non-significant trend toward less frequent call-back for a mass

on SM + DBT compared to FFDM + DBT. There was notably no significant difference in being recalled for architectural distortion between the SM + DBT and FFDM or the SM + DBT and FFDM + DBT.

### Cancer detection

The overall unadjusted actual CDR for SM + DBT was identical to that of FFDM (5.9/1000) and lower than that of FFDM + DBT (6.9/1000) (Table 2). When controlling for the presence of prior mammograms, patient age, breast density, and the effect of the interpreting radiologist, there remained no difference CDR for SM + DBT relative to FFDM [5.4/1000 vs. 5.0/1000; aRR 1.08 (95% CI (0.78, 1.49),  $p = 0.66$ )]. There was also no significant difference in the adjusted CDR for the SM + DBT versus FFDM + DBT [5.4/1000 vs. 5.7/1000; aRR 0.95 (95% CI (0.43, 2.10),  $p = 0.90$ )].

Additionally, both the invasive and in situ cancer detection rates were not significantly different between SM + DBT and either the FFDM + DBT or FFDM alone (adjusted invasive CDR: 4.3/1000 versus 3.4/1000 and 3.9/1000, respectively, and adjusted in situ CDR: 1.2/1000 vs 1.8/1000 and 1.2/1000, respectively) (Table 2). The adjusted percent of minimally invasive cancers and node-negative cancers also did not differ significantly between modalities.

There was a trend, although not significant, towards higher adjusted PPVs for PPV2 and PPV 3 for the SM + DBT relative to FFDM alone [36.4 vs. 30.9%; (aRR 0.85, 95% CI (0.61, 1.17),  $p = 0.32$ ) and 36.7 vs. 31%; (aRR 0.84, 95% CI (0.61, 1.16)  $p = 0.30$ ), respectively] with a statistically significant improvement in the adjusted PPV1 for SM + DBT relative to FFDM [9.1 vs. 6.2%; aRR 0.69, 95% CI (0.50, 0.95),  $p = 0.02$ ] (Table 5).

## Discussion

Our study supports the implementation of SM + DBT technology as a replacement for FFDM + DBT in clinical practice, as SM + DBT maintains the statistically significant decrease in recall rate for FFDM + DBT compared to FFDM. Further, we demonstrate that we decrease the false positives, while maintaining the cancer detection rate of FFDM. Additionally, although our comparison cohort of FFDM + DBT is small, and likely biased by selection bias, our recall rate and CDR using SM + DBT were not statistically significantly different from those of FFDM + DBT, although our cohort for this comparison is underpowered to detect such a change. We know from prior studies, that SM +DBT decreases the radiation dose from FFDM + DBT by approximately 45% on average,

**Table 2** Screening clinically relevant basic audit outcomes and adjusted outcomes (AO)

	Cohort 1	Cohort 2	Cohort 3	AO1	AO2	AO3	aRR1	aRR2
Recall Rate <sup>a</sup>	5.78 (551/9525)	6.97 (71/1019)	8.68 (1860/21435)	5.52	6.39	7.83	0.86 95% CI (0.67, 1.11) <i>p</i> = 0.25 MRR = 1.31	0.70 95% CI (0.64, 0.78) <i>p</i> < 0.001 MRR = 1.31
CDR <sup>b</sup>	5.9 (56/9525)	6.9 (7/1019)	5.9 (126/21435)	5.4	5.7	5.0	0.95 95% CI (0.43, 2.10) <i>p</i> = 0.90 MRR = 1.30	1.08 95% CI (0.78, 1.49) <i>p</i> = 0.66 MRR = 1.30
Invasive CDR <sup>c</sup>	4.6 (44/9525)	3.9 (4/1019)	4.3 (93/21435)	4.3	3.4	3.9	1.25 95% CI (0.45, 3.49) <i>p</i> = 0.67 MRR = 1.28	1.1 95% CI (0.77, 1.6) <i>p</i> = 0.59 MRR = 1.28
In situ CDR <sup>d</sup>	1.2 (11/9525)	2.0 (2/1019)	1.2 (26/21435)	1.2	1.8	1.2	0.66 95% CI (0.15, 3.00) <i>p</i> = 0.60 MRR = 1.00	0.97 95% CI (0.47, 1.99) <i>p</i> = 0.94 MRR = 1.00
Minimal Cancers <sup>e</sup>	63.6 (35/55)	50 (3/6)	56.3 (67/119)	60.6	50.9	57.6	1.19 95% CI (0.36, 3.97) <i>p</i> = 0.78 MRR = 1.00	1.05 95% CI (0.69, 1.60) <i>p</i> = 0.82 MRR = 1.00
Node-Negative Cancers <sup>f</sup>	70.5 (31/44)	50 (2/4)	79.6 (74/93)	67.1	52.2	81.4	1.29 95% CI (0.30, 5.47) <i>p</i> = 0.73 MRR = 1.00	0.82 95% CI (0.54, 1.27) <i>p</i> = 0.38 MRR = 1.00

Adjusted for patient age, mammographic breast density, presence of prior comparison mammograms, and the effect of the individual interpreting radiologist

AO1 Adjusted Outcome Percent (%) (or # / 1000 where indicated) controlling for priors, age, density, and effect of radiologist Cohort 1

AO2 Adjusted Outcome Percent (%) (or # / 1000 where indicated) controlling for priors, age, density, and effect of radiologist Cohort 2

AO3 Adjusted Outcome Percent (%) (or # / 1000 where indicated) controlling for priors, age, density, and effect of radiologist Cohort 3

aRR1 adjusted Relative Risk ratio, controlling for priors, age, density, and effect of radiologist of Cohort 1 relative to Cohort 2. It should be noted that Cohort 2 has small numbers and is underpowered to detect small changes

aRR2 adjusted Relative Risk ratio, controlling for priors, age, density, and effect of radiologist of Cohort 1 relative to Cohort 3

<sup>a</sup> Unless otherwise indicated, expressed as % (# of patients recalled / # of total screening mammograms)<sup>b</sup> Cancer Detection Rate: Unless otherwise indicated, expressed as #/1000 (# of total cancers / # of total screening mammograms)<sup>c</sup> Invasive Cancer Detection Rate: Unless otherwise indicated, expressed as #/1000 (# of invasive cancers / # of total screening mammograms)<sup>d</sup> In situ Cancer Detection Rate: Unless otherwise indicated, expressed as #/1000 (# of ductal carcinoma in situ cancers / # of total screening mammograms)<sup>e</sup> Unless otherwise indicated, expressed as % (# of invasive cancers <=10mm and ductal carcinoma in situ of any size / # of all cancers)<sup>f</sup> Unless otherwise indicated, expressed as % (# of node-negative cancers / # of all cancers)

**Table 3** Adjusted recall rates in dense versus non-dense breasts

	Adjusted recall rate cohort 1 (%)	Adjusted recall rate cohort 2 (%)	Adjusted recall rate cohort 3 (%)	aRR1	aRR2
Dense	7.11	6.63	8.92	1.07 95% CI (0.76, 1.51) $p = 0.69$ MRR = 1.37	<b>0.80</b> 95% CI (0.70, 0.91) $p = 0.001$ MRR = 1.37
Non-dense	4.50	6.92	7.34	<b>0.65</b> 95% CI (0.45, 0.94) $p = 0.021$ MRR = 1.24	<b>0.61</b> 95% CI (0.53, 0.71) $p < 0.001$ MRR = 1.24

Significant values are indicated in bold

Adjusted for patient age, mammographic breast density, presence of prior comparison mammograms, and the effect of the individual interpreting radiologist

*aRR1* adjusted Relative Risk ratio, controlling for priors, age, density, and effect of radiologist of Cohort 1 relative to Cohort 2

*aRR2* adjusted Relative Risk ratio, controlling for priors, age, density, and effect of radiologist of Cohort 1 relative to Cohort 3

averaging approximately 8% more than FFDM alone [13, 17]. Thus, being able to maintain our audit parameters of cancer detection, including minimal cancer percentage, invasive CDR, in situ CDR, and node-negative cancers, while improving our false positive rate at nearly the same radiation dose as FFDM alone demonstrates that the technology is able to be implemented in widespread clinical use as an improvement to standard FFDM screening or FFDM + DBT where accessible.

While SM + DBT was implemented in our practice with an abrupt change from FFDM + DBT, we found that our interpretation performance was maintained relative to FFDM + DBT with near no learning curve to adjust to the SM technology. Interestingly, we found that in women with non-dense breasts, there was a statistically significant reduction in recalls for the SM + DBT technology compared to FFDM + DBT technology, a result that to the best of our knowledge has not been described previously. The likelihood of being recalled with SM + DBT technology in women with dense breasts (adjusted recall rate 7.11%) is similar to the likelihood of recall in non-dense breast women on FFDM alone (adjusted recall rate 7.34%), meaning the experience of reading SM + DBT in women with dense breasts approaches that of FFDM with fatty breasts, which is well established to be more accurate.

A strength of our study is that in addition to reporting our actual audit data, we also controlled for the effects of confounders including age, density, presence of prior comparison mammograms, and for the effect of the individual interpreting radiologist as it has been previously shown that DBT technology may not have a uniform effect on recalls and cancer detection among all radiologists [2]. In fact, for many of the recalled findings, there was a marked effect with the chance for recall depending on the individual radiologist, most notable for calcifications with a MRR of 1.82. Therefore, our study is the first to demonstrate the power of the SM + DBT technology itself in a

real-world clinical U.S. population in a routine setting versus the FFDM + DBT or FFDM alone. To the best of our knowledge, this level of adjustment has not been performed in the clinical setting for SM + DBT and has only been adjusted for an occasion in a few of the studies comparing FFDM + DBT to FFDM [4, 5]. We noted in both the unadjusted and adjusted recalls, statistically significant decreases in asymmetries and focal asymmetries being recalled, and an increase in masses being recalled, for SM + DBT compared to FFDM, which likely accounts for increases in the PPV1, PPV2, and PPV3 for SM + DBT relative to FFDM. That is, while decreasing the overall false positives, the accuracy of the recalls is improved. This is not surprising given the DBT technology works by reducing superimposition of overlapping benign breast tissue. We noted a statistically significant decrease in calcifications being recalled from SM + DBT relative to FFDM, which may be explained by the fact that the SM is created from the low-dose (lower resolution, higher noise) DBT projection images, and thus some calcifications may not be well seen on the SM. However, this did not result in a decrease in the in situ CDR, again, suggesting the SM + DBT recalls are more accurate. Although there was a significant decrease in recall for calcifications for SM + DBT compared to FFDM, a non-significant increased recall rate was observed for SM + DBT compared to FFDM + DBT. This observation may partially be accounted for by the reconstruction algorithms for SM technology that over-emphasize high-attenuation structures such as calcifications in an attempt to not under-call calcifications. This trend in our data is different from the data found by Zuckerman et al. which showed a decrease in calcifications recalled with SM + DBT relative to FFDM + DBT [20]. Surprisingly, there was no significant difference in the recalls of architectural distortion between the modalities in our study, which may suggest a learning curve for our practice on recognizing architectural



**Table 4** Screening Recall by Finding Type with Adjusted Percent (AP)\*

	Total	Cohort 1	Cohort 2	Cohort 3	AP1	AP2	AP3	aRR1	aRR2
Asymmetry	2.05 (654/31, 979)	1.06 (101/9525)	1.28 (13/1019)	2.52 (540/21435)	0.89	1.1	2.1	0.79 95%CI (0.44, 1.41) $p = 0.42$ MRR = 1.77	<b>0.42</b> 95% CI (0.34, 0.53) $p < 0.005$ MRR = 1.77
Focal Asymmetry	1.59 (510/31,979)	0.56 (53/9525)	0.69 (7/1019)	2.10 (450/21435)	0.62	0.75	2.22	0.83 95% CI (0.38, 1.84) $p = 0.65$ MRR = 1.39	<b>0.28</b> 95% CI (0.21, 0.38) $p < 0.005$ MRR = 1.39
Mass	2.32 (741/31979)	2.54 (242/9525)	3.73 (38/1019)	2.15 (461/21435)	2.50	3.23	1.87	0.77 95% CI (0.55, 1.09) $p = 0.15$ MRR = 1.22	<b>1.33</b> 95% CI (1.13, 1.57) $p < 0.005$ MRR = 1.22
Architectural Distortion	0.44 (140/31979)	0.48 (46/9525)	0.39 (4/1019)	0.42 (90/21435)	0.40	0.30	0.35	1.33 95% CI (0.48, 3.72) $p = 0.59$ MRR = 1.55	1.16 95%CI (0.80, 1.68) $p = 0.43$ MRR = 1.55
Calcifications	1.37 (437/31979)	1.14 (109/9525)	0.88 (9/1019)	1.49 (319/21435)	0.83	0.61	1.06	1.34 95% CI (0.67, 2.66) $p = 0.40$ MRR = 1.82	<b>0.78</b> 95% CI (0.62, 0.98) $p = 0.034$ MRR = 1.82

Significant values are indicated in bold

Adjusted for patient age, mammographic breast density, presence of prior comparison mammograms, and the effect of the individual interpreting radiologist

Unless otherwise indicated for relative risks, each expressed as % (#recalled/#interpreted)

AP1 Adjusted Percent (%) controlling for priors, age, density, and effect of radiologist Cohort 1

AP2 Adjusted Percent (%) controlling for priors, age, density, and effect of radiologist Cohort 2

AP3 Adjusted Percent (%) controlling for priors, age, density, and effect of radiologist Cohort 3

aRR1 adjusted Relative Risk ratio, controlling for priors, age, density, and effect of radiologist of Cohort 1 relative to Cohort 2. Note that Cohort 2 is underpowered to detect small changes

aRR2 adjusted Relative Risk ratio, controlling for priors, age, density, and effect of radiologist of Cohort 1 relative to Cohort 3

**Table 5** Screening Positive Predictive Values (PPV) and Adjusted Percent (AP)

	Cohort 1	Cohort 2	Cohort 3	AP1	AP2	AP3	aRR1	aRR2
PPV1	10.2 (56/551)	9.86 (7/71)	6.77 (126/1860)	9.1	8.1	6.2	0.90	<b>0.69</b>
							95% CI (0.41, 1.98) $p = 0.79$	95% CI (0.50, 0.95) $p = 0.02$
PPV2	39.7 (56/141)	43.8 (7/16)	30.9 (126/408)	36.4	40.3	30.9	1.1	0.85
							95% CI (0.50, 2.45) $p = 0.80$	95% CI (0.61, 1.17) $p = 0.32$
PPV3	37.8 (56/148)	38.9 (7/18)	30.4 (126/415)	36.7	36.3	31.0	0.99	0.84
							95% CI (0.45, 2.18) $p = 0.98$	95% CI (0.61, 1.16) $p = 0.30$

Significant value is indicated in bold

Adjusted for patient age, mammographic breast density, presence of prior comparison mammograms, and the effect of the individual interpreting radiologist

AP1 Adjusted Percent (%) controlling for priors, age, density, and effect of radiologist Cohort 1

AP2 Adjusted Percent (%) controlling for priors, age, density, and effect of radiologist Cohort 2

AP3 Adjusted Percent (%) controlling for priors, age, density, and effect of radiologist Cohort 3

aRR1 adjusted Relative Risk ratio, controlling for priors, age, density, and effect of radiologist of Cohort 1 relative to Cohort 2

aRR2 adjusted Relative Risk ratio, controlling for priors, age, density, and effect of radiologist of Cohort 1 relative to Cohort 3

PPV1 Positive Predictive Value 1, expressed as % (# cancers / # of recalled exams)

PPV2 Positive Predictive Value 2, expressed as % (# cancers / # biopsy recommended)

PPV3 Positive Predictive Value 3, expressed as % (# cancers / # biopsy performed)

distortion on DBT as it has been shown that this finding is more conspicuous on DBT than FFDM. Alternatively, this lack of a difference may be accounted for by the relative rarity of the finding of pure architectural distortion in the screening population as we are likely underpowered to demonstrate a difference.

Our study results differ from the prospective STORM-2 trial performed in Europe that demonstrate that FFDM + DBT and SM + DBT showed a statistically significant incremental increase in CDR relative to FFDM alone (2.2/1000 and 2.5/1000, respectively) [19]. However, this increase in CDR was accompanied by a similar increase in recall rate that was statistically significant (3.97% and 4.45% for FFDM + DBT and SM + DBT, respectively, versus 3.42% for FFDM alone) [19]. Our results are likely quite different from the STORM-2 trial, and more similar to other U.S. studies of DBT, given the differences in screening between Europe with this study demonstrating a much lower recall rate in general (3.42% for FFDM), and a screening population that excludes women under 49 and is recommended every two years rather than many receiving it annually.

Additionally, our study adds to the total numbers in the literature for SM + DBT implementation which may be helpful in terms of proving an overall improvement in CDR when pooled with other studies on SM + DBT in an eventual meta-analysis, given the overall small numbers of cancers relative to negative screens in the routine population.

Our study is limited in that this is a retrospective review of the database and not a prospective collection. We are comparing cohorts to one another, as opposed to comparing the different technologies on the exact same patient. Our Cohort 2 of FFDM + DBT is relatively small compared to the other cohorts, is likely biased by selection bias as patients were presented with the option to decline the DBT if desired, and is underpowered, which all likely limit the interpretation of the audit data in this cohort. Additionally, as our practice is newer to DBT technology (2013), our results may be limited by experience and a learning curve of DBT. We do not have multi-year follow-up to be able to track our patients in the state tumor registry to allow for calculation of the true sensitivity and specificity of the technologies, although this tracking is planned for future follow-up. Additionally, there is a possible confounding factor that our practice sometimes used online screening with immediate results as opposed to delayed batch interpretation for some of these patients. It is not possible to control for this possible confounder in the database collected, although there is mixing within the cohorts of patients having immediate online results versus delayed batch reading.

In conclusion, our results demonstrate that SM + DBT is feasible in clinical implementation in the routine clinical



screening setting with a moderate (adjusted 29.5%) reduction in false positive exams relative to FFDM, while maintaining the cancer detection rate, at a similar radiation dose. Our results contribute substantially to the emerging literature that SM + DBT may serve as a replacement for FFDM + DBT and may emerge as the new standard of care for screening mammography in the general population.

#### Compliance with ethical standards

**Conflict of interest** All authors declare that he/she has no conflict of interest.

**Ethical standards** This study was carried out in accordance with the ethical standards of the institutional review board.

**Informed consent** The need for informed consent was waived secondary to the nature of the study.

## References

- Friedewald SM, Rafferty EA, Rose SL et al (2014) Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 311(24):2499–2507
- 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* 267(1):47–56
- Rose S, Tidwell A, Bujnoch L et al (2013) Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR Am J Roentgenol* 200(6):1401–1408
- 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* 106(11):316
- McDonald ES, Oustimov A, Weinstein SP, Synnestvedt M, Schnall M, Conant EC (2016) Effectiveness of digital breast tomosynthesis compared with digital mammography: outcomes analysis from 3 years of breast cancer screening. *JAMA Oncol* 2(6):737–743
- 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(7):583–589
- 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(3):694–700
- National Comprehensive Cancer Network Breast Cancer Screening and Diagnosis. [https://www.nccn.org/professionals/physician\\_gls/pdf/breast-screening.pdf](https://www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf) NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines). Version 1.2016, 7/27/16. Accessed 13 Dec 2016
- US Food and Drug Administration. Published February 11, 2011. [http://www.accessdata.fda.gov/cdrh\\_docs/pdf8/p080003a.pdf](http://www.accessdata.fda.gov/cdrh_docs/pdf8/p080003a.pdf). Accessed 13 Dec 2016
- US Food and Drug Administration. Published August 24, 2014. [http://www.accessdata.fda.gov/cdrh\\_docs/pdf13/P130020a.pdf](http://www.accessdata.fda.gov/cdrh_docs/pdf13/P130020a.pdf). Accessed 13 Dec 2016
- US Food and Drug Administration. Published April 21, 2015. [http://www.accessdata.fda.gov/cdrh\\_docs/pdf14/P140011a.pdf](http://www.accessdata.fda.gov/cdrh_docs/pdf14/P140011a.pdf). Accessed 13 Dec 2016
- Svahn TM, Houssami N, Sechopoulos I, Mattsson S (2015) Review of radiation dose estimates in digital breast tomosynthesis relative to those in two-view full-field digital mammography. *Breast* 24(2):93–99
- Feng SS, Sechopoulos I (2012) Clinical digital breast tomosynthesis system: dosimetric characterization. *Radiology* 263(1):35–42
- U.S. Food and Drug Administration. Published May, 16, 2013. [http://www.accessdata.fda.gov/cdrh\\_docs/pdf8/p080003s001a.pdf](http://www.accessdata.fda.gov/cdrh_docs/pdf8/p080003s001a.pdf). Accessed 14 Dec 2016
- Gur D, Zuley ML, Anello MI et al (2012) Dose reduction in digital breast tomosynthesis (DBT) screening using synthetically reconstructed projection images: an observer performance study. *Acad Radiol* 19(2):166–171
- Skaane P, Bandos AI, Eben EB et al (2014) Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. *Radiology* 271(3):655–663
- Zuley ML, Guo B, Catullo VJ et al (2014) Comparison of two-dimensional synthesized mammograms versus original digital mammograms alone and in combination with tomosynthesis images. *Radiology* 271(3):664–671
- Gilbert FJ, Tucker L, Gillan MG et al (2015) Accuracy of digital breast tomosynthesis for depicting breast cancer subgroups in a UK retrospective reading study (TOMMY Trial). *Radiology* 277(3):697–706
- Bernardi D, Macaskill P, Pellegrini M et al (2016) Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study. *Lancet Oncol* 17(8):1105–1113
- Zuckerman SP, Conant EF, Keller BM et al (2016) Implementation of synthesized two-dimensional mammography in a population-based digital breast tomosynthesis screening program. *Radiology* 281(3):730–736
- 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
- Williams R (2012) Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata J.* 12(2):308–331
- Muller CJ, MacLehose RF (2014) Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. *Int J Epidemiol* 43(3):962–970
- Merlo J, Chaix B, Ohlsson H et al (2006) A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epidemiol Commun Health* 60(4):290–297
- Rabe-Hesketh S, Skrondal A (2008) Multilevel and longitudinal modeling using stata, 2nd edn. Stata Press Publication, College Station