DOI:10.1093/jnci/dju316
First published online October 13, 2014

ARTICLE

Screening Outcomes Following Implementation of Digital Breast Tomosynthesis in a General-Population Screening Program

Anne Marie McCarthy, Despina Kontos, Marie Synnestvedt, Kay See Tan, Daniel F. Heitjan, Mitchell Schnall, Emily F. Conant

Manuscript received May 20, 2014; revised July 27, 2014; accepted August 22, 2014.

Correspondence to: Emily F. Conant, MD, Department of Radiology, University of Pennsylvania Perelman School of Medicine, 3400 Spruce St, Philadelphia, PA 19104 (e-mail: emily.conant@uphs.upenn.edu).

Background Early data on breast cancer screening utilizing digital breast tomosynthesis (DBT) combined with digital mammography (DM) have shown improvements in false-positive and false-negative screening rates compared with DM alone. However, these trials were performed at sites where conventional mammographic screening was concurrently performed, possibly leading to selection biases or with complex, multireader algorithms not reflecting general clinical practice. Our study reports the impact on screening outcomes for DBT screening implemented in an entire clinic population.

Methods Recall rates, cancer detection, and positive predictive values of screening were compared for 15571 women screened with DBT and 10728 screened with DM alone prior to DBT implementation at a single breast imaging center. Generalized linear mixed-effects models were used to estimate the odds ratio (OR) for recall rate adjusted for age, race, presence of prior mammograms, breast density and reader. All statistical tests were two-sided.

Results DBT screening showed a statistically significant reduction in recalls compared to DM alone. For the entire population, there were 16 fewer recalls (8.8% vs 10.4%, P <.001, adjusted OR = 0.80, 95% confidence interval [CI] = 0.74 to 0.88, P < .001) and 0.9 additional cancers detected per 1000 screened with DBT compared to DM alone. There was a statistically significant increase in PPV1 (6.2% vs 4.4%, P = .047). In women younger than age 50 years screened with DBT, there were 17 fewer recalls (12.3% vs 14.0%, P = .02) and 3.6 additional cancer detected per 1000 screened (5.7 vs 2.2 per 1000, P = .02).

Conclusions Our data support the clinical implementation of DBT in breast cancer screening; however, larger prospective trials are needed to validate our findings in specific patient subgroups.

JNCI J Natl Cancer Inst (2014) 106(11): dju316 doi:10.1093/jnci/dju316

Screening mammography, despite ongoing controversy regarding its risk-benefit ratio, remains the mainstay of early breast cancer detection (1,2). Digital breast tomosynthesis (DBT), a relatively new x-ray technology (3,4) that images the breast in 3-D, has shown promise in addressing some limitations of conventional mammography by alleviating the effect of superimposed structures that can lead to erroneous interpretations (5–8). This new technology is increasingly implemented in breast clinics across the country, despite relatively little data on its clinical outcomes and effectiveness in specific patient populations (9).

Early enthusiasm for DBT is based mostly on retrospective reader studies that compared DBT combined with digital mammography (DM) imaging vs DM alone. These studies demonstrated reductions of up to 30% to 40% in false-positives, with similar or slightly improved cancer detection (6,10–12). In the United States, Food and Drug Administration (FDA) approval was granted to a single vendor in 2011 (Hologic, Inc., Bedford, MA) based on a large multicenter retrospective reader study that demonstrated statistically significant improvement in performance when DBT was combined with DM (13,14).

More recently, two prospective European trials have also shown improvements in screening outcomes with DBT. An interim analysis from the Oslo Screening Trial (n = 12631) and the final results from the Italian Screening with Tomosynthesis or Standard Mammography Trial (STORM, n = 7292) have demonstrated reductions in recalls of 15% to 17% and improvements in cancer detection of 33% to 53% (15,16). Furthermore, in the Oslo trial there was a 40% increase in the detection of invasive cancers with a stable rate of in situ cancer detection. However, in both of these trials subject compliance with screening invitations was less than 100%, and there were complex reading protocols requiring at least two readers per case, which is uncommon in clinical practices in the United States.

More recently, two separate US centers have reported early results from DBT screening (17,18). These data again demonstrate improvements in outcomes with reductions in recalls up to 37% and increases in cancer detection up to 35%. At both sites, however, there was concurrent screening with DM alone and, therefore, a potential for bias in the selection of patients screened with DBT and potentially imaged with DBT at recall. In a recent report from a consortium of 13 US practices, a 15% reduction in recall rate and a 29% increase in cancer detection were seen with DBT screening compared with DM alone screening. However, no patient-level data was reported, and eleven of the thirteen sites had concurrent DM screening, leading to possible biases in selection of patients for DBT (19).

Here we report the patient-level outcomes of implementing DBT screening for the entire screening population at our institution beginning October 2011. We compare outcomes for the cohort screened with DBT over a period of 17 months to the cohort screened with DM alone during the 12 months prior to DBT implementation. All women presenting for routine screening were imaged with DBT after its implementation, and radiologist readers remained the same over the DM and DBT cohorts, providing a "pre and post" comparison of DBT to DM-alone screening.

Methods

Study Population

The institutional review board waived the requirement to obtain written consent for this Health Insurance Portability and Accountability Act–compliant observational study. All patients presenting for breast cancer screening at our institution in the 17-month period from October 1, 2011 to February 28, 2013 (n = 15 571) were imaged with DBT per the current FDA-approved protocol (Dimension, Hologic, Inc., Bedford, MA), consisting of two-view DBT and two-view DM of each breast. Henceforth, for brevity we will refer to this combined protocol as DBT. The comparison cohort was women imaged with DM alone from September 1, 2010 to August 30, 2011 (n = 10 728). The two time periods were defined so that the same group of six radiologists interpreted all screening studies, and there were at least 12 months of follow-up in electronic medical records for the DBT cohort. Since DBT screening was implemented during the middle of September 2011, this month was excluded from the analysis. Institutional cancer registry data were available up to six months after the end of screening of the DBT cohort. For both cohorts, screening was defined as imaging performed on women having no prior history of breast cancer and no clinical signs or symptoms of breast cancer, such as a new palpable lump or nipple discharge. We included women with breast implants and women with large breasts that required extra imaging or "tiling" of conventional views to optimally image their breasts.

Screening Interpretation and Data Collection

All studies were interpreted by one of six board-certified radiologists with three to 22 years of breast imaging experience. Reading assignments were according to clinical schedule and individual reader volumes varied by such assignments. All readers were formally trained in DBT interpretation per FDA guidelines. The same group of six radiologists also interpreted the entire volume of screening mammography studies during the prior year of DM screening.

The Report Information System (RIS) (GE Centricity, Milwaukee, WI) database was queried for all screening mammograms in the two time periods. Population demographics (age, race/ethnicity, body mass index (BMI), presence of prior mammographic studies for comparison), screening volumes, and imaging outcomes were obtained from the same database. Race/ethnicity was defined based on patient self-classification into the following categories: white/Caucasian, black/African American, Hispanic, Asian, Native American, or other race. Women with missing race data were included in the "other" category. For analysis, Hispanic, Asian, and Native American women were placed in the "other" category because of the small numbers in these groups. Race/ethnicity was assessed because breast cancer risk is known to differ for various groups. For a random sample of patients (n = 17 020), data on established breast cancer risk factors (reproductive factors, prior biopsy, hormone replacement, family history, BMI, etc.) were manually abstracted from hard-copy questionnaires completed at the time of screening (Supplementary Table 1, available online).

In addition, all screening reports were queried for breast density according to the Breast Imaging Reporting and Data System categories (BIRADS) 1 through 4: almost entirely fatty, scattered fibroglandular densities, heterogeneously dense, and extremely dense (20). For statistical analysis, breast density was dichotomized as nondense (BIRADS density categories 1 & 2) vs dense (BIRADS density categories 3 and 4).

Main Outcome Measures

Recalled cases were defined as patients called back from screening for further evaluation, a BIRADS assessment category 0 (additional imaging needed), or those with a final assessment of BIRADS category 4 (suspicious abnormality), or 5 (highly suggestive of malignancy). Cancer detection was determined from the outcomes of biopsies occurring within 180 days from screening recall. Outcomes of biopsies and/or definitive surgeries were obtained through queries of the RIS, the electronic medical record, and the pathology database through February 2014. Pathology records were abstracted for invasive cancer vs ductal carcinoma in situ (DCIS). Cancer outcomes were cross-checked with the institutional tumor registry, which was updated through September 2013, and with the Pennsylvania State Cancer Registry through June 2012, the latest date for which records were available. Positive predictive values were calculated as follows: PPV1, the number of cancers per number of recalls; PPV2, the number of cancers per biopsy recommended; and PPV3, the number of cancers per biopsy performed (20).

Statistical Analysis

We compared baseline characteristics and screening outcomes for the DBT and DM cohorts using the *t* test for continuous variables, the chi-squared test or Fisher's exact test for categorical variables; we estimated the effects of screening modality type on recall rate using logistic regression. Many women had multiple screening exams: there were 26 299 screening events among 18 220 women. Thirty-five percent of women (n = 6381) had both a DM screen and a DBT screen, 9% (n = 1693) had two DBT screens, and a

handful (n = 5) had two DM screens. Multiple screening exams were accounted for using generalized linear mixed-effects models (SAS PROC GLIMMIX) with individual as the unit of analysis. An identifier for individual patient was included as a G-side random effect to account for within-person correlation. Analyses were performed by unadjusted models and by multivariable models adjusting for age, race, presence of prior mammography films at the time of interpretation, breast density, and radiologist reader. Analysis was repeated adjusting for additional risk factors in the subset of women with available data. All statistical tests were two-sided and were performed using STATA Version 12 (StataCorp LP, College Station, TX) or SAS Version 9.3 (SAS Institute, Inc., Cary, NC). A *P* value of less than .05 was considered statistically significant.

Results

The DM and DBT cohorts had similar age distributions (Table 1). Compared with DM, women screened with DBT had a slightly lower proportion of black patients, a slightly higher proportion of women with lowest breast density (almost entirely fatty), and a slightly lower proportion of women with prior mammograms available. Table 2 compares screening outcomes for DM and DBT. The recall rate was 10.4% (95% CI = 9.8% to 10.9%) for DM compared with 8.8% (95% CI = 8.3% to 9.2%) for DBT ($P < .001$), representing a 15% reduction or 16 fewer recalls per 1000 screened with DBT. Biopsy rates were similar for DM and DBT (1.8%, 95% CI = 1.5% to 2.0% vs 2.0%, 95% CI = 1.8% to 2.2%, $P = .14$) as was cancer yield of both biopsies recommended and those actually performed (PPV2 and PPV3). PPV1 increased by 45% for DBT compared with DM (6.2%, 95% CI = 4.9% to 7.5% vs 4.4%, 95% CI = 3.2% to 5.6%, $P = .047$).

The cancer detection rate was 4.6 per 1000 (95% CI = 3.3 to 5.8 per 1000) for DM compared with 5.5 per 1000 (95% CI = 4.3 to 6.6 per 1000) for DBT, an absolute difference of 0.9 per 1000 screened, though this difference was not statistically significant ($P = .32$) (Table 3). For the DM cohort, 69% of cancers diagnosed were invasive and 32% were DCIS. Similarly, in the DBT cohort, 71% of cancers were invasive and 27% were DCIS.

Table 1. Patient characteristics, digital mammography vs digital breast tomosynthesis

Characteristic	Cohort 1: DM n = 10728	Cohort 2: DBT n = 15571	*P**
Age, mean (SD), y	56.9 (11.0)	56.7 (11.0)	.23
Age categories, no. (%), y			
<40	254 (2.4)	366 (2.4)	
40–49	2925 (27.3)	4365 (28.0)	
50–59	3563 (33.2)	5035 (32.3)	.27
60–69	2538 (23.7)	3783 (24.3)	
≥70	1448 (13.5)	2022 (13.0)	
Race, no. (%)			
White	4360 (40.6)	6329 (40.7)	
Black	5473 (51.0)	7822 (50.2)	
Hispanic	92 (0.9)	148 (1.0)	.005
Asian	357 (3.3)	477 (3.1)	
Other/unknown	446 (4.2)	795 (5.1)	
Breast density, no. (%)			
1 Almost entirely fatty	1149 (10.7)	1861 (12.0)	
2 Scattered fibroglandular densities	6090 (56.8)	8654 (55.6)	.02
3 Heterogeneously dense	3287 (30.6)	4752 (30.5)	
4 Extremely dense	202 (1.9)	304 (2.0)	
Prior mammogram, no. (%)	9524 (88.8)	13712 (88.1)	.08

* Student's *t* test was used for continuous variables, and the chi-squared test for categorical variables. All tests were two-sided. DBT = digital breast tomosynthesis; DM = digital mammography.

Table 2. Screening outcomes for digital mammography and digital breast tomosynthesis

Metric	Cohort 1: DM (95% CI) n = 10728	Cohort 2: DBT (95% CI) n = 15571	Difference per 1000 screens	*P**
Recall	1112	1366		
Recall, %	10.4 (9.8 to 10.9)	8.8 (8.3 to 9.2)	-16	<.001
Biopsy performed	190	315		
Biopsy performed, %	1.8 (1.5 to 2.0)	2.0 (1.8 to 2.2)	2	.14
Cancers detected	49	85		
Cancers per 1000 screened	4.6 (3.3 to 5.8)	5.5 (4.3 to 6.6)	0.9	.32
PPV1 (cancers/recall), %	4.4 (3.2 to 5.6)	6.2 (4.9 to 7.5)	—	.047
PPV2 (cancers/biopsy recommended), %	22.4 (16.8 to 28.0)	24.7 (20.1 to 29.3)	—	.54
PPV3 (cancers/biopsy performed), %	24.7 (18.6 to 30.9)	25.4 (20.6 to 30.2)	—	.87

* *P* values were estimated using the Chi-square test. All tests were two-sided. CI = confidence interval; DBT = digital breast tomosynthesis; DM = digital mammography.

Table 3. Cancer detection rates by type of cancer and screening method

Outcome	Cohort 1: DM n = 10728	Cohort 2: DBT n = 15571	Difference per 1000 screens	% Difference	P*
Total cancers†	49	85			
Rate per 1000 screened (95% CI)	4.6 (3.3 to 5.8)	5.5 (4.3 to 6.6)	0.9	19.6	.32
Invasive cancers					
N (% of total cancers)	34 (69)	60 (71)			
Rate per 1000 screened (95% CI)	3.2 (2.1 to 4.2)	3.9 (2.9 to 4.8)	0.7	21.9	.36
Ductal carcinoma in situ					
N (% of total cancers)	15 (32)	23 (27)			
Rate per 1000 screened (95% CI)	1.4 (0.7 to 2.1)	1.5 (0.8 to 2.1)	0.1	7.1	.87

* *P* values were estimated using the Chi-square test. All tests were two-sided. CI = confidence interval; DBT = digital breast tomosynthesis; DM = digital mammography.

† Includes one lymphoma and one lung metastasis in DBT cohort, not counted as invasive or ductal carcinoma in situ.

Table 4. Association of screening methods, age, race, breast density, and prior mammogram with odds of recall at screening*

Characteristic	Unadjusted† n = 26299		Multivariable 1‡ n = 26299		Multivariable 2§ n = 17020	
	OR (95% CI)	P‖	OR (95% CI)	P	OR (95% CI)	P
Screening modality						
DM	1.00 (referent)		1.00 (referent)		1.00 (referent)	
DBT	0.82 (0.75 to 0.89)	<.001	0.80 (0.74 to 0.88)	<.001	0.77 (0.68 to 0.87)	<.001
Age, y		<.001		<.001		.002
<40	1.92 (1.41 to 2.62)	<.001	1.14 (0.83 to 1.57)	.43	0.78 (0.49 to 1.23)	.28
40–49	2.45 (2.07 to 2.90)	<.001	1.85 (1.56 to 2.19)	<.001	1.45 (1.12 to 1.88)	.006
50–59	1.56 (1.32 to 1.84)	<.001	1.41 (1.19 to 1.67)	<.001	1.23 (0.99 to 1.53)	.07
60–69	1.25 (1.05 to 1.50)	.01	1.20 (1.00 to 1.43)	.048	1.10 (0.87 to 1.37)	.43
≥70	1.00 (referent)		1.00 (referent)		1.00 (referent)	
Race		.03		.01		.21
White	1.00 (referent)		1.00 (referent)		1.00 (referent)	
Black	1.14 (1.03 to 1.25)	.008	1.14 (1.04 to 1.26)	.008	1.11 (0.95 to 1.30)	.17
Other	1.12 (0.94 to 1.32)	.20	0.97 (0.82 to 1.14)	.70	0.93 (0.74 to 1.17)	.54
Prior mammogram						
Yes	1.00 (referent)		1.00 (referent)		1.00 (referent)	
No	2.51 (2.24 to 2.82)	<.001	2.31 (2.04 to 2.61)	<.001	2.58 (2.17 to 3.06)	<.001
Breast density		<.001		<.001		<.001
Almost entirely fatty	1.00 (referent)		1.00 (referent)		1.00 (referent)	
Scattered fibroglandular densities	1.78 (1.49 to 2.13)	<.001	1.66 (1.39 to 1.99)	<.001	1.63 (1.30 to 2.06)	<.001
Heterogeneously dense	2.46 (2.05 to 2.96)	<.001	2.11 (1.75 to 2.55)	<.001	2.10 (1.63 to 2.71)	<.001
Extremely dense	1.73 (1.20 to 2.52)	.004	1.41 (0.97 to 2.06)	.07	1.43 (0.87 to 2.36)	.16

* n = number of unique screening events. CI = confidence interval; DBT = digital breast tomosynthesis; DM = digital mammography.

† Mixed-effects models with random effect for individual.

‡ Mixed-effects model including random effect for individual, adjusted for age, race, prior mammogram, breast density, and radiologist.

§ Mixed-effects model including random effect for individual, age, race, prior mammogram, breast density, radiologist, prior atypical hyperplasia, prior biopsy, age at menarche, age at first birth, first degree family history of breast or ovarian cancer, Jewish ancestry, menopause status, hormone replacement therapy use, and body mass index.

‖ Overall *P* values from type-3 F-tests, *P* values for individual categories from *t* tests for fixed effects parameters. All tests were two-sided.

Table 4 displays the results of logistic regression for odds of recall by screening modality. In the unadjusted model, the odds of recall were 18% lower for DBT compared with DM (OR = 0.82, 95% CI = 0.75 to 0.89, $P < .001$). Younger age, absence of a prior mammogram, and higher breast density were associated with increased odds of recall in univariate analyses. In addition, black women had 14% higher likelihood of recall compared with white women (OR = 1.14, 95% CI = 1.03 to 1.25, $P = .008$). The likelihood of a recall was 20% lower for the DBT group than the DM group after adjusting for age, race, presence of prior mammograms, breast density, and radiologist (OR = 0.80, 95% CI = 0.74 to 0.88, $P < .001$). Age, race, presence of prior mammogram, and breast density remained statistically significantly associated with recall in the multivariable model. There was no statistically significant interaction between age (<50 years vs ≥50 years) and intervention (DM vs DBT) for recall. When adjusted for reproductive factors, prior breast biopsies, family history of breast or ovarian cancer, hormone replacement therapy, and BMI for the subset of women

Table 5. Screening outcomes for digital mammography and digital breast tomosynthesis by breast density and age

Characteristic	Total, n	Recall, n	Recall, % (95% CI)	Cancers, n	Cancers per 1000 screened (95% CI)	Invasive cancers, n	Invasive cancers per 1000 screened (95% CI)	DCIS, n	DCIS per 1000 screened (95% CI)	PPV1 cancers/ callback (95% CI)
By breast density*										
Nondense										
DM	7239	667	9.2 (8.5 to 9.9)	31	4.3 (2.8 to 5.8)	22	3.0 (1.8 to 4.3)	9	1.2 (0.4 to 2.1)	4.7 (3.0 to 6.2)
DBT	10515	819	7.8 (7.3 to 8.3)	50	4.8 (3.4 to 6.1)	36	3.4 (2.3 to 4.5)	13	1.2 (0.6 to 1.9)	6.1 (4.5 to 7.7)
P†			.001		.73		.69		.99	.25
Dense										
DM	3489	445	12.8 (11.6 to 13.9)	18	5.2 (2.8 to 7.5)	12	3.4 (1.5 to 5.4)	6	1.7 (0.3 to 3.1)	4.0 (0.2 to 5.9)
DBT	5056	547	10.8 (10.0 to 11.7)	35	6.9 (4.6 to 9.2)	24	4.7 (2.9 to 6.6)	10	2.0 (0.8 to 3.2)	6.4 (4.3 to 8.5)
P†			.006		.33		.40		.99	.12
By age <50, y										
DM	3179	445	14.0 (12.8 to 15.2)	7	2.2 (0.6 to 3.8)	4	1.3 (0.03 to 2.5)	3	0.9 (-0.1 to 2.0)	1.6 (0.4 to 2.7)
DBT	4731	580	12.3 (11.3 to 13.2)	27	5.7 (3.5 to 7.9)	17	3.6 (1.9 to 5.3)	9	1.9 (0.7 to 3.1)	4.7 (2.9 to 6.4)
P†			.02		.022		.072		.38	.007
≥50, y										
DM	7549	667	8.8 (8.2 to 9.5)	42	5.6 (3.9 to 7.2)	30	4.0 (2.6 to 5.4)	12	1.6 (0.7 to 2.5)	6.3 (4.4 to 8.1)
DBT	10840	786	7.3 (6.8 to 7.7)	58	5.4 (4.0 to 6.7)	43	4.0 (2.8 to 5.2)	14	1.3 (0.6 to 2.0)	7.4 (5.5 to 9.2)
P†			<.001		.84		.99		.69	.47

* Nondense BIRADS density categories 1 and 2, Dense BIRADS density categories 3 and 4. CI = confidence interval; DBT = digital breast tomosynthesis; DCIS = ductal carcinoma in situ; DM = digital mammography.

† P values from Chi Square test for Recall, Fisher's Exact test for total cancers, invasive cancers, DCIS, and PPV1. All tests were two-sided.

with available data, the DBT group had a 23% lower recall rate compared with DM (OR = 0.77, 95% CI = 0.68 to 0.87, $P < .001$).

We also examined screening outcomes stratified by breast density and age at screening (Table 5). The magnitude of the reduction in recall rate was similar for women with dense and nondense breasts. Among women with nondense breasts, the recall rate was 9.2% (95% CI = 8.5 to 9.9%) for DM compared with 7.8% (95% CI = 7.3% to 8.3%) for DBT ($P = .001$). Similarly, among women with dense breasts, the recall rate was lower for DBT compared with DM (10.8%, 95% CI = 10.0% to 11.7% vs 12.8%, 95% CI = 11.6% to 13.9%, $P = .006$). For women with dense breasts, the cancer detection rate was 6.9 per 1000 (95% CI = 4.6 to 9.2 per 1000) with DBT compared with 5.2 per 1000 (95% CI = 2.8 to 7.5 per 1000) with DM screening, a 33% increase ($P = .33$). The PPV1 was not statistically significantly different for DBT and DM when analyzed by subgroups of women with nondense (DBT 6.1%, 95% CI = 4.5% to 7.7% vs DM 4.7%, 95% CI = 3.0% to 6.2%, $P = .22$) and dense breasts (DBT 6.4%, 95% CI = 4.3 to 8.5 vs DM 4.0%, 95% CI = 0.2 to 5.9, $P = .10$). Screening outcomes for all four density categories are displayed in Supplementary Table 2 (available online).

When analyses were performed stratified by age at screen, DBT had a statistically significantly lower recall rate irrespective of age; however, the impact was greater in women aged 50 years and older. In women younger than 50 years, the recall rate for DM was 14.0% (95% CI = 12.8 to 15.2) vs 12.3% (95% CI = 11.3 to 13.2) for DBT ($P = .02$), a reduction of 12% (Supplementary Table 2, available online) or 17 fewer women recalled per 1000 screened. For women aged 50 years and older, the recall rate for DM was 8.8% (95% CI = 8.2 to 9.5) vs 7.3% (95% CI = 6.8 to 7.7) for DBT, a 17% reduction ($P < .001$) (Table 5). Notably, the cancer detection rate nearly tripled with DBT screening compared with DM in women under age 50 years, equating to an additional 3.6 cancers detected per 1000 screened (2.2 vs 5.7, $P = .02$). Though the number of cancer cases was small in the invasive and DCIS subgroups, in women younger than 50, the invasive cancers detection rate was 3.6 per 1000 (95% CI = 1.9 to 5.3 per 1000) for DBT compared with 1.3 per 1000 (95% CI = 0.03 to 2.5 per 1000) for DM ($P = .072$). There was no difference in DCIS detection for DBT compared with DM in women younger than 50 years, though the number of DCIS cases was small (0.9 vs 1.9 per 1000, $P = .28$). The PPV1 tripled for DBT screening for women under age 50 years, increasing from 1.6% for DM to 4.7% with DBT ($P = .007$). Screening outcomes for smaller categories of age are displayed in Supplementary Table 2 (available online).

Discussion

In our experience with an entire site's population exposed to DBT screening, we observed a statistically significant reduction in recall rates and an increase in cancers detected per recall (PPV1) for DBT compared with DM screening. In multivariable analysis, adjusted for age, race, prior mammogram, radiologist, and breast density, the recall rate was 20% lower for DBT, translating to 16 fewer recalls per 1000 women screened. This reduction in recall with DBT is understandably less than that demonstrated in earlier reader studies, which were enriched with

cancer cases, therefore creating an artificially high proportion of positive cases. In addition, our recall reduction is less than that demonstrated in the two recent US studies, where there may have been a selection bias in who was screened with DBT. Our reduction in recall of approximately 20% is similar to the results of the larger prospective Oslo screening trial (16) and the site-level data recently reported from 13 US screening facilities (19), suggesting that similar results can be achieved with routine clinical implementation of DBT.

In our study, the recall reduction with DBT was independent of density, with similar reductions for women with dense and nondense breasts. While there was no statistically significant difference in cancer detection between DBT and DM overall, there was a statistically significant increase in cancer detection among women under age 50 years who were screened with DBT. This has not been reported in the other trials. Some of the previous studies were smaller, with fewer cancers detected, making a rigorous statistical evaluation by age difficult. Though the increase in cancer detection with DBT was not statistically significant in women with dense breasts, the absolute increase was greater for women with dense breasts than for women with lower density breasts. This trend combined with the statistically significant improvements in cancer detection for women under age 50 years is similar to the improved cancer detection rate observed for both younger and denser-breasted women in the Digital Mammography Imaging Screening Trial comparing DM with film mammography (21). Our results of fewer recalls coupled with improvements in cancer detection supports earlier evidence that DBT is a promising new mammography screening technique, particularly for the controversial subgroup of women screened under the age of 50 years.

Several limitations should be considered when interpreting our findings. First, assignment to screening modality was not randomized. However, our institution transitioned from entirely DM to entirely DBT screening in a very short time frame, and therefore there was no possibility of selecting screening modality based on patient characteristics, as evidenced by the similarity of our two comparison cohorts. To further account for potential underlying systemic population differences, we also performed multivariable analysis of recall adjusted for age, race, prior mammogram, and breast density. We additionally adjusted for breast cancer risk factors such as family history of cancer, BMI, and reproductive risk factors in a subset of women for whom data were available and results were unchanged. We chose our two study periods based on a stable pool of radiologists interpreting all screening studies in both cohorts and to ensure at least 12 months of follow-up for all patients. Specifically for DBT, the 17 months of accrual were chosen so that a steady state of recall would be reached with this new technology (22). A steady state of recall and cancer detection had already been established for the six readers with DM screening, hence only one year of sampling. However, given the relatively short follow-up period, particularly for DBT, we were not able to link the entire study population to the state cancer registry, and therefore we were unable to fully evaluate sensitivity, specificity, and false-negative rates for DBT compared with DM; this analysis is ongoing.

The interpretation of our findings regarding cancer detection with DBT should be treated with caution. Our study was not powered to detect statistically significant changes in cancer detection, especially for subgroup analysis. As such, while our observed increase in cancer detection in women under age 50 years is interesting, it should not be treated as fully conclusive. On the other hand, this observation may be confounded by the effect of breast density, which is greater in younger women. Unfortunately, we did not have a large enough sample size to investigate this effect. Therefore, considering these limitations, our findings should be further investigated and replicated with additional, larger prospective trials.

Despite these limitations, our study is currently the largest to evaluate the clinical implementation of DBT in the screening of an entire clinical practice. We have shown that DBT screening can result in statistically significantly reduced recalls while also increasing the detection rate of breast cancers. These findings are particularly pronounced in women under the age of 50 years, potentially tipping the risk-benefit ratio towards incorporating tomosynthesis in screening for this group.

References

1. U.S. Preventative Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2009;151(10):716–726, W236.
2. Nelson HD, Tyne K, Naik A, et al. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2009;151(10):727–737, W237-W242.
3. Sechopoulos I. A review of breast tomosynthesis. Part I. The image acquisition process. *Med Phys*. 2013;40(1):014301.
4. Sechopoulos I. A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications. *Med Phys*. 2013;40(1):014302.
5. Niklason LT, Christian BT, Niklason LE, et al. Digital tomosynthesis in breast imaging. *Radiology*. 1997;205(2):399–406.
6. Baker JA, Lo JY. Breast tomosynthesis: state-of-the-art and review of the literature. *Acad Radiol*. 2011;18(10):1298–1310.
7. Helvie MA. Digital mammography imaging: breast tomosynthesis and advanced applications. *Radiol Clin North Am*. 2010;48(5):917–929.
8. Gennaro G, Toledano A, di Maggio C, et al. Digital breast tomosynthesis versus digital mammography: a clinical performance study. *Eur Radiol*. 2010;20(7):1545–1553.
9. Lee CI, Lehman CD. Digital breast tomosynthesis and the challenges of implementing an emerging breast cancer screening technology into clinical practice. *J Am Coll Radiol*. 2013;10(12):913–917.
10. Gur D, Abrams GS, Chough DM, et al. Digital breast tomosynthesis: observer performance study. *AJR Am J Roentgenol*. 2009;193(2):586–591.
11. Poplack SP, Tosteson TD, Kogel CA, et al. Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. *AJR Am J Roentgenol*. 2007;189(3):616–623.
12. Rafferty EA. Digital mammography: novel applications. *Radiol Clin North Am*. 2007;45(5):831–843, vii.
13. US Food and Drug Administration. *Summary of safety and effectiveness data*. Available at:www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/RadiologicalDevicesPanel/UCM324866.pdf. Accessed May 15, 2014.
14. Rafferty EA, Park JM, Philpotts LE, et al. Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial. *Radiology*. 2013;266(1):104–113.
15. 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(7):583–589.
16. 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(1):47–56.
17. Rose SL, Tidwell AL, Bujnoch LJ, et al. Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR Am J Roentgenol*. 2013;200(6):1401–1408.
18. Haas BM, Kalra V, Geisel J, et al. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology*. 2013;269(3):694–700.
19. Friedwald AM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA*. 2014;311(24):2499–2507.
20. D'Orsi CJ, Bassett LW, Berg WA, et al. BI-RADS: Mammography, 4th edition. In: D'Orsi CJ, Mendelson EB, Ikeda DM, et al: *Breast Imaging Reporting and Data System: ACR BI-RADS – Breast Imaging Atlas*. Reston, VA: American College of Radiology; 2003.
21. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med*. 2005;353(17):1773–1783.
22. Dang PA, Freer PE, Humphrey KL, et al. Addition of tomosynthesis to conventional digital mammography: effect on image interpretation time of screening examinations. *Radiology*. 2014;270(1):49–56

Funding

This work was supported by a U54 grant from the National Cancer Institute at the National Institutes of Health: Population-based Research Optimizing Screening through Personalized Regimens (PROSPR) Network (U54CA163313).

Notes

The study funder had no role in the design of the study, the collection, analysis, or interpretation of the data, the writing of the manuscript, nor the decision to submit the manuscript for publication.
Special thanks to Ms. Kathy Finnegan for help in data collecting.

Affiliations of authors: Division of General Internal Medicine (AMM, MS), Department of Radiology (DK, MS, EFC), and Department of Biostatistics and Epidemiology (KST, DFH), University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; General Medicine Division, Massachusetts General Hospital and Harvard Medical School, Boston, MA (AMM).