

BI-RADS Category 3 Comparison: Probably Benign Category after Recall from Screening before and after Implementation of Digital Breast Tomosynthesis¹

Elizabeth S. McDonald, MD, PhD
 Anne Marie McCarthy, PhD, ScM
 Susan P. Weinstein, MD
 Mitchell D. Schnall, MD, PhD
 Emily F. Conant, MD

Purpose:

To evaluate Breast Imaging Reporting and Data System (BI-RADS) category 3 assessment at diagnostic examination after recall from screening in a large urban population after implementation of digital breast tomosynthesis (DBT) by focusing both on overall use and use stratified by recalled finding type and outcome at 2 years.

Materials and Methods:

This was an intuitional review board–approved and HIPAA-compliant retrospective review of 10728 digital mammography (DM) examinations from September 1, 2010, to August 30, 2011, and 15571 screening DBT examinations from October 1, 2011, to February 28, 2013. The recall populations for DM and DBT were 1112 of 10728 (10.4% of women screened) and 1366 of 15571 (8.8% of women screened), respectively. Recall examinations were classified according to finding type: calcifications, asymmetry or focal asymmetry, mass, and architectural distortion. Differences between groups were compared by using the χ^2 test.

Results:

Although there was no significant change in the utilization rate of BI-RADS category 3 in those patients screened with DM compared with DBT (168 of 10728, 1.6% for DM vs 206 of 15571, 1.3% for DBT; $P = .102$), there was a mean overall reduction of 2.4 women per 1000 (95% confidence interval [CI]: $-0.5, 5.4$) recommended for short-term follow-up. Lesion types given a BI-RADS category 3 assessment after diagnostic work-up did not change. The distribution of recalled finding types significantly changed with DBT, with increased recall examinations for architectural distortion and mass ($P < .001$) and decreased recall examinations for asymmetries ($P \leq .001$). There was no change in recall examinations for calcifications ($P = .977$).

Conclusion:

Screening with DBT did not significantly change the utilization rate of BI-RADS category 3 classification; however, the overall number of patients recommended for short-interval follow-up decreased by a mean of 2.4 women per 1000 (95% CI: $-0.5, 5.4$).

©RSNA, 2017

Online supplemental material is available for this article.

¹ From the Department of Radiology, Breast Imaging Section, University of Pennsylvania, Perelman School of Medicine, Hospital of the University of Pennsylvania, 3400 Spruce St, 1 Silverstein, Philadelphia, PA 19104-4283 (E.S.M., S.P.W., M.D.S., E.F.C.); and General Medicine Division, Massachusetts General Hospital, Harvard Medical School, Boston, Mass (A.M.M.). Received January 3, 2017; revision requested February 22; revision received March 27; accepted April 14; final version accepted May 4. **Address correspondence to** E.F.C. (e-mail: Emily.Conant@uphs.upenn.edu).

Supported by Division of Cancer Prevention, National Cancer Institute (U54CA163313).

The addition of digital breast tomosynthesis (DBT) to digital mammography (DM) has changed mammography screening by decreasing false-positive results (1–8) while increasing cancer detection when compared with screening with DM alone (1,3,6–8). When a patient is recalled from a screening examination, a diagnostic evaluation including mammography and/or ultrasonography (US) is performed, with potential outcomes determined based on imaging features. After recall and diagnostic work-up, the studies are reported by using assessment categories recommended by the American College of Radiology Breast Imaging and Data Reporting System (BI-RADS) (9). These categories are as follows: resolution of the recalled finding (BI-RADS 1 or 2), recommendation

for biopsy (BI-RADS 4 or 5), or a more nebulous category that involves short-term follow-up by utilizing sequential diagnostic examinations, usually at 6-month or 1-year intervals and for up to 2 years (BI-RADS 3). Cases of BI-RADS category 3 assigned after diagnostic evaluation should have a probability of malignancy of less than 2%. Eventually, either stability is documented and a finding is resolved as benign, or a worrisome change is detected and biopsy is recommended. The additional economic impact of repeated diagnostic examinations is cost-effective (vs immediate biopsy) for US alone (10) and for mammography alone (11). However, repeated examinations may lead to patient anxiety because a finding representing a potential cancer may be unresolved, under current standards, for up to 2 years. Patient compliance is known to decrease with each subsequent diagnostic examination, limiting the utility of the “wait and see” approach (12,13).

Evidence exists for the cost-effectiveness of screening tomosynthesis on an annual (14) and biannual basis (15). Although a comprehensive evaluation of the impact of screening tomosynthesis on the cost of diagnostic evaluation has not been performed, there are fewer mammographic views at recall in patients screened with tomosynthesis and subsequently diagnosed with cancer, with an increase in patients needing no mammographic views from 6.3% to 35.1% (16) in one study and from 2.6% to 28.3% in another (5). In many cases due to the improved localization and characterization of lesions achievable with DBT, the diagnostic mammogram might be safely eliminated in recall

examinations for noncalcified findings and may lead to cost savings (17,18).

Other factors should be considered in population screening. One initial concern was an increase in radiation dose with combination DM and DBT imaging; however, reconstruction of synthetic two-dimensional images from the DBT acquisition has reduced the dose of DBT screening by 39%, approaching that of a two-dimensional examination while maintaining the benefit of reduced false-positive examinations (19). Another early finding was an increase in the biopsy rate with DBT (3), which could impact downstream cost. Also, if DBT increases the rate of BI-RADS category 3 use, as suggested in two early reports (5,20), then this increase could offset cost savings and reduce enthusiasm for this technology.

The purpose of our study was to evaluate BI-RADS category 3 assessment at diagnostic examination after recall from screening in a large urban population after implementation of DBT by focusing both on overall use and use stratified by recalled finding type and outcome at 2 years.

Advances in Knowledge

- The rate of use of Breast Imaging Reporting and Data System (BI-RADS) category 3 after recall from screening with digital mammography (DM) and digital breast tomosynthesis (DBT) was not significantly different than screening with DM alone.
- Because fewer women were recalled with DM and DBT, the downstream effect was that a mean of 2.4 per 1000 fewer women were placed in short-term follow-up after being screened with DBT compared with those screened with DM alone.
- Distribution of recalled finding types significantly changed with DBT, with increased recall examinations for architectural distortions (75 of 1231 [6.1%] to 156 of 1560 [9.9%]) and masses (261 of 1231 [21.2%] to 420 of 1560 [26.9%]) and decreased recall examinations for asymmetries (652 of 1231 [53.0%] to 701 of 1560 [44.9%]).
- There was no change in recall examinations for calcifications or in the types of lesions placed in BI-RADS category 3.

Implications for Patient Care

- The replacement of DM with combination DM and DBT screening is beneficial to women by reducing the number of recalled patients.
- As a result of the reduction in recall, fewer women from the total screened population were recommended for short-interval follow-up.

Materials and Methods

E.S.M. and A.M.M. were primarily responsible for the data analysis and had

<http://doi.org/10.1148/radiol.2017162837>

Content code: **BR**

Radiology 2017; 285:778–787

Abbreviations:

BI-RADS = Breast Imaging Reporting and Data System
 CI = confidence interval
 DBT = digital breast tomosynthesis
 DM = digital mammography

Author contributions:

Guarantors of integrity of entire study, E.S.M., E.F.C.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, E.S.M., S.P.W., E.F.C.; clinical studies, E.S.M., S.P.W., M.D.S., E.F.C.; experimental studies, S.P.W.; statistical analysis, A.M.M.; and manuscript editing, all authors

Conflicts of interest are listed at the end of this article.

control of the data and material submitted for publication. E.F.C. has served as a consultant to Hologic and S.P.W. has served as a consultant to Hologic and Siemens. There was no industry support for this trial; however, A.M.M., S.P.W., M.D.S., and E.F.C. were supported by a U54 grant from Population-based Research Optimizing Screening through Personalized Regimens Network (U54CA163313). E.S.M. and A.M.M. have no conflicts of interest.

Study Population

This was an institutional review board–approved and Health Insurance Portability and Accountability Act–compliant retrospective audit of population screening data obtained through routine clinical practice in a natural outcomes experiment incorporating more than 25 000 examinations before and after complete screening practice conversion to DBT. We evaluated the population given a BI-RADS category 3 assessment at diagnostic examination after recall from screening by focusing both on overall use and use stratified by recalled finding type and outcome at 2 years. The study population consisted of all patients presenting for DM from September 1, 2010, to August 30, 2011, ($n = 10728$) and for DBT from October 1, 2011, to February 28, 2013, ($n = 15571$). The entire practice converted to screening DBT and DM on September 19, 2011, adding a two-view DBT examination to the two-view full-field DM examination for each breast (Dimensions; Hologic, Bedford, Mass). DBT conversion occurred in a single day and all patients were offered the same service at no additional charge. Our screening population consisted of women without symptoms or physical examination findings and who had no prior history of breast cancer. Outcomes data were queried through August 17, 2016. Previous multivariate analysis of a subset of women from both cohorts (5626 of 10728 from cohort 1 and 11 394 of 15571 from cohort 2; 596 of 1112 recalled in cohort 1 and 996 of 1366 recalled in cohort 2; 85 of 168 of BI-RADS category 3 in cohort 1 and 138 of 206 of BI-RADS category 3

in cohort 2) with available data showed that the distribution of risk factors (age, race, prior mammogram, breast density, interpreting 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) in the DBT and DM cohorts was similar and did not explain differences in recall rates between the DBT and DM cohorts (6).

In different analyses, our institution previously published results comparing recall and cancer detection rates from 14 014 (3), 11 910 (21), and 15 571 of the DM and DBT screening studies and from 10 728 DM-only screening studies (6,19,22,23). In this article, we report on the outcomes of the subpopulation of patients placed in short-interval follow-up separated by specific imaging finding type.

Screening Interpretation

Examinations were interpreted by one of six radiologists with specialization in breast imaging. The radiologist staffing was constant and unchanged during the study period. Five of the interpreting radiologists were fellowship trained in breast imaging with 3–22 years of clinical practice (including E.F.C., S.P.W.). One radiologist was not fellowship trained but had 22 years of practice experience at the start of the study. All radiologists received 8 hours of training in DBT interpretation prior to practice conversion. Examinations were cataloged according to finding type: calcifications, asymmetry or focal asymmetry, mass, architectural distortion, and technical recall. Some recall examinations occurred for more than one finding. Breast density was evaluated by using the definitions in the fourth edition of the American College of Radiology BI-RADS atlas, which is defined as percent density: 1 is 0%–25%, 2 is 26%–50%, 3 is 51%–75%, and 4 is 76%–100%.

Diagnostic Examinations

After screening implementation of DBT, DBT imaging was also immediately

available for diagnostic examinations. Generally, only DM imaging was performed for the recalled finding type of calcifications (craniocaudal and mediolateral magnification views) before and after implementation of DBT; US only was performed for the recalled finding type of masses after implementation of DBT because the mass margins could be evaluated from the screening examination. Before implementation of DBT, mammographic spot compression was often performed for patients recalled for masses. Asymmetries and architectural distortion were worked up with additional mammographic views, which generally included at least one DBT view after implementation of DBT. Magnetic resonance (MR) imaging is rarely requested after a diagnostic work-up. If an MR imaging is recommended after a full-diagnostic work-up, then the diagnostic examination is given a final BI-RADS assessment category of 1 to 5 before the additional imaging is performed.

Data Collection

Mammographic studies were reported by using a reporting system that utilizes the assessment categories recommended by the American College of Radiology BI-RADS (Centricity; GE Healthcare, Milwaukee, Wis) (15). BI-RADS assessment categories and demographics were entered at the time of interpretation.

Screening Outcome Measures

Screening outcome measures were evaluated including imaging volume, recall rates (percentage), recalled finding type, and final BI-RADS assessment category. A recalled patient was any screening patient given an assessment of BI-RADS category 0 (additional imaging needed), category 4, or category 5. We individually evaluated cases of architectural distortion given a final assessment after recall of BI-RADS category 3 because of a high probability of malignancy for this finding type (24). Diagnostic outcome was tracked for at least 2 years after the screening recall and included matching with state tumor registry data. If a patient died during

the study period without evidence of breast cancer, then this was counted as completed follow-up, even if the death occurred before 2 years after the screening mammographic examination ($n = 7$). A cancer in the same breast given a designation of BI-RADS category 3 was counted as missed cancer, even if in a different quadrant than in the original recalled finding.

Statistical Analysis

Differences in the baseline characteristics, final assessment categories, and finding types between the DBT- and DM-screened cohorts were compared by using χ^2 tests, and 95% confidence intervals (CIs) were calculated by using standard methods (25). Percent change in outcomes between DBT and DM were calculated as follows: [(percent DBT – percent DM)/percent DM] \times 100. Sensitivity analyses were performed by comparing differences in final assessment categories, finding types, and percent change in outcomes between DBT and DM excluding women with no comparison examination and women with two DM examinations, and separately evaluating the population of women with a first and second DBT examination. A P value of less than .05 was considered to indicate a statistically significant difference. All statistical tests were two sided and performed by using Stata 13.1 software (Stata, College Station, Tex).

Results

Baseline characteristics of recalled patients and patients recommended for short-interval follow-up are presented in Table 1. Recalled populations for DM and DBT were 1112 and 1366, respectively. Recalled patients were stratified by finding type and final assessment category. There was no difference in the proportion of those patients who had no prior study available (Table 1).

There were no significant differences in patients assigned to short-interval follow-up in the DM cohort and the DBT cohort with respect to the total number of women screened (168 of 10728, 1.6% for DM vs 206

Table 1

Baseline Characteristics of Recalled Patients and Patients Recommended for Short-Interval Follow-up

A: Characteristics of Recalled Patients

Characteristic	Cohort 1: DM ($n = 10728$)	Cohort 2: DBT ($n = 15571$)	P Value
No. of women recalled	1112 (10.4)	1366 (8.8)	<.001
Age (y)*	54.2 \pm 10.4	53.8 \pm 10.6	.350
Age categories (y)			
<40	30 (2.7)	37 (2.7)	.630
40–49	415 (37.3)	543 (39.8)	...
50–59	350 (31.5)	421 (30.8)	...
60–69	224 (20.1)	246 (18.0)	...
≥ 70	93 (8.4)	119 (8.7)	...
Body mass index*	29.5 \pm 7.4	29.6 \pm 7.2	.792
Race			
White	429 (38.6)	516 (37.8)	.289
Black	589 (53.0)	720 (52.7)	...
Hispanic	11 (1.0)	22 (1.6)	...
Asian	37 (3.3)	35 (2.6)	...
Other/Unknown	46 (4.1)	73 (5.3)	...
Breast density			
BI-RADS category 1 (0%–25%)	69 (6.2)	94 (6.9)	.609
BI-RADS category 2 (26%–50%)	598 (53.8)	725 (53.1)	...
BI-RADS category 3 (51%–75%)	429 (38.6)	519 (38.0)	...
BI-RADS category 4 (76%–100%)	16 (1.4)	28 (2.1)	...
Prior study			
No	247 (22.2)	298 (21.8)	.813
Yes	865 (77.8)	1068 (78.2)	...

B: Characteristics of Patients Recommended for Short-Interval Follow-up

Characteristic	Cohort 1: DM ($n = 10728$) [†]	Cohort 2: DBT ($n = 15571$) [†]	P Value
No. of women screened	168 (1.6)	206 (1.3)	.102
Age (y)*	52.4 \pm 9.7	53.9 \pm 10.8	.189
Age categories (y)			
<40	4 (2.4)	4 (1.9)	.402
40–49	76 (45.2)	83 (40.3)	...
50–59	46 (27.4)	56 (27.1)	...
60–69	35 (20.8)	44 (21.4)	...
≥ 70	7 (4.2)	19 (9.2)	...
Body mass index*	32.0 \pm 8.9	31.1 \pm 8.0	.348
Race			
White	48 (28.6)	66 (32.0)	.266
Black	109 (64.9)	123 (59.7)	...
Hispanic	1 (0.6)	4 (1.9)	...
Asian	5 (3.0)	2 (1.0)	...
Other/Unknown	5 (3.0)	11 (5.3)	...
Breast density			
BI-RADS category 1 (0%–25%)	15 (8.9)	28 (13.6)	.337
BI-RADS category 2 (26%–50%)	96 (57.1)	118 (57.3)	...
BI-RADS category 3 (51%–75%)	56 (33.3)	57 (27.7)	...
BI-RADS category 4 (76%–100%)	1 (0.6)	3 (1.5)	...

Table 1 (continues)

of 15571, 1.3% for DBT; $P = .102$) or with respect to the number of women recalled (168 of 1112, 15.1% for DM vs 206 of 1366, 15.1% for DBT; $P = .981$) (Table 1). Numbers of women in short-interval follow-up were slightly lower after DBT screening: 2.4 fewer women per 1000 (95% CI: $-0.5, 5.4$). An additional analysis was performed excluding women with no comparison examination and women with two DM examinations (Table E1 [online]). This also separately evaluated the population of women with a first and second DBT examination so that a similar DM and DBT population could be compared, isolating the effect of new technology from the confounder of no prior studies or the presence of prior DBT examinations (Table 2). This analysis also demonstrated no difference in BI-RADS 3 use when comparing the entire screening population and the recalled patients. There was also no significant decrease in BI-RADS category 3 use at second DBT examination when comparing with DM or first DBT examination.

There was no significant difference in the use of BI-RADS category 3 for any finding type. For the DM cohort, the finding types given a BI-RADS category 3 assessment were calcifications (67 of 184, 36.4%), mass (41 of 184, 22.3%), asymmetry or focal asymmetry (71 of 184, 38.6%), and architectural distortion (five of 184, 2.7%). For the DBT cohort, the finding types were calcifications (67 of 227, 29.5%), mass (61 of 227, 26.9%), asymmetry or focal asymmetry (93 of 227, 40.5%) and architectural distortion (seven of 227, 3.1%) (Table 3). As mentioned previously, we further evaluated cases given a BI-RADS 3 assessment for architectural distortion because of a high probability of malignancy for this finding type (24). In the DM cohort, five patients were recalled for architectural distortion and given a BI-RADS category 3 assessment after a diagnostic evaluation. For two of these cases, the architectural distortion resolved with diagnostic imaging and the BI-RADS category 3 assessment was given for a different finding type perceived at the diagnostic examination. In

Table 1 (continued)

Baseline Characteristics of Recalled Patients and Patients Recommended for Short-Interval Follow-up

B: Characteristics of Patients Recommended for Short-Interval Follow-up

Characteristic	Cohort 1: DM ($n = 10728$) [†]	Cohort 2: DBT ($n = 15571$) [†]	P Value
Prior study			
No	69 (41.1)	79 (38.4)	.592
Yes	99 (58.9)	127 (61.6)	...

Note.—Unless otherwise indicated, data are the number of patients and data in parentheses are percentages.

* Data are means \pm standard deviation.

[†] In the DM cohort, the number of women recalled was 168 of 1112 (15.1%) and the in DBT cohort, the number of women recalled was 206 of 1366 (15.1%); $P = .981$.

a third case, the distortion correlated with scars from breast reduction. In a fourth case, the distortion resolved. In the final case, the distortion was thought to be questionable; a 6-month follow-up examination was requested only after negative findings at US and MR imaging (26). In the DBT cohort, seven patients were recalled for architectural distortion and given a BI-RADS category 3 assessment after diagnostic evaluation. In four of these cases, the architectural distortion resolved with diagnostic imaging, and the BI-RADS category 3 assessment was given for a different finding type at the diagnostic examination. In a fifth case, the distortion correlated with scars from prior placement of a chest wall catheter. In two cases, the distortion was felt to completely resolve; however, a follow-up examination was recommended “to assess stability of the mammogram.”

Overall, recall examinations for calcifications (196 of 1231, 15.9% for DM and 249 of 1560, 16.0% for DBT; $P = .977$) remained the same. Recall examinations for masses significantly increased by 26.9% (261 of 1231, 21.2% for DM and 420 of 1560, 26.9% for DBT; $P \leq .001$). Recall examinations for architectural distortion significantly increased by 62.3% (75 of 1231, 6.1% for DM to 156 of 1560, 9.9% for DBT; $P \leq .001$). Recall rate for asymmetries significantly decreased by 15.2% (652 of 1231, 53.0% for DM to 701 of 1560, 44.9% for DBT; $P \leq .001$). Technical

callbacks significantly decreased (47 of 1231, 3.8% for DM to 34 of 1560, 2.2% for DBT; $P = .010$) (Table 3). We calculated BI-RADS category 3 outcomes as a percentage of total recalled findings, stratified by recalled finding type, and no significant differences were found (Table 3).

Among 168 patients placed in short-interval follow-up in the DM population, 2 years of follow-up was available for 155 of 168 (92.3%). One year of follow-up was available for 162 of 168 (96.4%) patients. Among 206 patients placed in short-interval follow-up in the DBT population, 2 years of follow-up was available for 175 of 206 (85.0%). One year of follow-up was available for 189 of 206 (91.7%) patients. Four cancers were found in the DM cohort with BI-RADS category 3 (2.4%; 95% CI: 0.6%, 6.0%) and two cancers were found in the DBT cohort with BI-RADS category 3 (1.0%; 95% CI: 0.1%, 3.5%). The DM group had a delayed cancer detection rate above the 2% threshold as based on the recommended performance metrics of the American College of Radiology BI-RADS (15). Two of the four cancers in the DM cohort manifested as calcifications (ductal carcinoma in situ) and invasive ductal carcinoma (Fig 1), and another each as calcifications plus distortion (invasive ductal carcinoma), and mass (invasive ductal carcinoma). The two patients with cancer in the DBT cohort were recalled for calcifications alone (representing ductal carcinoma

Table 2

BI-RADS Category 3 Percentages and Rates per 1000, Excluding Women with No Available Comparison and Women with Two DM Examinations

Variable	DM	First DBT Examination	P Value (First DBT vs DM)	Second DBT Examination	P Value (Second DBT vs DM)	P Value (Second DBT vs First DBT)
Total screened	9518	12 017	...	1692
No. of BI-RADS category 3	99	116	...	11
Percent of BI-RADS category 3	1.0*	1.0*	.583	0.7*	.134	.205
BI-RADS category 3 per 1000 women screened	10.4 (8.5–12.6)	9.7 (8.0–11.6)	.583	6.5 (3.2–11.6)	.134	.205
Total recalled	865	968	...	100
No. of BI-RADS category 3	99	116	...	11
Percent of BI-RADS category 3	11.5*	12.0*	.721	11.0*	.895	.772
BI-RADS category 3 per 1000 women screened	114.5 (94.0–137.6)	119.8 (100.0–142.0)	.721	110 (56.2–188.3)	.895	.772

Note.—Unless otherwise noted, data are the number of patients and data in parentheses are the range.

* Data are percentages.

in situ and invasive ductal carcinoma) (Fig 2).

Discussion

Mammographic screening has a mortality benefit in the 40–74 year-old age group (27). Still, recommendations for screening have varied, with the U.S. Preventive Services Task Force and American Cancer Society (27,28) citing the need to weigh benefits with harms that include a high rate of false-positive examinations. Included with false-positive examinations is the small subgroup of women with unresolved findings after diagnostic evaluation who are placed in the “probably benign” category. These women are left to wait and wonder: when will resolution come, and what will be the ultimate outcome? A decrease in this population could have substantial clinical and economic impact.

The proportion of recalled women who were given a BI-RADS category 3 assessment remained the same in the DM and DBT cohorts. However, because fewer women were recalled with DM and DBT, the downstream effect was that 2.4 per 1000 fewer women were placed in short-term follow-up after being screened with DBT compared with those screened with DM alone. This result is clinically

Table 3

Changes in Recalled Finding Type and Findings Given a BI-RADS 3 Category Assessment Following Implementation of Screening DBT

A: Recalled Finding Type

Variable	Cohort 1: DM (n = 1231)	Cohort 2: DBT (n = 1560)	P Value	Change in Outcomes (%)
Calcifications	196 (15.9)	249 (15.9)	.977	0
Masses	261 (21.2)	420 (26.9)	<.001	+26.9
Asymmetries	652 (53.0)	701 (44.9)	<.001	–15.2
Architectural distortion	75 (6.1)	156 (9.9)	<.001	+62.3
Technical recall	47 (3.8)	34 (2.2)	.010	–42.1

B: Recalled Findings Given BI-RADS Category 3

Variable	Cohort 1: DM (n = 184)	Cohort 2: DBT (n = 227)	P Value	Change DM to DBT (%)
Calcifications	67 (36.4)	67 (29.5)	.138	–19.0
Masses	41 (22.3)	61 (26.9)	.284	+20.6
Asymmetries	71 (38.6)	92 (40.5)	.689	+4.9
Architectural distortion	5 (2.7)	7 (3.1)	.826	+14.8
Technical recall	0 (0)	0 (0)

C: Findings Given BI-RADS Category 3

Variable	Cohort 1 (DM): BI-RADS 3 Findings/Total Findings	Cohort 2 (DBT): BI-RADS 3 Findings/Total Findings	P Value	Change DM to DBT (%)
Calcifications	67/196 (34.2)	67/249 (26.9)	.097	–21.3
Masses	41/261 (15.7)	61/420 (14.5)	.674	–7.6
Asymmetries	71/652 (10.9)	92/701 (13.1)	.207	+20.2
Architectural distortion	5/75 (6.7)	7/156 (4.5)	.486	–32.8
Technical recall	0/47 (0)	0/34 (0)	1.0	0

Note.—Unless otherwise indicated, data are numbers of patients, with percentages in parentheses.

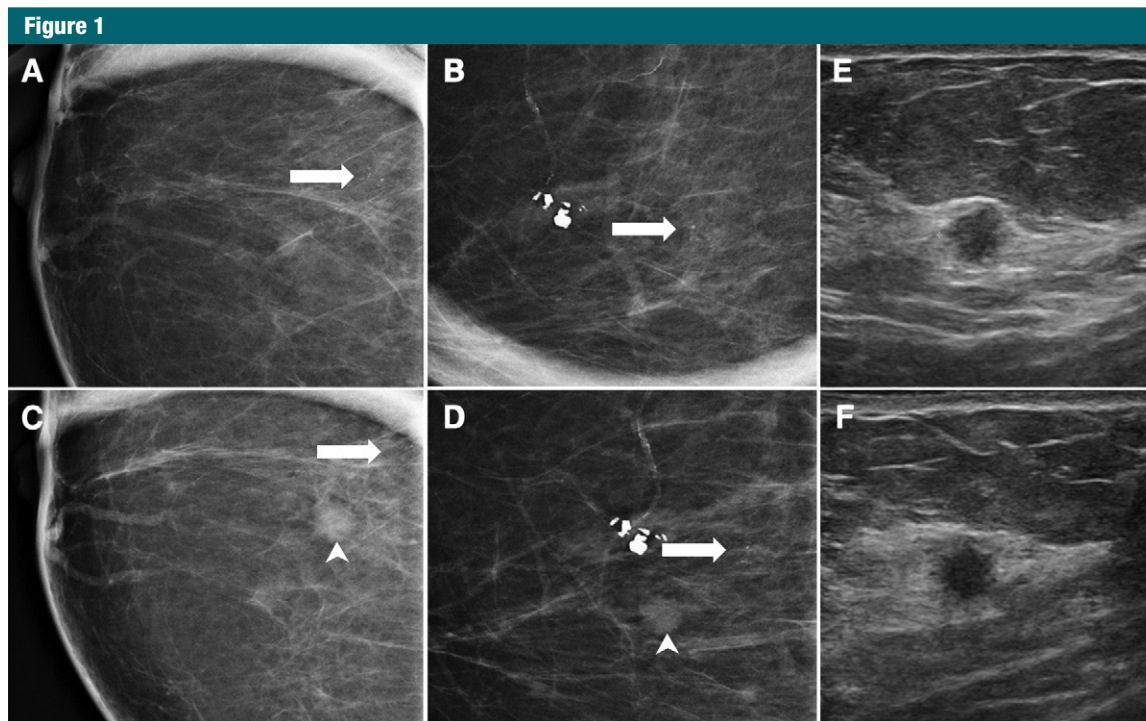


Figure 1: DM images show cancer depicted at 6-month follow-up. A 54-year-old woman was recalled from baseline screening examination for calcifications. She was given a BI-RADS category 3 assessment after additional work-up. *A*, craniocaudal and, *B*, mediolateral magnification views at first diagnostic study demonstrate punctate calcifications (arrow). At 6 months, calcifications were stable but a new mass (arrowhead) was noted on, *C*, craniocaudal and, *D*, mediolateral magnification views. Biopsy was recommended. Corresponding, *E*, radial and, *F*, antiradial orthogonal US images from 6-month diagnostic examination demonstrate a $5 \times 5 \times 5$ -mm round hypoechoic mass with indistinct margins. Both calcifications and mass were excised (invasive ductal cancer and ductal carcinoma in situ [T1N0M0]).

meaningful considering the cost of additional imaging events, as well as the potential for anxiety associated with short-term follow-up. Our results are in contrast to those of other studies that evaluated BI-RADS category 3 use after implementation of tomosynthesis. In one study, the use of category 3 increased to over 25% of recalled lesions; however, the increase was not statistically significant (5). Another study found the likelihood of BI-RADS category 3 was 80% higher after implementation of screening DBT and adjusting for patient characteristics (20). A 2015 Radiological Society of North America presentation reported that implementation of DBT increased the use of category 3 findings by nearly eightfold compared with pre-DBT screening (29). Another study evaluating a diagnostic population found that BI-RADS category 3 recommendations decreased over time, with an initial DM

rate of 33.3% before DBT implementation (26). In that study, the 3rd year of DBT use yielded a BI-RADS category 3 rate of 16.4% after diagnostic examination, which is similar to our results in patients recalled from screening. Our use of BI-RADS category 3 in the DM population was similar to that in other groups. Yasmeen and colleagues performed an analysis of over 58 000 women receiving baseline DM as part of the Women's Health Initiative yielding a utilization rate of 5% for short-interval follow-up on study enrollment (our overall utilization rate was 1.6% in the DM group) (30). Another study evaluating DM found that 5.2% of first screenings and 1.7% of subsequent screenings were placed into short-term follow-up (with or without a diagnostic evaluation) (31).

Some might question our use, although infrequent, of BI-RADS category 3 in recall examinations for

architectural distortion, given the high probability of malignancy for this finding type. As mentioned previously, this finding was only followed when postoperative changes were strongly suspected or when the distortion resolved at diagnostic imaging and there was an additional finding to follow. We do not advocate following architectural distortion depicted at either DBT or DM imaging unless a benign, pathognomonic etiology can be determined (such as scarring from a benign biopsy). Architectural distortion from both benign and malignant etiologies is more conspicuous at DBT imaging (24,32,33). In the case of persistent architectural distortion after diagnostic evaluation without evidence of prior surgical intervention and without a sonographic correlate, we generally recommend biopsy due to the high probability of malignancy (47% [76% invasive cancers] in one study [24]). In these cases, the interpreting radiologist

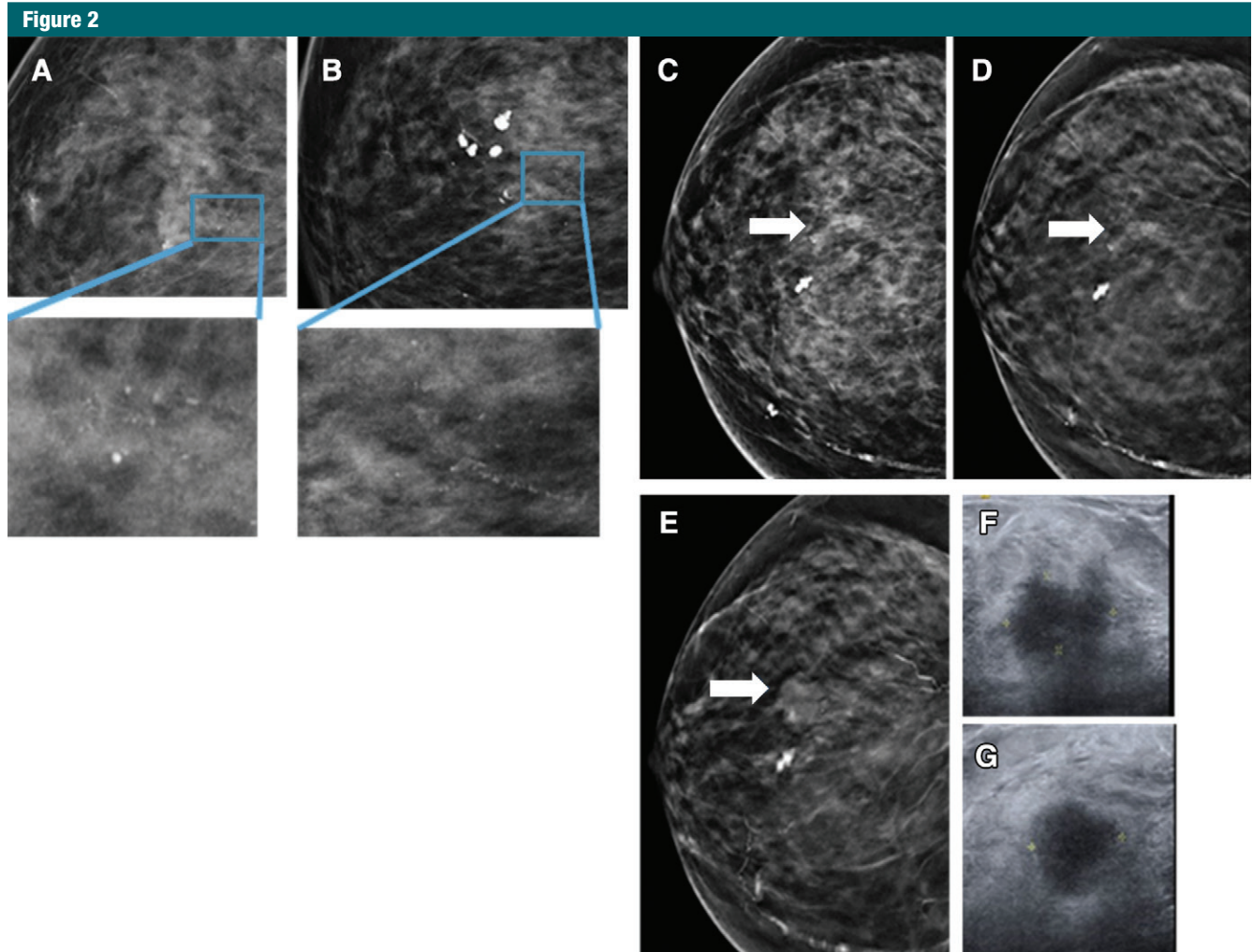


Figure 2: DBT images depict cancer at 12-month follow-up. A 74-year-old woman was recalled from a screening mammogram for calcifications. She was given a BI-RADS category 3 assessment after additional work-up. *A*, craniocaudal and *B*, mediolateral magnification views at first diagnostic study demonstrate punctate calcifications. At 6-month follow-up (not shown), calcifications were thought to be stable. At 1 year, a mass was noted and biopsy was recommended. In retrospect, the mass (arrow) was likely present on the original screen, *C*, DM craniocaudal view, but only visible on *D*, corresponding DBT craniocaudal view. *E*, Mass increased in size at 1-year follow-up when it was first detected. Corresponding, *F*, radial and *G*, antiradial orthogonal US images from 1-year diagnostic examination demonstrate a $19 \times 12 \times 15$ mm irregular hypoechoic mass with angular and indistinct margins that represented invasive ductal cancer, 2.5 cm at surgery (T2N0M0).

may also recommend MR imaging after a negative finding at diagnostic US. If an MR imaging correlate is found to the architectural distortion, then MR imaging-guided biopsy is performed. If no correlate is found or an MR imaging examination is not performed, then the patient proceeds to DBT-guided core biopsy or wire localization for surgical excision.

There was early prediction that implementation of DBT would influence recalled finding types. For example, masses or architectural distortion might

be more easily detected with a quasi-three-dimensional data set compared with a two-dimensional examination, leading to an increase in recall. Asymmetries due to overlapping fibroglandular tissue might be recognized as such and recalled less often. Some studies have supported this with greater recall for masses and architectural distortions (5), but others have found no change (2). The predominant recalled finding consistently decreased with implementation of DBT is asymmetries (2,5). In this study, implementation of DBT

resulted in a 26.9% increase in recalled masses and a 62.3% increase in architectural distortion. These increases were offset by a 15.2% reduction in recall examinations for asymmetries and nearly half the number of technical recall examinations, a small portion of the recalled population. Despite an overall decrease, a similar number of asymmetries were given BI-RADS category 3 assessment in both groups (38.6% for DM vs 40.5% for DBT; $P = .69$). This may be because asymmetries that could easily be resolved as overlapping

tissue after diagnostic imaging were not initially recalled after implementation of DBT. Improved evaluation of superimposed breast tissue also allows increased depiction and recall of masses at DBT that may not have been depicted at DM alone. Fortunately, although significantly more masses are depicted with DBT, the percentage of recalled masses for which biopsy is eventually recommended is lower when DBT screening is utilized (18.9% vs 30.5% in DM-screened patients), possibly indicating that proportionally more benign masses are depicted (5). The increased recall of architectural distortion with DBT is also expected due to the multiple projections and reconstruction, making both benign and malignant lesions more conspicuous. The improved depiction of architectural distortion is, in general, a clear benefit because the finding is associated with a high probability of malignancy (16%, which is higher than any other finding type) and a biopsy positive predictive value of approximately 50% (5,24,33). Our distribution of recalled finding types after implementation of DBT is consistent with prior studies, except that we did not find any change in recall examinations for calcifications whereas others have noted an increase in this finding type (5).

There were limitations to this study. The patient population was not randomized, which could have introduced bias. Also, no cost or psychological analysis was performed to estimate impact on these outcomes. The study was performed early in our DBT experience and later results could be influenced by a learning curve after implementation of DBT. Finally, we used a state tumor registry and patients may have left the state to be treated. In summary, after the implementation of DBT, there was no change in the overall utilization rate of BI-RADS category 3 after recall from screening. However, there was an absolute mean reduction of 2.4 women per 1000 placed in short-term follow-up after DBT screening due to overall reduction of women recalled compared with those screened with DM alone.

Acknowledgment: We would like to thank Marie Synnestevedt, PhD, for data collection related to this article.

Disclosures of Conflicts of Interest: E.S.M. disclosed no relevant relationships. A.M.M. disclosed no relevant relationships. S.P.W. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: has received personal fees from Siemens. Other relationships: disclosed no relevant relationships. M.D.S. Activities related to the present article: has received grants from National Institutes of Health. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. E.F.C. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: has received personal fees from Hologic. Other relationships: disclosed no relevant relationships.

References

- 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.
- Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology* 2015;274(1):85–92.
- Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 2014;311(24):2499–2507.
- 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.
- 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(2):337–342.
- McCarthy AM, Kontos D, Synnestevedt M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *J Natl Cancer Inst* 2014;106(11):dju316.
- 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.
- 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.
- Sickles EA, D'Orsi CJ, Bassett LW, et al. ACR BI-RADS Mammography. In: ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Reston, Va: American College of Radiology, 2013.
- Lee CI, Wells CJ, Bassett LW. Cost minimization analysis of ultrasound-guided diagnostic evaluation of probably benign breast lesions. *Breast J* 2013;19(1):41–48.
- Brenner RJ, Sickles EA. Surveillance mammography and stereotactic core breast biopsy for probably benign lesions: a cost comparison analysis. *Acad Radiol* 1997;4(6):419–425.
- Chung CS, Giess CS, Gombos EC, et al. Patient compliance and diagnostic yield of 18-month unilateral follow-up in surveillance of probably benign mammographic lesions. *AJR Am J Roentgenol* 2014;202(4):922–927.
- Helvie MA, Pennes DR, Rebner M, Adler DD. Mammographic follow-up of low-suspicion lesions: compliance rate and diagnostic yield. *Radiology* 1991;178(1):155–158.
- Kalra VB, Wu X, Haas BM, Forman HP, Philpotts LE. Cost-effectiveness of tomosynthesis in annual screening mammography. *AJR Am J Roentgenol* 2016;207(5):1152–1155.
- 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.
- 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 Am J Roentgenol* 2014;203(3):687–693.
- Brandt KR, Craig DA, Hoskins TL, et al. Can digital breast tomosynthesis replace conventional diagnostic mammography views for screening without calcifications? a comparison study in a simulated clinical setting. *AJR Am J Roentgenol* 2013;200(2):291–298.
- Noroozian M, Hadjiiski L, Rahnema-Moghadam S, et al. Digital breast tomosynthesis is comparable to mammographic spot views for mass characterization. *Radiology* 2012;262(1):61–68.
- Zuckerman SP, Conant EF, Keller BM, et al. Implementation of synthesized two-dimensional mammography in a population-based digital breast tomosynthesis screening program. *Radiology* 2016;281(3):730–736.
- Powell JL, Hawley JR, Lipari AM, Yildiz VO, Erdal BS, Carkaci S. Impact of the addition

- of digital breast tomosynthesis (DBT) to standard 2D digital screening mammography on the rates of patient recall, cancer detection, and recommendations for short-term follow-up. *Acad Radiol* 2017;24(3):302–307.
21. 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(1):109–116.
 22. 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.
 23. 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(6):737–743.
 24. Freer PE, Niell B, Rafferty EA. Preoperative tomosynthesis-guided needle localization of mammographically and sonographically occult breast lesions. *Radiology* 2015;275(2):377–383.
 25. Fleiss JL, Levin B, Paik MC. *Statistical methods for rates and proportions*. Hoboken, NJ: Wiley, 2003.
 26. Raghu M, Durand MA, Andrejeva L, et al. Tomosynthesis in the diagnostic setting: changing rates of BI-RADS final assessment over time. *Radiology* 2016;281(1):54–61.
 27. Recommendation Summary. U.S. Preventive Services Task Force. November 2009. <http://www.uspreventiveservicestaskforce.org/Page/Topic/recommendation-summary/breast-cancer-screening>. Accessed December 8, 2015.
 28. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. *JAMA* 2015;314(15):1599–1614.
 29. Lippert A, Peppard HR, Rochman CM, Harvey JA, Patrie J, Nicholson B. Tomosynthesis increases the use of BI-RADS category 3 [abstr]. In: Radiological Society of North America Scientific Assembly and Annual Meeting Program. Oak Brook, Ill: Radiological Society of North America, 2015; 225.
 30. Yasmeen S, Romano PS, Pettinger M, et al. Frequency and predictive value of a mammographic recommendation for short-interval follow-up. *J Natl Cancer Inst* 2003;95(6):429–436.
 31. Kerlikowske K, Smith-Bindman R, Abraham LA, et al. Breast cancer yield for screening mammographic examinations with recommendation for short-interval follow-up. *Radiology* 2005;234(3):684–692.
 32. Partyka L, Lourenco AP, Mainiero MB. Detection of mammographically occult architectural distortion on digital breast tomosynthesis screening: initial clinical experience. *AJR Am J Roentgenol* 2014;203(1):216–222.
 33. Ray KM, Turner E, Sickles EA, Joe BN. Suspicious findings at digital breast tomosynthesis occult to conventional digital mammography: imaging features and pathology findings. *Breast J* 2015;21(5):538–542.