Multicenter Evaluation of Breast Cancer Screening with Digital Breast Tomosynthesis in Combination with Synthetic versus Digital Mammography

Samantha P. Zuckerman, MD • Brian L. Sprague, PhD • Donald L. Weaver, MD • Sally D. Herschorn, MD • Emily F. Conant, MD

From the Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce St, 1 Silverstein Place, Philadelphia, PA 19104 (S.P.Z., E.F.C.); and Departments of Surgery (B.L.S.), Pathology (D.L.W.), and Radiology (S.D.H.), University of Vermont, University of Vermont Cancer Center, Burlington, Vt. Received January 30, 2020; revision requested March 17; revision received August 11; accepted August 21. **Address correspondence to** S.P.Z. (e-mail: *Samantha.zuckerman@uphs.upenn.edu*).

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

See also the editorial by Lång in this issue.

Background: Digital breast tomosynthesis (DBT) combined with digital mammography (DM) is increasingly used in the United States instead of DM alone for breast cancer screening. Early screening outcomes incorporating synthetic mammography (SM) with DBT have suggested that SM is an acceptable non–radiation dose alternative to DM.

Purpose: To compare multicenter outcomes from breast cancer screening with SM/DBT versus DM/DBT.

Materials and Methods: This was a retrospective study of consecutive screening mammograms obtained at two institutions. Eligible studies consisted of 34 106 DM/DBT examinations between October 3, 2011, and October 31, 2014, and 34 180 SM/DBT examinations between January 7, 2015, and February 2, 2018, at the University of Pennsylvania and 51 148 DM/DBT examinations between January 1, 2012, and May 31, 2016, and 31 929 SM/DBT examinations between June 1, 2016, and March 30, 2018, at the University of Vermont. Demographics of women who attended screening and results from screening were recorded. Recall rate, biopsy rate, false-negative rate, cancer detection rate, positive predictive value, sensitivity, and specificity were calculated according to modality and institution. Descriptive statistics, χ^2 tests, and logistic regression were used in analysis.

Results: The study included 151 363 screening examinations among 151 363 women (mean age, 58.1 years ± 10.9 [standard deviation]). The unadjusted recall rate was lower with SM/DBT than with DM/DBT (7.0% [4630 of 66 109 examinations] for SM/DBT vs 7.9% [6742 of 85 254 examinations] for DM/DBT; $P < .01$). However, after multivariable adjustment, SM/DBT was associated with a slightly higher recall rate compared with DM/DBT (adjusted odds ratio [OR], 1.06; adjusted 95% CI: 1.01, 1.11; $P = .02$). Similarly, after multivariable adjustment, SM/DBT was associated with slightly lower specificity compared with DM/DBT (adjusted OR, 0.95; adjusted 95% CI: 0.90, 0.99; $P = .02$). There was no statistically significant difference in biopsy rate ($P = .54$), false-negative rate ($P = .38$), cancer detection rate ($P = .55$), invasive or in situ cancer detection rate ($P = .52$ and $P = .98$, respectively), positive predictive value ($P = .78$), or sensitivity ($P = .33$) for SM/DBT versus DM/DBT overall or within either institution ($P > .05$ for all).

Conclusion: Breast cancer screening performance is maintained within benchmarks when synthetic mammography replaces digital mammography in digital breast tomosynthesis imaging.

©RSNA, 2020

Online supplemental material is available for this article.

Since U.S. Food and Drug Administration approval in 2011, digital breast tomosynthesis (DBT) has increasingly been used for breast cancer screening and diagnostic purposes (1–3). DBT has been shown in European prospective screening studies to improve cancer detection rates and in American retrospective screening studies to lower recall rates in women of all ages and breast densities compared with screening with digital mammography (DM) alone (4–13). However, the combined DM/DBT examination results in a twofold increase in radiation dose to women compared with screening with DM alone (14–16).

Synthetic mammography (SM) was developed to decrease the radiation dose to women undergoing DBT. SM is not meant to be an exact replica of DM; the reconstruction algorithm purposefully enhances areas, such as calcifications and potential areas of distortion, to make them more conspicuous to the reader (Figs 1, 2) (17,18). The synthetic reconstruction has led some imagers to anecdotally express concerns about SM image quality, specifically in regard to calcification detection and characterization (19). Concerns include both missing lower optical density, faint calcifications that represent ductal carcinoma in situ, and overcalling pseudocalcifications, lesions that are normal benign structures such as ligaments or fibroglandular tissue that appear to represent calcified lesions due to the reconstruction (Fig 1) (17–21). Although these concerns persist, a recent reader study (22) found no difference in specificity or sensitivity in SM/DBT versus DM detection and characterization of calcifications. An additional reader study (23)

Abbreviations

BI-RADS = Breast Imaging Reporting and Data System, BMI = body mass index, DBT = digital breast tomosynthesis, DM = digital mammography, OR = odds ratio, SM = synthetic mammography

Summary

In this multicenter retrospective review, synthetic mammography combined with digital breast tomosynthesis (DBT) maintained acceptable recall rates and cancer detection rates compared with digital mammography combined with DBT.

Key Results

- Synthetic mammography (SM) with digital breast tomosynthesis (DBT) maintained screening outcomes within acceptable benchmarks (eg, recall rate of 7%, biopsy rate of 1.4%, and cancer detection rate of 5.6 per 1000 women screened).

- In an adjusted analysis, SM with DBT had a slightly higher recall rate (odds ratio, 1.06; *P* = .02) and slightly decreased specificity (odds ratio, 0.95; *P* = .02) compared to digital mammography with DBT.

found no difference in SM/DBT versus DM/DBT calcification detection and characterization.

Retrospective studies, prospective studies, and reader studies from the early clinical implementation of SM/DBT have thus far demonstrated no significant difference in overall screening outcomes compared with DM/DBT (14,24–28). However, these studies have some limitations. Early retrospective studies were from single sites and included a smaller number of SM/DBT examinations than DM/DBT examinations (Zuckerman et al [14]: 5366 SM/DBT examinations vs 15 571 DM/DBT examinations; Aujero et al (24): 16 173 SM/DBT examinations vs 30 561 DM/DBT and 32 076 DM examinations). Simon et al (26) published a cancer-enriched reader study comparing the performance of SM/DBT with that of DM/DBT but only included 189 women. Caumo et al (27) published a prospective study that compared 16 666 SM/DBT examinations with 14 423 DM examinations but did not compare SM/DBT with DM/DBT. The largest prospective studies had all women undergo DM, SM, and DBT and then performed a reader study to determine outcomes of combination imaging (Bernardi et al [25] included 9672 women and compared DM alone, DM/DBT, and SM/DBT; Skaane et al (28) included 24 301 women and compared DM, DM plus computer-aided detection, DM/DBT, and SM/DBT).

The purpose of this study was to compare multicenter outcomes from SM/DBT with those from DM/DBT in breast cancer screening.

Materials and Methods

This retrospective Health Insurance Portability and Accountability Act–compliant study was approved by the human subjects institutional review boards at both the Hospital of the University of Pennsylvania (center 1) and the University of Vermont (center 2). The requirement to obtain informed consent was waived.

Subgroups of data from center 1 have been previously published (4–6,14,29–32). These studies did not specifically evaluate the performance of SM/DBT versus DM/DBT and instead examined recall rates and cancer detection rates of DM versus DBT with DM. The studies included 24 767 examinations (5), 26 299 examinations (6,29,30), 44 468 examinations (31), 23 206 examinations (32), and 11 623 examinations (4). An additional study examined early outcomes of SM/DBT versus DM/DBT with the SM/DBT population limited to January 7, 2015, to June 30, 2015, a study overlap of 31 666 examinations (14). Our current study represents an expanded analysis with almost three additional years of outcomes.

No previous studies have used the SM/DBT data from center 2. Data from the 51 562 DM/DBT examinations from center 2 included in this study were previously included within a larger analysis of malignant and benign lesion detection rates with DBT using data pooled from multiple institutions (33), as well as a pooled analysis of learning curve in DBT performance within the Breast Cancer Surveillance Consortium (34). Data from 18 983 DM/DBT examinations from center 2 included in this study were also used in pooled analyses of DBT performance within the Population-based Research Optimizing Screening through Personalized Regimens, or PROSPR, Consortium (4,32).

Results from consecutive screening examinations from both institutions were reviewed. Center 1 performed 37 615 DM/DBT examinations between October 3, 2011, and October 31, 2014, and 37 854 SM/DBT examinations between January 7, 2015, and February 2, 2018. The gap between DM/DBT and SM/DBT screening was secondary to an overlap period where DM and SM were performed with DBT in November and December 2014; these studies were not included in the analysis. Center 2 performed 51 562 DM/DBT examinations between January 5, 2012, and May 31, 2016, and 37 227 SM/DBT examinations between June 1, 2016, and March 30, 2018. The difference in studies performed over time at center 2 is secondary to less DBT capability during the DM/DBT period versus the SM/DBT period. All women during all time points underwent DBT screening at both institutions. If a woman declined a DBT study during the study period for personal reasons, she was not included in our analysis. At each site, one faculty member finalized each study; trainees were often involved in preliminary interpretations. The time between studies varied across women.

Examinations were excluded if they were missing Breast Imaging Reporting and Data Systems (BI-RADS) final assessments (four were excluded from center 1 and 14 from center 2). Examinations interpreted by radiologists who had interpreted fewer than 100 examinations in each category were also excluded (7179 were excluded from center 1 and 5698 from center 2). The final number of examinations for center 1 was 34 106 DM/DBT studies and 34 180 SM/DBT studies; the final number of examinations for center 2 was 51 148 DM/DBT studies and 31 929 SM/DBT studies. The total number of studies was 85 254 DM/DBT and 66 109 SM/DBT. Examinations performed at center 1 were interpreted by six radiologists with 6–31 years of experience (including author E.F.C., with 31 years of experience). Examinations performed at center 2 were interpreted by six radiologists with 12–33 years of experience (including author S.D.H., with
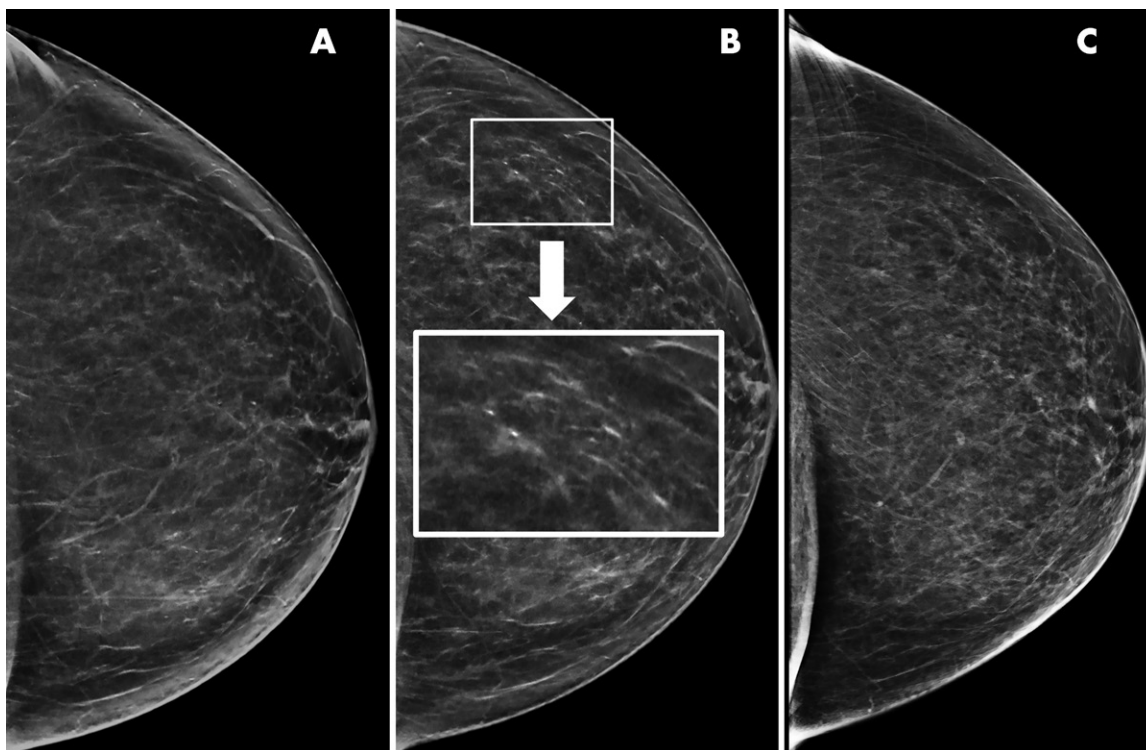
Figure 1: Example of artifactual pseudocalcifications on synthetic images in a 66-year-old woman who presented for screening. *A,* Screening mammogram obtained with synthetic imaging in 2019 (craniocaudal view). Finding was classified as Breast Imaging Reporting and Data System (BI-RADS) category 1. *B,* Screening mammogram obtained with synthetic imaging in 2020 (craniocaudal view). New calcifications in a linear distribution are seen in the outer breast. Finding was classified as BI-RADS category 0. Inset image was obtained at ×3 magnification. *C,* Diagnostic digital mammogram obtained in 2020 (craniocaudal view) reveals no calcifications in the outer breast. Previously seen calcification-like findings represent artifact from enhancement of ligaments or fibroglandular tissue on synthetic reconstruction. Finding was classified as BI-RADS category 1.

32 years of experience). All radiologists in both groups are qualified according to the American College of Radiology Mammography Quality and Standards Act and trained in DBT interpretation as mandated by the Food and Drug Administration.

All studies were performed with the same imaging system (Selenia Dimensions 2D/3D system; Hologic, Marlborough, Mass), and all SM images were generated from the DBT images by using industry-specific processing software (C-view software module, Hologic).

For each study, age, race and/or ethnicity, body mass index (BMI), first-degree relative with breast cancer, breast density using BI-RADS terminology (35), baseline examination status, and final BI-RADS assessment were recorded. If a woman was recalled from screening, the outcome of the subsequent diagnostic examination, including if a biopsy was performed and final histologic findings from biopsy or final excision, were recorded. The final histologic finding was classified as either benign (including atypia) or malignant (invasive or in situ). Recall rate (percentage), biopsy rate, and cancer detection rate per 1000 women screened, positive predictive value of recall, and positive predictive value of biopsies performed were calculated for each institution. Sensitivity (proportion of screening-detected cancers among all breast cancers) and specificity (proportion of negative screening studies among all women without breast cancer) of DM/DBT and SM/DBT for each institution

and in aggregate were also calculated. Each hospital's electronic medical record or radiology information system was used to extract data (center 1: Epic Systems, Madison, Wis; center 2: GE Centricity RIS-IC, GE Healthcare, Burlington, Vt).

Statistical Analysis

Descriptive statistics and χ^2 tests were used to characterize and compare population characteristics according to imaging modality and institution. Unadjusted screening performance measures (including recall rate, biopsy rate, cancer detection rate, false-negative rate, positive predictive value, sensitivity, and specificity) were calculated according to standard BI-RADS definitions (36). Cancer diagnoses for all outcomes were based on a 1-year period following the screening examination. Logistic regression was used to examine the association between imaging modality (SM/DBT vs DM/DBT) and various screening outcomes. We calculated unadjusted odds ratios (ORs) and 95% CIs, as well as adjusted ORs and 95% CIs, from separate models, including covariates for age, race, BMI, breast density, family history of breast cancer, baseline examination status, DBT round, radiologist, and institution. Tests of statistical significance were two-sided, with $P < .05$ considered indicative of a statistically significant difference. Cross-product interaction terms were evaluated to test for effect modification by institution. Interactions with $P < .05$ were considered statistically significant. SAS software
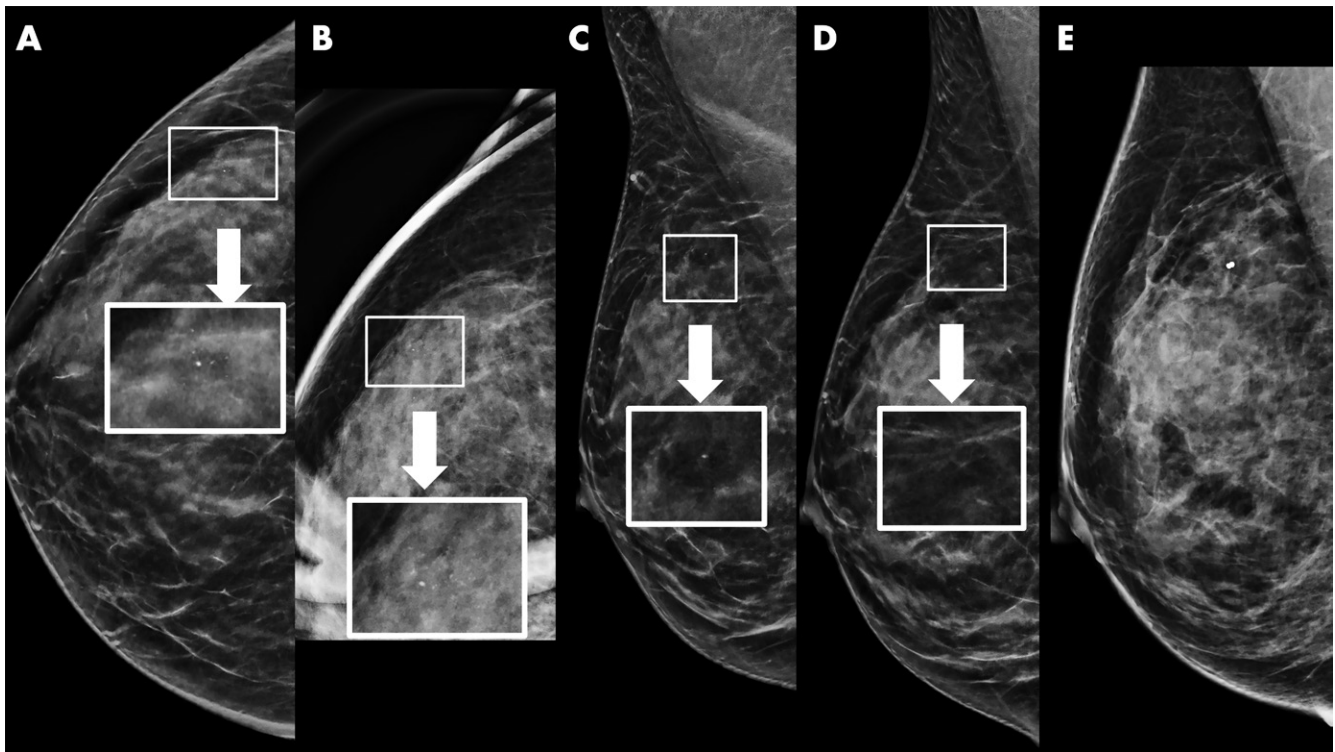
Figure 2: Example of synthetic imaging calcification enhancement in a 43-year-old woman who presented for screening. *A,* Screening mammogram obtained with synthetic imaging (craniocaudal view). New faint amorphous calcifications are seen in the outer breast. Finding was classified as Breast Imaging Reporting and Data System (BI-RADS) category 0. Inset image was obtained at ×3 magnification. *B,* Diagnostic mammogram, two-dimensional magnification view (craniocaudal view), confirms the presence of calcifications; however, they are of equal or less conspicuity compared with that seen with synthetic imaging. Inset image was obtained at ×3 magnification. Finding was classified as BI-RADS category 4. *C,* Screening mammogram obtained with synthetic imaging (mediolateral oblique view). New faint amorphous calcifications are seen in the superior breast. They are less conspicuous compared with those on the craniocaudal view in *A.* Finding was classified as BI-RADS category 0. Inset image was obtained at ×3 magnification. *D,* Diagnostic digital mammogram obtained at recall (mediolateral view). Calcifications are much less conspicuous on two-dimensional digital mediolateral view. Inset image was obtained at ×3 magnification. Finding was classified as BI-RADS category 4. *E,* Diagnostic mammogram obtained with two-dimensional imaging after biopsy (digital mediolateral view). The clip is in appropriate location. Pathologic examination showed ductal carcinoma in situ, intermediate grade.

Results

(version 9.4; SAS Institute, Cary, NC) was used for all statistical analyses.

We included 151 363 screening examinations in 151 363 women (mean age, 58.1 years ± 10.9 [standard deviation]) (Table 1). There were small but statistically significant differences in age distribution, family history of breast cancer, breast density, and baseline examination status between both institutions. However, there was a substantial difference in reported race and/or ethnicity, with center 1 having a larger percentage of Black women (33 520 of 68 286, 49.7%) and center 2 having a larger percentage of White women (78 550 of 83 077, 95.0%) (*P* < .01). There was also a difference in BMI across sites, with center 1 having a larger proportion (24 106 of 68 286, 35.3%) of obese women (BMI ≥30 kg/m²) compared with center 2 (19 335 of 83 077, 23.3%) (*P* < .01). When all DM/DBT examinations from both institutions were compared with all SM/DBT examinations from both institutions, there were small but statistically significant differences in age distribution, race and/or ethnicity, BMI, family history of breast cancer, breast density, and baseline examination status (any mammographic examination, in-

cluding DM or DBT examination). As expected, compared with DM/DBT examinations, SM/DBT examinations were less likely to be an individual's first DBT examination (*P* < .01) because SM/DBT was implemented after DM/DBT at each institution.

The unadjusted recall rate was lower with SM/DBT than with DM/DBT (7.0% [4630 of 66 109] for SM/DBT vs 7.9% [6742 of 85 254] for DM/DBT; *P* < .01) (Table 2). However, after multivariable adjustment for population characteristics, SM/DBT was associated with a slightly higher recall rate compared with DM/DBT (adjusted OR: 1.06; 95% CI: 1.01, 1.11; *P* = .02), indicating that the likelihood of recall for diagnostic imaging was approximately 6% higher in SM/DBT examinations compared with DM/DBT examinations after controlling for differences in population characteristics. There was a statistically significant interaction by institution (*P* < .01), such that the multivariable-adjusted increase in recall rate with SM/DBT was only experienced at center 2 (adjusted OR: 1.18; 95% CI: 1.11, 1.26).

The biopsy rate for SM/DBT examinations was 1.4% (924 of 66 109). The biopsy rate for DM/DBT examinations was 1.7% (1417 of 85 254). There was no statistically significant difference in overall biopsy rate for SM/DBT versus DM/DBT

Table 1: Characteristics of Women Undergoing Screening according to Modality and Site

| Characteristic | All Examinations (n = 151 363) | Modality | | | Site | | |
		DM/DBT (n = 85 254)	SM/DBT (n = 66 109)	P Value	Center 1 (n = 68 286)	Center 2 (n = 83 077)	P Value
Age (y)*				<.01			<.01
<40	2203 (1.5)	1275 (1.5)	928 (1.4)		1448 (2.1)	755 (0.9)	
40–49	36 654 (24.2)	21 956 (25.8)	14 698 (22.2)		18 026 (26.4)	18 628 (22.4)	
50–59	48 773 (32.2)	28 342 (33.2)	20 431 (30.9)		21 637 (31.7)	27 136 (32.7)	
60–69	41 403 (27.4)	22 272 (26.1)	19 131 (28.9)		18 033 (26.4)	23 370 (28.1)	
70–79	18 383 (12.1)	9378 (11.0)	9005 (13.6)		7659 (11.2)	10 724 (12.9)	
80+	3947 (2.6)	2031 (2.4)	1916 (2.9)		1483 (2.2)	2464 (3.0)	
Race and/or ethnicity*				<.01			<.01
White	106 779 (71.1)	62 504 (73.8)	44 275 (67.6)		28 229 (41.9)	78 550 (95.0)	
Black	33 853 (22.5)	17 167 (20.3)	16 686 (25.5)		33 520 (49.7)	333 (0.4)	
Asian	3194 (2.1)	1624 (1.9)	1570 (2.4)		2215 (3.3)	979 (1.2)	
Other	3713 (2.5)	2039 (2.4)	1674 (2.6)		2165 (3.2)	1548 (1.9)	
Hispanic	2663 (1.8)	1366 (1.6)	1297 (2.0)		1378 (2.0)	1285 (1.6)	
Missing	1161 (0.8)	554 (0.7)	607 (0.9)		779 (1.1)	382 (0.5)	
BMI (kg/m²)				<.01			<.01
<18.5	2024 (1.5)	1221 (1.5)	803 (1.4)		894 (1.5)	1130 (1.4)	
18.5–24.9	53 279 (38.5)	32 394 (39.9)	20 885 (36.6)		18 297 (30.7)	34 982 (44.5)	
25.0–29.9	39 569 (28.6)	23 386 (28.8)	16 183 (28.4)		16 322 (27.4)	23 247 (29.5)	
30.0–34.9	23 739 (17.2)	13 476 (16.6)	10 263 (18.0)		11 867 (19.9)	11 872 (15.1)	
≥35.0	19 702 (14.2)	10 809 (13.3)	8893 (15.6)		12 239 (20.5)	7463 (9.5)	
Missing	13 050 (8.6)	3968 (4.7)	9082 (13.7)		8667 (12.7)	4383 (5.3)	
First-degree relative with breast cancer				<.01			<.01
No	107 462 (77.7)	62 928 (79.1)	44 534 (75.8)		49 190 (79.0)	58 272 (76.7)	
Yes	30 805 (22.3)	16 577 (20.9)	14 228 (24.2)		13 085 (21.0)	17 720 (23.3)	
Missing	13 096 (8.7)	5749 (6.7)	7347 (11.1)		6011 (8.8)	7085 (8.5)	
Breast density				<.01			<.01
Non-dense	103 668 (68.5)	58 793 (69.0)	44 875 (67.9)		47 723 (70.0)	55 945 (67.3)	
Dense	47 631 (31.4)	26 403 (31.0)	21 228 (32.1)		20 520 (30.0)	27 111 (32.6)	
Missing	64 (0.04)	58 (0.07)	6 (0.009)		43 (0.06)	21 (0.03)	
Baseline examination				<.01			<.01
No	133 423 (97.2)	75 369 (97.9)	58 054 (96.2)		58 203 (97.5)	75 220 (96.9)	
Yes	3911 (2.8)	1640 (2.1)	2271 (3.8)		1503 (2.5)	2408 (3.1)	
Missing	14 029 (9.3)	8245 (9.7)	5784 (8.8)		8580 (12.6)	5449 (6.6)	
DBT round				<.01			<.01
1	60 090 (39.7)	46 395 (54.4)	13 695 (20.7)		27 984 (41.0)	32 106 (38.7)	
2	40 784 (26.9)	26 770 (31.4)	14 014 (21.2)		17 518 (25.7)	23 266 (28.0)	
3+	50 489 (33.4)	12 089 (14.2)	38 400 (58.1)		22 784 (33.4)	27 705 (33.4)	

Note.—Except where indicated, data are numbers of women, with percentages in parentheses. Center 1: University of Pennsylvania, Philadelphia, Pa; center 2: University of Vermont, Burlington, Vt. BMI = body mass index, DBT = digital breast tomosynthesis, DM = digital mammography, SM = synthetic mammography.

* Percentages were calculated based on non-missing data.

(P = .54) or within either institution (Table 2). There was also no statistically significant difference in overall benign biopsy rate for SM/DBT versus DM/DBT (P = .25). However, there was a statistically significant interaction by institution (P < .01), with a lower benign biopsy rate for SM/DBT compared with DM/DBT at center 1 (adjusted OR: 0.81; 95% CI: 0.67, 0.98; P = .03), indicating that the likelihood of a benign biopsy after SM/DBT was approximately 19% lower than that with DM/DBT after controlling for differences in population characteristics. No statistically significant difference was observed in the benign biopsy rate between SM/DBT and DM/DBT at center 2 (P = .24).

The cancer detection rate for SM/DBT examinations was 5.6 per 1000 woman screened (368 of 66 109). The cancer detection rate for DM/DBT examinations was 5.9 per 1000 women screened (499 of 85 254). There was no statistically significant difference in cancer detection rate (P = .55) for SM/DBT versus DM/DBT in aggregate or at either

Table 2: Recall Rate, Overall Biopsy Rate, Benign Biopsy Rate, and False-Negative Rate of Screening with SM/DBT versus DM/DBT

Center and Modality	No. of Screening Examinations	Recall Rate			Overal Biopsy Rate			Benign Biopsy Rate			False-Negative Rate		
		No. of Recalls	OR	AOR*	No. of Biopsies	OR	AOR*	No. of Benign Biopsies	OR	AOR*	No. of False-Negative Findings (per 1000)	OR	AOR*
Combined													
DM/DBT	85 254	6742 (7.9) [7.7, 8.1]	Ref	Ref	1417 (1.7) (1.6, 1.7)	Ref	Ref	918 (1.1) [1, 1.1]	Ref	Ref	70 (0.8) [0.6, 1]	Ref	Ref
SM/DBT	66 109	4630 (7.0) [6.8, 7.2]	0.88 [0.84, 0.91]	1.06 [1.01, 1.11]	924 (1.4) [1.3, 1.5]	0.84 [0.77, 0.91]	0.97 [0.88, 1.07]	556 (0.8) [0.8, 0.9]	0.78 [0.70, 0.87]	0.93 [0.82, 1.05]	49 (0.7) [0.5, 0.9]	0.90 [0.63, 1.30]	0.82 [0.53, 1.28]
Center 1													
DM/DBT	34 106	3040 (8.9) [8.6, 9.2]	Ref	Ref	639 (1.9) [1.7, 2]	Ref	Ref	434 (1.3) [1.2, 1.4]	Ref	Ref	19 (0.6) [0.3, 0.8]	Ref	Ref
SM/DBT	34 180	2384 (7) [6.7, 7.2]	0.77 [0.73, 0.81]	0.97 [0.91, 1.04]	472 (1.4) [1.3, 1.5]	0.73 [0.65, 0.83]	0.89 [0.77, 1.04]	280 (0.8) [0.7, 0.9]	0.64 [0.55, 0.75]	0.81 [0.67, 0.98]	27 (0.8) [0.5, 1.1]	1.42 [0.79, 2.55]	1.04 [0.51, 2.13]
Center 2													
DM/DBT	51 148	3702 (7.2) [7, 7.5]	Ref	Ref	778 (1.5) [1.4, 1.6]	Ref	Ref	484 (1) [0.9, 1]	Ref	Ref	51 (1) [0.7, 1.3]	Ref	Ref
SM/DBT	31 929	2246 (7) [6.8, 7.3]	0.97 [0.92, 1.02]	1.18 [1.11, 1.26]	452 (1.4) [1.3, 1.5]	0.93 [0.83, 1.05]	1.09 [0.95, 1.25]	276 (0.9) [0.8, 1]	0.91 [0.79, 1.06]	1.11 [0.93, 1.32]	22 (0.7) [0.4, 1]	0.69 [0.42, 1.14]	0.65 [0.36, 1.19]

Note.—Numbers in parentheses are percentages, and numbers in brackets are 95% confidence intervals. There was a statistically significant interaction between institution (center 1: University of Pennsylvania, Philadelphia, Pa; center 2: University of Vermont, Burlington, Vt) and screening type in the adjusted model for both recall and biopsy (*P* < .01). AOR = adjusted OR, DBT = digital breast tomosynthesis, DM = digital mammography, OR = odds ratio, Ref = reference, SM = synthetic mammography.

* Adjusted for age, race, body mass index, breast density, first-degree family history, baseline versus subsequent examination, DBT round, practice, and radiologist.

institution (Table 3). There was also no statistically significant difference in screening-detected invasive cancers (*P* = .52) or in situ cancers (*P* = .98) for SM/DBT versus DM/DBT in aggregate or at either institution. At each institution as well as in aggregate, there was no statistically significant difference in the false-negative rates between SM/DBT and DM/DBT (*P* = .38) (Table 3).

Among SM/DBT examinations, the positive predictive value of recall was 8% (368 of 4630) and the positive predictive value of biopsies performed was 39.8% (368 of 924). There was no statistically significant difference in positive predictive value of recall (*P* = .78) or positive predictive value of biopsies performed (*P* = .24) for SM/DBT versus DM/DBT in aggregate or at either institution (Table 4).

Overall sensitivity was unchanged between DM/DBT (87.7%, 499 of 569) and SM/DBT (88.3%, 368 of 417) (unadjusted OR: 1.05 [95% CI: 0.71, 1.56], *P* = .79; adjusted OR: 1.27 [95% CI: 0.79, 2.06], *P* = .33). Similarly, there was no difference between DM/DBT and SM/DBT sensitivity at either institution (*P* = .82 for center 1 and *P* = .17 for center 2) (Table E1 [online]).

Paralleling results seen in recall rate, overall unadjusted specificity was higher for SM/DBT (93.5%, 51 430 of 65 692 examinations) than for DM/DBT (92.6%, 78 442 of 84 685 examinations) (unadjusted OR: 1.15; 95% CI: 1.10, 1.19; *P* <

.01). However, when this result was controlled for differences in population characteristics, the specificity of SM/DBT was lower than that of DM/DBT (adjusted OR: 0.95; 95% CI: 0.90, 0.99; *P* = .02). No difference was seen in specificity at center 1 (*P* = .34). At center 2, there was no difference in the unadjusted specificity of DM/DBT versus SM/DBT (*P* = .30), but there was a lower specificity of SM/DBT compared with DM/DBT when the result was adjusted for population characteristics (adjusted OR: 0.84; 95% CI: 0.79, 0.90; *P* < .01) (Table E1 [online]).

Discussion

To our knowledge, our study is the largest study and only multicenter study in the United States to compare screening outcomes using synthetic mammography (SM) with digital breast tomosynthesis (DBT) to those from digital mammography (DM) with DBT. As shown in smaller retrospective population studies and in reader studies, breast cancer screening performance is similar when SM replaces DM in combination DBT screening. All major screening outcomes with SM/DBT remained within acceptable benchmarks (37) at both institutions, such as recall rate (7% [4630 of 66 109 examinations]; adjusted odds ratio [OR]: 1.06; 95% CI: 1.01, 1.11), biopsy rate (1.4% [924 of 66 109 examinations]; adjusted OR: 0.97; 95% CI: 0.88, 1.07), and cancer detection rate (5.6 per 1000 women screened [368 of 66 109 examinations]; adjusted OR:

Table 3: Cancer Detection at Screening with SM/DBT versus DM/DBT

Center and Modality	No. of Screening Examinations	Screening-detected Cancers			Screening-detected Invasive Cancers			Screening-detected DCIS		
		No. of Cancers (per 1000)	OR	AOR*	No. of Cancers (per 1000)	OR	AOR*	No. of Cancers (per 1000)	OR	AOR*
Combined										
DM/DBT	85 254	499 (5.9) [5.3, 6.4]	Ref	Ref	374 (4.4) [3.9, 4.8]	Ref	Ref	125 (1.5) [1.2, 1.7]	Ref	Ref
SM/DBT	66 109	368 (5.6) (5, 6.1)	0.95 [0.83, 1.09]	1.05 [0.89, 1.23]	278 (4.2) [3.7, 4.7]	0.96 [0.82, 1.12]	1.06 [0.88, 1.28]	90 (1.4) [1.1, 1.6]	0.93 [0.71, 1.22]	1.00 [0.72, 1.38]
Center 1										
DM/DBT	34 106	205 (6) [5.2, 6.8]	Ref	Ref	150 (4.4) [3.7, 5.1]	Ref	Ref	55 (1.6) [1.2, 2]	Ref	Ref
SM/DBT	34 180	192 (5.6) [4.8, 6.4]	0.93 [0.77, 1.14]	1.07 [0.84, 1.35]	131 (3.8) [3.2, 4.5]	0.87 [0.69, 1.10]	0.99 [0.74, 1.31]	61 (1.8) [1.3, 2.2]	1.11 [0.77, 1.59]	1.30 [0.84, 2.01]
Center 2										
DM/DBT	51 148	294 (5.7) [5.1, 6.4]	Ref	Ref	224 (4.4) (3.8, 5)	Ref	Ref	70 (1.4) [1, 1.7]	Ref	Ref
SM/DBT	31 929	176 (5.5) [4.7, 6.3]	0.96 [0.80, 1.16]	1.05 [0.84, 1.31]	147 (4.6) [3.9, 5.3]	1.05 [0.85, 1.30]	1.12 [0.87, 1.45]	29 (0.9) [0.6, 1.2]	0.66 [0.42, 1.02]	0.79 [0.48, 1.32]

Note.—Numbers in parentheses are percentages, and numbers in brackets are 95% confidence intervals. There was no statistically significant interaction between institution (center 1: University of Pennsylvania, Philadelphia, Pa; center 2: University of Vermont, Burlington, Vt) and screening type in the adjusted models for cancer detection ($P = .65$), invasive cancer detection ($P = .21$), or DCIS detection ($P = .17$). AOR = adjusted OR, DBT = digital breast tomosynthesis, DCIS = ductal carcinoma in situ, DM = digital mammography, OR = odds ratio, Ref = reference, SM = synthetic mammography.

* Adjusted for age, race, body mass index, breast density, first-degree family history of breast cancer, baseline versus subsequent examination, DBT round, practice, and radiologist.

1.05; 95% CI: 0.89, 1.23). There were no statistically significant differences in screening performance measures for SM/DBT and DM/DBT ($P > .05$ for all), with the exception of a slightly increased recall rate of SM/DBT compared with DM/DBT (OR, 1.06; $P = .02$) and slightly decreased specificity of SM/DBT compared with DM/DBT (OR, 0.95; $P = .02$).

In this study, the unadjusted recall rate of SM/DBT examinations was lower than that of DM/DBT examinations. This appeared to be due to a higher propensity for DM/DBT examinations to be a woman's first DBT examination (at which a higher recall rate is to be expected). This is most likely related to the incremental proficiency in using DBT over the years, as all DM/DBT examinations were performed and interpreted in the years before implementation of SM/DBT. After adjusting for covariates, including previous DBT studies, SM/DBT was associated with a slightly higher recall rate than DM/DBT. Similar to recall rate, unadjusted specificity of SM/DBT examinations was higher than that of DM/DBT examinations. However, when adjusted for population characteristics, the specificity with SM/DBT was lower than that with DM/DBT. This trend is not unexpected as specificity (the ratio of true-negative examinations and false-positive examinations) is closely related to recall rate.

Both differences in recall rate and specificity are entirely attributable to center 2; there was no difference in the adjusted recall rate or specificity between SM/DBT and DM/DBT at center 1. Notably, center 2 achieved low recall rates and high specificity with both DM/DBT and SM/DBT. Specific reasons for the slightly higher recall rate and lower specificity with SM/DBT at center 2 are unclear. Although we did not evaluate the specific finding type that prompted recall, reader experience with SM and calcification interpretation, specifically calling pseudocalcifications (Fig 1), may have contributed to the increase in recall rate. A limitation of our study is that we did not evaluate the mammographic finding type that prompted recall (specifically, calcified vs noncalcified lesions). Although we have previously published on this topic after early implementation of synthetic imaging (14), additional studies are recommended.

Some previous studies have observed a lower recall rate when SM replaces DM and questioned if the change in recall rate between SM and DM results in more false-negative screening examinations (14,19). Our study has shown no change in the rate of false-negative findings with SM/DBT compared with DM/DBT at either institution ($P = .43$); both recall rates and false-negative rates continue to be within national benchmarks (37). Further research with larger sample sizes is needed to confirm that false-negative rates are maintained within acceptable levels with SM/DBT.

Our study also demonstrates similar true-positive rates, false-positive rates, and sensitivity with SM/DBT compared with DM/DBT screening. All major screening outcomes remained within acceptable benchmarks (37): Overall cancer detection rate (including in situ and invasive cancer), positive predictive value of recall, and positive predictive value of

Table 4: Positive Predictive Value of SM/DBT versus DM/DBT Screening Examinations

Center and Modality	No. of Positive Examinations	No. of Biopsies	No. of Cancers Detected	PPV1 Percentage	PPV1 OR	PPV1 AOR*	PPV3 Percentage	PPV3 OR	PPV3 AOR*
Combined									
DM/DBT	6742	1417	499	7.4 (6.8, 8)	Ref	Ref	35.2 (32.7, 37.7)	Ref	Ref
SM/DBT	4630	924	368	8.0 (7.2, 8.7)	1.08 (0.94, 1.24)	1.02 (0.87, 1.21)	39.8 (36.7, 43)	1.22 (1.03, 1.45)	1.14 (0.92, 1.41)
Center 1									
DM/DBT	3040	639	205	6.7 (5.9, 7.6)	Ref	Ref	32.1 (28.5, 35.7)	Ref	Ref
SM/DBT	2384	472	192	8.1 (7, 9.1)	1.21 (0.99, 1.49)	1.11 (0.86, 1.43)	40.7 (36.2, 45.1)	1.45 (1.13, 1.86)	1.34 (0.96, 1.86)
Center 2									
DM/DBT	3702	778	294	7.9 (7.1, 8.8)	Ref	Ref	37.8 (34.3, 41.2)	Ref	Ref
SM/DBT	2246	452	176	7.8 (6.7, 8.9)	0.99 (0.81, 1.20)	0.94 (0.74, 1.19)	38.9 (34.4, 43.4)	1.05 (0.83, 1.33)	0.97 (0.72, 1.32)

Note.—Numbers in parentheses are 95% confidence intervals. There was no statistically significant interaction between institution (center 1: University of Pennsylvania, Philadelphia, Pa; center 2: University of Vermont, Burlington, Vt) and screening type in the adjusted models for PPV1 (P = .29) or PPV3 (P = .13). AOR = adjusted OR, DBT = digital breast tomosynthesis, DM = digital mammography, OR = odds ratio, PPV1 = positive predictive value of recall, PPV3 = and positive predictive value of biopsies performed, Ref = reference, SM = synthesized mammography.

* Adjusted for age, race, body mass index, breast density, first-degree family history, baseline versus subsequent examination, DBT round, and practice.

biopsies performed remained comparable at both institutions and in aggregate. Overall biopsy rate was also unchanged in aggregate as well as at the institutional level. The benign biopsy rate was also unchanged overall, although there was a small decrease in benign biopsy rate at center 1 when comparing SM/DBT to DM/DBT; this decrease in benign biopsies is not well understood. Further evaluation including specific lesion types at recall and biopsy may help explain this finding.

There was a significant difference in the distribution of population characteristics (race and/or ethnicity and BMI) between institutions. This reflects a difference in the communities served by each institution: Center 1 serves an urban population and center 2 serves a more suburban and/or rural population. Despite these differences in populations, overall screening outcomes were similar between both institutions. Our results show that screening outcomes with SM/DBT can be maintained in a variety of practice settings with diverse populations.

Our study is not without limitations. All examinations were performed with equipment from one vendor. Future research should be conducted to determine if our results are translatable to other vendors with different synthetic reconstruction algorithms. In addition, our study was essentially a natural experiment in which SM/DBT replaced DM/DBT at each institution. All SM/DBT studies were performed after each site had experience with DM/DBT studies; some of the results may be confounded by our readers' increased experience interpreting DBT examinations in the SM/DBT population versus the DM/DBT population. We were able to control for measured confounders (including DBT round) with rich woman-level data. However, we cannot rule out all sources of bias as radiologists who have had more experience with DBT may have evolved in how they interpret DBT examinations, with a change in their recall threshold. Nevertheless, our results should reflect the experiences of practices as they transition from DM/DBT to SM/DBT. Our results may not reflect outcomes when practices transition directly from DM to SM/DBT.

In summary, our study represents the first multicenter retrospective review of population-based outcomes using synthetic mammography (SM) combined with digital breast tomosynthesis (DBT) versus digital mammography (DM) combined with DBT. We found that SM/DBT maintains acceptable recall rates and cancer detection rates compared with DM/DBT.

References

1. Hardesty LA, Kreidler SM, Glueck DH. Digital Breast Tomosynthesis Utilization in the United States: A Survey of Physician Members of the Society of Breast Imaging. J Am Coll Radiol 2016;13(11S):R67–R73.

2. Gao Y, Babb JS, Toth HK, Moy L, Heller SL. Digital Breast Tomosynthesis Practice Patterns Following 2011 FDA Approval: A Survey of Breast Imaging Radiologists. Acad Radiol 2017;24(8):947–953.

3. Richman IB, Hoag JR, Xu X, et al. Adoption of Digital Breast Tomosynthesis in Clinical Practice. JAMA Intern Med 2019;179(9):1292.

4. 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.

5. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA 2014;311(24):2499–2507.

6. 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):dju316.

7. Skaane P, Bandos AI, Gullien R, et al. Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. Eur Radiol 2013;23(8):2061–2071.

8. 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.

9. 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.

10. Østerås BH, Martinsen ACT, Gullien R, Skaane P. Digital Mammography versus Breast Tomosynthesis: Impact of Breast Density on Diagnostic Performance in Population-based Screening. Radiology 2019;293(1):60–68.

11. Conant EF, Zuckerman SP, McDonald ES, et al. Five Consecutive Years of Screening with Digital Breast Tomosynthesis: Outcomes by Screening Year and Round. Radiology 2020;295(2):285–293.

12. Chong A, Weinstein SP, McDonald ES, Conant EF. Digital Breast Tomosynthesis: Concepts and Clinical Practice. Radiology 2019;292(1):1–14.

13. Marinovich ML, Hunter KE, Macaskill P, Houssami N. Breast Cancer Screening Using Tomosynthesis or Mammography: A Meta-analysis of Cancer Detection and Recall. J Natl Cancer Inst 2018;110(9):942–949.

14. 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):730–736https://doi.org/10.1148/radiol.2016160366.

15. Gennaro G, Bernardi D, Houssami N. Radiation dose with digital breast tomosynthesis compared to digital mammography: per-view analysis. Eur Radiol 2018;28(2):573–581.

16. Svahn TM, Houssami N, Sechopoulos I, Mattsson S. Review of radiation dose estimates in digital breast tomosynthesis relative to those in two-view full-field digital mammography. Breast 2015;24(2):93–99.

17. Ratanaprasatporn L, Chikarmane SA, Giess CS. Strengths and Weaknesses of Synthetic Mammography in Screening. RadioGraphics 2017;37(7):1913–1927.

18. Zuckerman SP, Maidment ADA, Weinstein SP, McDonald ES, Conant EF. Imaging With Synthesized 2D Mammography: Differences, Advantages, and Pitfalls Compared With Digital Mammography. AJR Am J Roentgenol 2017;209(1):222–229.

19. Gilbert FJ, Tucker L, Gillan MG, et al. Accuracy of Digital Breast Tomosynthesis for Depicting Breast Cancer Subgroups in a UK Retrospective Reading Study (TOMMY Trial). Radiology 2015;277(3):697–706.

20. Durand MA. Synthesized Mammography: Clinical Evidence, Appearance, and Implementation. Diagnostics (Basel) 2018;8(2):22.

21. Zuckerman SP, Sprague BL, Weaver DL, Herschorn SD, Conant EF. Survey Results Regarding Uptake and Impact of Synthetic Digital Mammography with Tomosynthesis in the Screening Setting. J Am Coll Radiol 2020;17(1 Pt A):31–37https://doi.org/10.1016/j.jacr.2019.07.020.

22. Lai YC, Ray KM, Lee AY, et al. Microcalcifications Detected at Screening Mammography: Synthetic Mammography and Digital Breast Tomosynthesis versus Digital Mammography. Radiology 2018;289(3):630–638.

23. Dodelzon K, Simon K, Dou E, et al. Performance of 2D Synthetic Mammography Versus Digital Mammography in the Detection of Microcalcifications at Screening. AJR Am J Roentgenol 2020;214(6):1436–1444https://doi.org/10.2214/AJR.19.21598.

24. Aujero MP, Gavenonis SC, Benjamin R, Zhang Z, Holt JS. Clinical Performance of Synthesized Two-dimensional Mammography Combined with Tomosynthesis in a Large Screening Population. Radiology 2017;283(1):70–76.

25. Bernardi D, Macaskill P, Pellegrini M, et al. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study. Lancet Oncol 2016;17(8):1105–1113.

26. Simon K, Dodelzon K, Drotman M, et al. Accuracy of Synthetic 2D Mammography Compared With Conventional 2D Digital Mammography Obtained With 3D Tomosynthesis. AJR Am J Roentgenol 2019;212(6):1406–1411.

27. Caumo F, Zorzi M, Brunelli S, et al. Digital Breast Tomosynthesis with Synthesized Two-Dimensional Images versus Full-Field Digital Mammography for Population Screening: Outcomes from the Verona Screening Program. Radiology 2018;287(1):37–46https://doi.org/10.1148/radiol.2017170745.

28. Skaane P, Bandos AI, Niklason LT, et al. Digital Mammography versus Digital Mammography Plus Tomosynthesis in Breast Cancer Screening: The Oslo Tomosynthesis Screening Trial. Radiology 2019;291(1):23–30.

29. 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.

30. McDonald ES, McCarthy AM, Weinstein SP, Schnall MD, Conant EF. BI-RADS Category 3 Comparison: Probably Benign Category after Recall from Screening before and after Implementation of Digital Breast Tomosynthesis. Radiology 2017;285(3):778–787.

31. 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.

32. Conant EF, Barlow WE, Herschorn SD, et al. Association of Digital Breast Tomosynthesis vs Digital Mammography With Cancer Detection and Recall Rates by Age and Breast Density. JAMA Oncol 2019;5(5):635–642.

33. Fujii MH, Herschorn SD, Sowden M, et al. Detection Rates for Benign and Malignant Diagnoses on Breast Cancer Screening With Digital Breast Tomosynthesis in a Statewide Mammography Registry Study. AJR Am J Roentgenol 2019;212(3):706–711.

34. Miglioretti DL, Abraham L, Lee CI, et al. Digital Breast Tomosynthesis: Radiologist Learning Curve. Radiology 2019;291(1):34–42.

35. Sickles E, 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.

36. D'Orsi CJ, Sickles E, Mendelson EB, et al. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Reston, Va: American College of Radiology, YEAR.

37. Lehman CD, Arao RF, Sprague BL, et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. Radiology 2017;283(1):49–58.