

## Original Investigation

# Effectiveness of Digital Breast Tomosynthesis Compared With Digital Mammography Outcomes Analysis From 3 Years of Breast Cancer Screening

Elizabeth S. McDonald, MD, PhD; Andrew Oustimov, MPH; Susan P. Weinstein, MD; Marie B. Synnestvedt, PhD; Mitchell Schnall, MD, PhD; Emily F. Conant, MD

**IMPORTANCE** Breast cancer screening with digital breast tomosynthesis (DBT) combined with digital mammography (DM) decreases false-positive examinations and increases cancer detection compared with screening with DM alone. However, the longitudinal performance of DBT screening is unknown.

**OBJECTIVES** To determine whether the improved outcomes observed after initial implementation of DBT screening are sustainable over time at a population level and to evaluate the effect of more than 1 DBT screening at the individual level.

**DESIGN, SETTING, AND PARTICIPANTS** Retrospective analysis of screening mammography metrics was performed for all patients presenting for screening mammography in an urban, academic breast center during 4 consecutive years (DM, year 0; DBT, years, 1-3). The study was conducted from September 1, 2010, to September 30, 2014 (excluding September 2011, which was the transition period from DM to DBT), for a total of 44 468 screening events attributable to a total of 23 958 unique women. Differences in screening outcomes between each DBT year and the DM year, as well as between groups of women with only 1, 2, or 3 DBT screenings, were assessed, and the odds of recall adjusted for age, race/ethnicity, breast density, and prior mammograms were estimated. Data analysis was performed between February 16 and October 26, 2015.

**EXPOSURE** Digital mammography screening supplemented with DBT.

**MAIN OUTCOMES AND MEASURES** Recall rates, cancer cases per recalled patients, and biopsy and interval cancer rates were determined.

**RESULTS** Screening outcome metrics were evaluated for a total of 44 468 examinations attributable to 23 958 unique women (mean [SD] age, 56.8 [11.0] years) over a 4-year period: year 0 cohort (DM0), 10 728 women; year 1 cohort (DBT1), 11 007; year 2 cohort (DBT2), 11 157; and year 3 cohort (DBT3), 11 576. Recall rates rose slightly for years 1 to 3 of DBT (88, 90, and 92 per 1000 screened, respectively) but remained significantly reduced compared with the DMO rate of 104 per 1000 screened. Reported as odds ratios (95% CIs), the findings were DM vs DBT1, 0.83 (0.76-0.91,  $P < .001$ ); DM vs DBT2, 0.85 (0.78-0.93,  $P < .001$ ); and DM vs DBT3, 0.87 (0.80-0.95,  $P = .003$ ). The cancer cases per recalled patients continued to rise from DMO rate of 4.4% to 6.2% ( $P = .06$ ), 6.5% ( $P = .03$ ), and 6.7% ( $P = .02$ ) for years 1 to 3 of DBT, respectively. Outcomes assessed for the most recent screening for individual women undergoing only 1, 2, or 3 DBT screenings during the study period demonstrated decreasing recall rates of 130, 78, and 59 per 1000 screened, respectively ( $P < .001$ ). Interval cancer rates, determined using available follow-up data, decreased from 0.7 per 1000 women screened with the use of DM to 0.5 per 1000 screened with the use of DBT1.

**CONCLUSIONS AND RELEVANCE** Digital breast tomosynthesis screening outcomes are sustainable, with significant recall reduction, increasing cancer cases per recalled patients, and a decline in interval cancers.

← Editorial

+ Supplemental content at [jamaoncology.com](http://jamaoncology.com)

**Author Affiliations:** Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia (McDonald, Weinstein, Schnall, Conant); Computational Breast Imaging Group, University of Pennsylvania, Philadelphia (Oustimov); Division of General Internal Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Synnestvedt).

**Corresponding Author:** Emily F. Conant, MD, Department of Radiology, Perelman School of Medicine, University of Pennsylvania, 3400 Spruce St, Philadelphia, PA 19104 ([emily.conant@uphs.upenn.edu](mailto:emily.conant@uphs.upenn.edu)).

*JAMA Oncol.* doi:10.1001/jamaoncol.2015.5536  
Published online February 18, 2016. Corrected on March 17, 2016.

There is growing evidence that screening women with digital breast tomosynthesis (DBT) in addition to digital mammography (DM) leads to an increase in cancer detection<sup>1-5</sup> and reduction in women recalled for additional imaging.<sup>1-8</sup> However, similar to any new technology, the earliest adoption of DBT was based on enriched reader studies and small, single-site retrospective studies.<sup>9-14</sup> More recently, improved screening outcomes have been replicated in a large, multisite retrospective US study<sup>15</sup> and in 3 prospective European trials.<sup>5,16,17</sup> The multisite study demonstrated a 16% recall reduction and a 41% increase in invasive cancer detection with DBT screening compared with screening with DM alone. However, despite encouraging initial outcomes, there are few data from consecutive years of DBT screening. Specifically, the sustainability of cancer detection, recall rates, and the rate of false-negative results in consecutive years after implementation of DBT screening is unknown.

The issues involved include whether the increased specificity and sensitivity demonstrated after implementation of DBT screening are sustainable or whether the benefits will occur only in the first round of DBT screening. There is evidence that DBT has additional benefit in the baseline subset of patients without prior DMs for comparison.<sup>18</sup> In addition, the possibility of reduced performance for patients with prior DBT examinations is unclear. The largest performance change is often seen after the introduction of a new technology, and some hypothesize that cancer detection may return to baseline levels with repeated DBT examinations.<sup>19</sup>

The goals of this study were to determine whether the improved outcomes of DBT screening are sustainable in a natural experiment incorporating nearly 45 000 routine screening examinations. Outcome data from 3 years of DBT screening of an entire population at a large, urban academic practice were evaluated at a population level (all patients presenting for screening) and at the individual level (patients with only a single round, only 2 rounds, and 3 rounds of DBT screening).

## Methods

### Study Population

The University of Pennsylvania institutional review board approved the retrospective analysis of screening mammography examinations performed during the year prior and the 3 consecutive years after complete practice conversion from DM to DBT screening in September 2011. The study population consisted of all women undergoing screening mammography at our institution from September 1, 2010, to September 30, 2014, excluding the month of DBT transition (September 2011) for a total of 44 468 screening events attributable to 23 958 unique women.

The patients included had no history or clinical symptoms of breast cancer. Those presenting for breast cancer screening from September 1, 2010, to August 30, 2011, underwent imaging with DM alone (Dimensions; Hologic Inc). Patients presenting for breast cancer screening from October 1, 2011, to September 30, 2014, received imaging with DBT (Dimension, Hologic), in accordance with the then current US Food and Drug Administration-approved protocol consisting of 2-view DM and 2-view

### Key Points

**Question:** Can it be determined whether improved outcomes after implementation of digital breast tomosynthesis (DBT) compared with digital mammography are sustainable at both the population and individual levels?

**Findings:** In this study of screening mammography metrics, DBT screening outcomes were sustainable, with significant recall reduction, increasing cancer cases per recalled patients, and a decline in interval cancers.

**Meaning:** Sustained and even improved performance is possible with consecutive DBT screening, which is an important initial step toward informing policies for possibly integrating this technology into population-screening programs.

DBT examination of each breast. Screening volumes remained stable during the time of the study: year 0 cohort (DMO [n = 10 728]), year 1 cohort (DBT1 [n = 11 007]), year 2 cohort (DBT2 [n = 11 157]), and year 3 cohort (DBT3 [n = 11 576]).

### Screening Interpretation

All examinations were interpreted by 1 of 7 board-certified radiologists, (including E.S.M., S.P.W., and E.F.C.) with specialization in breast imaging ranging from 8 to 26 years (median, 17; mean, 16.5 years). Before implementation of DBT, all readers received the US Food and Drug Administration-mandated 8 hours of training in DBT interpretation. Individual reader volumes varied by the radiologists' clinical schedule. Five of 7 radiologists (including E.S.M., S.P.W., and E.F.C.) were involved for the entire study, accounting for interpretation of 37 691 (84%) of all imaging examinations.

### Data Collection

All screening mammograms were evaluated using structured reporting through the Report Information System (GE Centricity) and using American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) assessment categories.<sup>20</sup> Demographics, breast density, and BI-RADS categories were documented at the time of interpretation and retrieved for outcomes analysis. Breast density was characterized according to BI-RADS categories: (1) almost entirely fatty, (2) scattered fibroglandular densities, (3) heterogeneously dense, and (4) extremely dense.<sup>20</sup> During statistical analysis, breast density was classified into 2 groups: nondense (BI-RADS categories 1 and 2) and dense (BI-RADS categories 3 and 4). Race/ethnicity was defined in accordance with patient self-classification.

### Screening Outcome Measures

Imaging volumes, recall rate, and cancer detection rates per 1000 screened women were evaluated. The number of cancer cases per number of recalled patients to undergo biopsy (PPV<sub>1</sub>), the number of cancers per biopsy recommended (PPV<sub>2</sub>), and the number of cancers per biopsy performed (PPV<sub>3</sub>) were calculated.<sup>20</sup> Patients recalled from screening examinations were counted as those given a BI-RADS assessment category of 0 (incomplete; additional imaging needed), 4 (suspicious; biopsy recommended), or 5 (highly suspicious; biopsy recommended). The screening results of patients assigned to short-term follow-up

(BI-RADS assessment category 3) were considered normal. Surgical excisional or percutaneous biopsy results based on screening recommendations were evaluated within 12 months of the screening examination through the electronic medical record, pathology laboratory database, and the Report Information System. The Pennsylvania State Cancer Registry was queried, through June 24, 2014, to determine the interval cancer rate (defined as symptomatic cancers presenting within 1 year). To assess the effect of prevalence and incidence screening, we compared recall and cancer detection rates at the most recent screening event across women who were participating in their first, second, and third round of DBT screening.

### Statistical Analysis

Baseline characteristics by year were compared via analysis of variance for continuous variables and the  $\chi^2$  test for categorical variables. For the population-level analysis, we compared differences in screening outcomes (ie, recall, biopsy, and cancer detection rates as well as positive predictive values) across the 3 DBT years, as well as between each DBT year with the baseline DM year. This analysis was also performed on subgroups defined by breast density (nondense [BI-RADS 1 and 2] and dense [BI-RADS 3, 4]) and age (<50 years and  $\geq$ 50 years). Pearson  $\chi^2$  and Fisher exact tests were used to assess statistical significance of differences.

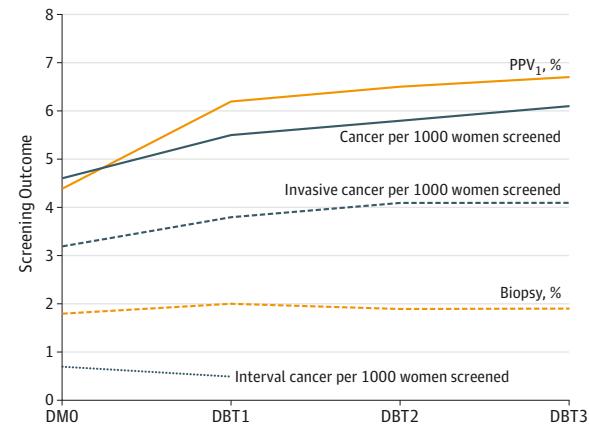
For the assessment of prevalence and incidence screening, we compared recall, cancer detection rates, and PPV<sub>1</sub> for groups of women undergoing only 2 and 3 DBT screenings, respectively, with those undergoing only 1 DBT screening. In addition, the 1-DBT screening group was further restricted to women who had prior DM screenings available. The Pearson  $\chi^2$  test was used to assess statistical significance.<sup>21</sup>

For the individual-level analysis of recall rates across 4 years of our study, we used generalized estimating equations,<sup>22</sup> with logistic link function, robust SE, and individual women as units of analysis. Parameters were created with this GEE model, focusing on estimating the main effects of each of the DBT years compared with DM, with screening year as a categorical variable (DM year as reference). The models were adjusted for age, race/ethnicity, breast density, and presence of a prior mammogram. A similar, but separate, generalized estimating equation model adjusted for race/ethnicity, breast density, and prior mammogram was fit to assess the interaction effect between breast density and screening year on the odds of recall. All statistical tests were 2-sided, and  $P < .05$  was considered statistically significant. The analyses were performed using SAS, version 9.4 (SAS Institute Inc). Data analysis was conducted from February 16 to October 26, 2015.

## Results

Among 44 468 examinations attributable to 23 958 unique women (mean [SD] age, 56.8 [11.0] years), there was no significant difference in patient characteristics including age, density, race/ethnicity, and screening volumes from year 0 to year 3. A previous study<sup>3</sup> demonstrated no statistically significant difference in calculated breast cancer risk between DM year 0

**Figure. Population-Level Analysis Comparing Screening Outcomes**



Analysis of screening outcomes. DMO indicates the reference digital mammography screening year; DBT1, DBT2, and DBT3, the consecutive years of screening with digital breast tomosynthesis; and PPV<sub>1</sub>, cancer cases per recalled patients.

and the first 18 months of screening with DBT. There was a slight increase in the number of patients without a previous mammogram for comparison over the study period (eTable 1 in the Supplement).

At the population level, recall rate, biopsies performed, cancer detection, and PPV 1 to 3 were compared between the DM cohort (year 0) and years 1 to 3 of DBT screening (Figure and eTable 2 in the Supplement). Recall rates rose slightly for years 1 to 3 of DBT (88, 90, and 92 per 1000 screened, respectively) but remained significantly reduced compared with the DMO rate of 104 per 1000. Reported as odds ratio (95% CI), the findings were DM vs DBT1, 0.83 (0.76-0.91,  $P < .001$ ); DM vs DBT2, 0.85 (0.78-0.93,  $P < .001$ ); and DM vs DBT3, 0.87 (0.80-0.95,  $P = .003$ ). There was no significant difference in recall across 3 DBT years ( $P = .55$ ). The rate of biopsies performed in each DBT year did not differ significantly from that of DM (DM vs DBT1, 1.05 [0.87-1.28],  $P = .17$ ; DM vs DBT2, 1.15 [0.94-1.39],  $P = .61$ ; and DM vs DBT3, 1.05 [0.86-1.29],  $P = .60$ ).

At the population level, the cancer detection rate continued to increase at 4.6, 5.5, 5.8, and 6.1 per 1000 women screened for years 0, 1, 2, and 3, respectively, but was not significantly different from the rate of DM (reported as OR [95% CI], DM vs DBT1, 1.35 [0.93-1.94],  $P = .37$ ; DM vs DBT2, 1.28 [0.88-1.85],  $P = .20$ ; and DM vs DBT3, 1.35 [0.93-1.94],  $P = .11$ ) and was not significantly different across 3 DBT years ( $P = .80$ ) (Figure and eTable 2 in the Supplement). The PPV<sub>1</sub> continued to rise from DMO rate of 4.4% to 6.2%, 6.5%, and 6.7% for years 1 to 3 of DBT and was significantly different from DM in the second and third DBT years (DM vs DBT1, 1.44 [0.98-2.12],  $P = .06$ ; DM vs DBT2, 1.51 [1.03-2.21],  $P = .03$ ; and DM vs DBT3, 1.56 [1.07-2.26],  $P = .02$ ). The PPV<sub>1</sub> was not significantly different across the 3 DBT years ( $P = .92$ ). The PPV<sub>2</sub> and PPV<sub>3</sub> were not significantly different from DM in any of the 3 DBT years and did not differ across the 3 DBT years (PPV<sub>2</sub>,  $P = .38$ ; PPV<sub>3</sub>,  $P = .37$ ). State cancer registry data for calculation of interval cancer rates were available only for DM and the first DBT years. The change in

Table. Odds of Recall Among Screened Patients

Characteristic	Unadjusted <sup>a</sup>		Adjusted <sup>a</sup>	
	OR (95% CI)	P Value	OR (95% CI)	P Value
Screening year				
DM0	1 [Reference]		1 [Reference]	
DBT1	0.83 (0.76-0.91)	<.001	0.81 (0.74-0.89)	<.001
DBT2	0.85 (0.78-0.93)	<.001	0.84 (0.77-0.92)	<.001
DBT3	0.87 (0.80-0.95)	.002	0.84 (0.77-0.92)	<.001
Age, y				
40	NA		1.24 (0.99-1.55)	.07
40-49	NA		1.73 (1.53-1.97)	<.001
50-59	NA		1.38 (1.22-1.56)	<.001
60-69	NA		1.22 (1.07-1.39)	.003
70	NA		1 [Reference]	
Race/ethnicity				
White	NA		1 [Reference]	
Black	NA		1.11 (1.03-1.19)	.004
Other	NA		0.88 (0.78-1.00)	.050
Breast density <sup>b</sup>				
BI-RADS 1, 2	NA		1 [Reference]	
BI-RADS 3, 4	NA		1.45 (1.35-1.56)	<.001
Prior mammogram				
Yes	NA		1 [Reference]	
No	NA		2.18 (2.01-2.37)	<.001
Interaction <sup>b</sup>				
DBT1, for BI-RADS 1, 2	NA		0.82 (0.73-0.93)	
DBT1, for BI-RADS 3, 4	NA		0.82 (0.75-0.95)	.99
DBT2, for BI-RADS 1, 2	NA		0.80 (0.71-0.90)	
DBT2, for BI-RADS 3, 4	NA		0.95 (0.83-1.09)	.07
DBT3, for BI-RADS 1, 2	NA		0.81 (0.72-0.91)	
DBT3, for BI-RADS 3, 4	NA		0.99 (0.86-1.13)	.03

interval cancer rates per 1000 women screened across these years (DM, 0.7; and DBT1, 0.5) was not statistically significant ( $P = .60$ ). Although the rate of invasive cancers detected per 1000 women screened increased slightly over time (DM, 3.2; DBT1, 3.8; DBT2, 4.1; and DBT3, 4.1), the increase in any DBT year compared with DM or across the DBT years was not significant. Cancer detection rates were compared in a similar manner in subgroups characterized by breast density (dense and nondense) and age (<50 and  $\geq 50$  years) (eTable 2 in the *Supplement*). The increase in cancer detection per 1000 women screened in the subgroup of women younger than 50 years between DM and the first DBT year was not significant (DM, 2.2 and DBT1, 5.0;  $P = .06$ ). Increases in cancer detection across the period were observed in the dense breast and 50 years or older subgroups but were not statistically significant.

To compare the odds of recall at the individual level, for each DBT screening year with DM, 3 generalized estimating equation models with the individual woman as the unit of analysis were used (2 main effects models: 1 with and 1 without adjustment for age, race/ethnicity, breast density, and prior mammogram) as well as an adjusted model containing terms to model interactions between screening year and breast density (Table). Results of the unadjusted main-effects model (reported as OR [95% CI]) indicate that the odds of recall re-

mained lower with DBT than with DM during 3 years (0.83 [0.76-0.91],  $P < .001$ ; 0.85 [0.78-0.93],  $P < .001$ ; and 0.87 [0.80-0.95],  $P = .002$  during years 1, 2, and 3, respectively). Results from the adjusted main-effects model similarly suggest that the odds of recall were lower with DBT than with DM (0.81 [0.74-0.89],  $P < .001$ ; 0.84 [0.77-0.92],  $P < .001$ ; and 0.84 [0.77-0.92],  $P < .001$ , in years 1, 2, and 3, respectively) and that the odds of recall were 2.18 times higher if no prior mammogram was available ( $P < .001$ ), higher in women aged 40 to 49 years (1.73 [1.53-1.97],  $P < .001$ ), and higher for dense breasts (1.45 [1.35-1.56],  $P < .001$ ). Results from the adjusted interaction model indicate that, although the odds of recall for the first and second DBT years compared with DM were similar across the dense and nondense breast subgroups, the odds of recall in the third DBT year compared with DM were significantly lower in the nondense (0.81 [0.72-0.91]) compared with the dense (0.99 [0.86-1.13]) breast subgroup ( $P = .03$ ).

To examine the effects of the prevalence and incidence of DBT screening, we compared the 21 395 unique women screened with DBT in our study population. Among these women, 12 079 had only 1 DBT screen (8170 of these women had previous DM screening), 6293 had only 2 DBT screens, and 3023 had 3 DBT screens (eTable 3 in the *Supplement*). Compared with the entire group receiving only 1 DBT screening

Abbreviations: DBT, digital breast tomosynthesis; BI-RADS, Breast Imaging Reporting and Data System; NA, not applicable; OR, odds ratio.

<sup>a</sup> Odds ratios, 95% CIs, and P values comparing the odds of recall across years 0 to 4 were calculated via a generalized estimating equations model, with robust SE, unstructured correlation matrix, and the individual woman as the unit of analysis. Empty cells indicate that no analysis was performed for the corresponding variable-model combinations.

<sup>b</sup> BI-RADS categories are explained in the Data Collection subsection of the Methods section.

(ie, including women with and without prior DM screening), recall rates continued to decrease with number per 1000 equaling 130, 78, and 59, for the only 1-screen, only 2-screen, and 3-screen women, respectively. The decreases were statistically significant for women undergoing only 2 (0.56 [0.51–0.63],  $P < .001$ ) and only 3 (0.42 [0.35–0.49]),  $P < .001$  DBT screens. Cancer detection rates were also significantly lower for the only 2-screen group (0.55 [0.39–0.79],  $P < .001$ ) but were not significantly lower for those with 3 DBT screens (0.65 [0.41–1.02],  $P = .06$ ). Similar results were observed when the reference group was restricted to only 1-time screeners with available DM screens. The PPV<sub>1</sub> for the only 2-screen group was 7.9% lower than that for the only 1-screen group with available prior mammograms, but it was not significantly lower than that for the entire group (PPV<sub>1</sub>: 11.7%,  $P = .03$ ; and 8.6%,  $P = .66$ , respectively). However, the PPV<sub>1</sub> for the 3-screen group (12.4%) was somewhat higher than that for the entire ( $P = .09$ ) and the restricted ( $P = .78$ ) 1-screen groups.

## Discussion

The controversy surrounding mammographic screening largely revolves around the “harms” of a false-positive examination.<sup>23</sup> Initial excitement for DBT reflected an apparent reduction in patients recalled for additional imaging (reduced harm) with equivalent or even increased breast cancer detection (increased benefit). Many practices moved to implement this new technology even though evidence of sustainable patient benefit was lacking. Three critical evidence gaps regarding imaging with DBT have been proposed: (1) detection measures at subsequent screening, (2) incremental mortality benefit, and (3) cost-effectiveness.<sup>24</sup>

In this study, we addressed the first evidence gap by analyzing data from 3 consecutive years of DBT screening and including analysis of women recalled for screening examinations. We found that reduction in recall was sustainable at a population level (Figure and Table), with additional reduction in recall as women returned for a second and third DBT examination (eTable 3 in the *Supplement*). Because false-positive examinations rather than low sensitivity for cancer detection is the primary criticism of screening mammography, the reduced recall obtained with DBT alone provides substantial benefit to support the continued evaluation of this new technology.

The DM outcomes from year 0 are remarkably similar to recommended benchmarks for screening mammography interpretations from the Breast Cancer Surveillance Consortium using 2 061 691 mammograms from 2004 to 2008.<sup>25</sup> The consortium-recommended recall rate per 1000 women screened was 99 (DM0, 104). The Breast Cancer Surveillance Consortium benchmarks for PPV<sub>1</sub>, PPV<sub>2</sub>, and PPV<sub>3</sub> were 4.3%, 23.6%, and 26.7% (DM0, 4.4%, 23.9%, and 25.8%), respectively. Although the overall population-based recall rate in our study remained below the recall for the DM cohort, there was a slight but nonsignificant increase in recall with each DBT year in the overall population. However, in women with more than 1 round of DBT screening, the recall rate continued to decrease with each additional DBT examination.

Other studies<sup>1–8</sup> have reported cancer detection rates based on initial use of DBT, which may indicate prevalence rather than incidence screening. In our study, the proportion of cancers detected in recalled patients decreased from those with only 1 DBT screen with prior comparisons from 13 to 6.2 per 1000 screened for those with only 2 screens, but then this increased significantly in women at the third DBT round of screening (eTable 3 in the *Supplement*). This finding suggests a possible prevalence-screening effect in the first round of screening with decreased cancer detection at the second incidence round. However, the decrease in the cancer detection in the women with only 2 screens to 6.2 per 1000 screened is still higher than published data for an incidence screening round with DM (4.6 per 1000).<sup>18</sup> Of note, at the third round, the cancer detection rate again increased.

Although some<sup>26</sup> have suggested that DBT may detect insignificant cancers at an earlier stage, our invasive cancer detection remained constant over the study period. Recent discussion about the implementation of DBT screening by Gur et al<sup>19</sup> has suggested that the introduction of DBT could result in a shift of cancer detection to an earlier time point, after which the detection rate might again reach a steady state, equivalent to pre-DBT screening. We used a similar model to examine our data without addressing possible underlying cancer incidence changes in the population. Our data show a mean time shift to earlier detection of at least 15 months and no suggestion of a decline to steady state, pre-DBT rates. Thus, the actual time shift will likely be longer than 15 months if a return to the steady state is ever observed.

Although we did not test for reduced morbidity and mortality, we addressed the second evidence gap by tracking the number of interval cancers as a surrogate for screening benefit.<sup>27,28</sup> Reducing the rate of interval cancers has been deemed “crucial, representing the potential benefit of early detection rather than overdiagnosis.”<sup>28(p679)</sup> Furthermore, there was no change in invasive cancer rates, indicating the continued detection of clinically significant cancers in our population.

There are limitations to our study: our natural experiment of an entire population with screening converted from DM to DBT screening was not randomized, which could introduce bias when comparing the methods. Unfortunately, data regarding risk-related characteristics (eg, family history) were not available for the entire population, thereby introducing a level of uncertainty regarding the similarity of the groups being compared, which could possibly confound the results. However, because our site fully converted to DBT screening in a single day, this is unlikely. In addition, all patients at our site with a history of breast cancer receive a diagnostic examination, which removes intermediate- to high-risk patients, possibly inflating the detected cancer counts. Finally, risk assessment data were available for the DM group and the first 18 months of DBT screening (previously published<sup>3</sup>) and multivariate analysis demonstrated no significant difference. Some might wonder whether decreased recall and increased cancer detection over the study period represent a learning curve, but this is difficult to evaluate. Finally, this study was designed to test for reduction in false-positive examinations—reducing the primary harm of screening when performed with

DM alone. We did not test for cost-effectiveness, although others<sup>29</sup> have suggested that initial DBT screening is cost-effective for a population undergoing biennial screening. We are currently performing this analysis based on our actual patient outcomes and services rendered.

## Conclusions

This study addresses issues regarding the sustainability of DBT screening outcomes: whether there are initial benefits in recall reduction and increased cancer detection achievable on both

a population basis as well as on an individual basis for women returning for further screening. There was significantly lower recall through 3 years of DBT screening, with greatly reduced recall in women presenting for consecutive screenings. Although DBT was initially implemented without knowledge of long-term performance, this is, to our knowledge, the first evidence that sustained and even improved performance is possible with consecutive DBT screening. Despite limitations, we believe this represents the first longitudinal analysis of women recalled for further DBT screening and is an important initial step toward informing policies for possibly integrating this technology into population-screening programs.

### ARTICLE INFORMATION

**Correction:** This article was corrected on March 17, 2016, to fix an error in the Abstract.

**Accepted for Publication:** November 5, 2015.

**Published Online:** February 18, 2016.  
doi:10.1001/jamaoncol.2015.5536.

**Author Contributions:** Drs McDonald and Conant had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

**Study concept and design:** McDonald, Schnall, Conant.

**Acquisition, analysis, or interpretation of data:** All authors.

**Drafting of the manuscript:** McDonald, Oustimov, Synnestvedt, Conant.

**Critical revision of the manuscript for important intellectual content:** McDonald, Weinstein, Schnall, Conant.

**Statistical analysis:** McDonald, Oustimov, Synnestvedt.

**Obtained funding:** Schnall, Conant.

**Administrative, technical, or material support:** McDonald, Synnestvedt, Schnall, Conant.

**Study supervision:** McDonald, Schnall, Conant.

**Conflict of Interest Disclosures:** Drs Weinstein and Conant were paid consultants for Siemens' Healthcare. Dr Conant is a paid scientific advisor and lecturer for Hologic Inc. No other disclosures were reported.

**Funding/Support:** This work was supported by grant U54CA163313 from the National Cancer Institute at the National Institutes of Health: Population-Based Research Optimizing Screening through Personalized Regimens Network.

**Role of the Funder/Sponsor:** The funding organization had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Additional Contributions:** Lauren Pantelone, BS (Perelman School of Medicine, University of Pennsylvania), assisted with data collection; she was supported by grant U54CA163313.

### REFERENCES

1. 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.
2. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA*. 2014;311(24):2499-2507.
3. 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;106(11):106.
4. 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 Am J Roentgenol*. 2013;200(6):1401-1408.
5. 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.
6. Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology*. 2015;274(1):85-92.
7. 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(3):694-700.
8. Lourenco AP, Barry-Brooks M, Baird G, et al. Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. *Radiology*. 2015;274(2):337-342.
9. Andersson I, Ikeda DM, Zackrisson S, et al. Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and BI-RADS classification in a population of cancers with subtle mammographic findings. *Eur Radiol*. 2008;18(12):2817-2825.
10. Bernardi D, Ciatto S, Pellegrini M, et al. Prospective study of breast tomosynthesis as a triage to assessment in screening. *Breast Cancer Res Treat*. 2012;133(1):267-271.
11. Gur D, Abrams GS, Chough DM, et al. Digital breast tomosynthesis: observer performance study. *AJR Am J Roentgenol*. 2009;193(2):586-591.
12. 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.
13. Teertstra HJ, Loo CE, van den Bosch MA, et al. Breast tomosynthesis in clinical practice: initial results. *Eur Radiol*. 2010;20(1):16-24.
14. Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution X-ray imaging observer study. *Radiology*. 2012;262(3):788-796.
15. Friedewald SM, Rafferty EA, Conant EF. Breast cancer screening with tomosynthesis and digital mammography—reply. *JAMA*. 2014;312(16):1695-1696.
16. Caumo F, Bernardi D, Ciatto S, et al. Incremental effect from integrating 3D-mammography (tomosynthesis) with 2D-mammography: Increased breast cancer detection evident for screening centres in a population-based trial. *Breast*. 2014;23(1):76-80.
17. Gilbert FJ, Tucker L, Gillan MG, et al. The TOMMY trial: A Comparison of Tomosynthesis With Digital Mammography in the UK NHS Breast Screening Programme: a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. *Health Technol Assess*. 2015;19(4):i-xxv, 1-136.
18. McDonald ES, McCarthy AM, Akhtar AL, Synnestvedt MB, Schnall M, Conant EF. Baseline screening mammography: performance of full-field digital mammography versus digital breast tomosynthesis. *AJR Am J Roentgenol*. 2015;205(5):1143-1148.
19. Gur D, Nishikawa RM, Sumkin JH. New screening technologies and practices: a different approach to estimation of performance improvement by using data from the transition period. *Radiology*. 2015;275(1):9-12.
20. Sickles EA, D'Orsi CJ, Bassett LW, et al. ACR BI-RADS Mammography. In: D'Orsi CJ, Sickles EA, Mendelson EB, et al. *ACR BI-RADS Atlas, Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology; 2013.
21. Rohatgi VK, Ehsanes Saleh AKM. *An Introduction to Probability and Statistics*. Vol 910. New York, NY: John Wiley & Sons, Inc; 2001:500-509.
22. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988;44(4):1049-1060.
23. US Preventive Services Task Force. Breast cancer: screening: recommendation summary. <http://www.uspreventiveservicestaskforce.org/Page/Topic/recommendation-summary/breast-cancer-screening>. Published November 2009; updated July 2015. Accessed December 18, 2015.
24. Houssami N. Digital breast tomosynthesis (3D-mammography) screening: data and implications for population screening. *Expert Rev Med Devices*. 2015;12(4):377-379.

- 25.** Breast Cancer Surveillance Consortium. Benchmarks for abnormal screening mammography interpretations from 2004–2008—based on BCSC data through 2009 (HHSN261201100031C). <http://breastscreening.cancer.gov/statistics/benchmarks/screening/2009/table5.html>. Accessed June 20, 2015.
- 26.** 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 [published online October 9, 2015]. *Radiology*.
- 27.** Houssami N. STORM, a new dimension for mammography screening. *Med J Aust*. 2013;199(5):308–309.
- 28.** Irwig L, Houssami N, Armstrong B, Glasziou P. Evaluating new screening tests for breast cancer. *BMJ*. 2006;332(7543):678–679.
- 29.** Lee CI, Cevik M, Alagoz O, et al. Comparative effectiveness of combined digital mammography and tomosynthesis screening for women with dense breasts. *Radiology*. 2015;274(3):772–780.