

Effect of the Availability of Prior Full-Field Digital Mammography and Digital Breast Tomosynthesis Images on the Interpretation of Mammograms¹

Christiane M. Hakim, MD
Victor J. Catullo, MD
Denise M. Chough, MD
Marie A. Ganott, MD
Amy E. Kelly, MD
Dilip D. Shinde, MD
Jules H. Sumkin, DO
Luisa P. Wallace, MD
Andriy I. Bandos, PhD
David Gur, ScD

Purpose:

To assess the effect of and interaction between the availability of prior images and digital breast tomosynthesis (DBT) images in decisions to recall women during mammogram interpretation.

Materials and Methods:

Verbal informed consent was obtained for this HIPAA-compliant institutional review board–approved protocol. Eight radiologists independently interpreted twice deidentified mammograms obtained in 153 women (age range, 37–83 years; mean age, 53.7 years \pm 9.3 [standard deviation]) in a mode by reader by case-balanced fully crossed study. Each study consisted of current and prior full-field digital mammography (FFDM) images and DBT images that were acquired in our facility between June 2009 and January 2013. For one reading, sequential ratings were provided by using (a) current FFDM images only, (b) current FFDM and DBT images, and (c) current FFDM, DBT, and prior FFDM images. The other reading consisted of (a) current FFDM images only, (b) current and prior FFDM images, and (c) current FFDM, prior FFDM, and DBT images. Fifty verified cancer cases, 60 negative and benign cases (clinically not recalled), and 43 benign cases (clinically recalled) were included. Recall recommendations and interaction between the effect of prior FFDM and DBT images were assessed by using a generalized linear model accounting for case and reader variability.

Results:

Average recall rates in noncancer cases were significantly reduced with the addition of prior FFDM images by 34% (145 of 421) and 32% (106 of 333) without and with DBT images, respectively ($P < .001$). However, this recall reduction was achieved at the cost of a corresponding 7% (23 of 345) and 4% (14 of 353) reduction in sensitivity ($P = .006$). In contrast, availability of DBT images resulted in a smaller reduction in recall rates (false-positive interpretations) of 19% (76 of 409) and 26% (71 of 276) without and with prior FFDM images, respectively ($P = .001$). Availability of DBT images resulted in 4% (15 of 338) and 8% (25 of 322) increases in sensitivity, respectively ($P = .007$). The effects of the availability of prior FFDM images or DBT images did not significantly change regardless of the sequence in presentation ($P = .81$ and $P = .47$ for specificity and sensitivity, respectively).

Conclusion:

The availability of prior FFDM or DBT images is a largely independent contributing factor in reducing recall recommendations during mammographic interpretation.

© RSNA, 2015

¹From the Department of Radiology, Division of Breast Imaging, Magee-Womens Hospital of UPMC, 300 Halket St, Pittsburgh, PA 15213 (C.M.H., V.J.C., D.M.C., M.A.G., A.E.K., D.D.S., J.H.S., L.P.W.); Department of Biostatistics, Graduate School of Public Health (A.I.B.), and Department of Radiology, Division of Research Imaging (D.G.), University of Pittsburgh, Pittsburgh, Pa. Received August 22, 2014; revision requested October 15; revision received November 14; accepted December 10; final version accepted December 18. Address correspondence to C.M.H. (e-mail: chakim@magee.edu).

Digital breast tomosynthesis (DBT) is approved for clinical use with the expectation that it will improve the efficacy of screening and diagnostic mammography by increasing the conspicuity of imaged abnormalities (1). Publications have shown an effect on the interpretation of screening mammograms in terms of reducing recall rates while, in some studies, simultaneously increasing cancer detection rates, particularly in patients with invasive cancers (2–9). However, some studies did not include

the availability of prior full-field digital mammography (FFDM) images during the interpretation of DBT images and, to our knowledge, none have specifically reported on the effect of the availability of prior FFDM images or the interaction between the availability of prior FFDM and DBT images on the resulting interpretation. In clinical practice, the vast majority of women have prior FFDM studies available for comparison. It is well known that the availability of these prior studies during interpretation of screening mammograms reduces recall rates substantially by enabling the observer to simultaneously (a) assess change, if any, over time and (b) discard depicted abnormalities that clearly represent a variety of stable benign findings (10–13). Prospective studies have shown that even with the availability of prior studies in the majority of patients, recall rates decreased substantially when DBT images were included in the studies being interpreted (5,9,10,14,15). However, we are unaware of any studies designed to evaluate the relative contribution of prior FFDM studies and DBT studies on the decision of whether or not to recall women for diagnostic work-up of suspicious findings. There are also no data on the interaction, if any, between these two factors.

The purpose of this study was to assess the relative effect of and potential interaction between the availability of prior FFDM images and DBT images on the recall rate when interpreting mammography studies.

Materials and Methods

The enriched fully deidentified digital mammography image set used in this

study consisted of prior FFDM, current FFDM, and DBT images. The FFDM and DBT images were acquired at our facility between June 2009 and January 2013 under institutional review board–approved research protocols (written consent was required). Consenting women underwent both conventional FFDM as part of their routine care and DBT as a part of our research. All current studies were acquired with a DBT system (Dimensions; Hologic, Bedford, Mass) by using a standard tomosynthesis acquisition technique that generated 15 low-dose projection images, or frames, acquired for reconstruction of the three-dimensional DBT image sets. Although studies acquired prior to the U.S. Food and Drug Administration approval (February 2011) were considered experimental, the system used was technically the same as the one that was later approved for clinical use. After acquisition, data from the low-dose frames were used to reconstruct 50–90 parallel 1-mm-thick sections, the number of which depended on the thickness of the compressed breast. All prior FFDM examinations were performed with a clinical system (Selenia;

Advances in Knowledge

- Availability of prior full-field digital mammograms (FFDM) and digital breast tomosynthesis (DBT) images is a largely independent contributing component in reducing recall rate during interpretation of mammograms.
- When compared with interpretation of FFDM images exclusively, review of current and prior FFDM images led to a 34% (145 of 421) and 32% (106 of 333) reduction in the number of false-positive interpretations without and with availability of DBT images, respectively ($P < .001$); however, review of prior FFDM images also led to an actual reduction in sensitivity of 7% (23 of 345) and 4% (14 of 353) without and with availability of DBT images, respectively ($P = .006$).
- When compared with interpretations of current FFDM images exclusively, review of current FFDM and DBT images led to a reduction of false-positive interpretations by 19% (76 of 409) and 26% (71 of 276) without and with availability of prior FFDM images, respectively ($P = .001$); furthermore, review of current FFDM and DBT images also led to an increase in true-positive interpretations by 4% (15 of 338) and 8% (25 of 322) without and with availability of prior FFDM images, respectively ($P = .007$).

Implication for Patient Care

- Understanding the primary variables that effect recall rates and the relative magnitude of their effect during interpretation of mammograms could improve our ability to optimize clinical practices.

Published online before print

10.1148/radiol.15142009 Content codes: **BR** **DM**

Radiology 2015; 276:65–72

Abbreviations:

AUC = area under the ROC curve
CI = confidence interval
DBT = digital breast tomosynthesis
FFDM = full-field digital mammography
ROC = receiver operating characteristic

Author contributions:

Guarantors of integrity of entire study, C.M.H., J.H.S.; 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.G.; clinical studies, C.M.H., D.M.C., M.A.G., D.D.S., J.H.S.; statistical analysis, A.I.B.; and manuscript editing, C.M.H., D.M.C., M.A.G., D.D.S., J.H.S., D.G.

Funding:

This research was supported by the National Institutes of Health (grants CA163300, CA143019, and GM098253).

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

Hologic) as a part of standard care and were collected retrospectively with an institutional review board–approved research protocol (written informed consent was waived).

Study Population

FFDM and DBT images obtained in 153 women who ranged in age from 37 to 83 years (mean age, 53.7 years \pm 9.3 [standard deviation]) were sequentially selected for this study from a pool of 519 studies (predominantly from prior research studies), 36 of which were used in a pilot study (16) and were excluded from this study. The selection of cases was based on the availability of a prior FFDM study obtained between 1 and 3 years prior to the current FFDM and DBT studies of interest and the requirement to fulfill a desired distribution of verified outcomes, breast densities, and abnormality types. Cases with visually extremely obvious findings were excluded. Negative and benign cases were verified at least 1 year after the examination in question. All cases with findings positive for cancer were verified with pathologic examination. Acquisition time of prior FFDM images ranged from 11 to 37 months prior to acquisition of the current FFDM images (mean, 23.4 months \pm 5.7). To stress test the experiment in terms of obtaining higher-than-clinically-expected recall rates, 43 (42%) of the 103 women with negative benign findings were actually recalled for diagnostic work-up during clinical practice. In the remaining 60 (58%) women, the negative benign findings were rated as negative or benign (Breast Imaging Reporting and Data System category 1 or 2); hence, they were not recalled during the original clinical interpretations. Among the 43 recalled women with negative studies, biopsy was performed in eight, and 35 were determined to have benign findings during diagnostic work-up and later verified as such during subsequent imaging-based periodic follow-up for at least 1 year. Fifty cancer cases were included in the study. Of these, 16 were ductal carcinomas in situ, 31 were invasive ductal

carcinomas, two were invasive lobular carcinomas, and one was papillary carcinoma. The subjectively rated breast density distribution for these cases was eight of 153 (5.2%), 43 of 153 (28.1%), 95 of 153 (62.1%), and seven of 153 (4.6%) for breast tissue density almost entirely fat, scattered fibroglandular density, heterogeneously dense, and extremely dense, respectively. In this study, we focused on evaluating changes, if any, in patients with Breast Imaging Reporting and Data System category 2 and 3 breast density (particularly Breast Imaging Reporting and Data System category 3) that constitute the majority of screening cases.

Participating Radiologists

Eight Mammography Quality Standards Act–qualified radiologists (for FFDM and DBT) with 3–28 years of experience participated in this study (C.M.H., V.J.C., D.M.C., M.A.G., A.E.K., D.D.S., J.H.S., L.P.W.). The radiologists received a detailed Instruction to Observers document to review before beginning the study. The document informed the readers that this was a study to assess if and how the availability of different types of imaging-based information might affect decisions in a screening environment. The document also defined the type of examinations used in the study, described the general set-up and protocol for reviewing and rating each of the studies, informed the readers that computer-aided detection would not be provided, and requested that, if at all possible, readers should attempt to complete the readings in each mode (set) in approximately 2 months. No prevalence information in the data set was provided to the readers.

Observer Study

The reader study was conducted with a dual-monitor SecurView (Hologic) workstation that was remotely controlled via a laptop with an in-house–developed study management application (Study Manager; written in-house by two members of the research staff). This application communicated with the workstation via the SecurView AppSync

(Hologic) programming interface. This application was responsible for driving the study in terms of case randomization by reader and session and case presentation sequencing (worklist) within each session and was also used to record all ratings. The workstation includes two 5-megapixel liquid crystal displays that enabled viewing of one, two, or four images per display for each monitor. Each study consisted of two conventional views of each breast (craniocaudal and mediolateral oblique projections).

For cross-balancing purposes, the study involved two separate sequential readings of each study, and the radiologists retrospectively interpreted each study twice. During the first half of the cases, four of the radiologists read current four-view FFDM images (current FFDM only), followed by the DBT images (current FFDM + DBT), and then the prior FFDM images (current FFDM + DBT + prior FFDM). At midpoint, the order of presentation was switched so that radiologists first viewed current four-view FFDM images only (current FFDM only), then current and prior FFDM images (current FFDM + prior FFDM), then current FFDM, prior FFDM, and DBT images (current FFDM + prior FFDM + DBT). During the second reading of each study, the order of presentation between DBT and prior FFDM images for each case was reversed. The order of presentation for the remaining four radiologists was reversed from the first group; namely, they started with reading half the cases as current FFDM images only, followed by current and prior FFDM images, then current and prior FFDM images and DBT images. During each reading, the radiologist was asked to score or rate the study in question after each group (combination) of images was presented within the sequence; thus, each reader sequentially scored each case three times under each presentation sequence. Radiologists reported or scored their breast-based recommendations by using a screening BI-RADS rating scale (category 0, 1, or 2). The reader was prompted to provide a

Table 1

Recall Rates for 153 Mammography Cases Evaluated Twice by Each of Eight Radiologists in Two Sequential Evaluations

Reader	Total No. of Cases	FFDM				FFDM and DBT		Current and Prior FFDM		Current FFDM, DBT, and Prior FFDM			
		Reading 1		Reading 2		Reading 1		Reading 2		Reading 1		Reading 2	
		No. of Cases Recalled	Recall Rate	No. of Cases Recalled	Recall Rate	No. of Cases Recalled	Recall Rate	No. of Cases Recalled	Recall Rate	No. of Cases Recalled	Recall Rate	No. of Cases Recalled	Recall Rate
Noncancer Cases													
1	103	33	0.32	33	0.32	28	0.27	23	0.22	17	0.17	9	0.09
2	103	38	0.37	36	0.35	36	0.35	28	0.27	31	0.30	32	0.31
3	103	30	0.29	38	0.37	31	0.30	22	0.21	16	0.16	18	0.17
4	103	74	0.72	77	0.75	68	0.66	60	0.58	43	0.42	46	0.45
5	103	92	0.89	75	0.73	69	0.67	33	0.32	39	0.38	24	0.23
6	103	48	0.47	49	0.48	30	0.29	34	0.33	20	0.19	21	0.20
7	103	34	0.33	41	0.40	26	0.25	31	0.3	21	0.20	21	0.20
8	103	60	0.58	72	0.70	45	0.44	45	0.44	40	0.39	34	0.33
Total	824	409	...	421	...	333	...	276	...	227	...	205	...
Average	0.50	...	0.51	...	0.40	...	0.33	...	0.28	...	0.25
Cancer Cases													
1	50	36	0.72	39	0.78	43	0.86	38	0.76	40	0.80	41	0.82
2	50	42	0.84	41	0.82	46	0.92	40	0.80	46	0.92	44	0.88
3	50	43	0.86	45	0.90	48	0.96	43	0.86	47	0.94	46	0.92
4	50	46	0.92	47	0.94	46	0.92	42	0.84	43	0.86	46	0.92
5	50	43	0.86	43	0.86	42	0.84	37	0.74	39	0.78	38	0.76
6	50	40	0.8	42	0.84	40	0.80	40	0.80	39	0.78	44	0.88
7	50	43	0.86	42	0.84	43	0.86	39	0.78	42	0.84	44	0.88
8	50	45	0.9	46	0.92	45	0.90	43	0.86	43	0.86	44	0.88
Total	400	338	...	345	...	353	...	322	...	339	...	347	...
Average	0.85	...	0.86	...	0.88	...	0.81	...	0.85	...	0.87

Note.—Current FFDM-only images were evaluated in both sessions. Recall rate for noncancer cases represents a false-positive fraction or 1 – specificity; for cancer cases, it represents a true-positive fraction or sensitivity.

breast-based probability of malignancy rating (range, 0–100) only if the breast in question was scored as BI-BRADS category 0 (ie, recall). There was a minimum 8-week interval (time delay) between each of the reading modes and a delay of at least 100 days between a specific case being presented to a radiologist for the second time. All radiologists completed the reading sessions within the planned period.

Data Analysis

Recall recommendations were evaluated for each reader and for different subsets of cases. A noncancer case was considered a false-positive recall if either breast received a Breast Imaging Reporting and Data System score of 0. A

case with cancer was considered a true-positive recall only if the breast with a verified cancer received a Breast Imaging Reporting and Data System score of 0 (there were no bilateral cancer cases in the set). Analysis of recall rates, cancer detection rates, and possible interactions between the effect of prior studies and DBT images on decisions was performed by using the generalized linear mixed model for binary data (PROC GLIMMIX in SAS version 9.3; SAS Institute, Cary, NC) and adjusting for sources of variability due to cases and readers. Effects of availability of DBT images and prior FFDM images were adjusted for possible differences between the reading sequences (order of presentation) and evaluated by using Type III

tests with a significance level of $P < .05$. The 95% confidence interval (CI) was constructed by using the nonparametric bootstrap method, with 50000 resamples over cases and readers. To confirm the consistency of our findings, we also performed receiver operating characteristic (ROC) curve analysis using the probability of malignancy ratings (after assigning a probability of malignancy = 0 for all nonrecalled studies). The average area under the ROC curves (AUC) was compared by using multireader analysis (OR-DBM MRMC, version 3.0; SAS Institute) (17), accounting for sources of heterogeneity and correlations due to readers and cases. Pairwise comparisons of imaging modalities at the specific reading sequence were performed by

Table 2

Average Recall Rates for the Same Case for Different Subtypes of Cases over All Readers and for Both Readings

Case Type	Current FFDM Images		Current FFDM and DBT Images		Current FFDM and Prior FFDM Images		Current FFDM, DBT, and Prior FFDM Images	
	Recall Rate	No. of Studies (n×8×2)	Recall Rate	No. of Studies (n×8)	Recall Rate	No. of Studies (n×8)	Recall Rate	No. of Studies (n×8×2)
All noncancer	0.50	1648	0.40	824	0.33	824	0.27	1648
Not recalled in clinic	0.43	960	0.32	480	0.21	480	0.17	960
Benign	0.61	688	0.52	344	0.51	344	0.39	688
Cancer	0.85	800	0.88	400	0.81	400	0.86	800

Note.—The effect of availability of prior FFDM and DBT images was significant in the not-recalled-in-clinic subset, with $P = .03$ and $P < .001$, respectively, and in the benign subset, with $P = .002$ and $P < .001$, respectively.

using a two-sided test at a significance level of $P < .05$ and 95% CIs based on the subsets of data corresponding to the modalities being compared.

Results

The frequencies of recall recommendation for each reader individually and for all readers together are summarized in Tables 1 and 2. The combination of current FFDM, DBT, and prior FFDM images had the lowest average recall rate of noncancer cases for both reading sequences, with a rate of 0.28 (227 of 824) and 0.25 (205 of 824) during reading sequences 1 and 2, respectively. The addition of the DBT images led to a significant reduction of 22% in the noncancer recall rate (average reduction of 19% without and 26% with prior FFDM images) after adjusting for the availability of prior FFDM images and reading sequences ($P = .001$; 95% CI: 0.09, 0.34). During the first reading sequence, the noncancer recall rate was reduced by 19% (76 of 409) from 0.50 (409 of 824) for current FFDM images only during the first reading to 0.40 (333 of 824) for current FFDM and DBT images. Similarly, during the second reading when DBT images were made available after prior FFDM images, the recall rate for noncancer cases was reduced by 26% (71 of 276), from 0.33 (276 of 824) for current FFDM and prior FFDM images to 0.25 (205 of 824) for current FFDM, prior FFDM, and DBT images.

At the same time, the availability of DBT images resulted in a significant increase in cancer cases recall rate by 6% (averages of 4% [15 of 338] without and 8% [25 of 322] with prior images) after adjusting for the availability of prior images and reading sequence ($P = .007$; 95% CI: 0.001, 0.14). There was no meaningful interaction between the effects of the availability of DBT images and prior FFDM images (specificity, $P = .81$; sensitivity, $P = .47$); namely, the improvement associated with the addition of DBT images remained approximately the same regardless of the availability (or lack thereof) of prior FFDM images (Fig 1).

The availability of prior FFDM images also led, on average, to a significant 33% (averages of 34% without and 32% with DBT images) reduction in the recall rate of noncancer cases after adjusting for the availability of DBT image and reading session ($P < .001$; 95% CI: 0.21, 0.44) (Table 1). Recall rates were reduced by 32% (106 of 333) from 0.40 (333 of 824) for current FFDM and DBT images to 0.28 (227 of 824) for current FFDM, DBT, and prior FFDM images in the first reading and by 34% (145 of 421) from 0.51 (421 of 824) for current FFDM images only to 0.33 (276 of 824) for current and prior FFDM images during the second reading. However, the availability of prior FFDM images also led to a significant decrease in the recall rate of cancer cases of 5% (averages of 7% [23 of 345] without and

Figure 1

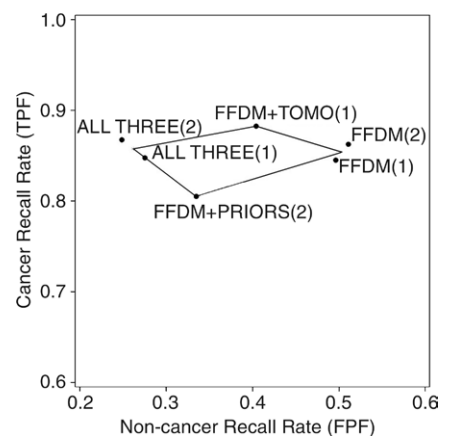


Figure 1: Graph shows average recall rates for reading modalities and sequences (reading sequence in parenthesis and lines connect the modality-specific averages for the specific reading sequence).

4% [14 of 353] with DBT images) after adjusting for availability of DBT images and reading sequence ($P = .006$; 95% CI: 0.02, 0.10).

There was no significant difference in recall rate for either cancer or noncancer cases between the comparable modes in the first and second reading sequence (specificity, $P = .64$; sensitivity, $P = .15$). Reduction in recall rates due to the availability of prior or DBT images remained significant in the benign ($P = .002$ and $P < .001$, correspondingly) and negative (nonrecalled in clinic) subsets of noncancer cases ($P = .03$ and $P < .001$, correspondingly) (Table 2). ROC analysis by using

Table 3

Average Probability of Malignancy–based AUC for Each Modality and Statistical Summaries of Differences Showing Effects of Availability of Prior FFDM Images Only, DBT Images Only, the Combination of Prior FFDM and DBT Images, and the Sequence of Reading

Effect	Compared Modalities		AUC*				
	Modality 1	Modality 2	Modality 1	Modality 2	Difference (1 vs 2)	95% CI for the Difference*	P Value†
Sequence							
Comparison A	Current FFDM (2)	Current FFDM (1)	0.827	0.808	0.019	−0.001, 0.039	.055
Comparison B	Current FFDM+prior FFDM +DBT (2)	Current FFDM+DBT+prior FFDM (1)	0.895	0.877	0.018	−0.004, 0.040	.109
DBT							
Comparison A	Current FFDM+DBT (1)	Current FFDM (1)	0.873	0.808	0.065	0.024, 0.107	.002
Comparison B	Current FFDM+ prior FFDM +DBT (2)	Current FFDM+prior FFDM (2)	0.895	0.834	0.061	0.023, 0.099	.002
Prior images							
Comparison A	Current FFDM+prior FFDM (2)	Current FFDM (2)	0.834	0.827	0.007	−0.017, 0.031	.543
Comparison B	Current FFDM+DBT+prior FFDM (1)	Current FFDM+DBT (1)	0.877	0.873	0.004	−0.018, 0.027	.692

* Each AUC provided represents the average of eight reader-specific AUCs for the specific reading sequence (1 or 2 in parentheses).

† P values for testing hypotheses of no difference and confidence intervals for computed differences were obtained by using OR-DBM software (MRMC, version 3.0; SAS Institute) based on the specific subsets of cases for the modalities being compared.

Figure 2

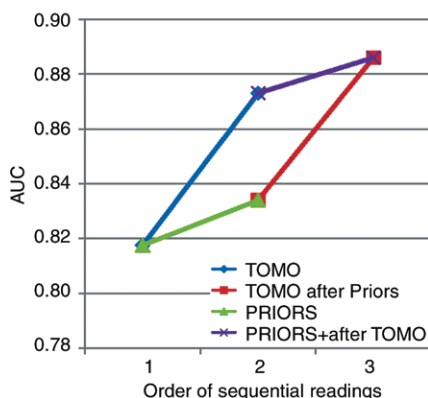


Figure 2: Graph shows the effect of the availability of prior FFDM and DBT images on average probability of malignancy (POM)-based AUCs. *TOMO* = digital breast tomosynthesis.

the probability of malignancy ratings showed similarly that the availability of DBT images led to improvement in the AUC of approximately 0.06, regardless of the availability of prior images (Table 3, Fig 2). In particular, with the addition of DBT images, the AUC increased from 0.81 to 0.87 ($P = .002$) when prior images were not available and from 0.83 to 0.90 ($P = .002$) when prior images were available (Tables 3, 4). The effect of the availability of prior

Table 4

Reader-specific Probability of Malignancy-based AUCs

Reader	Modality*					
	Current FFDM (1)	Current FFDM (2)	Current FFDM and DBT (1)	Current and Prior FFDM (2)	Current FFDM, DBT, and Prior FFDM (1)	Current FFDM, Prior FFDM, and DBT (2)
1	0.777	0.829	0.887	0.833	0.871	0.893
2	0.840	0.842	0.924	0.850	0.935	0.903
3	0.867	0.878	0.943	0.891	0.945	0.940
4	0.809	0.831	0.881	0.812	0.883	0.906
5	0.668	0.721	0.742	0.775	0.797	0.827
6	0.803	0.806	0.858	0.834	0.859	0.903
7	0.857	0.844	0.867	0.808	0.869	0.897
8	0.839	0.862	0.880	0.867	0.860	0.891
Average	0.808	0.827	0.873	0.834	0.877	0.895

* Numbers in parentheses represent the reading sequence.

images on AUC was negligible, regardless of the availability of DBT images (AUC change < 0.01 , $P = .54$ and $P = 0.69$ for changes without and with DBT images, respectively). The reading sequence did not have any substantial effect on the AUC (difference of 0.02, $P = .06$ for current FFDM images only; difference of 0.02, $P = .11$ when both prior FFDM and DBT images were available). Of note, the almost identical average changes in

performance levels shown in the similar and almost perfect parallelograms for binary responses and probability of malignancy-based AUCs in Figures 1 and 2 show not only the consistency in our results for the specific type of additional information made available to the interpreter (prior FFDM and/or DBT images) but also the magnitude of the average change in performance with each type of additional information provided.

Discussion

Our finding of no significant interaction between the two primary factors of prior FFDM and DBT images suggests that the effect of the two types of information is largely independent of each other. The almost identical slopes of average changes in performance levels shown in the similar and almost perfect parallelograms for binary responses and probability of malignancy-based AUCs show not only the consistency in our results for specific types of additional information made available to the interpreter (prior FFDM images and/or DBT images) but also the magnitude of the average change in performance with each type of additional information provided to the readers. In our study, the availability of DBT images had an effect on radiologists' performance levels comparable with that of the availability of prior images. While availability of prior images resulted in a larger decrease in recall rates, it also reduced sensitivity, whereas availability of DBT images simultaneously decreased recall rate and increased sensitivity. To our knowledge, no studies have been performed to assess the interaction, if any, between the availability of prior FFDM and DBT images on decisions to recall (or not) a woman for additional imaging on the basis of diagnostic work-up. Our results suggest that in cases in which prior images are not readily available (eg, baseline studies, first time in a clinic without prior images for some reason, loss of prior images because of archiving errors) the use of DBT as a primary modality should yield accuracy that is comparable to, if not better than, that of current and prior FFDM images. In the context of this study, the consistency in our results in terms of the lack of significant interaction between the effect of the availability of prior FFDM or DBT images on recall rates and sensitivity, regardless of the availability of the other supplementary type of information or the order of presentation, support the general validity of our conclusion.

Our study had several limitations. First, this was a single-institutional

retrospective reader study and, as such, the results may not be directly generalized to actual clinical practice. Second, our data set was enriched and is not representative of the distribution of cases in the clinical environment; therefore, our results could potentially lead to under- or overestimation of the expected effect on recall rates in actual clinical practice. We also chose to increase the number of cases of Breast Imaging and Reporting Data Systems category 3 breast density during clinical interpretation. This use of a stress test is important, as it generally helps in assessing an upper limit of the effect being investigated. Third, our study included a combination of FFDM and DBT images; hence, the effect of DBT images in an environment where synthesized images (eg, C-View; Hologic) are used in lieu of FFDM images could not be assessed (18,19).

In conclusion, the availability of prior FFDM and DBT images during interpretation of screening mammograms is a largely independent contributing factor leading to a reduction in the frequency of recommendations to recall a woman without cancer for diagnostic work-up. The availability of DBT images had a larger effect on radiologists' performance levels than did the availability of prior FFDM images.

Disclosures of Conflicts of Interest: C.M.H. disclosed no relevant relationships. V.J.C. disclosed no relevant relationships. D.M.C. disclosed no relevant relationships. M.A.G. disclosed no relevant relationships. A.E.K. disclosed no relevant relationships. D.D.S. disclosed no relevant relationships. J.H.S. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: received a grant from Hologic. Other relationships: disclosed no relevant relationships. L.P.W. disclosed no relevant relationships. A.I.B. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: provided statistical analysis of another study for Hologic. Other relationships: disclosed no relevant relationships. D.G. disclosed no relevant relationships.

References

1. U.S. Food and Drug Administration. Selenia Dimensions 3D Systems – P080003. <http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/Recently-ApprovedDevices/ucm246400.htm>. Published 2013. Accessed August 1, 2014.
2. Poplack SP, Tosteson TD, Kogel CA, Nagy HM. Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. *AJR Am J Roentgenol* 2007;189(3):616–623.
3. Gur D, Abrams GS, Chough DM, et al. Digital breast tomosynthesis: observer performance study. *AJR Am J Roentgenol* 2009;193(2):586–591.
4. Skaane P, Gullien R, Bjørndal H, et al. Digital breast tomosynthesis (DBT): initial experience in a clinical setting. *Acta Radiol* 2012;53(5):524–529.
5. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology* 2013;267(1):47–56.
6. Rafferty EA, Park JM, Philpotts LE, et al. Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial. *Radiology* 2013;266(1):104–113.
7. Bernardi D, Ciatto S, Pellegrini M, et al. Application of breast tomosynthesis in screening: incremental effect on mammography acquisition and reading time. *Br J Radiol* 2012;85(1020):e1174–e1178.
8. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 2014;311(24):2499–2507.
9. Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *AJR Am J Roentgenol* 2014;203(3):687–693.
10. Sumkin JH, Holbert BL, Herrmann JS, et al. Optimal reference mammography: a comparison of mammograms obtained 1 and 2 years before the present examination. *AJR Am J Roentgenol* 2003;180(2):343–346.
11. Thurffjell MG, Vitak B, Azavedo E, Svane G, Thurffjell E. Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. *Acta Radiol* 2000;41(1):52–56.
12. Roelofs AA, Karssemeijer N, Wedekind N, et al. Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology* 2007;242(1):70–77.
13. Varela C, Karssemeijer N, Hendriks JH, Holland R. Use of prior mammograms in

- the classification of benign and malignant masses. *Eur J Radiol* 2005;56(2):248–255.
14. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol* 2013;14(7):583–589.
15. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology* 2013;269(3):694–700.
16. Hakim CM, Anello MI, Cohen CS, et al. Impact of and interaction between the availability of prior examinations and DBT on the interpretation of negative and benign mammograms. *Acad Radiol* 2014;21(4):445–449.
17. Hillis SL, Schartz KM, Berbaum KS. OR-DBM MRMC 3.0 for SAS. <http://perception.radiology.uiowa.edu>. Published 2013. Accessed August 13, 2014.
18. Skaane P, Bandos AI, Eben EB, et al. Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. *Radiology* 2014;271(3):655–663.
19. Zuley ML, Guo B, Catullo VJ, et al. Comparison of two-dimensional synthesized mammograms versus original digital mammograms alone and in combination with tomosynthesis images. *Radiology* 2014;271(3):664–671.