Detection Rates for Benign and Malignant Diagnoses on Breast Cancer Screening With Digital Breast Tomosynthesis in a Statewide Mammography Registry Study

Mayo H. Fujii[1]
Sally D. Herschorn[2,3]
Michelle Sowden[1,3]
Elise L. Hotaling[2,3]
Pamela M. Vacek[3,4]
Donald L. Weaver[3,5]
Brian L. Sprague[1,2,3]

Keywords: breast cancer screening, digital breast tomosynthesis, mammography

doi.org/10.2214/AJR.18.20255

Received June 14, 2018; accepted after revision August 16, 2018.

Supported in part by the National Cancer Institute (U54 CA163303, P01 CA154292, R03 CA223725), the Patient-Centered Outcomes Research Institute (PCS-1504-30370), and the University of Vermont Cancer Center with funds awarded by the Lake Champlain Cancer Research Organization (pilot grant 032800). None of the funders had a role in the design or conduct of the study or the reporting of results. The views in this work are solely the responsibility of the authors and do not necessarily represent the views of the National Cancer Institute or the Patient-Centered Outcomes Research Institute, its Board of Governors, or its Methodology Committee.

Based on a presentation at the American Society of Preventive Oncology 2017 annual meeting, Seattle, WA.

[1]Department of Surgery, University of Vermont, 1 S Prospect St, UHC Rm 4425, Burlington, VT 05401. Address correspondence to B. L. Sprague (bsprague@uvm.edu).

[2]Department of Radiology, University of Vermont, Burlington, VT.

[3]University of Vermont Cancer Center, University of Vermont, Burlington, VT.

[4]Department of Biostatistics, University of Vermont, Burlington, VT.

[5]Department of Pathology, University of Vermont, Burlington, VT.

AJR 2019; 212:1–6

0361–803X/19/2123–1

OBJECTIVE. The objective of our study was to determine whether detection rates of specific benign and malignant diagnoses differ for breast cancer screening with digital breast tomosynthesis (DBT) versus full-field digital mammography (FFDM) alone.

MATERIALS AND METHODS. We analyzed observational data from the Vermont Breast Cancer Surveillance System, including 86,349 DBT screening examinations and 97,378 FFDM screening examinations performed at eight radiology facilities in Vermont that adopted DBT screening during 2012–2016. We determined the most severe diagnosis made within 6 months after positive screening examinations. Multivariable-adjusted logistic regression was used to compare detection rates for specific diagnoses on DBT versus FFDM.

RESULTS. Compared with FFDM, DBT had a lower recall rate (adjusted odds ratio [OR], 0.81; 95% CI, 0.77–0.85) but comparable biopsy rate (OR = 1.05; 95% CI, 0.93–1.17), benign biopsy rate (OR = 1.12; 95% CI, 0.97–1.29), and cancer detection rate (OR = 0.94; 95% CI, 0.78–1.14). Among benign diagnoses, DBT and FFDM had comparable detection rates for nonproliferative lesions (OR = 1.19; 95% CI, 0.92–1.53), fibroepithelial proliferations (OR = 1.24; 95% CI, 0.85–1.81), proliferative lesions without atypia (OR = 1.13; 95% CI, 0.90–1.42), atypical lesions (OR = 0.77; 95% CI, 0.43–1.38), and lobular carcinoma in situ (LCIS) (OR = 0.92; 95% CI, 0.53–1.61). Among malignant diagnoses, DBT and FFDM had comparable detection rates for ductal carcinoma in situ (OR = 1.05; 95% CI, 0.70–1.57) and invasive breast cancer (OR = 0.92; 95% CI, 0.74–1.13), with no statistically significant differences in detection of invasive ductal carcinoma (OR = 0.83; 95% CI, 0.66–1.06), invasive lobular carcinoma (OR = 1.11; 95% CI, 0.59–2.07), or invasive mixed ductal-lobular carcinoma (OR = 1.49; 95% CI, 0.65–3.39).

CONCLUSION. Compared with FFDM, breast cancer screening with DBT has a lower recall rate while detecting a similar distribution of benign and malignant diagnoses.

Digital breast tomosynthesis (DBT) has emerged as a new breast cancer screening modality that could substantially improve the benefit-to-harm ratio for screening [1]. Observational studies in the United States suggest that DBT decreases recall rates and increases invasive cancer detection rates when added to conventional FFDM [2–9]. Although early data are promising, the U.S. Preventive Services Task Force concluded in early 2016 that there is insufficient evidence to evaluate the benefits and harms of DBT for screening [10]. One particular area of uncertainty is the effect DBT has on the detection and diagnosis of specific benign and malignant lesions.

The use of DBT results in the improved depiction of architectural distortion, which on 2D images often appears because of overlapping fibroglandular tissue [11]. Previous studies have reported that DBT increases the rate of recall for architectural distortion and masses [8, 12] while reducing the rate of recall for asymmetries [12, 13]. These various types of mammographic findings are associated with different types of benign and malignant diagnoses, and thus these properties could influence the type of benign and malignant lesions detected on screening.

Benign diagnoses are typically perceived as an undesired outcome of screening, representing scenarios in which cancer was suspected but is found not to be present. At least one study has suggested that the use of DBT may be associated with increased benign biopsy rates [2], although this has not been consistently observed [14]. To our knowledge, no prior studies have described the rate of spe-

cific benign diagnoses identified via screening with DBT compared with FFDM alone.

The purpose of this study was to determine whether detection rates of specific benign and malignant diagnoses differ for breast cancer screening with DBT versus FFDM alone. We used data from the Vermont Breast Cancer Surveillance System (VBCSS), which includes a statewide registry of all breast cancer screening mammography examinations performed in Vermont and is linked to patient risk factor data and pathology records for benign and malignant diagnoses.

Materials and Methods

Design

We conducted an analysis of observational data from the VBCSS [15]. The VBCSS has collected statewide mammography data in Vermont since 1994 and is a member of the Breast Cancer Surveillance Consortium [16]. The VBCSS includes a registry of all breast imaging examinations (mammography, ultrasound, and MRI) performed at radiology facilities in Vermont that is linked to statewide breast pathology reports and records from the Vermont Cancer Registry. This study was approved by the University of Vermont Institutional Review Board with a waiver of consent, and all study procedures were compliant with the HIPAA. However, women attending breast imaging examinations at radiology facilities in Vermont are given the option to opt out of participation in research via a health questionnaire they complete at each visit. This study was limited to women who did not opt out of participation in research (90% of women). A prior study reporting basic screening performance statistics for DBT and FFDM from the consortium for Population-Based Research Optimizing Screening Through Personalized Regimens included data on 18,983 DBT and 43,198 FFDM examinations [14], and those examinations are also included in this study. The current study includes a larger number of examinations from the VBCSS and evaluates detection rates for specific benign and malignant diagnoses.

Study Setting and Population

The analyses for this study were restricted to eight radiology facilities in Vermont that adopted DBT for breast cancer screening during 2012–2016. All facilities used DBT Selenia Dimensions mammography systems (Hologic). The DBT adoption date and implementation method varied by facility. At facilities that gradually transitioned from FFDM to DBT screening, DBT screening was not explicitly targeted to certain patient populations. Rather, screening modality (FFDM vs DBT) was generally assigned on the basis of room availability, although women were given the option to decline DBT screening if they preferred FFDM alone. All facilities used DBT combined with 2D FFDM views at the start of the study period. Six facilities replaced FFDM views with synthetic 2D views reconstructed from the DBT views during the course of the study period. Among DBT examinations included in the analyses, 67% did not include synthetic 2D views, 13% included synthetic 2D views but also obtained conventional 2D FFDM views, and 20% obtained synthetic 2D views without conventional 2D FFDM views.

FFDM and DBT screening examinations from January 2012 through December 2016 were identified for women who had not opted out of participation in research (n = 201,523 screening examinations among 70,276 women). Screening examinations among women with a history of breast cancer (n = 15,679 examinations) or breast implants (n = 2117 examinations) were excluded because screening performance metrics differ markedly among women in these populations compared with the general screening population [17, 18]. A total of 86,349 DBT and 97,378 FFDM examinations among 66,003 women and interpreted by 49 radiologists met the final eligibility criteria.

Data Collection

Patient demographic and risk factor data (including age, race and ethnicity, and family history of breast cancer) were obtained from standardized questionnaires completed by subjects at each breast imaging examination.

Radiologic information including date of examination, modality (FFDM vs DBT), indication for examination (i.e., screening vs diagnostic), assessment category, and breast density category was provided by the radiology facility. Assessments and breast density were categorized per standard clinical practice according to BI-RADS [19].

The VBCSS obtains copies of pathology reports for all breast specimens evaluated at pathology facilities in the state of Vermont. Malignant and benign diagnoses were abstracted from pathology reports by a trained abstractor. The VBCSS also obtains consolidated breast cancer diagnosis data, including date of diagnosis, histologic

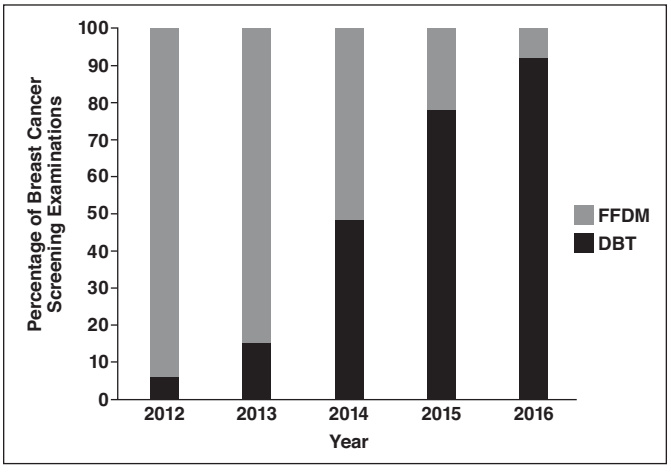

Fig. 1—Bar graph shows adoption of digital breast tomosynthesis (DBT) for breast cancer screening at eight radiology facilities in Vermont, Vermont Breast Cancer Surveillance System, for period of 2012–2016. FFDM = full-field digital mammography.

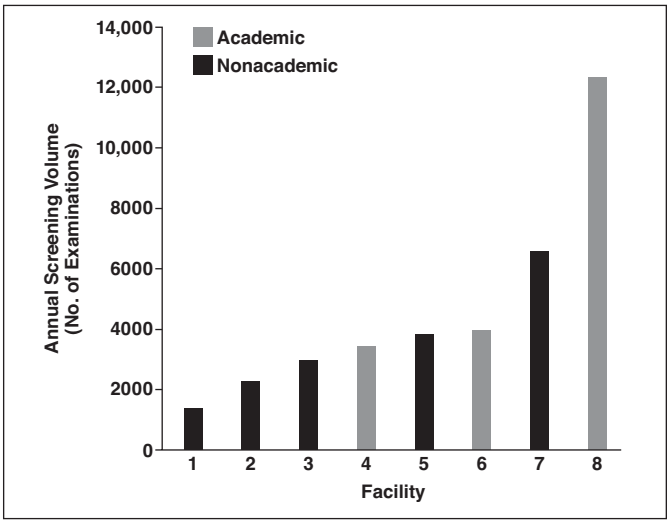

Fig. 2—Bar graph shows annual screening volume by facility and academic affiliation at eight radiology facilities in Vermont, Vermont Breast Cancer Surveillance System, for period of 2012–2016.

subtype, and stage at diagnosis, via linkage to the statewide Vermont Cancer Registry.

Measures and Definitions

A positive screening examination was defined as one with an initial BI-RADS assessment of 0, 3, 4, or 5 [19]. The recall (abnormal interpretation) rate was defined as the number of positive screening examinations divided by the total number of screening examinations. The biopsy rate was determined as the proportion of screening examinations that were positive and followed by a biopsy within 6 months. Screen-detected lesions were defined as those that were diagnosed within 6 months of a positive screening examination. Positive predictive value of recall (PPV1) was defined as the proportion of positive screening examinations that resulted in a screen-detected cancer (i.e., ductal carcinoma in situ [DCIS] or invasive). Positive predictive value of biopsy (PPV3) was defined as the proportion of positive examinations with biopsy that resulted in a screen-detected cancer.

Each breast pathology diagnosis occurring within 6 months of a positive screening examination was categorized as nonproliferative benign changes (fibrosis, cysts, adenosis, and apocrine metaplasia), fibroepithelial proliferations (fibroadenoma, adenomyoepithelioma, and phyllodes tumor), proliferative lesions without atypia (intraductal papilloma without atypia, usual ductal hyperplasia, columnar cell change, columnar cell hyperplasia, sclerosing adenosis, and complex sclerosing lesion), atypical lesions (atypical ductal hyperplasia, atypical lobular hyperplasia, flat epithelial atypia, columnar cell change or hyperplasia with atypia, and atypical papilloma), LCIS, DCIS, and invasive breast cancer.

For women with multiple screen-detected diagnoses after a single screening examination, the most severe diagnosis was determined according to the following hierarchy: nonproliferative benign changes, fibroepithelial proliferations, proliferative lesions without atypia, atypical lesions, LCIS, DCIS, and invasive breast cancer. Invasive breast cancer diagnoses were further subdivided by histologic subtype into ductal, lobular, mixed ductal-lobular, and other or unknown (metaplastic, invasive not otherwise specified).

Statistical Analysis

We compared screening performance metrics and rates of specific types of screen-detected benign and malignant lesions for DBT versus FFDM examinations. Because screening modality was not randomly assigned, we controlled for potential differences in the risk profiles of women undergoing DBT versus FFDM using multivariable logistic regression. We selected potential confounding factors a priori for inclusion in the model on the basis of their known association with screening performance metrics. Logistic regression models were adjusted for examination year, age group, breast density, family history of breast cancer, and facility. Odds ratios (ORs) and 95% CIs were used to describe the strength of the associations and evaluate statistical significance. The use of 95% CIs helps to avoid the limitations of reliance on *p* values by quantifying the precision of the OR point estimates and indicating the range of possible associations that are reasonably compatible with the observed data [20, 21].

We conducted stratified analyses to examine whether the cancer detection results varied according to academic versus nonacademic affiliation and tested for a statistical interaction using cross-product terms in the regression model. To evaluate the potential influence of a learning curve on DBT diagnoses,

TABLE 1: Characteristics of the Study Population Undergoing Breast Cancer Screening at Eight Radiology Facilities in Vermont, Vermont Breast Cancer Surveillance System, 2012–2016

Characteristic	FFDM (*n* = 97,378 Examinations)		DBT (*n* = 86,349 Examinations)	
	No.	%	No.	%
Age at mammography (y)				
< 40	988	1.0	805	0.9
40–49	21,368	21.9	20,133	23.3
50–59	31,684	32.5	28,639	33.2
60–69	26,654	27.4	24,162	28.0
≥ 70	16,684	17.1	12,610	14.6
Breast density				
Almost entirely fat	16,782	17.5	12,115	14.1
Scattered fibroglandular density	48,759	50.8	45,550	52.9
Heterogeneously dense	26,478	27.6	24,565	28.5
Extremely dense	4028	4.2	3932	4.6
Unknown	1331		187	
Family history of breast cancer				
No first-degree relative with a history of breast cancer	71,344	80.7	61,232	78.3
First-degree relative with a history of breast cancer	17,057	19.3	16,922	21.7
Unknown	8977		8195	

Note—FFDM = full-field digital mammography alone, DBT = digital breast tomosynthesis.

TABLE 2: Screening Performance Metrics for Full-Field Digital Mammography (FFDM) and Digital Breast Tomosynthesis (DBT) at Eight Radiology Facilities in Vermont, Vermont Breast Cancer Surveillance System, 2012–2016

Performance Metric Based on Diagnosis at Pathology	FFDM (*n* = 97,378)		DBT (*n* = 86,349)		Unadjusted		Adjusted[a]	
	No.	Rate	No.	Rate	OR	95% CI	OR	95% CI
Abnormal interpretation rate (recall)	10,608	10.9%	6798	7.9%	0.70	0.68–0.72	0.81	0.77–0.85
Biopsy rate	1477	15.2[b]	1201	13.9[b]	0.92	0.85–0.99	1.05	0.93–1.17
Cancer detection rate	548	5.6[b]	436	5.0[b]	0.90	0.79–1.02	0.94	0.78–1.14
Benign disease detection rate	929	9.5[b]	765	8.9[b]	0.93	0.84–1.02	1.12	0.97–1.29
PPV1 (cancer/recall)	548	5.2%	436	6.4%	1.26	1.11–1.43	1.12	0.92–1.36
PPV3 (cancer/biopsy)	548	37.1%	436	36.3%	0.97	0.83–1.14	0.84	0.65–1.08

Note—OR = odds ratio, PPV = positive predictive value.
[a]Adjusted for examination year, age group, breast density, family history of breast cancer, and facility.
[b]Rate is reported as value per 1000 examinations.

TABLE 3: Benign and Malignant Diagnoses Detected Via Screening With Full-Field Digital Mammography (FFDM) Alone Versus Digital Breast Tomosynthesis (DBT) at Eight Radiology Facilities in Vermont, Vermont Breast Cancer Surveillance System, 2012–2016

	FFDM (*n* = 97,378 Examinations)		DBT (*n* = 86,349 Examinations)		Unadjusted		Adjusted[b]	
Most Severe Diagnosis[a]	No.	Rate	No.	Rate	OR	95% CI	OR	95% CI
Benign								
Nonproliferative lesions	325	3.3	232	2.7	0.81	0.68–0.95	1.19	0.92–1.53
Fibroepithelial proliferations	112	1.2	120	1.4	1.21	0.93–1.56	1.24	0.85–1.81
Proliferative lesions without atypia	379	3.9	329	3.8	0.98	0.84–1.14	1.13	0.90–1.42
Atypical lesions	55	0.6	42	0.5	0.86	0.58–1.29	0.77	0.43–1.38
Lobular carcinoma in situ	15	0.2	9	0.1	0.72	0.49–1.05	0.92	0.53–1.61
Benign, other or NOS	43	0.4	33	0.4	0.87	0.55–1.36	0.80	0.40–1.59
Malignant								
Ductal carcinoma in situ	117	1.2	91	1.1	0.88	0.67–1.15	1.05	0.70–1.57
Invasive breast cancer	431	4.4	345	4.0	0.90	0.78–1.04	0.92	0.74–1.13
Ductal carcinoma	335	3.4	265	3.1	0.89	0.76–1.05	0.83	0.66–1.06
Lobular carcinoma	42	0.4	42	0.5	1.13	0.74–1.73	1.11	0.59–2.07
Mixed ductal-lobular carcinoma	30	0.3	26	0.3	0.98	0.58–1.65	1.49	0.65–3.39
Invasive, other or NOS	24	0.2	12	0.1	0.56	0.28–1.13	1.47	0.53–4.02

Note—OR = odds ratio, NOS = not otherwise specified.

[a]For women with multiple screen-detected diagnoses after a single screening examination, the most severe diagnosis was determined according to the following hierarchy: nonproliferative benign changes, fibroepithelial proliferations, proliferative lesions without atypia, atypical lesions, lobular carcinoma in situ, ductal carcinoma in situ, and invasive breast cancer.

[b]Adjusted for examination year, age group, breast density, family history of breast cancer, and facility.

we conducted a sensitivity analysis in which the first year of DBT screening data at each facility was excluded from the analyses. SAS software (version 9.4, SAS Institute) was used for all analyses.

Results

The use of DBT for breast cancer screening at the eight included facilities increased steadily throughout the study period, with 92% of screening examinations using DBT views by 2016 (Fig. 1). Approximately half of the examinations (54%) were conducted at facilities associated with an academic medical center, and the remainder were conducted at community hospitals. Mean annual facility volumes ranged from 1378 to 12,307 examinations per year (Fig. 2). Patient characteristics were very similar between the FFDM and DBT screening groups (Table 1). The mean age was 59.1 years in the FFDM group and 58.3 years in the DBT group.

Screening with DBT had a lower recall rate than screening with FFDM alone (7.9% vs 10.9%; adjusted OR = 0.81; 95% CI, 0.77–0.85) (Table 2). The biopsy rate was very similar on DBT and FFDM screening (adjusted OR = 1.05; 95% CI, 0.93–1.17). A total of 1694 benign diagnoses and 984 malignant diagnoses were screen-detected (Table 3). The benign biopsy rate and cancer detection rate on DBT were effectively equivalent to FFDM after statistical adjustment for covariates. Similarly, there were no statistically significant differences in PPV1 or PPV3 after statistical adjustment. In stratified analyses, we found that the cancer detection rates were similar for DBT vs FFDM among both academic (OR = 1.04; 95% CI, 0.83–1.32) and nonacademic (OR = 0.85; 95% CI, 0.61–1.19) facilities ($p = 0.74$ for interaction).

Compared with FFDM, DBT had slightly elevated rates after covariate adjustment of screen-detected nonproliferative lesions (OR = 1.19; 95% CI, 0.92–1.53), fibroepithelial proliferations (OR = 1.24; 95% CI, 0.85–1.81), and proliferative lesions without atypia (OR = 1.13; 95% CI, 0.90–1.42) and had a lower rate of atypical lesions (OR = 0.77; 95% CI, 0.43–1.38), but none of these differences was statistically significant. DBT and FFDM had comparable detection rates for LCIS (OR = 0.92; 95% CI, 0.53–1.61).

Detection rates for DCIS and invasive breast cancer were similar on DBT and FFDM (Table 3). Among invasive cancers, there was a slight decrease in detection of invasive ductal cancer on DBT compared with FFDM (OR = 0.83; 95% CI, 0.66–1.06), but it was not statistically significant. There was no evidence that detection rates for lobular or mixed ductal-lobular invasive cancers were different on DBT vs FFDM, although the CIs were wide.

In sensitivity analyses in which the first year of DBT screening data at each facility was excluded, there were a total of 561 benign diagnoses and 336 malignant diagnoses made on 66,940 DBT screens. The results remained essentially the same. There was a similar reduction in recall rate on DBT (OR = 0.82; 95% CI, 0.77–0.87), and there were very small changes in the other screening performance metrics. Similarly, there were small changes in OR estimates for any of the categories of benign and malignant diagnoses, and there remained no statistically significant differences between DBT and FFDM.

Discussion

In this study, the adoption of DBT for breast cancer screening was associated with reduced recall rate but did not appear to substantially change the distribution of specific screen-detected benign and malignant diagnoses in this sample of eight academic and community-practice Vermont facilities. Our study provides the first evidence, to our knowledge, regarding detection rates for

specific types of benign diagnoses on DBT, and the comparable benign diagnosis detection rates for DBT and FFDM suggest that long-standing evidence on detection rates for specific benign diagnoses for breast cancer screening with FFDM can likely be expected to apply to screening with DBT.

Similar to the investigators of a previous study in the literature [3], we observed that DBT screening was associated with a reduced recall rate compared with FFDM alone. In prior studies, investigators have reported elevated biopsy rates on DBT screening [2, 7], but others have reported no difference [5, 6]. We observed no difference in overall or benign biopsy rates on DBT versus FFDM screening. A majority of prior studies have observed elevated cancer detection rates with DBT compared with FFDM alone [2, 5–7, 14, 22], although other studies have provided exceptions [8, 13, 23].

It is unclear why increased cancer detection with DBT screening was not experienced in Vermont. One potential contributing factor is the relatively high cancer detection rate on FFDM screening examinations in this study (5.6 per 1000 examinations). Studies reporting increased cancer detection with DBT in the United States have had comparison FFDM cancer detection rates of less than 5 per 1000 examinations [2, 5–7, 14]. Our findings suggest that it may be difficult for DBT to increase cancer detection rates among providers who are already achieving high cancer detection rates with FFDM. Our results are consistent with those in a recent study in which Bahl et al. [23] also reported a relatively high cancer detection rate on FFDM (5.0 per 1000 examinations) that was not improved by DBT. However, randomized trials in European settings have achieved elevated cancer detection rates on DBT in settings with high FFDM cancer detection rates [24, 25]. Our study group included low-volume community hospitals with relatively limited experience with DBT. However, stratified analyses revealed no increase in cancer detection with DBT at both the academic and nonacademic facilities. Further research is needed to identify facility, radiologist, and patient factors associated with differences in the impact of DBT on cancer detection rates and other screening performance metrics, including the potential for a learning curve effect with increasing experience in DBT interpretation.

We are aware of only two prior studies reporting on DBT detection of specific categories of benign disease. Lourenco et al. [8] noted that 19.6% of screen-detected diagnoses with DBT were high-risk benign lesions, compared with 11.7% for FFDM screening. Lamb et al. [26] described the distribution of high-risk benign lesions detected after FFDM and DBT screening: They reported that atypical hyperplasia constituted a lower proportion of all high-risk lesions in the DBT group, whereas radial scar, papilloma, and atypical lobular hyperplasia made up a higher percentage. Absolute detection rates for benign and high-risk benign lesions were not reported in either study. We found no evidence in our study that detection rates of high-risk benign lesions (i.e., atypical lesions or LCIS) were elevated on DBT. There was a small decrease in detection of atypical lesions on DBT, but the CI was wide and was not statistically significant. CIs for other benign diagnoses were more narrow, although we could not exclude small increases in the detection of nonproliferative changes, fibroepithelial proliferations, and proliferative lesions without atypia.

As an observational study, the results of our study must be interpreted with caution and the potential influence of selection bias must be considered. Half of the facilities included in the study transitioned gradually from FFDM to DBT screening. Although DBT was not targeted on the basis of patient characteristics at any facility, patients were permitted to undergo screening with FFDM alone if they preferred. We used statistical adjustment to control for the modest differences in measured patient factors and additionally controlled for secular trends and variation by facility by including calendar year and facility identification in the regression models. Although there is little racial or ethnic diversity in Vermont (96% of women in the study were white), socioeconomic diversity is prevalent—with particularly high representation of rural women. Additional studies will be needed to confirm our findings in other populations and examine potential differences in racial and ethnic subgroups. Finally, all the facilities in our study used Hologic mammography systems and thus our results may not be generalizable to other mammography systems.

In this statewide registry-based observational study, breast cancer screening with DBT was associated with reduced recall rate and did not substantially change the distribution of specific benign and malignant diagnoses compared with screening with FFDM alone. Our results provide the first evidence to our knowledge regarding detection of specific types of benign diagnoses on DBT versus FFDM screening and suggest that the introduction of DBT improves screening performance by reducing recall rates but has little influence on the benefits or harms of breast cancer screening through an impact on benign diagnoses. Our finding of no elevated cancer detection on DBT screening provides motivation for further research on factors associated with variability in the impact of DBT on cancer detection and other performance metrics.

References

1. Houssami N, Miglioretti DL. Digital breast tomosynthesis: a brave new world of mammography screening. *JAMA Oncol* 2016; 2:725–727
2. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 2014; 311:2499–2507
3. Melnikow J, Fenton JJ, Miglioretti D, Whitlock EP, Weyrich MS. Screening for breast cancer with digital breast tomosynthesis. Rockville, MD: Agency for Healthcare Research and Quality, 2015: AHRQ Publication No. 14-05201-EF-2
4. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology* 2013; 269:694–700
5. McCarthy AM, Kontos D, Synnestvedt M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *J Natl Cancer Inst* 2014; 13:106
6. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R Jr. Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR* 2013; 200:1401–1408
7. Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *AJR* 2014; 203:687–693
8. Lourenco AP, Barry-Brooks M, Baird GL, Tuttle A, Mainiero MB. Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. *Radiology* 2015; 274:337–342
9. Sharpe RE Jr, Venkataraman S, Phillips J, et al. 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* 2016; 278:698–706
10. Siu AL; U.S. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2016; 164:279–296

11. Dibble EH, Lourenco AP, Baird GL, Ward RC, Maynard AS, Mainiero MB. Comparison of digital mammography and digital breast tomosynthesis in the detection of architectural distortion. *Eur Radiol* 2018; 28:3–10
12. McDonald ES, Oustimov A, Weinstein SP, Synnestvedt MB, Schnall M, Conant EF. Effectiveness of digital breast tomosynthesis compared with digital mammography: outcomes analysis from 3 years of breast cancer screening. *JAMA Oncol* 2016; 2:737–743
13. Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology* 2015; 274:85–92
14. Conant EF, Beaber EF, Sprague BL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: a cohort study within the PROSPR Consortium. *Breast Cancer Res Treat* 2016; 156:109–116
15. Sprague BL, Bolton KC, Mace JL, et al. Registry-based study of trends in breast cancer screening mammography before and after the 2009 U.S. Preventive Services Task Force recommendations. *Radiology* 2014; 270:354–361
16. Lehman CD, Arao RF, Sprague BL, et al. National performance benchmarks for modern screening digital mammography: update from the Breast Cancer Surveillance Consortium. *Radiology* 2017; 283:49–58
17. Houssami N, Abraham LA, Miglioretti DL, et al. Accuracy and outcomes of screening mammography in women with a personal history of early-stage breast cancer. *JAMA* 2011; 305:790–799
18. Miglioretti DL, Rutter CM, Geller BM, et al. Effect of breast augmentation on the accuracy of mammography and cancer characteristics. *JAMA* 2004; 291:442–450
19. Sickles EA, D'Orsi CJ, Bassett LW, et al. ACR BI-RADS Mammography, 5th ed. In: D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA, et al. *ACR BI-RADS Atlas, Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology, 2013
20. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)* 1986; 292:746–750
21. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007; 147:W163–W194
22. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology* 2013; 267:47–56
23. Bahl M, Gaffney S, McCarthy AM, Lowry KP, Dang PA, Lehman CD. Breast cancer characteristics associated with 2D digital mammography versus digital breast tomosynthesis for screening-detected and interval cancers. *Radiology* 2018; 287:49–57
24. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol* 2013; 14:583–589
25. Skaane P, Sebuodegard S, Bandos AI, et al. Performance of breast cancer screening using digital breast tomosynthesis: results from the prospective population-based Oslo Tomosynthesis Screening Trial. *Breast Cancer Res Treat* 2018; 169:489–496
26. Lamb LR, Bahl M, Hughes KS, Lehman CD. Pathologic upgrade rates of high-risk breast lesions on digital two-dimensional vs tomosynthesis mammography. *J Am Coll Surg* 2018; 226:858–867