UC Davis

UC Davis Previously Published Works

Title

Digital Breast Tomosynthesis versus Digital Mammography Screening Performance on Successive Screening Rounds from the Breast Cancer Surveillance Consortium.

Permalink

https://escholarship.org/uc/item/4ss2t4kc

Journal

Radiology, 307(5)

Authors

Sprague, Brian

Coley, Rebecca

Lowry, Kathryn

et al.

Publication Date

2023-06-01

DOI

10.1148/radiol.223142

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Radiology

ORIGINAL RESEARCH • BREAST IMAGING

Digital Breast Tomosynthesis versus Digital Mammography Screening Performance on Successive Screening Rounds from the Breast Cancer Surveillance Consortium

Brian L. Sprague, PhD • Rebecca Yates Coley, PhD • Kathryn P. Lowry, MD • Karla Kerlikowske, MD • Louise M. Henderson, PhD • Yu-Ru Su, PhD • Christoph I. Lee, MD • Tracy Onega, PhD • Erin J. A. Bowles, MPH • Sally D. Herschorn, MD • Roberta M. diFlorio-Alexander, MD • Diana L. Miglioretti, PhD

From the Departments of Surgery (B.L.S.) and Radiology (B.L.S., S.D.H.), University of Vermont Cancer Center, University of Vermont Larner College of Medicine, UHC Bldg, Room 4425, 1 S Prospect St, Burlington, VT 05401; Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington, Seattle, Wash (R.Y.C., Y.R.S., E.J.A.B., D.L.M.); Department of Radiology, University of Washington, Fred Hutchinson Cancer Center, Seattle, Wash (K.P.L., C.I.L.); Departments of Medicine and Epidemiology and Biostatistics, University of California–San Francisco, San Francisco, Calif (K.K.); Department of Radiology, University of North Carolina, Chapel Hill, NC (L.M.H.); Department of Population Health Sciences, University of Utah, Huntsman Cancer Institute, Salt Lake City, Utah (T.O.); Department of Radiology, Giesel School of Medicine at Dartmouth, Lebanon, NH (R.M.d.F.A.); and Division of Biostatistics, Department of Public Health Sciences, University of California–Davis, Davis, Calif (D.L.M.). Received December 12, 2022; revision requested January 17, 2023; revision received March 23; accepted March 29. **Address correspondence to** B.L.S. (email: *bsprague@uvm.edu*).

Research reported in this work was funded by the Breast Cancer Surveillance Consortium, with funding from the National Cancer Institute (NCI) (P01CA154292, R01CA248068) and the Patient-Centered Outcomes Research Institute (PCS-1504-30370). Data collection for this research was additionally supported by the NCI (U54CA163303), the Agency for Healthcare Research and Quality (R01 HS018366-01A1), and the University of Vermont Cancer Center with funds generously awarded by the Lake Champlain Cancer Research Organization (grant 032800). B.L.S. was also supported in part by the National Institute of General Medical Sciences (P20GM103644). K.P.L. is also supported in part by the American Cancer Society (CSDG-21-078-01-CPSH). E.J.A.B. was also supported in part by the NCI (R50CA211115). Cancer and vital status data collection was supported by several state public health departments and cancer registries. For a full description of these sources, please see *http://www.bcsc-research.org/work/acknowledgement.html*.

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

See also the editorial by Skaane in this issue.

Radiology 2023; 307(5):e223142 • https://doi.org/10.1148/radiol.223142 • Content code: BR

Background: Prior cross-sectional studies have observed that breast cancer screening with digital breast tomosynthesis (DBT) has a lower recall rate and higher cancer detection rate compared with digital mammography (DM).

Purpose: To evaluate breast cancer screening outcomes with DBT versus DM on successive screening rounds.

Materials and Methods: In this retrospective cohort study, data from 58 breast imaging facilities in the Breast Cancer Surveillance Consortium were collected. Analysis included women aged 40–79 years undergoing DBT or DM screening from 2011 to 2020. Absolute differences in screening outcomes by modality and screening round were estimated during the study period by using generalized estimating equations with marginal standardization to adjust for differences in women's risk characteristics across modality and round.

Results: A total of 523 485 DBT examinations (mean age of women, 58.7 years ± 9.7 [SD]) and 1 008 123 DM examinations (mean age, 58.4 years ± 9.8) among 504 863 women were evaluated. DBT and DM recall rates decreased with successive screening round, but absolute recall rates in each round were significantly lower with DBT versus DM (round 1 difference, −3.3% [95% CI: −4.6, −2.1] [$P < .001$]; round 2 difference, −1.8% [95% CI: −2.9, −0.7] [$P = .003$]; round 3 or above difference, −1.2% [95% CI: −2.4, −0.1] [$P = .03$]). DBT had significantly higher cancer detection (difference, 0.6 per 1000 examinations [95% CI: 0.2, 1.1]; $P = .009$) compared with DM only for round 3 and above. There were no significant differences in interval cancer rate (round 1 difference, 0.00 per 1000 examinations [95% CI: −0.24, 0.30] [$P = .96$]; round 2 or above difference, 0.04 [95% CI: −0.19, 0.31] [$P = .76$]) or total advanced cancer rate (round 1 difference, 0.00 per 1000 examinations [95% CI: −0.15, 0.19] [$P = .94$]; round 2 or above difference, −0.06 [95% CI: −0.18, 0.11] [$P = .43$]).

Conclusion: DBT had lower recall rates and could help detect more cancers than DM across three screening rounds, with no difference in interval or advanced cancer rates.

© RSNA, 2023

Supplemental material is available for this article.

Digital breast tomosynthesis (DBT) has been widely adopted for breast cancer screening, comprising 46% of the U.S. Food and Drug Administration–accredited mammography units as of October 2022 (1). DBT dissemination has been supported by studies reporting a higher cancer detection rate and lower recall rate compared with digital mammography (DM) screening examinations (2–4). Most DBT examinations included in these studies have been the woman's first DBT examination. Evaluation of screening performance on successive DBT examinations is important for several reasons. First, evidence for sustained reductions in recall rate over multiple screening rounds may further support dissemination of DBT. Second, the initially observed increase in cancer detection rate from the first DBT screening examination is consistent with the performance expectations of newly introduced improved screening technology (5). However, the magnitude and duration of the increase in cancer

This copy is for personal use only. To order copies, contact *reprints@rsna.org*

Abbreviations

BCSC = Breast Cancer Surveillance Consortium, BI-RADS = Breast Imaging Reporting and Data System, DBT = digital breast tomosynthesis, DM = digital mammography

Summary

Digital breast tomosynthesis had lower recall rates and could help detect more cancers than digital mammography across three screening rounds, with no difference in interval or advanced cancer rates.

Key Results

- In this retrospective study of 504 863 women, both digital breast tomosynthesis (DBT) and digital mammography (DM) had lower recall rates on successive screening rounds (DBT round 1, 7.5%; DBT round 2, 6.9%; DBT round 3, 6.1%; DM round 1, 10.9%; DM round 2, 8.6%; DM round 3, 7.3%), but absolute recall rates were significantly lower with DBT (round 1 difference, −3.3% [$P < .001$]; round 2 difference, −1.8% [$P = .003$]; round 3 or above difference, −1.2% [$P = .03$]).
- DBT had significantly higher cancer detection than DM on the third screening round (difference, 0.6 per 1000 examinations; $P = .009$).
- No differences were observed for DBT versus DM on any screening round for rates of interval cancer (round 1 difference, 0.00 [$P = .96$]; round 2 or above difference, 0.04 [$P = .76$]) or advanced cancer (round 1 difference, 0.00 [$P = .94$]; round 2 or above difference, −0.06 [$P = .43$]).

detection over multiple screening rounds is required to understand the effectiveness of DBT in shifting toward earlier cancer detection relative to DM (5). Finally, there is little evidence to date that DBT screening reduces the incidence of screening failures, including interval or advanced breast cancer (6–12). Multiple screening rounds may be required to observe reductions in screening failures if DBT can help detect aggressive cancers earlier than DM can (13).

A handful of studies have evaluated DBT performance by screening round, including two small European studies (14,15) and two single-institution studies (16,17) from the United States, all of which reported persistently low recall rates on successive DBT screening rounds but declines in cancer detection. The purpose of the current study was to evaluate DBT screening outcomes across multiple screening rounds in a large geographically diverse sample from multiple breast imaging facilities in the United States. We examined a comprehensive set of screening performance metrics, including rates of interval and advanced cancers, and included a comparator group of consecutive DM screening examinations during the same study period.

Materials and Methods

Study Setting and Data Sources

We conducted a retrospective analysis using observational clinical data collected by five breast imaging registries within the Breast Cancer Surveillance Consortium (BCSC) (*http://www.bcsc-research.org*): the Carolina Mammography Registry, Metropolitan Chicago Breast Cancer Registry, New Hampshire Mammography Network, San Francisco Mammography Registry, and Vermont Breast Cancer Surveillance System. The registries collect clinical data on women undergoing breast imaging at radiology facilities within their catchment areas. Each BCSC registry and the Statistical Coordinating Center received institutional review board approval for all study procedures, including passive consenting processes or a waiver of consent. All procedures were compliant with the Health Insurance Portability and Accountability Act.

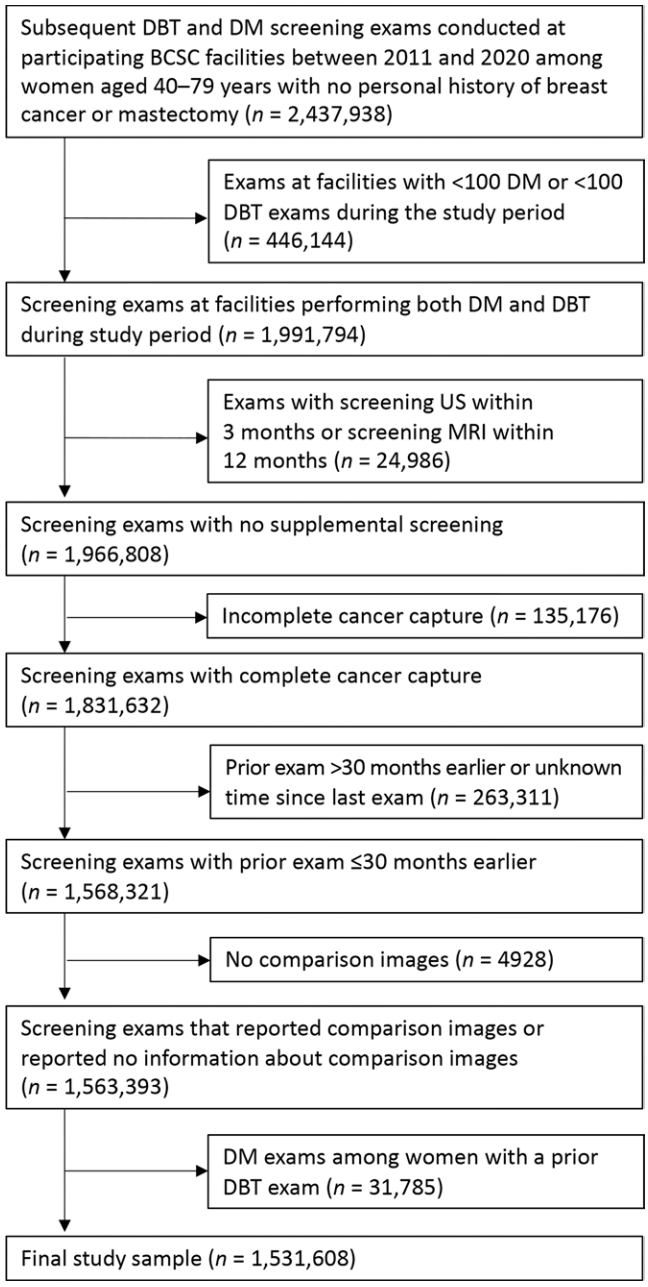

Figure 1: Study flow diagram. BCSC = Breast Cancer Surveillance Consortium, DBT = digital breast tomosynthesis, DM = digital mammography.

Participants and Examinations

We included subsequent DBT and DM screening examinations conducted between 2011 and 2020 among women aged 40–79 years with no personal history of breast cancer or mastectomy (Fig 1). We excluded examinations from facilities that did not perform at least 100 DBT and at least 100 DM

screening examinations during the study period. Examinations were excluded if screening US was performed within 3 months of the mammogram, if screening MRI was performed within 12 months of the mammogram, or if less than 1 year of follow-up data for cancer diagnoses were available. To reflect regular participation in screening, analyses were limited to subsequent screening mammograms obtained 30 months or less after a woman's prior mammogram. Thus, a woman's first-ever screening mammogram was not included, and any mammograms more than 30 months since their most recent mammogram were excluded. Examinations were also excluded if no comparison images from prior mammographies were available during clinical interpretation, and DM examinations were restricted to women with no known prior history of DBT imaging. All participating facilities started using DBT in 2011 or later; the vast majority of DBT examinations were performed on Hologic machines (96%), with the remainder on GE Healthcare and Philips machines. Many of the women in this study were included in prior BCSC publications regarding DBT performance (2,10,18,19); the current analysis is the first BCSC study to examine DBT performance by screening round among subsequent (nonbaseline) screening examinations and includes a larger sample size of DBT examinations than in previous studies.

Data Collection

Participating radiology facilities provided imaging modality, indication, breast density, and assessment data for each examination using Breast Imaging Reporting and Data System (BI-RADS) nomenclature (20). Demographic and risk factor information at each examination were self-reported by women or extracted from electronic health records. Self-reported race and ethnicity were included as a social construct that could potentially explain differences in screening performance due to social determinants of health, including inequities in access to high-quality screening. Categories for this variable included Hispanic/Latina and non-Hispanic/Latina, Asian, Black, White, or multiracial/other (Native Hawaiian or other Pacific Islander, American Indian or Alaska Native, self-reported "other" race, or multiple races). Data on breast cancer diagnoses (invasive cancer or ductal carcinoma in situ) diagnosed 1 year or less after each screening mammogram and before the next screening mammogram were ascertained by linking women's imaging data to BCSC registry pathology databases, regional Surveillance Epidemiology and End Results Program programs, and state tumor registries (10).

Outcome Measures and Definitions

We used the BCSC standard definition of a screening mammogram, based on the radiologist classification of the clinical indication and the type of views obtained (2,21). Each eligible DBT examination was classified by screening round according to the number of prior DBT screening examinations for the individual within the BCSC database during the study period (2011–2020). As a comparator group, DM screening examinations were classified by screening round according to the number of prior DM screening examinations for the individual within the BCSC database during the study period.

Recall was defined by an initial positive assessment at screening (BI-RADS category 0, 3, 4, or 5). Short-interval follow-up recommendation was defined as a final assessment of BI-RADS category 3 after diagnostic work-up of a positive screening examination. Biopsy recommendation was defined as a final positive assessment (BI-RADS category 4 or 5) after diagnostic work-up of a positive screening examination. False-positive recall, false-positive short-interval follow-up, and false-positive biopsy recommendation were defined as recall, short-interval follow-up, and biopsy recommendation, respectively, with no cancer diagnosis within 1 year.

Analyses of screening-detected cancer rates included all examinations with a 1-year or greater capture of pathology or cancer registry data from participating BCSC registries. Metrics relying on capture of interval cancers were limited to examinations with a 1-year or greater capture of cancer diagnoses from population-based tumor registries to ensure ascertainment of cancer diagnoses occurring outside of the screening facility. If multiple breast cancer diagnoses occurred within 1 year after the screening mammogram, the diagnosis with the most severe stage was selected. Cancers detected at screening examination were defined as those occurring within 1 year of a screening mammogram with a final BI-RADS category 3, 4, or 5 assessment (22). Cancer detection rates were calculated overall and separately for advanced-stage (pathologic prognostic stage II, III, or IV) and early-stage invasive cancer (pathologic prognostic stage I) (10). If pathologic prognostic stage was missing, we used the American Joint Committee on Cancer anatomic stage to classify stage I and IIa as early invasive cancer and stage IIb, III, and IV as advanced cancer, as these groupings have a similar risk of breast cancer death as those defined by pathologic prognostic stage (23). Interval cancer was defined as cancer occurring within 1 year of a screening mammogram with a final BI-RADS category 1 or 2 assessment.

Positive predictive value of recall and positive predictive value of biopsy recommendation were defined as the proportion of examinations with recall and biopsy recommendation, respectively, that were followed by a cancer detected at screening examination. Sensitivity of recall and sensitivity of positive final assessment were defined as the number of mammograms with a positive initial assessment (BI-RADS category 0, 3, 4, or 5) and positive final assessment (BI-RADS category 3, 4, or 5), respectively, divided by the number of mammograms with a cancer diagnosis within 1 year. Specificity was defined as the number of mammograms with a negative initial assessment (BI-RADS category 1 or 2) and no cancer diagnosed during 1 year of follow-up divided by the number of mammograms with no cancer diagnosed during 1 year of follow-up.

Statistical Analysis

All analyses used the screening mammogram as the unit of analysis. We used descriptive statistics to characterize screening mammograms by imaging modality and screening round. BCSC 5-year invasive breast cancer risk was calculated using the BCSC risk calculator version 2 *(https://tools.bcsc-scc.org/BC5yearRisk/calculator.htm)* (24). We estimated screening outcome rates, absolute differences in screening outcomes within modality by screening round, and absolute differences in screening

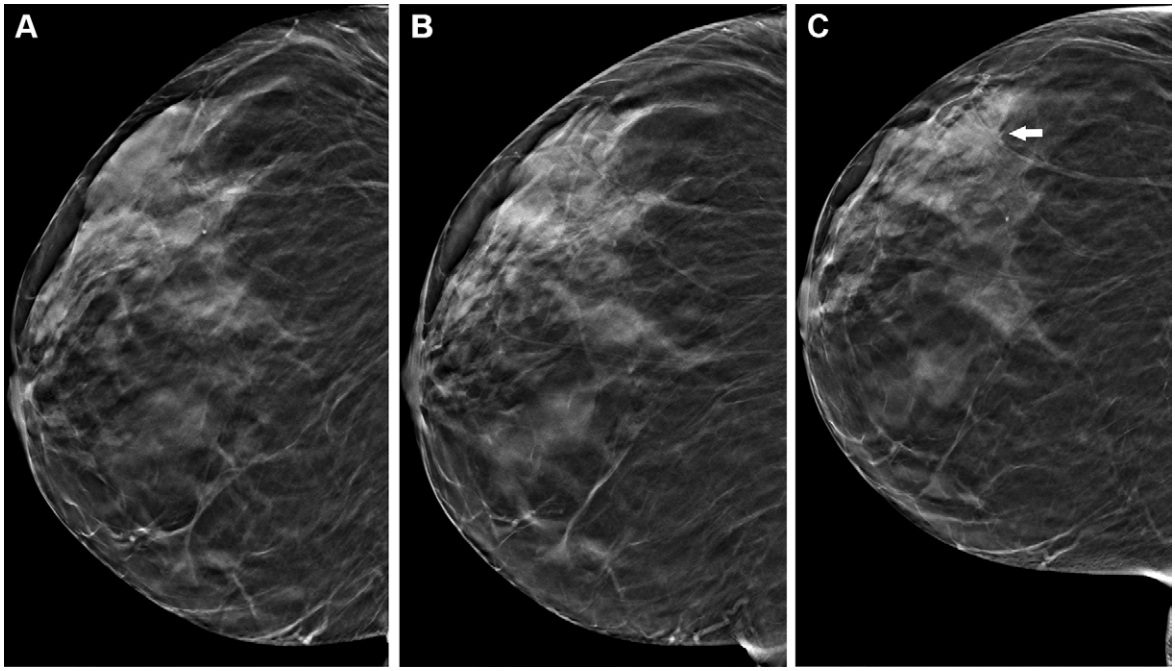

Figure 2: Example images from three successive rounds of digital breast tomosynthesis. Right craniocaudal tomosynthesis section in a 66-year-old woman screened in **(A)** 2020, **(B)** 2021, and **(C)** 2022 with a new area of architectural distortion identified in 2022 (arrow), which led to a cancer diagnosis.

outcomes within screening round by modality using generalized estimating equations (25) with a log link and Poisson distribution. This robust approach estimates variances empirically to obtain valid estimates for binary outcomes without assuming that outcomes follow a Poisson distribution and has been shown to be a valid alternative to log-binomial regression for clustered binary outcome data (26). Robust covariance estimates were used to account for nonnested clustering of examinations within women, radiologists, and facilities (27). All models were adjusted for age, breast density, race or ethnicity, time since last mammogram, BCSC 5-year invasive breast cancer risk, benign breast disease history, first-degree family history of breast cancer, and examination year, which were selected a priori based on previously demonstrated associations with mammography performance (28). For rare outcomes where sample size was limited, including interval and advanced cancer, round 2 and round 3 and above examinations were combined.

Absolute screening outcome rates were estimated with use of marginal standardization to adjust for differences in risk characteristics across modality and screening round (29). Differences in adjusted screening outcome rates between rounds were estimated within modality. Differences in screening outcome rates between modality were estimated within the screening round by marginalizing over the distribution of examination characteristics within that round. Parametric bootstrap resampling was used to estimate standard errors and construct 95% CIs for all outcome rates. Two-sided *P* values were calculated by identifying the minimum α level for which the 95% CI would not include the null value of 0 difference in risk and multiplying this α level by 2 (30). All statistical analyses were conducted by a biostatistician (R.Y.C., with 15 years of experience) using SAS, version 9.4 (SAS Institute) and R, version 4.0.2 (R Foundation for Statistical Computing). Two-sided $P < .05$ indicated a statistically significant difference.

Results

Examination-level Characteristics of Women Undergoing Mammography

After the exclusion of 446 144 examinations at facilities performing less than 100 DM or less than 100 DBT examinations during the study period, 24 986 examinations with supplemental screening, 135 176 examinations with incomplete cancer capture, 263 311 examinations without prior imaging in the past 30 months, 4928 examinations with no comparison images, and 31 785 DM examinations with a prior DBT examination (Fig 1), a total of 523 485 DBT (Fig 2) and 1 008 123 DM (Fig 3) screening examinations among 504 863 women undergoing imaging at 58 health care facilities were available for analysis. Compared with DM examinations, DBT examinations were performed more frequently among non-Hispanic White women and women with a first-degree family history of breast cancer (Table 1). Older age, women self-identifying as White and non-Hispanic, first-degree family history of breast cancer, and annual screening frequency were more common in later DBT screening rounds compared with first-round DBT examinations. DBT examinations were more frequently performed in women at high BCSC 5-year invasive breast cancer risk compared with DM in all screening rounds. Women at high risk were more numerous in later DBT screening rounds compared with first-round DBT.

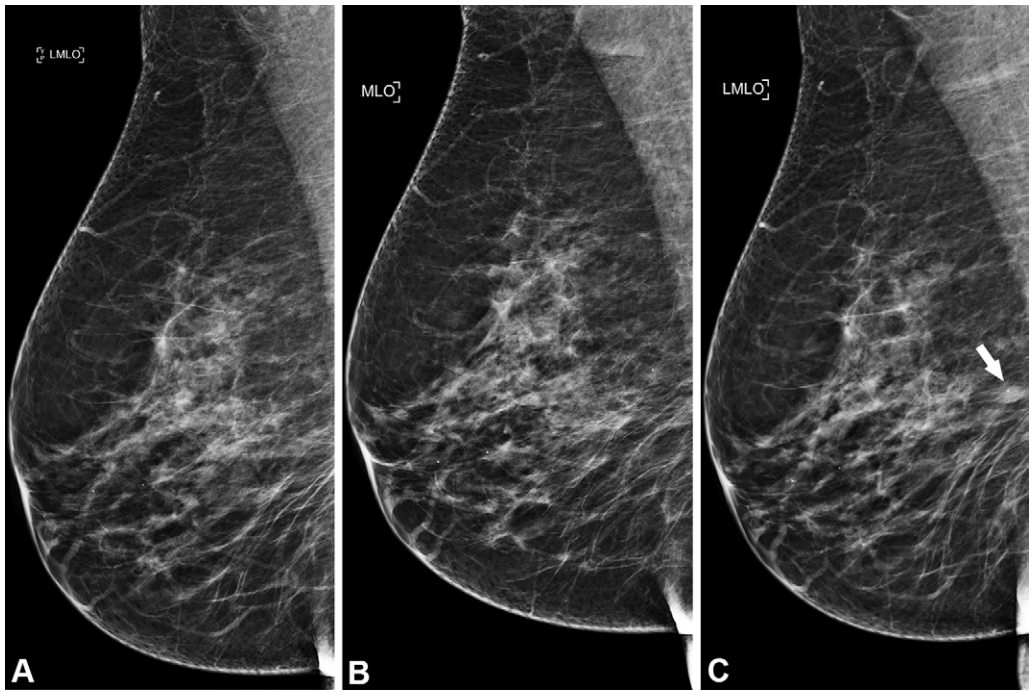

Figure 3: Example images from three successive rounds of digital mammography screening. Left mediolateral oblique (LMLO) view in a 64-year-old woman screened in **(A)** 2014, **(B)** 2015, and **(C)** 2016 with a new irregular mass identified in 2016 (arrow), which led to a cancer diagnosis. MLO = mediolateral oblique.

Comparison of Outcome Measures by Screening Round

Multivariable-adjusted screening outcome measures for DM and DBT examinations are shown in Table 2 (see Table S1 for unadjusted outcome measures). For screening round 3 or above, recall rate was 6.1% (95% CI: 5.4, 6.8) for DBT and 7.3% (95% CI: 6.6, 8.1) for DM, while the cancer detection rates were 3.9 per 1000 examinations (95% CI: 3.6, 4.4) and 3.4 per 1000 examinations (95% CI: 3.1, 3.8), respectively. The recall, biopsy recommendation, and cancer detection rates declined significantly on screening round 2 and round 3 or above compared with screening round 1 for both DBT and DM (Table 3). Specificity of recall increased significantly over successive screening rounds for both modalities (difference, 0.7% [95% CI: 0.1, 1.3] [$P = .01$] for DBT round 2 vs DBT round 1; difference, 1.2% [95% CI: 0.1, 2.3] [$P = .03$] for DBT round 3 or above vs DBT round 1; difference, 2.5% [95% CI: 1.6, 3.4] [$P < .001$] for DM round 2 vs DM round 1; difference, 3.8% [95% CI: 2.5, 5.2] [$P < .001$] for DM round 3 or above vs DM round 1), whereas the positive predictive value of recall and positive predictive value of biopsy recommendation increased significantly with successive screening rounds for DM (positive predictive value of recall difference, 0.4% [95% CI: 0.0, 0.8] [$P = .04$] for DM round 2 vs DM round 1; positive predictive value of recall difference, 0.8% [95% CI: 0.3, 1.3] [$P = .002$] for DM round 3 or above vs DM round 1; positive predictive value of biopsy recommendation difference, 3.3% [95% CI: 1.3, 5.4] [$P = .002$] for DM round 2 vs DM round 1; positive predictive value of biopsy recommendation difference, 4.2% [95% CI: 1.9, 6.6] [$P < .001$] for DM round 3 or above vs DM round 1) but not for DBT (positive predictive value of recall difference, −0.2% [95% CI: −0.8, 0.4] [$P = .43$] for DBT round 2 vs DBT round 1; positive predictive value of recall difference, 0.3% [95% CI: −0.7, 1.4] [$P = .53$] for DBT round 3 or above vs DBT round 1; positive predictive value of biopsy recommendation difference, −0.3% [95% CI: −3.2, 2.7] [$P = .87$] for DBT round 2 vs DBT round 1; positive predictive value of biopsy recommendation difference, 0.7% [95% CI: −4.5, 6.6] [$P = .80$] for DBT round 3 or above vs DBT round 1).

Comparison of Outcome Measures by Modality and Screening Round

Within each screening round, the recall rate was significantly lower for DBT versus DM examinations (round 1, $P < .001$; round 2, $P = .003$; round 3 or above, $P = .03$) (Table 4). For screening round 3 or above, the absolute difference in recall rate for DBT versus DM was −1.2% (95% CI: −2.4, −0.1) ($P = .03$). For false-positive recall rate, no difference was observed between DBT and DM in round 3 or above. Short-interval follow-up rates were significantly lower within each screening round for DBT versus DM examinations (round 1, $P < .001$; round 2, $P = .048$; round 3 or above, $P = .03$). False-positive short-interval follow-up rates were significantly lower for DBT versus DM examinations in round 1 ($P < .001$) and round 3 ($P = .01$) but not in round 2 ($P = .08$). Biopsy recommendation rate and false-positive biopsy recommendation rate were significantly lower for DBT versus DM examinations on screening round 1 ($P = .01$ for biopsy recommendation rate; $P < .001$ for false-positive biopsy recommendation rate) but not in rounds 2 ($P = .64$ for biopsy recommendation rate; $P = .05$ for false-positive biopsy recommendation rate) or 3 and above ($P =$

Table 1: Examination-Level Characteristics for Digital Breast Tomosynthesis or Digital Mammography Examinations by Screening Round during the Study Period

Characteristic	Digital Breast Tomosynthesis			Digital Mammography		
	Round 1 (*n* = 207 280)	Round 2 (*n* = 152 328)	Round ≥3 (*n* = 163 877)	Round 1 (*n* = 355 944)	Round 2 (*n* = 297 777)	Round ≥3 (*n* = 354 402)
Age						
40–49 years	46 146 (22.3)	34 194 (22.4)	27 909 (17.0)	99 988 (28.1)	67 315 (22.6)	51 905 (14.6)
50–59 years	68 385 (33.0)	48 763 (32.0)	54 330 (33.2)	116 967 (32.9)	100 461 (33.7)	114 966 (32.4)
60–69 years	61 726 (29.8)	45 865 (30.1)	52 752 (32.2)	93 208 (26.2)	87 201 (29.3)	118 558 (33.5)
70 years or older	31 023 (15.0)	23 506 (15.4)	28 886 (17.6)	45 781 (12.9)	42 800 (14.4)	68 973 (19.5)
Race and ethnicity						
Asian	9910 (4.9)	6310 (4.2)	5884 (3.6)	30 756 (8.9)	29 372 (10.2)	31 915 (9.2)
Black	22 929 (11.4)	10 609 (7.1)	5622 (3.5)	48 229 (14.0)	40 964 (14.2)	43 363 (12.6)
Hispanic/Latina	10 291 (5.1)	5785 (3.9)	4422 (2.7)	22 222 (6.5)	18 122 (6.3)	18 509 (5.4)
White	155 245 (77.1)	124 391 (83.2)	144 012 (88.6)	237 397 (68.9)	195 999 (67.8)	246 249 (71.3)
Multiracial or other*	3006 (1.5)	2436 (1.6)	2551 (1.6)	5911 (1.7)	4730 (1.6)	5224 (1.5)
Missing†	5899 (2.8)	2797 (1.8)	1386 (0.8)	11 429 (3.2)	8590 (2.9)	9142 (2.6)
First-degree family history of breast cancer						
Absent	158 295 (80.8)	116 871 (80.0)	122 856 (77.0)	284 432 (83.3)	241 203 (82.9)	283 812 (80.9)
Present	37 555 (19.2)	29 244 (20.0)	36 754 (23.0)	57 012 (16.7)	49 587 (17.1)	66 919 (19.1)
Missing†	11 430 (5.5)	6213 (4.1)	4267 (2.6)	14 500 (4.1)	6987 (2.3)	3671 (1.0)
History of breast biopsy						
None	157 942 (76.2)	116 594 (76.5)	122 125 (74.5)	277 518 (78.0)	230 781 (77.5)	268 791 (75.8)
Prior biopsy, diagnosis unknown	26 302 (12.7)	17 420 (11.4)	17 483 (10.7)	47 434 (13.3)	40 309 (13.5)	50 794 (14.3)
Nonproliferative lesion	16 196 (7.8)	12 848 (8.4)	16 376 (10.0)	21 855 (6.1)	18 754 (6.3)	24 876 (7.0)
Proliferative changes without atypia	5468 (2.6)	4372 (2.9)	6290 (3.8)	7487 (2.1)	6516 (2.2)	8088 (2.3)
Proliferative changes with atypia	1178 (0.6)	930 (0.6)	1349 (0.8)	1371 (0.4)	1206 (0.4)	1602 (0.5)
Lobular carcinoma in situ	194 (0.1)	164 (0.1)	254 (0.2)	279 (0.1)	211 (0.1)	251 (0.1)
BI-RADS breast density						
Almost entirely fatty	19 175 (9.5)	14 028 (9.3)	16 286 (10.0)	36 289 (10.6)	29 318 (10.3)	38 012 (11.0)
Scattered fibroglandular densities	95 049 (47.2)	70 032 (46.4)	76 704 (46.9)	144 082 (42.1)	129 766 (45.5)	165 165 (47.7)
Heterogeneously dense	73 620 (36.5)	56 215 (37.2)	60 508 (37.0)	132 843 (38.8)	105 172 (36.9)	122 065 (35.2)
Extremely dense	13 722 (6.8)	10 674 (7.1)	10 129 (6.2)	28 778 (8.4)	21 123 (7.4)	21 326 (6.2)
Missing†	5714 (2.8)	1379 (0.9)	250 (0.2)	13 952 (3.9)	12 398 (4.2)	7834 (2.2)
BCSC 5-year risk						
Low (<1.00%)	47 390 (24.9)	34 551 (24.2)	29 049 (18.9)	103 209 (31.8)	75 351 (28.0)	66 753 (20.9)
Average (1.00%–1.66%)	75 697 (39.7)	54 648 (38.3)	58 434 (38.0)	124 705 (38.4)	108 727 (40.4)	133 559 (41.7)
Intermediate (1.67%–2.49%)	43 921 (23.0)	33 991 (23.8)	39 639 (25.8)	65 000 (20.0)	57 022 (21.2)	77 811 (24.3)
High (2.50%–3.99%)	20 223 (10.6)	16 534 (11.6)	21 961 (14.3)	27 556 (8.5)	24 664 (9.2)	36 177 (11.3)
Very high (≥4.00%)	3341 (1.8)	2860 (2.0)	4499 (2.9)	4008 (1.2)	3666 (1.4)	5813 (1.8)
Missing†	16 708 (8.1)	9744 (6.4)	10 295 (6.3)	31 466 (8.8)	28 347 (9.5)	34 289 (9.7)
Time since last mammogram						
≤15 months	134 219 (64.8)	111 394 (73.1)	137 641 (84.0)	222 954 (62.6)	209 526 (70.4)	283 803 (80.1)
16–30 months	73 061 (35.2)	40 934 (26.9)	26 236 (16.0)	132 990 (37.4)	88 251 (29.6)	70 599 (19.9)

Note.—Data are numbers of examinations, with percentages in parentheses. For parameters with missing data, percentages were calculated among records with nonmissing values. BCSC = Breast Cancer Surveillance Consortium, BI-RADS = Breast Imaging Reporting and Data System.

* Other race includes Native Hawaiian or Pacific Islander, American Indian or Alaska Native, and self-reported "other" race.

† Percentages were calculated based on the total number of examinations for the corresponding round and modality.

Table 2: Multivariable-adjusted Performance Metrics according to Modality and Screening Round

A: Performance Outcomes

Metric	Digital Breast Tomosynthesis			Digital Mammography		
	Round 1 (*n* = 207 280)	Round 2 (*n* = 152 328)	Round ≥3 (*n* = 163 877)	Round 1 (*n* = 355 944)	Round 2 (*n* = 297 777)	Round ≥3 (*n* = 354 402)
Recall rate (%)	7.5 (6.9, 8.2)	6.9 (6.2, 7.7)	6.1 (5.4, 6.8)	10.9 (9.6, 12.3)	8.6 (7.8, 9.6)	7.3 (6.6, 8.1)
False-positive recall rate (%)*	6.6 (5.9, 7.5)	6.0 (5.3, 7.0)	5.5 (4.7, 6.6)	10.1 (8.9, 11.5)	7.8 (7.1, 8.7)	6.6 (6.0, 7.4)
Short-interval follow-up rate (%)	1.6 (1.3, 1.9)	1.4 (1.2, 1.8)	1.1 (0.9, 1.4)	2.5 (2.0, 3.1)	1.9 (1.5, 2.3)	1.6 (1.3, 2.0)
False-positive short-interval follow-up rate (%)*	1.4 (1.2, 1.8)	1.3 (1.0, 1.7)	0.9 (0.7, 1.2)	2.4 (1.9, 3.0)	1.7 (1.4, 2.1)	1.5 (1.2, 1.8)
Biopsy recommendation rate (%)	1.5 (1.4, 1.6)	1.3 (1.2, 1.4)	1.2 (1.1, 1.4)	1.8 (1.6, 2.0)	1.4 (1.2, 1.5)	1.3 (1.1, 1.4)
False-positive biopsy recommendation (%)*	1.0 (0.9, 1.1)	0.8 (0.7, 0.9)	0.8 (0.6, 0.9)	1.3 (1.1, 1.5)	1.0 (0.8, 1.1)	0.8 (0.7, 1.0)
Cancer detection rate (per 1000 examinations)	4.8 (4.4, 5.2)	4.1 (3.8, 4.5)	3.9 (3.6, 4.4)	4.4 (3.8, 5.2)	3.7 (3.3, 4.2)	3.4 (3.1, 3.8)
Screening-detected early-stage invasive cancer rate (per 1000 examinations)	3.5 (3.2, 3.8)	2.9 (2.7, 3.3)	3.0 (2.8, 3.4)	3.0 (2.5, 3.5)	2.5 (2.2, 2.9)	2.4 (2.2, 2.7)
Positive predictive value of recall (%)	6.7 (5.9, 7.6)	6.4 (5.7, 7.4)	7.0 (6.1, 8.1)	4.0 (3.6, 4.4)	4.4 (3.9, 5.0)	4.7 (4.1, 5.5)
Positive predictive value of biopsy recommendation (%)	31.2 (28.7, 34.4)	30.9 (28.0, 34.5)	31.9 (27.0, 38.2)	22.5 (20.5, 24.9)	25.8 (23.4, 28.8)	26.7 (24.3, 29.5)
Sensitivity of recall (%)*	87.6 (85.3, 90.1)	89.9 (85.7, 94.5)	85.7 (82.0, 89.6)	87.3 (84.9, 89.9)	88.5 (85.7, 91.4)	85.9 (83.6, 88.3)
Sensitivity of positive final assessment (%)	86.1 (83.4, 89.2)	87.2 (82.8, 92.1)	84.5 (80.8, 88.6)	85.8 (83.2, 88.6)	86.9 (84.0, 90.0)	84.7 (82.3, 87.3)
Specificity of recall (%)*	93.5 (92.7, 94.4)	94.2 (93.2, 95.2)	94.7 (93.5, 95.9)	89.7 (88.3, 91.1)	92.2 (91.4, 93.0)	93.5 (92.8, 94.2)

B: Rare Outcomes[†]

Metric	Digital Breast Tomosynthesis		Digital Mammography	
	Round 1 (*n* = 207 280)	Round ≥2 (*n* = 316 205)	Round 1 (*n* = 355 944)	Round ≥2 (*n* = 652 179)
Screening-detected advanced cancer rate (per 1000 examinations)	0.21 (0.16, 0.33)	0.13 (0.09, 0.23)	0.26 (0.21, 0.39)	0.20 (0.17, 0.25)
Interval cancer (DCIS or invasive) rate (per 1000 examinations)*	0.79 (0.65, 1.02)	0.66 (0.49, 0.94)	0.79 (0.66, 0.99)	0.62 (0.55, 0.75)
Interval invasive cancer rate (per 1000 examinations)*	0.69 (0.56, 0.89)	0.57 (0.45, 0.77)	0.71 (0.59, 0.90)	0.56 (0.50, 0.66)
Interval advanced cancer rate* (per 1000 examinations)	0.20 (0.15, 0.32)	0.15 (0.10, 0.29)	0.18 (0.13, 0.28)	0.12 (0.10, 0.18)
Total advanced cancer rate (per 1000 examinations)*	0.45 (0.37, 0.61)	0.26 (0.17, 0.45)	0.45 (0.36, 0.60)	0.33 (0.29, 0.40)

Note.—Data in parentheses are 95% CIs. Models are adjusted for age, breast density, race and ethnicity, time since last mammogram, Breast Cancer Screening Consortium 5-year invasive breast cancer risk, benign biopsy history, family history of breast cancer, and examination year. DCIS = ductal carcinoma in situ.

* Restricted to examinations with complete cancer capture for 1-year follow-up. The total number of examinations for digital breast tomosynthesis round 1 was 144 848; round 2, 100 159; and round 3 or above, 82 396. The total number of examinations for digital mammography round 1 was 342 867; round 2, 288 211; and round 3 or above, 309 812.

[†] Due to limited sample size, screening round 2 and round 3 and above were combined for rare outcomes.

Table 3: Multivariable-adjusted Absolute Differences in Performance Metrics by Screening Round for Digital Breast Tomosynthesis and Digital Mammography

A: Differences in Performance Outcomes

	Digital Breast Tomosynthesis		Digital Mammography	
Metric	Round 2 vs Round 1	Round ≥3 vs Round 1	Round 2 vs Round 1	Round ≥3 vs Round 1
Recall rate (%)	−0.6 (−1.0, −0.2)*	−1.5 (−2.2, −0.7)*	−2.3 (−3.1, −1.5)*	−3.6 (−4.8, −2.4)*
False-positive recall rate (%)†	−0.6 (−1.1, −0.1)*	−1.1 (−2.0, −0.1)*	−2.2 (−3.1, −1.5)*	−3.4 (−4.7, −2.4)*
Short-interval follow-up rate (%)	−0.2 (−0.3, 0.0)*	−0.5 (−0.8, −0.2)*	−0.6 (−1.0, −0.4)*	−0.9 (−1.3, −0.5)*
False-positive short-interval follow-up rate (%)†	−0.2 (−0.3, 0.0)	−0.5 (−0.8, −0.3)*	−0.7 (−1.0, −0.4)*	−0.9 (−1.4, −0.5)*
Biopsy recommendation rate (%)	−0.2 (−0.3, −0.1)*	−0.3 (−0.4, −0.1)*	−0.4 (−0.6, −0.3)*	−0.6 (−0.8, −0.3)*
False-positive biopsy recommendation (%)†	−0.2 (−0.3, −0.1)*	−0.2 (−0.4, 0.0)*	−0.4 (−0.5, −0.3)*	−0.5 (−0.6, −0.3)*
Cancer detection rate (per 1000 examinations)	−0.7 (−1.1, −0.3)*	−0.8 (−1.3, −0.3)*	−0.7 (−1.3, −0.2)*	−1.0 (−1.8, −0.3)*
Screening-detected early-stage invasive cancer rate (per 1000 examinations)	−0.5 (−0.9, −0.1)*	−0.4 (−0.9, 0.0)	−0.5 (−0.9, 0.0)*	−0.6 (−1.2, 0.0)
Positive predictive value of recall (%)	−0.2 (−0.8, 0.4)	0.3 (−0.7, 1.4)	0.4 (0.0, 0.8)*	0.8 (0.3, 1.3)*
Positive predictive value of biopsy recommendation (%)	−0.3 (−3.2, 2.7)	0.7 (−4.5, 6.6)	3.3 (1.3, 5.4)*	4.2 (1.9, 6.6)*
Sensitivity of recall (%)†	2.3 (−1.9, 5.5)	−1.9 (−5.2, 2.2)	1.2 (−2.2, 4.5)	−1.4 (−5.4, 2.6)
Sensitivity of positive final assessment (%)	1.1 (−3.1, 5.5)	−1.6 (−5.3, 2.2)	1.0 (−2.4, 4.6)	−1.1 (−5.1, 3.1)
Specificity of recall (%)†	0.7 (0.1, 1.3)*	1.2 (0.1, 2.3)*	2.5 (1.6, 3.4)*	3.8 (2.5, 5.2)*

B: Differences in Rare Outcomes‡

	Digital Breast Tomosynthesis	Digital Mammography
Metric	Round ≥2 vs Round 1	Round ≥2 vs Round 1
Screening-detected advanced cancer rate (per 1000 examinations)	−0.08 (−0.17, 0.00)	−0.07 (−0.19, 0.02)
Interval cancer (DCIS or invasive) rate (per 1000 examinations)*	−0.13 (−0.33, 0.09)	−0.17 (−0.37, 0.02)
Interval invasive cancer rate (per 1000 examinations)†	−0.12 (−0.30, 0.07)	−0.15 (−0.35, 0.02)
Interval advanced cancer rate (per 1000 examinations)†	−0.05 (−0.18, 0.09)	−0.06 (−0.16, 0.02)
Total advanced cancer rate (per 1000 examinations)†	−0.19 (−0.33, −0.02)	−0.12 (−0.28, 0.01)

Note.—Data in parentheses are 95% CIs. Models are adjusted for age, breast density, race and ethnicity, time since last mammogram, Breast Cancer Screening Consortium 5-year invasive breast cancer risk, benign biopsy history, family history of breast cancer, and examination year. Results are standardized with respect to the distribution of covariates in the entire study sample. DCIS = ductal carcinoma in situ.

* Statistically significant difference ($P < .05$).

† Restricted to examinations with complete cancer capture for 1-year follow-up. The total number of examinations for digital breast tomosynthesis round 1 was 144 848; round 2, 100 159; and round 3 or above, 82 396. The total number of examinations for digital mammography round 1 was 342 867; round 2, 288 211; and round 3 or above, 309 812.

‡ Due to limited sample size, screening round 2 and round 3 and above were combined for rare outcomes.

.85 for biopsy recommendation rate; $P = .38$ for false-positive biopsy recommendation rate).

No difference in the overall cancer detection rate was observed between DBT versus DM examinations in screening round 1 or 2 (Table 4). However, for screening round 3 or above, the cancer detection rate was significantly higher for DBT versus DM examinations (difference, 0.6 per 1000 examinations [95% CI: 0.2, 1.1]; $P = .009$). A significant difference between DBT and DM examinations was also observed in the rate of early-stage invasive cancers detected in screening round 3 and above (difference, 0.7 [95% CI: 0.3, 1.0]; $P < .001$). No difference was observed in the rates of advanced cancer detected at screening (round 1 difference, −0.05 [95% CI: −0.19, 0.09] [$P = .43$]; round 2 or above difference, −0.06 [95% CI: −0.14, 0.03] [$P = .17$]), interval cancer (round 1 difference, 0.00 [95% CI: −0.24, 0.30]

Table 4: Multivariable-adjusted Absolute Differences in Performance Metrics for DBT versus DM according to Screening Round

A: Differences in Performance Outcomes

Metric	Round 1 DBT vs Round 1 DM	Round 2 DBT vs Round 2 DM	Round ≥3 DBT vs Round ≥3 DM
Recall rate (%)	−3.3 (−4.6, −2.1)*	−1.8 (−2.9, −0.7)*	−1.2 (−2.4, −0.1)*
False-positive recall rate (%)	−3.4 (−4.7, −2.2)*	−1.8 (−3.0, −0.7)*	−1.1 (−2.3, 0.2)
Short-interval follow-up rate (%)	−0.9 (−1.4, −0.5)*	−0.5 (−0.9, 0.0)*	−0.5 (−1.0, 0.0)*
False-positive short-interval follow-up rate (%)	−0.9 (−1.4, −0.4)*	−0.4 (−0.9, 0.4)	−0.5 (−1.0, −0.1)*
Biopsy recommendation rate (%)	−0.3 (−0.5, −0.1)*	0.0 (−0.3, 0.2)	0.0 (−0.2, 0.3)
False-positive biopsy recommendation (%)	−0.3 (−0.5, −0.2)*	−0.2 (−0.4, 0.0)	−0.1 (−0.3, 0.1)
Cancer detection rate (per 1000 examinations)	0.4 (−0.4, 1.2)	0.4 (−0.2, 0.9)	0.6 (0.2, 1.1)*
Screening-detected early-stage invasive cancer rate (per 1000 examinations)	0.5 (−0.1, 1.1)	0.5 (0.0, 0.9)	0.7 (0.3, 1.0)*
Positive predictive value of recall (%)	2.7 (1.9, 3.8)*	2.0 (1.1, 3.0)*	2.3 (1.1, 3.5)*
Positive predictive value of biopsy recommendation (%)	8.8 (5.8, 12.2)*	5.0 (1.4, 8.9)*	5.2 (−0.4, 11.5)
Sensitivity of recall (%)†	0.3 (−3.8, 4.5)	1.4 (−4.1, 6.8)	−0.2 (−4.3, 3.9)
Sensitivity of positive final assessment (%)	0.3 (−4.2, 5.1)	0.4 (−5.3, 6.2)	−0.2 (−4.4, 4.0)
Specificity of recall (%)†	3.9 (2.3, 5.4)*	2.0 (0.7, 3.3)*	1.2 (−0.2, 2.6)

B: Differences in Rare Outcomes‡

Metric	Round 1 DBT vs Round 1 DM	Round ≥2 DBT vs Round ≥2 DM
Screening-detected advanced cancer rate (per 1000 examinations)	−0.05 (−0.19, 0.09)	−0.06 (−0.14, 0.03)
Interval cancer (DCIS or invasive) rate (per 1000 examinations)†	0.00 (−0.24, 0.30)	0.04 (−0.19, 0.31)
Interval invasive cancer rate (per 1000 examinations)†	−0.02 (−0.26, 0.24)	0.01 (−0.16, 0.20)
Interval advanced cancer rate (per 1000 examinations)†	0.03 (−0.06, 0.14)	0.03 (−0.05, 0.15)
Total advanced cancer rate (per 1000 examinations)†	0.00 (−0.15, 0.19)	−0.06 (−0.18, 0.11)

Note.—Data in parentheses are 95% CIs. Models are adjusted for age, breast density, race and ethnicity, time since last mammogram, Breast Cancer Screening Consortium 5-year invasive breast cancer risk, benign biopsy history, family history of breast cancer, and examination year. Results are standardized with respect to the distribution of covariates within each screening round. DBT = digital breast tomosynthesis, DCIS = ductal carcinoma in situ, DM = digital mammography.

* Statistically significant difference ($P < .05$).

† Restricted to examinations with complete cancer capture for 1-year follow-up. The total number of examinations for digital breast tomosynthesis round 1 was 144 848; round 2, 100 159; and round 3 or above, 82 396. The total number of examinations for digital mammography round 1 was 342 867; round 2, 288 211; and round 3 or above, 309 812.

‡ Due to limited sample size, screening round 2 and round 3 and above were combined for rare outcomes.

[$P = .96$]; round 2 or above difference, 0.04 [95% CI: −0.19, 0.31] [$P = .76$]), interval invasive cancer (round 1 difference, −0.02 [95% CI: −0.26, 0.24] [$P = .86$]; round 2 or above difference, 0.01 [95% CI: −0.16, 0.20] [$P = .92$]), interval advanced cancer (round 1 difference, 0.03 [95% CI: −0.06, 0.14] [$P = .57$]; round 2 or above difference, 0.03 [95% CI: −0.05, 0.15] [$P = .49$]), or total advanced cancer (round 1 difference, 0.00 [95% CI: −0.15, 0.19] [$P = .94$]; round 2 or above difference, −0.06 [95% CI: −0.18, 0.11] [$P = .43$]) between DBT and DM in any screening round.

The positive predictive value of recall was significantly higher for DBT versus DM examinations for all screening rounds (Table 4). For screening round 3 or above, the absolute difference in positive predictive value of recall for DBT versus DM was 2.3% (95% CI: 1.1, 3.5) ($P < .001$). Positive predictive value of biopsy recommendation and specificity of recall were significantly

higher for DBT versus DM in screening round 1 (positive predictive value of biopsy recommendation difference, 8.8% [95% CI: 5.8, 12.2] [$P < .001$]; specificity of recall difference, 3.9% [95% CI: 2.3, 5.4] [$P < .001$]) and round 2 (positive predictive value of biopsy recommendation difference, 2.0% [95% CI: 1.1, 3.0] [$P = .008$]; specificity of recall difference, 2.0% [95% CI: 0.7, 3.3] [$P = .002$]) but not in screening round 3 and above (positive predictive value of biopsy recommendation difference, 5.2% [95% CI: −0.4, 11.5] [$P = .07$]; specificity of recall difference, 1.2% [95% CI: −0.2, 2.6] [$P = .09$]). There were no differences found in sensitivity of recall or sensitivity of positive final assessment for DBT versus DM in any screening round.

Discussion

The objective of our study was to evaluate breast cancer screening outcomes with digital breast tomosynthesis (DBT) versus digital mammography (DM) on successive screening rounds. Our results demonstrate that DBT shows reduced recall rate and increased cancer detection over multiple screening rounds compared with DM. No differences in interval or advanced cancer rates or sensitivity were observed between DBT and DM. These findings provide new evidence for women, health care providers, and policymakers evaluating the benefits, harms, and limitations of multiple rounds of breast cancer screening with DBT compared with DM.

Several prior studies analyzing cross-sectional data have reported an elevated cancer detection rate and reduced false-positive rate with DBT versus DM (2–4,6–12,31). Our study is, to our knowledge, the first multicenter report in a U.S. population on DBT performance stratified by screening round and includes methodologic advances over prior single-institution studies afforded by our large sample size. Conant et al (16) conducted a single-institution study of DBT performance by screening round and found that recall, biopsy, and cancer detection rates declined at subsequent versus first DBT screening examinations. We observed similar changes in DBT performance metrics over successive subsequent screening rounds. Conant et al (16) did not conduct formal comparisons of DBT versus DM performance based on screening round. In another single-institution study, Bahl et al (17) compared DBT performance by round to DM performance without stratifying DM examinations by screening round. Consistent with our results, they found that the recall rate at initial and successive DBT examinations was lower than within the DM group. In contrast to our results, Bahl et al (17) observed no difference in cancer detection rate between DBT and DM. This difference may be due to our larger study sample size or study design, which stratified both the DBT and DM groups by study round; notably, we found differences in screening outcomes across screening rounds for DM, indicating the importance of controlling for screening round in DBT versus DM comparisons.

A prior report from the BCSC compared DBT and DM according to whether it was the woman's first-ever ("baseline") screening examination or a subsequent screening examination, finding that the differences in recall and cancer detection rates between DBT and DM were most pronounced at baseline examinations (2). Our current study excludes baseline examinations, which have notably different performance characteristics compared with subsequent examinations, and instead stratifies subsequent screening examinations by screening round and examines a comprehensive set of screening outcomes beyond recall and cancer detection rate. In our study, the difference in recall rate between the two imaging modalities was greatest in round 1 (difference, −3.3% [95% CI: −4.6, −2.1]; $P < .001$). Although the difference in recall rate remained significantly different between modalities, it decreased over subsequent screening rounds. This attenuating pattern may be explained by the relatively high DM recall rate in round 1, which decreased on successive screening rounds—likely reflecting the value of comparison images from multiple prior screening rounds in reducing recall rate (32). Similarly, differences in the false-positive recall rate and specificity of recall between DBT and DM attenuated on successive rounds and were no longer significant at round 3 and above (false-positive recall rate difference, −1.1% [95% CI: −2.3, 0.2] [$P = .08$]; specificity of recall difference, 1.2% [95% CI: −0.2, 2.6] [$P = .09$]). These results are consistent with a recent analysis showing modest differences in cumulative 10-year risk of a false-positive result with DBT versus DM screening (33).

There is little evidence to date that DBT reduces the rate of screening failures, including interval or advanced breast cancer. Many studies have reported no difference in interval cancer rates for DBT versus DM (6–10,12). A notable exception is the Malmö Breast Tomosynthesis Screening Trial (11), which reported a reduced interval cancer rate with DBT versus DM, though the DBT interval cancer rate (1.6 per 1000 examinations) was much higher than typically observed with DM in clinical practice in the United States (<1 per 1000) (34). We sought to investigate whether differences in screening failure rates become apparent when stratifying by screening round. Consistent with prior studies (6–12), we found that at round 3 and above, DBT had a higher overall cancer detection rate (difference, 0.6 per 1000 examinations [95% CI: 0.2, 1.1] [$P = .009$]) and higher early-stage cancer detection rate (difference, 0.7 per 1000 examinations [95% CI: 0.3, 1.0] [$P < .001$]) compared with DM. We did not find lower interval or advanced cancer rates or improvement in sensitivity with DBT compared with DM. It is possible that the elevated cancer detection observed with DBT may predominantly represent slow-growing cancers that would not have arisen as interval or advanced cancers before the next screening examination. Alternatively, additional follow-up examinations and increased sample size may be required to detect small reductions in interval or advanced cancer rates between DBT and DM, particularly if the differences are limited to subgroups of women, such as those with extremely dense breasts and high breast cancer risk (10). Gur et al (5) have argued that the impact of a new screening technology on advancing cancer detection cannot be fully appreciated until the temporary rise in cancer detection rate with the new technology returns to the same steady-state cancer detection rate of the old technology. Our observation that the cancer detection rate was higher with DBT versus DM on round 3 or above suggests that the transition to a new steady state with DBT screening takes more than three screening cycles.

Our study had several limitations. First, women were not randomly assigned to DBT or DM. While we collected women's risk factor information and controlled for differences in risk characteristics across imaging modality and screening round, it is possible that residual confounding could have still impacted our estimates. Second, despite our large study population, the sample size for rare outcomes, including interval and advanced cancer, was limited, and round 2 and round 3 and above examinations were combined for these outcomes. Small differences in interval and advanced cancer rates that may be clinically significant cannot be excluded. Third, we did not have examination-level information on the use of DBT synthetic two-dimensional views, which can be constructed from the DBT images sections and eliminate the need for concurrent two-dimensional DM views. Therefore, we were unable to evaluate differences in DBT performance according to whether synthetic two-dimensional views were obtained. Prior studies suggest similar outcomes for DBT with synthetic versus DM two-dimensional views (35,36); thus, DBT with synthetic views would be preferred to avoid the extra radiation exposure associated with combined DBT and DM examinations. Fourth, we did not evaluate patterns in diagnostic work-up following screening examinations or the potential impact of diagnostic imaging technologies on screening performance metrics. Some women with positive DM screening examinations may have undergone diagnostic DBT imaging, which could have influenced the DM performance metrics.

In summary, we found that digital breast tomosynthesis (DBT) had lower recall rates and could help detect more cancers than digital mammography (DM) across three screening rounds, with no difference in interval or advanced cancer rates. DBT has disseminated widely, but not completely, into clinical practice in the United States. The limited cost-effectiveness of DBT at current reimbursement rates (37) and the lack of evidence to date for a reduction in interval or advanced cancer (6–12) with DBT has likely limited dissemination somewhat. Screening outcomes improved with both modalities over successive screening rounds. Notably, the DM recall rate in our study was within the acceptable range (5%–12%) stipulated by American College of Radiology guidelines (20) at each screening round. DBT was not associated with lower interval or advanced cancer rates; thus, one cannot exclude the possibility that the increase in early-stage cancer detection associated with DBT may predominantly add to overdiagnosis. However, DBT provides benefits over DM, with a lower recall rate at each screening round and an increased positive predictive value across three screening rounds. Overall, our results provide support for further dissemination of DBT, tempered by continued uncertainty regarding the effect of DBT on screening failure rates, overdiagnosis, and breast cancer mortality. Additional studies with a larger sample size and longer duration of follow-up are needed to further evaluate potential differences in screening failure rates between DBT and DM.

Acknowledgments: We thank the participating women, mammography facilities, and radiologists for the data they have provided for this study.

Author contributions: Guarantors of integrity of entire study, **B.L.S., R.Y.C., K.K.**; 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, **B.L.S., K.K., T.O.**; clinical studies, **K.K., T.O., D.L.M.**; statistical analysis, **R.Y.C., D.L.M.**; and manuscript editing, **B.L.S., R.Y.C., K.P.L., K.K., L.M.H., Y.R.S., C.I.L., T.O., E.J.A.B., S.D.H., R.M.d.F.A.**

Data sharing: Data generated or analyzed during the study, as well as the associated computer code, are available from the corresponding author by request.

Disclosures of conflicts of interest: B.L.S. Grants and contracts paid to institution from the American College of Radiology and the National Cancer Institute (NCI). **R.Y.C.** Co-investigator on grants from the National Institute of Mental Health (NIMH), National Institute on Aging (NIA), NCI, Agency for Healthcare Research and Quality, and Patient-Centered Outcomes Research Institute (PCORI); co-investigator on contract with the Food and Drug Administration; principal investigator on grant from the NIMH; honoraria from the University of Washington and the American Statistical Association (ASA); payment for grant review from PCORI; support to attend conference from ASA; communications committee chair of the ASA JEDI Outreach Group. **K.P.L.** Grant from the American Cancer Society. **K.K.** No relevant relationships. **L.M.H.** No relevant relationships. **Y.R.S.** Grants to institution from the NIA, Syneos Health, and the University of Washington. **C.I.L.** Textbook royalties from McGraw Hill, Oxford University Press, and UpToDate; personal fees for participation on a data safety monitoring board or advisory board for Grail; personal fees for journal editorial board work from the American College of Radiology. **T.O.** No relevant relationships. **E.J.A.B.** No relevant relationships. **S.D.H.** Participation on the medical advisory board of DenseBreast-Info.org. **R.M.d.F.A.** Grants from the National Institutes of Health; continuing medical education funds from department; membership on the American College of Radiology Appropriateness Criteria Committee (2015–2020). **D.L.M.** No relevant relationships.

References

1. U.S. Food & Drug Administration. MQSA National Statistics. https://www.fda.gov/radiation-emitting-products/mqsa-insights/mqsa-national-statistics. Updated April 3, 2022. Accessed April 20, 2023.
2. Lowry KP, Coley RY, Miglioretti DL, et al. Screening performance of digital breast tomosynthesis vs digital mammography in community practice by patient age, screening round, and breast density. JAMA Netw Open 2020;3(7):e2011792.
3. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA 2014;311(24):2499–2507.
4. 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.
5. 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.
6. Houssami N, Zackrisson S, Blazek K, et al. Meta-analysis of prospective studies evaluating breast cancer detection and interval cancer rates for digital breast tomosynthesis versus mammography population screening. Eur J Cancer 2021;148:14–23.
7. Pattacini P, Nitrosi A, Giorgi Rossi P, et al. A randomized trial comparing breast cancer incidence and interval cancers after tomosynthesis plus mammography versus mammography alone. Radiology 2022;303(2):256–266.
8. Durand MA, Friedewald SM, Plecha DM, et al. False-negative rates of breast cancer screening with and without digital breast tomosynthesis. Radiology 2021;298(2):296–305.
9. 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.
10. Kerlikowske K, Su YR, Sprague BL, et al. Association of screening with digital breast tomosynthesis vs digital mammography with risk of interval invasive and advanced breast cancer. JAMA 2022;327(22):2220–2230.
11. Johnson K, Lång K, Ikeda DM, Åkesson A, Andersson I, Zackrisson S. interval breast cancer rates and tumor characteristics in the prospective population-based Malmö Breast Tomosynthesis Screening Trial. Radiology 2021;299(3):559–567.
12. Hofvind S, Moshina N, Holen AS, et al. Interval and subsequent round breast cancer in a randomized controlled trial comparing digital breast tomosynthesis and digital mammography screening. Radiology 2021;300(1):66–76.

13. Pisano ED. Is tomosynthesis the future of breast cancer screening? Radiology 2018;287(1):47–48.
14. Hovda T, Brandal SHB, Sebuødegård S, et al. Screening outcome for consecutive examinations with digital breast tomosynthesis versus standard digital mammography in a population-based screening program. Eur Radiol 2019;29(12):6991–6999.
15. Caumo F, Montemezzi S, Romanucci G, et al. Repeat screening outcomes with digital breast tomosynthesis plus synthetic mammography for breast cancer detection: results from the prospective Verona Pilot Study. Radiology 2021;298(1):49–57.
16. 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.
17. Bahl M, Mercaldo S, Dang PA, McCarthy AM, Lowry KP, Lehman CD. Breast cancer screening with digital breast tomosynthesis: are initial benefits sustained? Radiology 2020;295(3):529–539.
18. Sprague BL, Coley RY, Kerlikowske K, et al. Assessment of radiologist performance in breast cancer screening using digital breast tomosynthesis vs digital mammography. JAMA Netw Open 2020;3(3):e201759.
19. Miglioretti DL, Abraham L, Lee CI, et al. Digital breast tomosynthesis: radiologist learning curve. Radiology 2019;291(1):34–42.
20. American College of Radiology. ACR BI-RADS - Mammography. In: ACR BI-RADS Atlas: Breast Imaging Reporting and Data System. 5th ed. Reston, Va: American College of Radiology, 2013.
21. Breast Cancer Surveillance Consortium. BCSC Standard Definitions. https://www.bcsc-research.org/data/bcsc_standard_definitions. Published 2020. Accessed February 6, 2023.
22. Sprague BL, Miglioretti DL, Lee CI, Perry H, Tosteson AAN, Kerlikowske K. New mammography screening performance metrics based on the entire screening episode. Cancer 2020;126(14):3289–3296.
23. Kerlikowske K, Bissell MCS, Sprague BL, et al. Advanced breast cancer definitions by staging system examined in the Breast Cancer Surveillance Consortium. J Natl Cancer Inst 2021;113(7):909–916.
24. Tice JA, Miglioretti DL, Li CS, Vachon CM, Gard CC, Kerlikowske K. Breast density and benign breast disease: risk assessment to identify women at high risk of breast cancer. J Clin Oncol 2015;33(28):3137–3143.
25. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13–22.
26. Yelland LN, Salter AB, Ryan P. Performance of the modified Poisson regression approach for estimating relative risks from clustered prospective data. Am J Epidemiol 2011;174(8):984–992.
27. Miglioretti DL, Heagerty PJ. Marginal modeling of multilevel binary data with time-varying covariates. Biostatistics 2004;5(3):381–398.
28. Nelson HD, O'Meara ES, Kerlikowske K, Balch S, Miglioretti D. Factors associated with rates of false-positive and false-negative results from digital mammography screening: an analysis of registry data. Ann Intern Med 2016;164(4):226–235.
29. le Cessie S. Bias formulas for estimating direct and indirect effects when unmeasured confounding is present. Epidemiology 2016;27(1):125–132.
30. Thulin M. Modern Statistics with R. Uppsala, Sweden: Eos Chasma Press, 2021.
31. Conant EF, Talley MM, Parghi CR, et al. Mammographic screening in routine practice: multisite study of digital breast tomosynthesis and digital mammography screenings. Radiology 2023. 10.1148/radiol.221571. Published online March 14, 2023.
32. Hayward JH, Ray KM, Wisner DJ, et al. Improving screening mammography outcomes through comparison with multiple prior mammograms. AJR Am J Roentgenol 2016;207(4):918–924.
33. Ho TH, Bissell MCS, Kerlikowske K, et al. Cumulative probability of false-positive results after 10 years of screening with digital breast tomosynthesis vs digital mammography. JAMA Netw Open 2022;5(3):e222440.
34. 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.
35. Weigel S, Heindel W, Hense HW, et al. Breast density and breast cancer screening with digital breast tomosynthesis: a TOSYMA Trial subanalysis. Radiology 2023;306(2):e221006.
36. Zuckerman SP, Sprague BL, Weaver DL, Herschorn SD, Conant EF. Multicenter evaluation of breast cancer screening with digital breast tomosynthesis in combination with synthetic versus digital mammography. Radiology 2020;297(3):545–553.
37. Lowry KP, Trentham-Dietz A, Schechter CB, et al. Long-term outcomes and cost-effectiveness of breast cancer screening with digital breast tomosynthesis in the United States. J Natl Cancer Inst 2020;112(6):582–589.