Title

Digital Breast Tomosynthesis: Radiologist Learning Curve.

Journal

Radiology, 291(1)

ISSN

0033-8419

Authors

Miglioretti, Diana L
Abraham, Linn
Lee, Christoph I
et al.

Publication Date

2019-04-01

DOI

10.1148/radiol.2019182305

Peer reviewed

Radiology

Digital Breast Tomosynthesis: Radiologist Learning Curve

Diana L. Miglioretti, PhD • Linn Abraham, MS • Christoph I. Lee, MD, MS • Diana S. M. Buist, PhD • Sally D. Herschorn, MD • Brian L. Sprague, PhD • Louise M. Henderson, PhD • Anna N. A. Tosteson, ScD • Karla Kerlikowske, MD • for the Breast Cancer Surveillance Consortium

From the Division of Biostatistics, Department of Public Health Sciences, University of California, Davis School of Medicine, One Shields Ave, Med Sci 1C, Room 144, Davis, CA 95616 (D.L.M.); Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington, Seattle, Wash (D.L.M., L.A., D.S.M.B.); Department of Radiology, University of Washington School of Medicine; Department of Health Services, University of Washington School of Public Health; Hutchinson Institute for Cancer Outcomes Research, Seattle, Wash (C.I.L.); Department of Radiology (S.D.H.) and Department of Surgery, Office of Health Promotion Research (B.L.S.), Larner College of Medicine at the University of Vermont and University of Vermont Cancer Center, Burlington, Vt; Department of Radiology, University of North Carolina, Chapel Hill, NC (L.M.H.); The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth and Norris Cotton Cancer Center, Lebanon, NH (A.N.A.T.); and Departments of Medicine andEpidemiology and Biostatistics and the General Internal Medicine Section, Department of Veterans Affairs, University of California, San Francisco, Calif (K.K.). Received October 5, 2018; revision requested November 25; revision received January 4, 2019; accepted January 8. **Address correspondence to** D.L.M. (e-mail: *dmiglioretti@ucdavis.edu*).

This research was supported by a Patient-Centered Outcomes Research Institute Program Award (PCS-1504-30370). C.I.L's time was supported, in part, by the American Cancer Society (126947-MRSG-1416001CPHPS). Data collection was additionally supported by the Agency for Healthcare Research and Quality (R01 HS018366-01A1), National Cancer Institute (P01CA154292, U54CA163303), and the University of Vermont Cancer Center (Lake Champlain Cancer Research Organization grant 032800).

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

See also the editorial by Hooley in this issue.

Radiology 2019; 291:34–42 • https://doi.org/10.1148/radiol.2019182305 • Content code: BR

Background: There is growing evidence that digital breast tomosynthesis (DBT) results in lower recall rates and higher cancer detection rates when compared with digital mammography. However, whether DBT interpretative performance changes with experience (learning curve effect) is unknown.

Purpose: To evaluate screening DBT performance by cumulative DBT volume within 2 years after adoption relative to digital mammography (DM) performance 1 year before DBT adoption.

Materials and Methods: This prospective study included 106 126 DBT and 221 248 DM examinations in 271 362 women (mean age, 57.5 years) from 2010 to 2017 that were interpreted by 104 radiologists from 53 facilities in the Breast Cancer Surveillance Consortium. Conditional logistic regression was used to estimate within-radiologist effects of increasing cumulative DBT volume on recall and cancer detection rates relative to DM and was adjusted for examination-level characteristics. Changes were also evaluated by subspecialty and breast density.

Results: Before DBT adoption, DM recall rate was 10.4% (95% confidence interval [CI]: 9.5%, 11.4%) and cancer detection rate was 4.0 per 1000 screenings (95% CI: 3.6 per 1000 screenings, 4.5 per 1000 screenings); after DBT adoption, DBT recall rate was lower (9.4%; 95% CI: 8.2%, 10.6%; $P = .02$) and cancer detection rate was similar (4.6 per 1000 screenings; 95% CI: 4.0 per 1000 screenings, 5.2 per 1000 screenings; $P = .12$). Relative to DM, DBT recall rate decreased for a cumulative DBT volume of fewer than 400 studies (odds ratio [OR] = 0.83; 95% CI: 0.78, 0.89) and remained lower as volume increased (400–799 studies, OR = 0.8 [95% CI: 0.75, 0.85]; 800–1199 studies, OR = 0.81 [95% CI: 0.76, 0.87]; 1200–1599 studies, OR = 0.78 [95% CI: 0.73, 0.84]; 1600–2000 studies, OR = 0.81 [95% CI: 0.75, 0.88]; $P < .001$). Improvements were sustained for breast imaging subspecialists (OR range, 0.67–0.85; $P < .02$) and readers who were not breast imaging specialists (OR range, 0.80–0.85; $P < .001$). Recall rates decreased more in women with nondense breasts (OR range, 0.68–0.76; $P < .001$) than in those with dense breasts (OR range, 0.86–0.90; $P \leq .05$; P interaction $< .001$). Cancer detection rates for DM and DBT were similar, regardless of DBT volume ($P \geq .10$).

Conclusion: Early performance improvements after digital breast tomosynthesis (DBT) adoption were sustained regardless of DBT volume, radiologist subspecialty, or breast density.

There is growing evidence to suggest screening with digital breast tomosynthesis (DBT) in the United States results in lower recall and higher cancer detection rates when compared with those attained with digital mammography (1–6). The Food and Drug Administration requires 8 hours of additional training for Mammography Quality Standards Act (MQSA)-qualified radiologists to interpret DBT images in clinical practice (7). Prior studies have not evaluated whether DBT interpretive performance changes with clinical cumulative DBT interpretive volume (ie, whether there is a learning curve) or whether improvements in performance are sustained by radiologists over time.

A single-institution study from the University of Pennsylvania (2) evaluated screening DBT outcomes by year since

Abbreviations

BCSC = Breast Cancer Surveillance Consortium, BI-RADS = Breast Imaging Reporting and Data System, CI = confidence interval, DBT = digital breast tomosynthesis, MQSA = Mammography Quality Standards Act, OR = odds ratio

Summary

Recall rates for digital breast tomosynthesis (DBT) were reduced early after DBT adoption and were sustained for 2 years relative to preadoption digital mammography recall rates, with no change in cancer detection rate.

Key Points

- In a prospective study of 104 radiologists from 53 facilities, screening mammography recall rate was lower for digital breast tomosynthesis (DBT) (9.4%) than for digital mammography prior to DBT adoption (10.4%, $P = .02$), without a corresponding decrease in cancer detection rate (4.6 per 1000 screenings for DBT, 4.0 per 1000 screenings for digital mammography; $P = .12$).
- Radiologists improved their screening mammography recall rates shortly after DBT adoption relative to their performance with digital mammography prior to DBT adoption (odds ratio [OR] range, 0.78–0.83), without decreasing their cancer detection rates, regardless of DBT cumulative volume, radiologist subspecialty, or breast density.
- Improvements in recall rate were greater in women with nondense breasts (OR range, 0.68–0.76) as compared with women with dense breasts (OR range, 0.86–0.90).

adoption of DBT relative to digital mammography performance during the year before adoption. Overall, the study found the DBT recall rate increased slightly from year 1 to year 3 after the facility adopted DBT, but it was still lower than the recall rate with digital mammography; however, when evaluated separately by breast density, this reduction was sustained only among women with nondense breasts (2,8). The cancer detection rate of DBT increased during the 3 years after adoption but was not significantly different from that with digital mammography (2). This study did not evaluate within-radiologist changes in performance, nor did it evaluate performance with increasing cumulative DBT volume.

We used data from the Breast Cancer Surveillance Consortium (BCSC) (9) to evaluate whether radiologists experience a learning curve for DBT interpretive performance. We compared the performance of screening DBT by cumulative total (screening and diagnostic) DBT volume within 2 years after DBT adoption relative to performance of screening digital mammography during the year before DBT adoption. We also evaluated whether the learning curve might be different for breast imaging subspecialists compared with nonbreast imaging subspecialists or for women with dense breasts versus those with nondense breasts.

Materials and Methods

BCSC registries and the Statistical Coordinating Center received institutional review board approval for passive consenting processes (three registries) or waiver of written informed consent (two registries and the Statistical Coordinating Center) to enroll participants, link and pool data, and perform analysis. All procedures were compliant with the Health Insurance Portability and Accountability Act. Registries and the Statistical Coordinating Center received a federal Certificate of Confidentiality and other protections for the identities of women, physicians, and facilities.

Study Setting and Prospective Cohort

Prospective data were collected by five breast imaging registries in the BCSC (9,10): Carolina Mammography Registry, New Hampshire Mammography Network, Vermont Breast Cancer Surveillance System, San Francisco Mammography Registry, and Metropolitan Chicago Breast Cancer Registry. BCSC registries prospectively collect woman-level characteristics and clinical information from community radiology facilities with populations representing the U.S. population (11) and link to state or regional tumor registries and pathology databases for complete capture of breast cancer diagnoses (9). A total of 219 radiologists interpreted DBT screening examinations at BCSC facilities that adopted DBT between 2011 and 2017. We excluded 40 radiologists whose average interpretive volume during the DBT interpretation period was less than the MQSA minimum (960 mammograms over 2 years [12]) because they likely interpreted mammograms at non-BCSC facilities. We excluded 15 radiologists who interpreted fewer than 400 DBT studies of any type because we were interested in within-radiologist effects of cumulative DBT volume on screening performance and radiologists had to contribute outcome data for at least the first two volume subgroups to provide information about model parameters of interest. We excluded 22 radiologists who interpreted fewer than 100 screening DBT studies and 38 radiologists who interpreted fewer than 100 screening digital mammograms within 1 year prior to DBT adoption to ensure there would be a sufficient number of events to estimate within-radiologist effects. Most of the excluded radiologists were locums who provided coverage for radiologists who were away on vacation or professional travel, with low interpretive volumes at BCSC facilities. After these exclusions, 104 radiologists who interpreted mammograms at 53 facilities remained in the final sample. Screening DBT mammograms (n = 106 126) were included if they were interpreted within 2 years after the radiologist's adoption of DBT and before interpretation of more than 2000 DBT studies for any indication. Screening digital mammograms (n = 221 248) interpreted 1 year before DBT adoption were included for comparison. A screening examination was defined according to the BCSC strict definition *(http://www.bcsc-research.org/data/bcsc_data_definitions_version2_final__2017.pdf)*.

Radiologist Characteristics

For each of 104 eligible radiologists, we determined whether breast imaging was their main subspecialty by using facilities' websites or by directly contacting facilities. We were unable to determine subspecialties for 15 radiologists (14%). Radiologist-specific DBT interpretation periods were defined as the time between the first date of any DBT interpretation at a BCSC facility (ie, DBT adoption date) and the date radiologists interpreted 2000 DBT examinations for any indication or 2 years after start date (interpretation period end date), whichever was earliest. Radiologist's annual interpretive volume was calculated based on the total number of mammograms obtained with either modality (digital mammography or DBT) for any indication (screening or diagnostic) that were interpreted at a BCSC facility during the radiologist's

Table 1: Characteristics of 104 Radiologists Interpreting 106 126 Screening DBT Studies from 2011 to 2017 and 221 248 Screening Digital Mammograms 1 Year before DBT Adoption

Measurement	All Radiologists (*n* = 104)	Breast Imaging Subspecialist*		*P* Value Comparing Breast Imaging Subspecialties
		Yes (*n* = 25)	No (*n* = 64)	
Cumulative DBT volume over DBT interpretation period				
400–799 studies	28 (26.9)	3 (12.0)	19 (29.7)	…
800–1199 studies	17 (16.3)	7 (28.0)	7 (10.9)	…
1200–1599 studies	19 (18.3)	3 (12.0)	14 (21.9)	…
1600–2000 studies	40 (38.5)	12 (48.0)	24 (37.5)	…
Length of DBT interpretation period				
6 to <12 months	12 (11.5)	1 (4.0)	10 (15.6)	…
12 to <18 months	17 (16.3)	5 (20.0)	12 (18.8)	…
18 to ≤24 months	75 (72.1)	19 (70.6)	42 (65.6)	…
Annual interpretive volume of all mammograms†	2470 (1605, 4096)	3325 (2129, 5044)	2242 (1415, 3576)	…
Annual interpretive volume of DBT†	817 (515, 1444)	875 (496, 1455)	841 (565, 1376)	…
Percentage screening DBT (%)†	89.9 (81.3, 94.7)	86.9 (77.4, 92.4)	91.9 (83.2, 94.7)	…
Screening recall rate (%)				
Digital mammography 1 year prior to DBT adoption‡	10.4 (9.5, 11.4)	10.0 (9.0, 11.1)	11.0 (9.8, 12.4)	.19
DBT‡	9.4 (8.2, 10.6)	9.7 (8.5, 11.1)	9.6 (8.1, 11.4)	.90
P value comparing digital mammography to DBT	.017	.60	.025	
Screening cancer detection rate (per 1000 screens) ‡§				
Digital mammography 1 year prior to DBT adoption‡	4.0 (3.6, 4.5)	4.3 (3.5, 5.4)	3.9 (3.4, 4.4)	.30
DBT	4.6 (4.0, 5.2)	4.9 (3.8, 6.5)	4.3 (3.7, 4.9)	.28
P value comparing digital mammography to DBT	.12	.30	.26	…

Note.—Unless otherwise indicated, data are number of radiologists and data in parentheses are percentages. DBT = digital breast tomosynthesis.

* Radiologists were considered breast imaging subspecialists if breast imaging was their main subspecialty. Subspecialty was unknown for 15 (14%) of the 104 radiologists.

† Data are medians, and data in parentheses are 25th and 75th percentiles.

‡ Data are means, and data in parentheses are 95% confidence intervals.

§ Based on 182 568 screening digital studies and 74 983 screening DBT studies with at least 90 days of follow-up interpreted by 80 radiologists.

DBT interpretation period and the length of the interpretation period. Radiologist's cumulative DBT volume associated with each screening DBT examination was calculated based on the total number of DBT studies for any indication interpreted at a BCSC facility prior to the date of screening DBT mammography.

Risk Factors

Demographic and health history information (age, first-degree family history of breast cancer, time since last mammography, history of biopsy, and history of breast cancer) were gathered from self-administered questionnaires at the time of mammography, extraction from electronic medical records, or both. Radiologists categorized Breast Imaging Reporting and Data System (BI-RADS) breast density at the time of clinical interpretation as almost entirely fat, scattered fibroglandular densities, heterogeneously dense, or extremely dense (13). History of benign breast disease was obtained from pathology databases, and diagnoses were grouped into four categories by the diagnosis with the highest grade (14,15). BCSC 5-year risk of invasive cancer was calculated by using the BCSC risk calculator, version 2 (16,17).

Screening Outcomes

Breast cancer diagnoses and tumor characteristics were obtained by linking with pathology databases and regional surveillance, epidemiology, and end results programs or state tumor registries (18). Recall and cancer detection rates were calculated by using initial BI-RADS screening assessment (considered positive if 0, 3, 4, or 5 or negative if 1 or 2 based on American College of Radiology guidelines) (13). Recall rate was calculated as the number of screenings with positive initial assessment divided by the total number of screenings. Cancer detection rate was calculated as the number of screenings with positive initial assessment and invasive carcinoma or ductal carcinoma in situ diagnosed within 90 days divided by the total number of screenings.

Table 2: Examination-level Characteristics of 106 126 Screening DBT Studies Interpreted by 104 Eligible Radiologists from 2011 to 2017 and 221 248 Screening Digital Mammograms Interpreted 1 Year before DBT Adoption

	Overall		Breast Imaging Subspecialist*			
			Yes		No	
Characteristic	Digital Mammography (*n* = 221 248)	DBT (*n* = 106 126)	Digital Mammography (*n* = 64 903)	DBT (*n* = 25 885)	Digital Mammography (*n* = 123 957)	DBT (*n* = 67 440)
Age at screening (y)†	57 (49, 66)	56 (48, 65)	57 (49, 65)	54 (47, 62)	57 (49, 65)	57 (49, 65)
<40	3227 (1.5)	1892 (1.8)	947 (1.5)	500 (1.9)	1843 (1.5)	1164 (1.7)
40–49	55 091 (24.9)	28 425 (26.8)	16 159 (24.9)	8310 (32.1)	31 009 (25.0)	17 056 (25.3)
50–59	69 625 (31.5)	34 176 (32.2)	21 006 (32.4)	8704 (33.6)	39 083 (31.5)	21 545 (31.9)
60–69	56 619 (25.6)	27 342 (25.8)	16 649 (25.7)	6017 (23.2)	31 844 (25.7)	17 822 (26.4)
≥70	36 686 (16.6)	14 291 (13.5)	10 142 (15.6)	2354 (9.1)	20 178 (16.3)	9853 (14.6)
First-degree family history of breast cancer (missing 6%)						
No	179 385 (85.2)	83 002 (84.3)	51 054 (83.2)	18 653 (81.9)	100 151 (85.4)	53 253 (84.6)
Yes	31 248 (14.8)	15 403 (15.7)	10 342 (16.8)	4131 (18.1)	17 150 (14.6)	9658 (15.4)
BI-RADS breast density (missing 7%)						
Almost entirely fat	22 310 (10.8)	7949 (8.0)	9249 (15.8)	2044 (8.8)	10 751 (9.3)	4967 (7.7)
Scattered fibroglandular densities	92 411 (44.9)	44 518 (44.6)	26 955 (46.0)	10 419 (44.8)	51 974 (44.7)	28 463 (44.4)
Heterogeneously dense	74 432 (36.1)	39 507 (39.6)	17 525 (29.9)	8415 (36.2)	43 865 (37.8)	25 980 (40.5)
Extremely dense	16 752 (8.1)	7790 (7.8)	4904 (8.4)	2356 (10.1)	9564 (8.2)	4746 (7.4)
BCSC version 2 5-year risk (missing 15%)						
Low (<1.0%)	61 215 (32.8)	28 686 (30.9)	18 512 (34.7)	7790 (35.1)	33 244 (31.4)	17 101 (28.8)
Average 1.0% to <1.67%)	75 117 (40.2)	37 119 (39.9)	20 184 (37.9)	8037 (36.2)	43 371 (41.0)	24 215 (40.8)
Intermediate (1.67% to <2.5%)	36 227 (19.4)	19 035 (20.5)	10 159 (19.1)	4173 (18.8)	20 899 (19.8)	12 783 (21.5)
High (≥2.5%)	14 217 (7.6)	8110 (8.7)	4448 (8.3)	2187 (9.9)	8277 (7.8)	5306 (8.9)
Time since last mammography of any type (missing 3%)						
No previous mammography	15 491 (7.2)	6793 (6.7)	4077 (6.5)	1752 (7.2)	7834 (6.5)	3858 (5.9)
Within 2 years	180 737 (84.1)	85 969 (84.3)	53 689 (85.5)	20 403 (84.2)	101 875 (84.7)	55 207 (84.8)
3 or more years	18 785 (8.7)	9213 (9.0)	5041 (8.0)	2085 (8.6)	10 593 (8.8)	6004 (9.2)
First versus subsequent DBT						
First	…	96 323 (90.8)	…	24 024 (92.8)	…	60 501 (89.7)
Subsequent	…	9803 (9.2)	…	1861 (7.2)	…	6939 (10.3)

Note.—Unless otherwise indicated, data are numbers of radiologists, and data in parentheses are percentages. BI-RADS = Breast Imaging Reporting and Data System, BCSC = Breast Cancer Surveillance Consortium, DBT = digital breast tomosynthesis.

* Radiologists were considered breast imaging subspecialists if breast imaging was their main subspecialty. Subspecialty unknown for 15 of 104 radiologists (14%).

† Data are medians, and data in parentheses are 25th and 75th percentiles.

Statistical Analysis

We summarized characteristics of included radiologists and screening mammograms. The unit of analysis was the screening mammogram. Mean digital mammography and DBT recall and cancer detection rates with 95% confidence intervals (CIs) were estimated with a logistic regression model via generalized estimating equations with a working independence correlation structure to account for correlation among mammograms from the same woman, radiologists, facility, or a combination thereof (19,20). We used conditional logistic regression analysis to evaluate the average

Table 3: Odds Ratios Measuring the Association with Cumulative DBT Volume by Reference Group Measuring Average Within-Radiologist Changes in Performance with Cumulative DBT Volume at the Time of Screening Examination

		Model with Digital Mammography as Reference Group		Model with DBT Volume <400 Studies as Reference Group	
Measure or Subgroup	No. of Examinations	Odds Ratio*	*P* Value	Odds Ratio*	*P* Value
Recall rate	292 091				
Digital mammography	197 948	Reference	…	1.20 (1.13, 1.28)	<.001
Cumulative DBT volume					
<400 studies	27 892	0.83 (0.78, 0.89)	<.001	Reference	…
400–799 studies	24 379	0.80 (0.75, 0.85)	<.001	0.95 (0.90, 1.01)	.13
800–1199 studies	18 632	0.81 (0.76, 0.87)	<.001	0.98 (0.91, 1.04)	.46
1200–1599 studies	13 302	0.78 (0.73, 0.84)	<.001	0.94 (0.87, 1.01)	.10
1600–2000 studies	9938	0.81 (0.75, 0.88)	<.001	0.97 (0.89, 1.06)	.53
Cancer detection rate†	234 388				
Digital mammography	168 794	Reference		1.02 (0.72, 1.44)	.90
Cumulative DBT volume					
<400 studies	21 114	0.98 (0.69, 1.38)	.90	Reference	
400–799 studies	17 208	0.97 (0.68, 1.37)	.84	0.99 (0.72, 1.35)	.93
800–1199 studies	12 036	1.12 (0.79, 1.59)	.52	1.14 (0.81, 1.61)	.44
1200–1599 studies	8424	1.36 (0.95, 1.96)	.10	1.39 (0.96, 2.01)	.08
1600–2000 studies	6812	0.86 (0.54, 1.36)	.51	0.87 (0.55, 1.39)	.57

Note.—Unless otherwise indicated, data are odds ratios, and data in parentheses are 95% confidence intervals. Odds ratios characterize average within-radiologist changes in performance with cumulative digital breast tomosynthesis (DBT) volume at time of screening relative to screening digital mammograms interpreted 1 year before DBT adoption (model with digital mammography as reference group) and relative to cumulative DBT volume of fewer than 400 studies (model with DBT volume of fewer than 400 studies as the reference group).

* Separate conditional logistic regression models were fit for recall rate and cancer detection rate, adjusting for a woman's age, Breast Imaging Reporting and Data System breast density, family history, whether she had undergone mammography within the past 2 years, and whether it was her first or a subsequent DBT examination, and were limited to examinations with nonmissing covariates. Odds ratios from each model are reported with two different reference groups.

† Cancer detection rate restricted to studies with at least 3 months of complete cancer capture.

within-radiologist effects of increasing DBT volume subgroups (<400, 400–799, 800–1199, 1200–1599, and 1600–2000 studies) on the probability of recall and cancer detection relative to both screening digital mammograms 1 year before DBT adoption and lowest volume subgroup. Conditional logistic regression removes effects of between-radiologist heterogeneity, including effects of any potential confounding variables that vary among radiologists, such as clinical practice characteristics and BCSC registry. Thus, we adjusted only for potential confounders that vary within radiologists: a woman's age, BI-RADS density, first-degree family history, time since last mammography, and for DBT examinations, a woman's first DBT examination versus any subsequent DBT examination. Models were fit overall, by radiologist's main subspecialty, and by BI-RADS breast density categorized as not dense (almost entirely fat or scattered fibroglandular densities) or dense (heterogeneously dense or extremely dense).

Statistical analyses were performed with SAS software (SAS, version 9.4; SAS Institute, Cary, NC). Statistical significance tests were two sided, with $\alpha = .05$.

Results

Table 1 shows the characteristics of the 104 radiologists included in our study. More than half of the radiologists (59 of 104, 57%) had a cumulative DBT volume of at least 1200 studies during their interpretation period. Most radiologists (75 of 104, 72%) contributed at least 18 months of DBT interpretation data. Overall, the median annual interpretive volume of all mammograms (digital and DBT) was 2470 studies, and the median DBT volume was 817 studies. Breast imaging subspecialists had higher annual mammography and DBT volumes than did readers who were not breast imaging subspecialists (median annual volume, 3325 vs 2242 studies; median DBT volume, 875 vs 841 studies). A median of 89.9% of DBT mammograms were for a screening indication, with a slightly lower percentage for breast imaging specialists (86.9%) versus readers who were not breast imaging specialists (91.9%).

The mean recall rate was higher for screening digital mammography 1 year before DBT adoption (10.4%; 95% CI: 9.5%, 11.4%) than for DBT (9.4%; 95% CI: 8.2%, 10.6%) (P = .02). The difference in recall rates for digital mammography versus DBT was smaller for breast imaging specialists (10.0% vs 9.7%, P = .60) than for readers who were not breast imaging subspecialists (11.0% vs 9.6%, P = .02). The cancer detection rate was lower for digital mammography (4.0 per 1000 screenings; 95% CI: 3.6 per 1000 screenings, 4.5 per 1000 screenings) compared with DBT (4.6 per 1000 screenings; 95% CI: 4.0 per 1000 screenings, 5.2 per 1000 screenings) (P = .12). Recall and cancer detection rates were not significantly

Table 4: Odds Ratios Measuring Average Within-Radiologist Changes in Performance with Cumulative DBT Volume at Time of Screening Examination Relative to Screening Digital Mammograms Interpreted 1 Year before DBT Adoption

Measure or Subgroup	Breast Imaging Subspecialists*	Readers Who Were Not Breast Imaging Subspecialists*	Women with Not Dense Breasts	Women with Dense Breasts	Screening DBT Volume†
	Recall Rate				
No. of examinations	76994	172285	159659	132432	275948
No. of radiologists	25	64	104	103	96
Conditional logistic regression results					
Digital mammography	Reference	Reference	Reference	Reference	Reference
Cumulative DBT volume					
<400 studies	0.85 (0.75, 0.98)	0.82 (0.76, 0.89)	0.76 (0.69, 0.83)	0.90 (0.83, 0.98)	0.81 (0.76, 0.86)
400–799 studies	0.76 (0.67, 0.87)	0.80 (0.74, 0.86)	0.70 (0.64, 0.77)	0.88 (0.81, 0.96)	0.79 (0.74, 0.84)
800–1199 studies	0.84 (0.73, 0.96)	0.81 (0.75, 0.87)	0.76 (0.69, 0.84)	0.86 (0.79, 0.94)	0.80 (0.75, 0.85)
1200–1599 studies	0.67 (0.57, 0.79)	0.85 (0.78, 0.93)	0.68 (0.61, 0.76)	0.87 (0.79, 0.96)	0.83 (0.78, 0.90)
1600–2000 studies	0.78 (0.66, 0.91)	0.83 (0.75, 0.92)	0.73 (0.65, 0.82)	0.90 (0.80, 1.00)	0.69 (0.63, 0.75)
	Cancer Detection Rate‡				
No. of examinations	68389	136738	126840	107548	223425
No. of radiologists	21	51	80	79	74
Conditional logistic regression results					
Digital mammography	Reference	Reference	Reference	Reference	Reference
Cumulative DBT volume					
<400 studies	0.94 (0.47, 1.88)	0.83 (0.54, 1.28)	0.95 (0.58, 1.56)	0.94 (0.57, 1.55)	0.99 (0.72, 1.35)
400–799 studies	1.17 (0.61, 2.24)	0.88 (0.56, 1.37)	0.75 (0.43, 1.29)	1.04 (0.64, 1.69)	0.98 (0.71, 1.34)
800–1199 studies	1.11 (0.57, 2.15)	1.08 (0.70, 1.68)	1.29 (0.79, 2.11)	0.96 (0.57, 1.59)	1.23 (0.90, 1.69)
1200–1599 studies	0.90 (0.41, 1.97)	1.62 (1.04, 2.50)	1.55 (0.94, 2.58)	1.18 (0.69, 2.00)	1.25 (0.87, 1.79)
1600–2000 studies	0.62 (0.26, 1.50)	0.95 (0.52, 1.72)	1.15 (0.64, 2.07)	0.60 (0.28, 1.27)	1.01 (0.65, 1.55)

Note.—Unless otherwise indicated, data are odds ratios, and data in parentheses are 95% confidence intervals. Separate conditional logistic regression models were fit for each outcome, adjusting for a woman's age, BI-RADs breast density, family history, whether or not she had undergone mammography within the past 2 years, and whether it was the woman's first or a subsequent digital breast tomosynthesis (DBT) examination, and were limited to examinations with nonmissing covariates.

* Radiologists were considered breast imaging subspecialists if breast imaging was their main subspecialty. Subspecialty was unknown for 15 of 104 radiologists (14%).

† Screening DBT volume was used instead of overall DBT volume for this analysis.

‡ Cancer detection rate restricted to studies with at least 3 months of complete cancer capture.

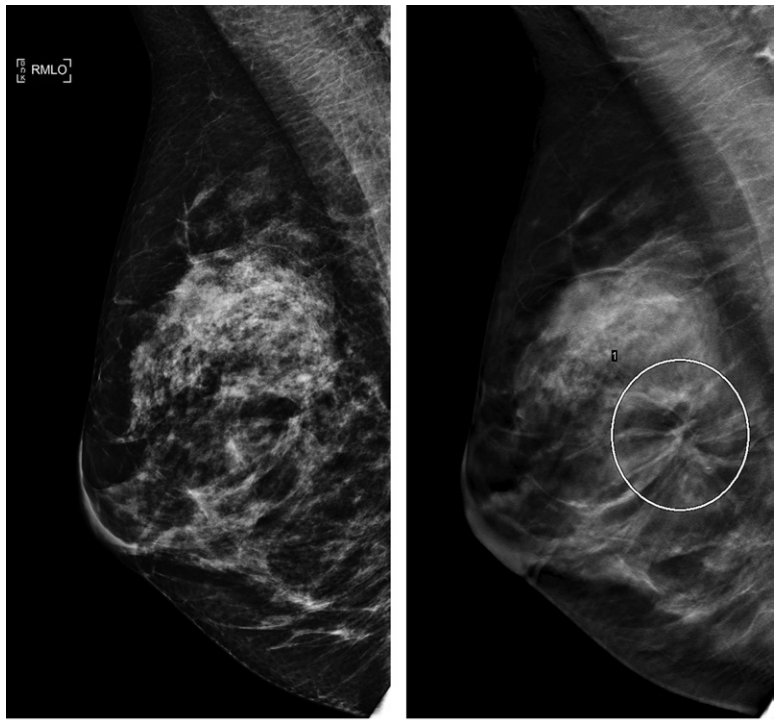
different between breast imaging subspecialists and readers who were not breast imaging subspecialists within each modality ($P > .18$ in all cases).

Table 2 shows examination-level characteristics of the 221248 screening digital mammograms and 106126 screening DBT examinations performed in 271362 women. The median age at screening was 54–57 years. Similar proportions of digital and DBT examinations (31248 of 221248 [14.8%] vs 15403 of 106126 [15.7%], respectively) were performed in women with a first-degree family history of breast cancer overall regardless of the radiologist's subspecialty. Women who underwent digital mammography were slightly less likely to have heterogeneously or extremely dense breasts (91184 of 205905 [44.3%]) compared with women who underwent DBT (47297 of 99764 [47.4%]), and this difference was larger for screenings interpreted by breast imaging subspecialists (22429 of 58633 [38.3%] vs 10771 of 23234 [46.4%]) than for those interpreted by readers who were not breast imaging subspecialists (53429 of 116154 [46.0%] vs 30726 of 64156 [47.9%]). One-third of screenings were performed in women with a BCSC 5-year risk of less than 1.0%. Most screenings (digital mammography, 180737 of 215013 [84.1%]; DBT, 85969 of 101975 [84.3%]) were performed in women who had undergone mammography in the prior 2 years, and 96323 of 106126 (90.8%) DBT screenings were the woman's first DBT examination.

Table 3 shows the conditional logistic regression results after adjusting for patient- and examination-level characteristics, with odds ratios (ORs) calculated both relative to digital mammography performance 1 year before DBT adoption and relative to the lowest cumulative DBT volume category (<400 studies). Relative to digital mammography, the recall rate for the lowest cumulative volume subgroup decreased within radiologists (OR = 0.83; 95% CI: 0.78, 0.89; $P < .001$) and continued to remain lower for all volume subgroups (400–799 studies: OR = 0.80; 95% CI: 0.75, 0.85) (800–1199 studies: OR = 0.81; 95% CI: 0.76, 0.87) (1200–1599 studies: OR = 0.78; 95% CI: 0.73, 0.84) (1600–2000 studies: OR = 0.81; 95% CI: 0.75, 0.88) ($P < .001$ for all subgroups). When compared with the lowest DBT volume

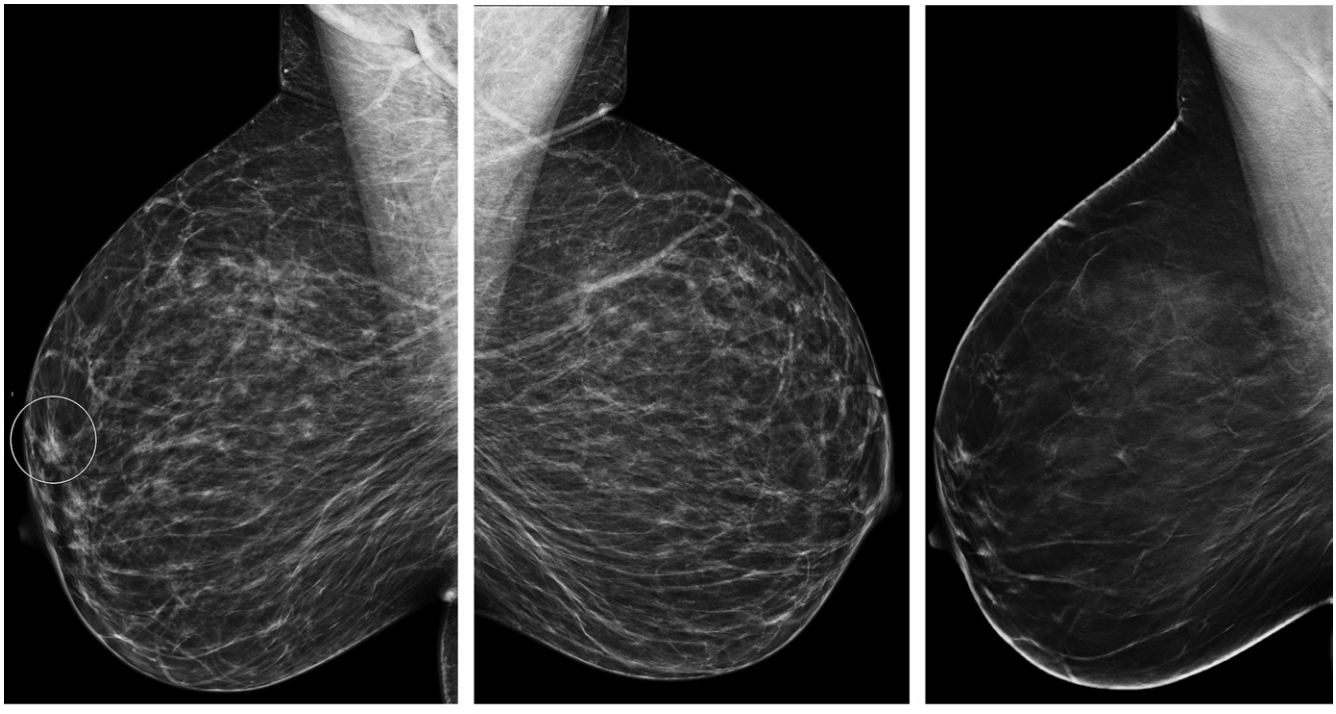
category, the within-radiologist effects of increasing cumulative DBT volume on recall rate were not significant ($P > .09$ for all subgroups). Cancer detection rates were similar for DBT and digital mammography examinations regardless of DBT volume ($P > .09$ for all subgroups); however, 95% CIs were wider than for recall rate analysis.

Table 4 shows the subgroup analyses by radiologist subspecialty and breast density, adjusting for patient- and examination-level characteristics. Improvements in performance were similar for breast imaging subspecialists and readers who were not breast imaging subspecialists and were sustained with increasing DBT volume in both groups. There was an interaction between increasing DBT volume and breast imaging subspecialty (P interaction = .06), with a trend toward decreasing recall rate with increasing DBT volume among breast imaging subspecialists (P trend = .09) but not among readers who were not breast imaging subspecialists (P = .50); however, differences among DBT volume subgroups are small relative to the differences when comparing DBT with digital mammography. Within-radiologist changes in DBT cancer detection rates relative to digital mammography were similar for breast imaging subspecialists and readers who were not breast imaging subspecialists (P interaction = .50). Relative decreases in recall rates were larger in women with nondense breasts (OR range, 0.68–0.76 across volume subgroups) than in



a. b.

Figure 1: Screening digital mammogram and digital breast tomosynthesis image in a 43-year-old woman. **(a)** Right mediolateral oblique digital mammogram shows a heterogeneously dense breast with no abnormality. **(b)** Right mediolateral oblique digital breast tomosynthesis section shows architectural distortion (○) that represents invasive ductal carcinoma that is occult at digital mammography.



a. b. c.

Figure 2: Screening digital mammogram and digital breast tomosynthesis image in a 52-year-old woman. **(a)** Right and **(b)** left mediolateral oblique digital mammograms show scattered fibroglandular densities with an asymmetry superior to the right nipple (○). **(c)** Right mediolateral oblique digital breast tomosynthesis image (representative section shown) shows normal overlapping fibroglandular tissue at the site of digital mammography asymmetry. The woman avoided a false-positive recall from screening.

women with dense breasts (OR range, 0.86–0.90; *P* interaction < .001). Cancer detection rates for DBT were higher than those for digital mammography only when cumulative DBT volume was 1200–1599 studies for radiologists who were not breast imaging subspecialists (OR = 1.62; 95% CI: 1.04, 2.50; *P* = .03). Results were similar when evaluating cumulative screening DBT volume instead of cumulative volume of all DBT studies.

Discussion

The Food and Drug Administration requires only 8 hours of additional training for Mammography Quality Standards Act (MQSA)-qualified radiologists to be able to interpret digital breast tomosynthesis (DBT) studies (7), which may be insufficient for radiologists to improve their DBT screening interpretative performance. We found no evidence of a learning curve for the clinical interpretation of screening DBT studies. The average within-radiologist recall rate decreased shortly after DBT adoption both for breast imaging subspecialists (OR range, 0.67–0.85) and for readers who are not breast imaging subspecialists (OR range, 0.80–0.85), without a corresponding decrease in average cancer detection rate. These performance improvements were sustained for at least 2 years after adoption.

Improvements in recall rate were greater in women with nondense breasts (OR range, 0.68–0.76) than in women with dense breasts (OR range, 0.86–0.90). This is an important finding, as three-dimensional mammography is considered a valuable adjunct screening test in women with dense breasts because it can be used to potentially improve detection of cancers masked by dense tissue at digital mammography (Fig 1) while eliminating focal asymmetries from overlapping fibroglandular tissue, thereby reducing the number of false-positive findings. However, our study suggests that women with nondense breasts benefit from DBT screening equally or possibly more than women with dense breasts through a reduced recall rate (Fig 2). This finding is mainly consistent with findings of the University of Pennsylvania study which showed reduced recall rates for DBT versus digital mammography in women with either dense or nondense breasts; however, this improvement was significantly sustained only in women with nondense breasts (2,8). In contrast, a study using aggregate data from 13 U.S. institutions reported larger absolute reductions in recall rates in women with dense breasts versus women with nondense breasts (1.8 per 100 women vs 1.2 per 100 women) (21). A meta-analysis including U.S. and European studies found no significant association between the proportion of women with dense breasts and the magnitude of recall rate reductions with DBT versus digital mammography; however, that study had relatively low power to test for this relationship (1).

Rapid improvements in screening performance after DBT adoption that persisted over time were not unexpected. While DBT is a relatively recent technology, interpretation of DBT images is very similar to interpretation of digital mammograms. DBT images are acquired with multiple two-dimensional mammograms taken in standard mammographic projections. DBT images can be scrolled through, similar to how three-dimensional CT scans are scrolled through, using a process familiar to radiologists. DBT unmasks small imaging findings that radiologists must learn to recognize as benign and subtle areas of architectural distortion that they must learn to recognize as suspicious. Our analysis across a large geographically diverse set of radiologists suggests that the learning curve for acquiring these skills is very short, on average, with sustained improved screening performance soon after DBT adoption; however, changes in performance with experience may vary across radiologists.

We evaluated average within-radiologist changes in DBT performance with increasing cumulative interpretive volume of screening and diagnostic DBT. BCSC studies suggest that radiologists' performance depends on both screening and diagnostic interpretive volumes (22,23). Similarly, radiologists' DBT screening performance is likely influenced by additional exposure to diagnostic DBT images, with interpretive confidence resulting from both screening and diagnostic experience. Our sensitivity analysis evaluating cumulative screening volume instead of cumulative interpretive volume of DBT of any type showed similar results.

Most participating facilities operated in a hybrid setting during the study period, meaning they offered both digital mammography and DBT. A strength of this study was our ability to adjust for patient risk factors and other potential confounders that may have been associated with the imaging modality and its performance. We also adjusted for whether screenings were the first or a subsequent DBT examination; this may be a confounder, given that more DBT examinations are subsequent examinations, as cumulative volume increases, and performance may differ if prior DBT images are available for comparison (24).

Another strength of our study is the use of BCSC data collected for 104 radiologists with a mix of experience and interpretative volumes from 53 academic and nonacademic facilities in five diverse U.S. states. Evaluation of the effectiveness of technologies in general community practice is important because most U.S. women do not have their mammograms interpreted by a breast imaging subspecialist, nor do they have access to academic medical centers (25,26).

Our study limitations included possible underestimation of cumulative volume for radiologists who interpreted mammograms at non-BCSC facilities; however, prior research suggests this is rare for radiologists who meet MQSA volume requirements based on mammograms interpreted at BCSC facilities (22). Also, not all radiologists were followed for a full 2 years or interpreted 2000 DBT studies within the first 2 years after DBT adoption; however, more than half of radiologists interpreted at least 1200 DBT studies, and most contributed at least 18 months of data.

In conclusion, we found within-radiologist improvement in screening interpretive performance early after DBT adoption that was sustained for at least 2 years. We observed an improvement in recall rate with no decrease in cancer detection rate, regardless of DBT volume, for both breast imaging subspecialists and readers who were not breast imaging subspecialists. Improvements in recall rates were observed regardless of breast density but were larger in women with nondense breasts than in those with dense breasts.

Acknowledgments: Chris Tachibana, PhD, provided scientific editing. We thank the participating women, mammography facilities, and radiologists for the data they provided for this study. Information about the Breast Cancer Surveillance Consortium (BCSC) may be found at *http://www.bcsc-research.org/*. Cancer and vital status data collection was supported by several state public health departments and cancer registries *(http://www.bcsc-research.org/work/acknowledgement.html)*.

Author contributions: Guarantors of integrity of entire study, D.L.M., 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, D.L.M., C.I.L., K.K.; clinical studies, D.L.M., D.S.M.B., K.K.; statistical analysis, D.L.M., L.A.; and manuscript editing, all authors

Disclosures of Conflicts of Interest: **D.L.M.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: was on the scientific advisory board for Hologic in 2017, receives royalties from Elsevier. Other relationships: disclosed no relevant relationships. **L.A.** disclosed no relevant relationships. **C.I.L.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: institution received a grant from GE Healthcare; received royalties from UpToDate, Oxford University Press, and McGraw-Hill. Other relationships: disclosed no relevant relationships. **D.S.M.B.** disclosed no relevant relationships. **S.D.H.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: was compensated for travel to the ABUS meeting by GE Healthcare. Other relationships: disclosed no relevant relationships. **B.L.S.** disclosed no relevant relationships. **L.M.H.** disclosed no relevant relationships. **A.N.A.T.** disclosed no relevant relationships. **K.K.** disclosed no relevant relationships.

References

1. 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.
2. 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.
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, 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. 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.
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. Lee CI, Gupta S, Sherry SJ, et al. Translating new imaging technologies to clinical practice. Acad Radiol 2018;25(1):3–8.
8. Houssami N, Miglioretti DL. Digital breast tomosynthesis: a brave new world of mammography screening. JAMA Oncol 2016;2(6):725–727.
9. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol 1997;169(4):1001–1008.
10. Breast Cancer Surveillance Consortium. Research Sites & Principal Investigators. http://www.bcsc-research.org/about/sites.html. Published 2018. Accessed September 4, 2018.
11. 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.
12. Mammography Quality Standards Act. 42 USC 102–539. 1992.
13. American College of Radiology. American College of Radiology Breast Imaging Reporting and Data System Atlas (BI-RADS Atlas). 5th ed. Reston, Va: American College of Radiology, 2013.
14. Tice JA, O'Meara ES, Weaver DL, Vachon C, Ballard-Barbash R, Kerlikowske K. Benign breast disease, mammographic breast density, and the risk of breast cancer. J Natl Cancer Inst 2013;105(14):1043–1049.
15. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Ann Intern Med 2008;148(5):337–347.
16. Breast Cancer Surveillance Consortium Risk Calculator V2. https://tools.bcsc-scc.org/bc5yearrisk/calculator.htm. Published 2018. Accessed September 4, 2018.
17. 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.
18. Ernster VL, Ballard-Barbash R, Barlow WE, et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. J Natl Cancer Inst 2002; 94(20):1546–1554.
19. Miglioretti DL, Heagerty PJ. Marginal modeling of nonnested multilevel data using standard software. Am J Epidemiol 2007;165(4):453–463.
20. Miglioretti DL, Heagerty PJ. Marginal modeling of multilevel binary data with time-varying covariates. Biostatistics 2004;5(3):381–398.
21. Rafferty EA, Durand MA, Conant EF, et al. Breast cancer screening using tomosynthesis and digital mammography in dense and nondense breasts. JAMA 2016; 315(16):1784–1786.
22. Buist DS, Anderson ML, Haneuse SJ, et al. Influence of annual interpretive volume on screening mammography performance in the United States. Radiology 2011; 259(1):72–84.
23. Buist DS, Anderson ML, Smith RA, et al. Effect of radiologists' diagnostic work-up volume on interpretive performance. Radiology 2014;273(2):351–364.
24. Hakim CM, Catullo VJ, Chough DM, et al. Effect of the availability of prior full-field digital mammography and digital breast tomosynthesis images on the interpretation of mammograms. Radiology 2015;276(1):65–72.
25. Lewis RS, Sunshine JH, Bhargavan M. A portrait of breast imaging specialists and of the interpretation of mammography in the United States. AJR Am J Roentgenol 2006;187(5):W456–W468.
26. Smith-Bindman R, Miglioretti DL, Rosenberg R, et al. Physician workload in mammography. AJR Am J Roentgenol 2008;190(2):526–532.