



Comparison of performance metrics with digital 2D versus tomosynthesis mammography in the diagnostic setting

Manisha Bahl¹ · Sarah Mercaldo² · Charmi A. Vijapura¹ · Anne Marie McCarthy³ · Constance D. Lehman¹

Received: 27 December 2017 / Revised: 13 May 2018 / Accepted: 5 June 2018 / Published online: 2 July 2018
© European Society of Radiology 2018

Abstract

Objectives To compare performance metrics between digital 2D mammography (DM) and digital breast tomosynthesis (DBT) in the diagnostic setting.

Methods Consecutive diagnostic examinations from August 2008 to February 2011 (DM group) and from January 2013 to July 2015 (DM/DBT group) were reviewed. Core biopsy and surgical pathology results within 365 days after the mammogram were collected. Performance metrics, including cancer detection rate (CDR), abnormal interpretation rate (AIR), positive predictive value (PPV) 2, PPV3, sensitivity, and specificity were calculated. Multivariable logistic regression models were fit to compare performance metrics in the DM and DM/DBT groups while adjusting for clinical covariates.

Results A total of 22,883 mammograms were performed before DBT integration (DM group), and 22,824 mammograms were performed after complete DBT integration (DM/DBT group). After adjusting for multiple variables, the CDR was similar in both groups (38.2 per 1,000 examinations in the DM/DBT group versus 31.3 per 1,000 examinations in the DM group, $p = 0.14$); however, a higher proportion of cancers were invasive rather than in situ in the DM/DBT group [83.7% (731/873) versus 72.3% (518/716), $p < 0.01$]. The AIR was lower in the DM/DBT group ($p < 0.01$), and PPV2, PPV3, and specificity were higher in the DM/DBT group (all $p = 0.01$ or $p < 0.01$).

Conclusions Complete integration of DBT into the diagnostic setting is associated with improved diagnostic performance. Increased utilization of DBT may thus result in better patient outcomes and lead to a shift in the benchmarks that have been established for DM.

Key Points

- Integration of tomosynthesis into the diagnostic setting is associated with improved performance.
- A higher proportion of cancers are invasive rather than in situ with digital breast tomosynthesis.
- Increased utilization of tomosynthesis may lead to a shift in established benchmarks.

Keywords Benchmark · Breast cancer · Breast carcinoma in situ · Digital breast tomosynthesis · Digital mammography

Abbreviations

AIR Abnormal interpretation rate
BCSC Breast Cancer Surveillance Consortium

BI-RADS Breast Imaging Reporting and Data System
CDR Cancer detection rate
DBT Digital breast tomosynthesis
DM Digital 2D mammography
PPV Positive predictive value

✉ Manisha Bahl
mbahl1@mgh.harvard.edu

Introduction

Digital breast tomosynthesis (DBT) has emerged as a valuable imaging modality involving the acquisition of projection images that are reconstructed into thin image slices of the breast [1]. Multiple studies have demonstrated that integration of DBT into the screening setting has led to increased cancer detection rates (CDR) with fewer false-positive examinations

¹ Division of Breast Imaging/Department of Radiology, Massachusetts General Hospital, 55 Fruit Street, WAC 240, Boston, MA, USA

² Department of Radiology, Massachusetts General Hospital, 101 Merrimac Street, STE 1010, Boston, MA, USA

³ Department of Medicine, Massachusetts General Hospital, 100 Cambridge Street, Boston, MA, USA

[2–6]. For example, in a retrospective analysis of screening performance metrics from 13 academic and nonacademic breast centers, the addition of DBT to digital 2D mammography (DM) was associated with an increase in the invasive CDR from 2.9 to 4.1 per 1,000 examinations and a decrease in the abnormal interpretation rate (AIR) from 10.7% to 9.1% [4].

The benefits of DBT have been primarily studied in the screening setting, with limited research on DBT performance metrics in the diagnostic setting [7]. A single-institution study found that the integration of DBT into the diagnostic setting led to a significant improvement in positive predictive value (PPV) 3, but no other performance metrics were reported [7]. The American College of Radiology (ACR) recommends separate auditing of screening and diagnostic mammograms, but there are no established benchmarks for diagnostic DBT [8].

Our institution first integrated DBT into clinical practice in early 2011 and transitioned completely to DBT in early 2013, and thus it is well positioned to study performance metrics of DBT in the diagnostic setting. The purpose of our study is to compare performance metrics between DM and DBT in the diagnostic setting.

Materials and methods

Study subjects

The institutional review board exempted this Health Insurance Portability and Accountability Act-compliant retrospective study from the requirement for informed consent. We retrospectively reviewed consecutive diagnostic mammograms from August 1, 2008, to February 28, 2011 (DM group, before DBT integration), and from January 1, 2013, to July 31, 2015 (DM/DBT group, after complete DBT integration), at a single large tertiary academic medical center. After complete integration of DBT in January 2013, all examinations were performed with combined DBT and conventional DM. Diagnostic mammograms performed from March 2011 to December 2012 were not included to avoid selection bias, since DM examinations only were performed in some patients and combined DBT and DM examinations were performed in others during this period.

We searched the mammography information system (MagView, Burtonsville, MD) for all diagnostic mammogram reports with BI-RADS final assessment categories of 0 to 5 from August 2008 to February 2011 and from January 2013 to July 2015, in addition to all image-guided core biopsy and surgical pathology results within 1 year after the diagnostic mammogram was performed. Indications for diagnostic examinations included primarily evaluation of a breast problem (such as a palpable lump or focal pain) and short-interval follow-up. A less common indication was mammography

performed after identification of a finding on MRI, CT, or PET/CT. Diagnostic examinations performed for additional evaluation of a recent screening mammogram were not included in the study cohort. The database generated from the mammography information system was linked to tumor registries for five hospitals within our health care system.

Imaging technique and interpretation

Mammographic examinations were performed using full-field DM or DM/DBT (Hologic, Bedford, MA). All DBT examinations combined tomosynthesis and conventional DM, without using reconstructed DM views. At our institution, diagnostic examinations typically are unilateral and include craniocaudal and mediolateral oblique or mediolateral views of the affected breast, with additional views (such as spot compression or spot magnification views) and ultrasound examinations obtained at the discretion of the interpreting radiologist. During the time period of this study, the ultrasound units used by our institution remained the same (Philips Healthcare, Amsterdam, The Netherlands), but the software was periodically updated.

All examinations were interpreted by 1 of 28 breast imaging radiologists, with experience ranging from 1 to 35 years, using the BI-RADS Atlas [8]. Sixteen of the 28 radiologists were the same in the DM and DM/DBT groups. In the DM group, there were 21 interpreting radiologists, 8 of whom interpreted more than 1,000 examinations each, accounting for 66.9% (15,300/22,883) of the examinations. The other 13 radiologists in the DM group interpreted 33.1% (7,583/22,883) of the examinations. There were 23 interpreting radiologists in the DM/DBT group, 8 of whom interpreted more than 1,000 examinations each, accounting for 69.1% (15,768/22,824) of the examinations. The other 15 radiologists in the DM/DBT group interpreted 30.9% (7,056/22,824) of the examinations.

Standard definitions

Mammograms were categorized as true positive, true negative, false positive, or false negative according to standard definitions in the BI-RADS Atlas [8]. Mammograms were considered “true-positive” examinations if there was a tissue diagnosis of cancer [invasive cancer of ductal carcinoma in situ (DCIS)] within 1 year after a positive examination (BI-RADS 4 or 5). Mammograms were considered “true-negative” examinations if there was no known tissue diagnosis of cancer within 1 year of a negative examination (BI-RADS 1, 2, or 3). Mammograms were considered “false-positive” examinations if there was no known tissue diagnosis of cancer within 1 year of a positive examination (BI-RADS 4 or 5). Mammograms were considered “false-negative” examinations if there was a tissue diagnosis of

cancer within 1 year of a negative examination (BI-RADS 1, 2, or 3). For diagnostic mammograms with a BI-RADS final assessment category of 0 [0.6% (144/22,883) in the DM group and 0.4% (89/22,824) in the DM/DBT group], the final assessment category (1–5) given after completion of the recommended workup was used for purposes of the analysis. (An assessment of “0” was used primarily when the patient was unable to stay for additional recommended workup, such as ultrasound, or when an MRI examination was recommended for problem-solving purposes.)

Performance metrics were calculated using standard formulas from the BI-RADS Atlas [8]. The CDR is the number of cancers detected (true positives) per 1,000 examinations. The AIR is the percentage of total examinations interpreted as positive (BI-RADS 4 or 5). PPV2 is the percentage of diagnostic examinations recommended for tissue diagnosis or surgical consultation (BI-RADS 4 or 5) that result in a tissue diagnosis of cancer within 1 year. PPV3 is the percentage of all known biopsies performed as a result of positive diagnostic examinations (BI-RADS 4 or 5) that resulted in a tissue diagnosis of cancer within 1 year. Sensitivity, which is the probability of interpreting an examination as positive when cancer exists, is measured as the number of positive examinations for which there is a tissue diagnosis of cancer within 1 year of the examination (true positives) divided by all cancers present in the population examined in the same time period (true positives and false negatives). Specificity, which is the probability of interpreting an examination as negative when cancer does not exist, is measured as the number of negative examinations for which there is no tissue diagnosis of cancer within 1 year of the examination (true negatives) divided by all examinations for which there is no diagnosis of cancer within the same time period (true negatives and false positives).

Data collection and statistical analysis

For each diagnostic examination report, the following information was extracted from the mammography information system (MagView, Burtonsville, MD): all image-guided core biopsy and surgical pathology results within 1 year after the diagnostic mammogram, patient age, patient race, breast density, history of breast cancer, and reader. The mammographic finding types (architectural distortion, asymmetry, calcifications, or architectural distortion) for all true- and false-positive examinations were recorded.

All data were analyzed with statistical software (R version 3.4.2). We evaluated the distribution of age, race, breast density, and history of breast cancer by modality type (DM versus DM/DBT) using the Wilcoxon test (for continuous variables) and the Pearson’s chi-squared test (for categorical variables). For each of the performance metrics, parameters of a multivariable regression model were estimated using generalized

estimating equations (GEE) with a logit link and an autoregressive (AR1) correlation structure to account for multiple examinations per subject. The advantage of a multivariable regression model is that the association between an independent variable and the outcome can be estimated holding all other variables constant, which allows us to account for potentially confounding variables. Each model was adjusted for age, race, breast density, history of breast cancer, and reader. For each model, we calculated adjusted odds ratios, 95% confidence intervals, and p values for the DM and DM/DBT groups. In addition, the mammographic finding types that led to true- and false-positive examinations were compared between the DM and DM/DBT groups using the Pearson’s chi-squared test. $P < 0.05$ was considered statistically significant.

Results

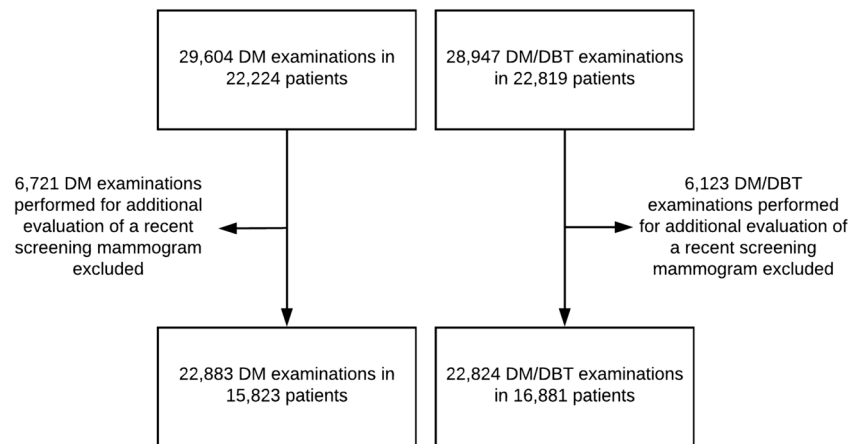
Characteristics of the study population

Over a 31-month period before DBT integration, 22,883 DM diagnostic examinations were performed in 15,823 women (DM group); over a 31-month period after complete DBT integration, 22,824 DBT diagnostic examinations were performed in 16,881 women (DM/DBT group) (Fig. 1). There was a small but statistically significant difference in age (52.8 versus 53.5 years in the DM and DM/DBT groups, respectively, $p < 0.01$) and proportion of prior history of breast cancer (11.9% versus 10.9% in the DM and DM/DBT groups, respectively, $p < 0.01$) (Table 1). There were no differences in race or breast density between the DM and DM/DBT groups ($p = 0.10$ and $p = 0.13$, respectively) (Table 1).

Of 45,707 total examinations performed in the DM and DM/DBT groups, 29,190 (63.9%) were given a final assessment category of 1 or 2, 11,776 (25.8%) were given a final assessment category of 3, and 4,741 (10.4%) were given a final assessment category of 4 or 5. The proportions of cases reported as BI-RADS 1 or 2, BI-RADS 3, and BI-RADS 4 or 5 were similar in the DM and DM/DBT groups ($p = 0.53$ – 0.73) (Table 2).

Performance metrics with DM versus DM/DBT

A total of 1,589 breast cancers were diagnosed within 1 year after 45,707 diagnostic examinations for an overall CDR of 34.8 per 1,000. After adjusting for age, race, breast density, prior history of breast cancer, and reader, the CDR was similar in both groups (adjusted OR = 1.11, 95% CI 0.97–1.27, $p = 0.14$) (Table 3). However, a higher proportion of cancers were invasive rather than in situ with DM/DBT compared with DM [83.7% (731/873) versus 72.3% (518/716), $p < 0.01$].

Fig. 1 Flow diagram of patient selection

The AIR was lower in the DM/DBT group (adjusted OR = 0.87, 95% CI 0.80-0.94, $p < 0.01$). PPV2 and PPV3 were higher in the DM/DBT group (adjusted OR = 1.34, 95% CI 1.13-1.60, $p < 0.01$, for PPV2; adjusted OR = 1.26, 95% CI 1.05-1.51, $p = 0.01$, for PPV3) (Table 3). Sensitivity was similar in the DM and DM/DBT groups (adjusted OR = 1.26, 95% CI 0.93-1.72, $p = 0.14$) (Table 3). Specificity was higher in the DM/DBT group (adjusted OR = 1.28, 95% CI 1.17-1.41, $p < 0.01$) (Table 3).

Mammographic finding types with DM versus DM/DBT

The mammographic finding of mass led to more true-positive examinations in the DM/DBT group than DM group [52.9% (462/873) vs. 41.3% (296/716), $p < 0.01$] (Table 4). The mammographic finding of calcifications led to fewer true-positive examinations in the DM/DBT group than DM group [23.8% (208/873) vs. 32.1% (230/716), $p < 0.01$] (Table 4).

Table 1 Comparison of patient characteristics with DM versus DM/DBT

	DM <i>n</i> = 22,883	DM/DBT <i>n</i> = 22,824	<i>p</i> value
Age ± 1 SD	52.8 ± 13.7	53.5 ± 13.6	< 0.01 ²
Race*			
White	91.1% (20,644/22,662)	91.1% (20,566/22,569)	0.10 ¹
African American	2.1% (468/22,662)	2.3% (523/22,569)	
Other	6.8% (1,550/22,662)	6.6% (1,480/22,569)	
Breast density**			
Almost entirely fatty or scattered areas of fibroglandular density	48.7% (10,733/22,036)	48.0% (10,639/22,167)	0.13 ¹
Heterogeneously dense or extremely dense	51.3% (11,303/22,036)	52.0% (11,528/22,167)	
Prior history of breast cancer			< 0.01 ¹
No prior history	88.1% (20,166/22,883)	89.1% (20,345/22,824)	
Prior history	11.9% (2,717/22,883)	10.9% (2,479/22,824)	

Tests used: ¹ Pearson test; ² Wilcoxon test

*Race was not reported in 221 cases in the DM group and 255 cases in the DM/DBT group

**Breast density was not reported in 847 cases in the DM group and 657 cases in the DM/DBT group

Table 2 Comparison of BI-RADS final assessment categories with DM versus DM/DBT

	DM	DM/DBT	<i>p</i> value
BI-RADS 1/2	63.9% (14,632/22,883)	63.8% (14,558/22,824)	0.72
BI-RADS 3	25.6% (5,866/22,883)	25.9% (5,910/22,824)	0.53
BI-RADS 4/5	10.4% (2,385/22,883)	10.3% (2,356/22,824)	0.73

The mammographic findings of architectural distortion and mass led to more false-positive examinations in the DM/DBT group than DM group [5.3% (78/1,485) vs. 1.4% (24/1,669), *p* < 0.01, for architectural distortion and 37.1% (551/1,485) vs. 20.8% (347/1,669), *p* < 0.01, for masses] (Table 5). The mammographic findings of asymmetry and calcifications led to fewer false-positive examinations in the DM/DBT group than DM group [11.6% (172/1,485) vs. 18.2% (303/1,669), *p* < 0.01, for asymmetries and 30.3% (450/1,485) vs. 36.8% (614/1,669), *p* < 0.01, for calcifications] (Table 5).

Discussion

In our study, integration of DBT into the diagnostic setting led to a significant decrease in AIR and significant increase in PPV2, PPV3, and specificity. Although there was no significant difference in CDR after adjusting for multiple variables, there was a higher proportion of invasive relative to in situ carcinomas with DM/DBT compared with DM. These results demonstrate that integration of DBT into the diagnostic setting leads to improved performance. In addition, our research supports guidelines

Table 3 Comparison of performance metrics with DM versus DM/DBT

	DM	DM/DBT	DM/DBT:DM aOR* (95% CI)	<i>p</i> value (adjusted)
Cancer detection rate (per 1,000 examinations)	31.3	38.2	1.11 (0.97, 1.27)	0.14
Number of cancers	716	873		
Total number of examinations	22,883	22,824		
Abnormal interpretation rate (%)	10.4	10.3	0.87 (0.80, 0.94)	< 0.01
Number of positive examinations (BI-RADS 4 or 5)	2,385	2,358		
Total number of examinations	22,883	22,824		
PPV2 (%)	30.0	37.0	1.34 (1.13, 1.60)	< 0.01
Number of cancers	716	873		
Number of biopsy recommendations	2,385	2,358		
PPV3 (%)	33.8	39.6	1.26 (1.05, 1.51)	0.01
Number of cancers	716	873		
Number of biopsies performed	2,121	2,207		
Sensitivity (%)	82.1	84.5	1.26 (0.93, 1.72)	0.14
Number of true-positive examinations	716	873		
Number of cancers (true-positive and false-negative examinations)	872	1,033		
Specificity (%)	92.4	93.2	1.28 (1.17, 1.41)	< 0.01
Number of true-negative examinations	20,342	20,306		
Number of non-cancers (true-negative and false-positive examinations)	22,011	21,791		

*aOR = adjusted odds ratio

Table 4 Comparison of mammographic finding types that led to true-positive examinations with DM versus DM/DBT

	DM*	DM/DBT*	<i>p</i> value
Architectural distortion	5.6% (40/716)	6.6% (58/873)	< 0.01
Asymmetry	10.5% (75/716)	9.4% (82/873)	
Calcifications	32.1% (230/716)	23.8% (208/873)	
Mass	41.3% (296/716)	52.9% (462/873)	

*No mammographic findings were reported in 75 DM examinations and 63 DM/DBT examinations

that recommend separate auditing of screening and diagnostic examinations [8].

Our results suggest that diagnostic DBT leads to increased positive predictive values with improved selection of patients recommended for biopsy. Our findings are consistent with the robust literature that supports the use of DBT in the screening setting. Through the acquisition of projection images that are reconstructed into thin image slices of the breast, DBT minimizes the masking effect of overlying tissue, enables improved cancer detection, and reduces false-positive findings [1]. In addition, we observed a higher proportion of invasive relative to in situ carcinomas with DM/DBT. This phenomenon has also been described with the use of DBT in the screening setting [3, 4, 9, 10]. The preferential ratio of invasive relative to in situ carcinomas with the transition to DBT may contribute to the optimization of patient outcomes from diagnostic mammography.

There is limited research on the performance metrics of diagnostic DBT. In a retrospective study of all diagnostic mammograms obtained during 1 year before DBT integration and during 3 consecutive years transitioning to full DBT integration (DBT1, DBT2, and DBT3), the authors found a significant increase in PPV3 from 29.6% (85/287) in the DM group to 50% (182/364) in the DBT3 group [7]. Of note, in the DBT1 and DBT2 groups, the patients received DBT or DM based on unit availability, so comparisons were primarily made between the DM group and DBT3 group. Other performance metrics were not calculated since limited follow-up data were available for the DBT3 group. Our study offers a comprehensive review of performance metrics in a large DBT practice and compares metrics during a 31-month period before the integration of DBT, when our institution was

performing only DM, to a 31-month period after complete integration of DBT, when our institution was performing combined DBT and DM in all patients, thus eliminating selection biases that may occur in a hybrid environment.

The aforementioned study reported a significant increase in examinations given a BI-RADS final assessment category of 1 or 2 and a significant decrease in examinations given a BI-RADS final assessment category of 3 after integration of DBT into the diagnostic setting [7]. (No significant change in examinations given a BI-RADS final assessment category of 4 or 5 was observed.) In our study, however, we observed no significant differences in the proportion of cases reported as BI-RADS 1 or 2 or BI-RADS 3 with integration of DBT. This finding may reflect differences in institutional practice regarding the use of BI-RADS 3. Of note, in the study that reported a decrease in BI-RADS 3 assessments, the baseline use of BI-RADS 3 with DM was higher than our institution's baseline use of BI-RADS 3 with DM (33.3% versus 25.6%) [7].

In a recent update from the Breast Cancer Surveillance Consortium (BCSC) on diagnostic DM, the CDR was 34.7 per 1,000 examinations (95% CI: 34.1, 35.2) [11]. The CDR for diagnostic DM in our study was 31.3 per 1,000. The lower CDR in our study may be due to patient selection: we included all diagnostic mammograms apart from those that were performed for additional evaluation of a recent screening mammogram. We chose to focus on this specific diagnostic population, since there is limited literature on diagnostic DBT for evaluation of findings that are not recalled from screening. If we exclude mammograms performed for additional evaluation of a recent screening mammogram from the BCSC study, the CDR would be 29.4 (8,006/272,572), which is more in keeping with the CDR in our study. As practices rapidly transition from DM to DBT, there is a need for benchmarks for facility auditing based on modern DBT performance in the diagnostic setting.

Our study also demonstrates that the types of mammographic findings that led to true- and false-positive examinations differed before and after complete integration of DBT. The mammographic finding of mass led to more true-positive examinations in the DM/DBT group than DM group, whereas the mammographic finding of calcifications led to fewer true-positive examinations in the DM/DBT group than DM group. These findings are consistent with our observation of a

Table 5 Comparison of mammographic finding types that led to false-positive examinations with DM versus DM/DBT

	DM*	DM/DBT*	<i>p</i> value
Architectural distortion	1.4% (24/1,669)	5.3% (78/1,485)	< 0.01
Asymmetry	18.2% (303/1,669)	11.6% (172/1,485)	
Calcifications	36.8% (614/1,669)	30.3% (450/1,485)	
Mass	20.8% (347/1,669)	37.1% (551/1,485)	

*No mammographic findings were reported in 381 DM examinations and 234 DM/DBT examinations

higher proportion of invasive relative to in situ carcinomas with DM/DBT, with invasive carcinomas often presenting as masses and in situ carcinomas often presenting as calcifications.

The types of mammographic findings that led to false-positive examinations differed before and after complete integration of DBT, with fewer false-positive examinations resulting from asymmetries and calcifications after complete integration of DBT. The types of findings that led to false-positive examinations with DBT provide insight into the advantages and possible disadvantages of this new imaging modality. More false-positive examinations resulted from architectural distortion after integration of DBT. These results are consistent with a recent study that found that architectural distortion is more readily detected on DBT than DM but is less likely to be malignant on DBT than DM; as such, the finding of architectural distortion can lead to more nonmalignant biopsies with DBT [12]. More false-positive examinations also resulted from masses with DBT, suggesting that short-term follow-up may be a reasonable option when mammographic features with DBT appear likely or probably benign after margin analysis. On the other hand, fewer false-positive examinations resulted from asymmetries with DBT, suggesting that fewer asymmetries are being recommended for biopsy with DBT or that asymmetries that persist on additional imaging are more likely to be significant.

Several studies have compared DM and DBT with regard to lesion characterization and image quality in the diagnostic setting and have concluded that the integration of DBT into the diagnostic setting can lead to more efficient evaluations and provide sufficient information for lesion localization [13–23]. For example, in a study of 146 women with 158 abnormalities who underwent DM and two-view DBT, with imaging reviewed by three radiologists, the authors reported that DBT could replace conventional DM views for the evaluation of non-calcified findings [18]. In contrast to masses, asymmetries, and architectural distortion, however, DM is more sensitive than DBT for the detection of calcifications [24].

In clinical practice, the results of this study suggest that the use of DBT is particularly helpful in the diagnostic setting and can reduce benign biopsies. DBT allows for better margin analysis and can increase the radiologist's confidence in determining the benignity of a lesion [1]. With subtle lesions, DBT can inform the decision for biopsy and increase overall accuracy [22]. Unnecessary ultrasound examinations and an increase in examination time can also be avoided in cases in which spot compression images with DBT demonstrate that an asymmetry represents normal superimposition of breast tissue. Currently, one of the disadvantages of DBT at our institution is the increased radiation dose to patients with combined DM and DBT; however, the radiation dose can be reduced using reconstructed DM views in place of the 2D DM acquisitions [25, 26].

Our study has several limitations. First, it was conducted at a large tertiary academic medical center that has been performing DBT since early 2011, and thus the results may not be generalizable to other institutions, especially those that have more recently transitioned to DBT. Second, the database generated from our mammography information system was linked to tumor registries for the five hospitals within our health care system; however, our database was not linked to a state registry. Third, although we avoided the selection bias inherent to other retrospective studies by comparing performance metrics in non-hybrid environments (DM in all patients versus combined DBT and DM in all patients), this study was not a randomized trial, and thus other factors may have influenced our results; however, we performed a multivariable regression and adjusted each model for age, race, breast density, history of breast cancer, and reader.

To our knowledge, our study is the largest and one of the first to report performance metrics for DBT fully integrated into the diagnostic setting. In a large DBT practice, complete integration of DBT into the diagnostic setting is associated with improved diagnostic performance. Increased utilization of DBT in the diagnostic setting may thus result in decreased health care costs and lessened patient anxiety, in addition to a shift in the benchmarks and outcomes that have been established for conventional DM.

Acknowledgements This study was presented as an electronic poster at the 2017 Society of Breast Imaging (SBI)/American College of Radiology (ACR) Breast Imaging Symposium.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Manisha Bahl, MD, MPH.

Conflict of interest The authors of this manuscript declare relationships with the following companies: Constance D. Lehman, MD, PhD, has a research grant from GE Healthcare and serves on an advisory board for GE Healthcare. The authors of this manuscript declare no other relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors (Sarah Mercaldo, PhD) has significant statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology

- retrospective
- observational
- performed at one institution

References

1. Hooley RJ, Durand MA, Philpotts LE (2017) Advances in digital breast tomosynthesis. *AJR Am J Roentgenol* 208(2):256–266
2. 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
3. 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
4. Friedewald SM, Rafferty EA, Rose SL et al (2014) Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 311(24):2499–2507
5. 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)
6. Sharpe RE Jr, Venkataraman S, Phillips J et al (2016) Increased cancer detection rate and variations in the recall rate resulting from implementation of 3D digital breast tomosynthesis into a population-based screening program. *Radiology* 278(3):698–706
7. Raghu M, Durand MA, Andrejeva L et al (2016) Tomosynthesis in the diagnostic setting: changing rates of BI-RADS final assessment over time. *Radiology* 281(1):54–61
8. Sickles EA, D'Orsi CJ, Bassett LW et al (2013) ACR BI-RADS mammography. In: *ACR BI-RADS atlas, breast imaging reporting and data system*. American College of Radiology, Reston, VA
9. 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(6):1401–1408
10. Bahl M, Gaffney S, McCarthy AM, Lowry KP, Dang PA, Lehman CD (2017) Breast cancer characteristics associated with 2D digital mammography versus digital breast tomosynthesis for screening-detected and interval cancers. *Radiology* 287(1):49–57
11. Sprague BL, Arao RF, Miglioretti DL et al (2017) National performance benchmarks for modern diagnostic digital mammography: update from the Breast Cancer Surveillance Consortium. *Radiology* 283(1):59–69
12. Bahl M, Lamb LR, Lehman CD (2017) Pathologic outcomes of architectural distortion on digital 2D versus tomosynthesis mammography. *AJR Am J Roentgenol* 209(5):1162–1167
13. Poplack SP, Tosteson TD, Kogel CA, Nagy HM (2007) Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. *AJR Am J Roentgenol* 189(3):616–623
14. Gennaro G, Toledano A, di Maggio C et al (2010) Digital breast tomosynthesis versus digital mammography: a clinical performance study. *Eur Radiol* 20(7):1545–1553
15. Hakim CM, Chough DM, Ganott MA, Sumkin JH, Zuley ML, Gur D (2010) Digital breast tomosynthesis in the diagnostic environment: a subjective side-by-side review. *AJR Am J Roentgenol* 195(2):W172–W176
16. Noroozian M, Hadjiiski L, Rahnama-Moghadam S et al (2012) Digital breast tomosynthesis is comparable to mammographic spot views for mass characterization. *Radiology* 262(1):61–68
17. Tagliafico A, Astengo D, Cavagnetto F et al (2012) One-to-one comparison between digital spot compression view and digital breast tomosynthesis. *Eur Radiol* 22(3):539–544
18. Brandt KR, Craig DA, Hoskins TL et al (2013) Can digital breast tomosynthesis replace conventional diagnostic mammography views for screening recalls without calcifications? A comparison study in a simulated clinical setting. *AJR Am J Roentgenol* 200(2):291–298
19. Waldherr C, Cerny P, Altermatt HJ et al (2013) Value of one-view breast tomosynthesis versus two-view mammography in diagnostic workup of women with clinical signs and symptoms and in women recalled from screening. *AJR Am J Roentgenol* 200(1):226–231
20. Zuley ML, Bandos AI, Ganott MA et al (2013) Digital breast tomosynthesis versus supplemental diagnostic mammographic views for evaluation of noncalcified breast lesions. *Radiology* 266(1):89–95
21. Lei J, Yang P, Zhang L, Wang Y, Yang K (2014) Diagnostic accuracy of digital breast tomosynthesis versus digital mammography for benign and malignant lesions in breasts: a meta-analysis. *Eur Radiol* 24(3):595–602
22. Peppard HR, Nicholson BE, Rochman CM, Merchant JK, Mayo RC 3rd, Harvey JA (2015) Digital breast tomosynthesis in the diagnostic setting: indications and clinical applications. *Radiographics* 35(4):975–990
23. Poplack S (2017) Breast tomosynthesis: clinical evidence. *Radiol Clin North Am* 55(3):475–492
24. Spangler ML, Zuley ML, Sumkin JH et al (2011) Detection and classification of calcifications on digital breast tomosynthesis and 2D digital mammography: a comparison. *AJR Am J Roentgenol* 196(2):320–324
25. 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
26. Gennaro G, Bernardi D, Houssami N (2018) Radiation dose with digital breast tomosynthesis compared to digital mammography: per-view analysis. *Eur Radiol* 28(2):573–581