Two-View Digital Breast Tomosynthesis Screening with Synthetically Reconstructed Projection Images: Comparison with Digital Breast Tomosynthesis with Full-Field Digital Mammographic Images[1]

Per Skaane, MD, PhD
Andriy I. Bandos, PhD
Ellen B. Eben, MD
Ingvild N. Jebsen, MD
Mona Krager, MD
Unni Haakenaasen, MD
Ulrika Ekseth, MD
Mina Izadi, MD
Solveig Hofvind, PhD
Randi Gullien, RT, MSc

Purpose: To compare the performance of two versions of reconstructed two-dimensional (2D) images in combination with digital breast tomosynthesis (DBT) versus the performance of standard full-field digital mammography (FFDM) plus DBT.

Materials and Methods: This trial had ethical committee approval, and all participants gave written informed consent. Examinations (n = 24901) in women between the ages of 50 and 69 years (mean age, 59.2 years) were interpreted prospectively as part of a screening trial that included independent interpretations of FFDM plus DBT and reconstructed 2D images plus DBT. Reconstructed 2D images do not require radiation exposure. Using analyses for binary data that accounted for correlated interpretations and were adjusted for reader-specific volume, two versions (initial and current) of reconstructed 2D images used during trial periods 1 (from November 22, 2010, to December 21, 2011; 12631 women) and 2 (from January 20, 2012, to December 19, 2012; 12270 women) were compared in terms of cancer detection and false-positive rates with the corresponding FFDM plus DBT interpretations.

Results: Cancer detection rates were 8.0, 7.4, 7.8, and 7.7 per 1000 screening examinations for FFDM plus DBT in period 1, initial reconstructed 2D images plus DBT in period 1, FFDM plus DBT in period 2, and current reconstructed 2D images plus DBT in period 2, respectively. False-positive scores were 5.3%, 4.6%, 4.6%, and 4.5%, respectively. Corresponding reader-adjusted paired comparisons of false-positive scores revealed significant differences for period 1 (P = .012) but not for period 2 (ratio = 0.99; 95% confidence interval: 0.88, 1.11; P = .85).

Conclusion: The combination of current reconstructed 2D images and DBT performed comparably to FFDM plus DBT and is adequate for routine clinical use when interpreting screening mammograms.

[1] From the Department of Radiology, Oslo University Hospital Ullevaal, University of Oslo, Kirkeveien 166, N-0407 Oslo, Norway (P.S., E.B.E., I.N.J., M.K., U.H., M.I., R.G.); Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pa (A.I.B.); Curato Roentgen Institute, Oslo, Norway (U.E.); and Institute of Population-based Cancer Research, the Cancer Registry, Oslo, Norway (S.H.). Received June 17, 2013; revision requested August 6; revision received September 20; accepted September 27; final version accepted November 13. Hologic sponsored the study by providing equipment and financial support for additional interpretations. **Address correspondence to** P.S. (e-mail: *PERSKA@ous-hf.no*).

Many breast imaging practices have implemented the combined procedure of full-field digital mammography (FFDM) and digital breast tomosynthesis (DBT). DBT can be used for both screening and diagnostic mammography (1–10). The benefit demonstrated in screening is the possibility of simultaneously increasing cancer detection rates—those for invasive cancers in particular—and reducing recall rates (6,9). The primary limitations of using FFDM plus DBT for screening are the increase in interpretation time (8) and the increase in the radiation dose to the breast being imaged, which approximately doubles (11,12). The primary reasons for performing FFDM plus DBT are the concern that comparisons with prior studies may be more difficult and that microcalcification clusters may not be as readily and easily detected and/or correctly interpreted on tomosynthesis images or sections alone (7). Therefore, it is important to investigate methods to address the limitations of FFDM plus DBT. Methods have been developed to reconstruct two-dimensional (2D) projection images from the information acquired during a DBT data acquisition procedure. Hence, the need for double exposure could potentially be eliminated if it could be demonstrated that synthetically reconstructed 2D images result in satisfactory diagnostic performance. Initial experimental reconstructions of synthesized 2D images suggest that radiologists' performance levels are lower when using synthesized 2D plus DBT images than when using the original (radiation dose–requiring) FFDM plus DBT images (13). However, recent improvements in the quality of 2D image reconstruction have resulted in approval for clinical use of software for generating synthesized 2D images (C-View; Hologic, Bedford, Mass) (14).

As an integral part of a large prospective screening trial, we used an early version of synthesized 2D images during period 1 of the trial and the current version of the synthesized 2D images during period 2 of the trial. Therefore, we compared the performance of two versions of reconstructed 2D images in combination with DBT with the performance of standard FFDM in combination with DBT.

Advances in Knowledge

- When using a new version of synthesized two-dimensional (2D) images plus digital breast tomosynthesis (DBT) data during the interpretation of screening mammograms, radiologists' overall performance levels were comparable to their performance levels when using standard full-field digital mammography (FFDM) plus DBT, with cancer detection rates of 7.8 and 7.7 per 1000 screening examinations for FFDM plus DBT and for current synthesized 2D images plus DBT, respectively (false-positive scores, 4.6% and 4.5%, respectively).
- The combination of synthesized 2D images and DBT reduced the average radiation dose to the breast by approximately 45% (from a total of 3.53 mGy for FFDM plus DBT to a total of 1.95 mGy for synthesized 2D images plus DBT), without compromising performance.

Materials and Methods

Hologic sponsored this study by providing equipment and financial support for additional interpretations. The authors had control of the data and the material submitted for publication.

Study Population

This clinical trial, which had ethical committee approval (Regional Committees for Medical and Health Research Ethics, South East) and for which written informed consent was obtained, has been previously described (6,15). All women (n = 24 901) included in this study (mean age, 59.2 years; range, 50–69 years) participated in the biennial Oslo breast cancer screening program between November 22, 2010, and December 19, 2012. The Oslo screening program is part of the Norwegian Breast Cancer Screening Program, which is administered by the Cancer Registry of Norway and has been described elsewhere (16,17). The study population of period 1 included a total of 12 631 women who participated in the Oslo breast cancer screening program between November 22, 2010, and December 21, 2011. Analyses of results during this time period have been previously reported (6,15). The study population of period 2 included 12 270 women who participated in the Oslo screening program between the time the new version of the synthesized 2D image software was installed on January 20, 2012, and December 19, 2012.

The selection of potential candidates, who were asked to participate at the time of their arrival for a scheduled examination, was based solely on the availability of technical staff members and DBT imaging systems during the date of the examination in question and

Implication for Patient Care

- Implementing synthesized 2D images plus DBT in breast cancer screening programs should result in similar performance levels to those of FFDM plus DBT while significantly reducing the radiation dose to screened women.

Published online before print
10.1148/radiol.13131391 **Content codes:** BR DM

Radiology 2014; 271:655–663

Abbreviations:
CI = confidence interval
DBT = digital breast tomosynthesis
FFDM = full-field digital mammography
PPV = positive predictive value
2D = two-dimensional

Author contributions:
Guarantor of integrity of entire study, P.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; manuscript final version approval, all authors; literature research, P.S., U.E., S.H.; clinical studies, P.S., E.B.E., I.N.J., M.K., U.H., U.E., M.I.; statistical analysis, A.I.B.; and manuscript editing, P.S., E.B.E., I.N.J., M.K., U.H., U.E., M.I., S.H., R.G.

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

See also the editorial by Kopans and the article by Zuley et al in this issue.

was not based on any personal information. Women who were unable to stand and women who had breast implants were not asked to consider participation. There were no other exclusion criteria. Recording the reasons for a woman's declining to participate was not permitted. Women who were not asked to participate and those who declined to participate underwent conventional FFDM imaging. There were 5329 and 3388 women who underwent only conventional FFDM in study periods 1 and 2, respectively. Most women participating in the study were undergoing DBT imaging for the first time. Because invitation letters were sent 6–8 weeks in advance, 992 women who were imaged with FFDM plus DBT in period 1 were also imaged in period 2. One woman with bilateral cancer who participated late in 2011 and who had had equal scores for both breasts in both reading modes underwent delayed work-up, and data for this woman are therefore not included in the analysis of period 1. Owing to the time gap between the two periods, not all women who participated in the Oslo screening program are included in this study. The results include all screening-detected breast cancers during periods 1 and 2. Interval cancers, metastases, and lymphomas were not included in the analysis.

The assessment of DBT data in this prospective clinical trial was incorporated into our routine biennial screening mammography practice. In period 1, we compared independent readings of FFDM plus DBT with independent readings of the initial version of the synthesized 2D images plus DBT, while in period 2, we compared independent readings of FFDM plus DBT with independent readings of the current version of synthesized 2D images plus DBT.

Imaging Technique

All screening examinations were performed in our outpatient clinic by using commercial systems (Selenia Dimensions; Hologic). Screening examinations included two-view (craniocaudal and mediolateral oblique) mammography of each breast. FFDM and DBT views of each breast were acquired during a single breast compression per view. This combined imaging procedure took approximately 10 seconds per view. The acquisition protocol resulted in fully registered FFDM and DBT images. The imaging system automatically computed (estimated) average glandular dose. After a quality assurance check by the technologist (to ensure that positioning was correct), all imaging studies were transferred to the Breast Imaging Center at the Oslo University Hospital Ullevaal for interpretation and treatment recommendation.

Generation of Synthesized Reconstructed 2D Images

The synthesized 2D images used in the first period of the trial (from November 22, 2010, to December 21, 2011; 12631 women) were reconstructed by using an early version of the image reconstruction software. An improved version that uses section-weighted summation and enhanced features was installed on January 20, 2012, and this version was used for examinations performed thereafter (14). Hence, period 2 of the trial was from January 20, 2012, to December 19, 2012, and included 12270 women. The synthesized 2D image is created by summing and filtering the stack of reconstructed DBT sections—similar to the process for generating a maximum intensity projection image. This image processing approach was developed by Hologic. A detailed description of the method is provided elsewhere (13,14).

Reading Modes

A description of our four-arm prospective study, with interim analyses of the results of using FFDM alone and FFDM plus DBT, as well as the results of double reading, has been presented elsewhere (6,15). The four interpretation modes included FFDM only, FFDM plus computer-aided detection, FFDM plus DBT, and synthesized 2D images plus DBT (with an early version of the 2D images in period 1 and a current version in period 2).

Participating Radiologists and Training

Prior to commencement of the trial, all radiographers and radiologists participating in the trial received specific training in the operation of the imaging system and in the interpretation of DBT studies (6,15). Eight radiologists with 2–31 years of experience in screening mammography (average, 16 years) participated in this study (including P.S., E.B.E., I.N.J., M.K., U.H., U.E., and M.I.). However, because one of the radiologists (who was not an author) interpreted only 80 examinations during the second period of the trial, we also assessed whether the inclusion or exclusion of this radiologist could have affected the study results and/or conclusions.

Each participating radiologist received training of approximately 4 hours in reviewing an enriched set of a minimum of 100 examinations with feedback, using the same workstations and respective hanging protocols used in the trial (15). There was no supplemental training after the installation of the current version of the synthesized 2D image software.

Interpretation Workflow

Each examination included in the trial was independently interpreted, in batch mode, by four radiologists, each using a different reading mode and dedicated workstation for each arm of the trial; however, only two of these modes are compared in the current article.

Reading assignments were made by a nonradiologist staff member who independently assigned each radiologist sets of cases to be interpreted in specific reading modes. The scheduler attempted to balance the number of cases interpreted by each radiologist in each mode as much as reasonably achievable in a busy clinical practice in which some of the radiologists were not present full time in the Breast Imaging Center. During the readings, hanging protocols for the specifically assigned modes were preset (15).

Image Interpretation and Consensus/ Arbitration Decision

After reviewing an examination, each radiologist independently rated his or her findings on a per-breast basis by using a standardized five-point ordinal rating scale of increasing probability of

suspicion for malignancy (6). A score of 1 by all study readers was regarded as indicating a negative examination. The decision to undertake additional actions other than dismissal of a case as negative was based solely on one or more of the radiologists recording a score of 2 or higher. In these instances, mammographic findings (features) had to be listed. Other indications for referral to a consensus meeting included the presence of clinical symptoms (eg, palpable lump) or technical insufficiency of the examination. Because women with self-reported clinical symptoms were included in our screening population, our population included 156 women with negative mammograms who reported symptoms (eg, lump, discharge, skin retraction) and thus would not be considered asymptomatic at screening. All of these cases were discussed in consensus, but only 30 of these women were actually recalled for a diagnostic work-up. Because the self-reporting of symptoms applied identically to the two study arms (ie, any findings or outcomes, whether positive or negative, were rated as "negative" in both study arms of interest), these cases (85 in period 1 and 71 in period 2) were included in the analyses. All 156 cases with clinical symptoms were included in both modes being analyzed. All scores (ratings) were recorded directly into the Norwegian Breast Cancer Screening Program database, and results were locked at the end of each reading session.

Review and consolidation of positive ratings during arbitration is the standard of practice in our screening program. All patients whose images received one or more scores of 2 or greater were discussed, with availability of all imaging and nonimaging information, by at least two radiologist participants. A consensus-based clinical management decision (to dismiss or to recall a patient for diagnostic work-up) was reached for all examinations that received at least one rating of 2 or 3. An examination that received a score of 4 or 5 could not be dismissed, and the patient was recalled. Diagnostic work-up for recalled women, which could potentially include additional views, ultrasonography, magnetic resonance imaging, and needle biopsy, if indicated, was performed by the same group of radiologists during a single visit to the Breast Imaging Center. A recommendation for short-term (6-month) follow-up was not used. Follow-up data for negative cases and for cases dismissed at the consensus meeting were not available at the time of preparation of this article.

Statistical Analysis

Statistical inferences about relative changes in rates, adjusted for between-radiologist variability in performance levels and for correlations between assessments of the same cases, were performed by using log-linear mixed models for binary data (proc glimmix in SAS, version 9.3; SAS Institute, Cary, NC). Ratios of false-positive scores and cancer detection rates between the compared modalities were estimated from the models, and 95% confidence intervals were used for statistical inferences. $P < .05$ was considered to indicate a statistically significant difference.

We compared false-positive rates prior to arbitration, attributable recall rates as a result of arbitration decisions, attributable cancer detection rates, and positive predictive values (PPVs) (number of verified attributable cancers per number of recalls) for paired independent readings by using the following outcome measure assignments: A cancer was considered to be "detected" with the specific reading mode (initial synthesized 2D images plus DBT, current synthesized 2D images plus DBT, or FFDM plus DBT in trial period 1 or 2) if the correct breast received a positive score (rating > 1) during interpretation of the screening examination with the given reading modality. A case with a positive score (>1) for at least one breast and without a verified cancer (ie, a case that was either dismissed at the arbitration meeting or found to be benign during the diagnostic work-up, which may have included a biopsy) was attributed as a false-positive finding for that reading mode. The false-positive rates and cancer detection rates were defined as the corresponding numbers of false-positive examinations and/or detected cancers per 1000 screening examinations. The PPV was computed as the percentage of cases with screening-detected cancer among all positively scored cases (for at least one breast) that were recalled at the arbitration meeting. After 10 exclusions from period 1 (6), only 12 621 cases from period 1 were included in the analysis.

Results

General Findings

The average glandular dose for a single mammographic view was 1.58 mGy ± 0.61 (standard deviation) (range, 0.74–4.51 mGy) for FFDM and 1.95 mGy ± 0.58 (range, 1.05–3.78 mGy) for DBT. These dose levels constitute an average dose reduction of 45% (1.58/3.53) for synthesized 2D images plus DBT as compared with FFDM plus DBT.

False-Positive Scores

The rates of false-positive examinations with FFDM plus DBT in period 1, with initial synthesized 2D images plus DBT in period 1, with FFDM plus DBT in period 2, and with current synthesized 2D images plus DBT in period 2 were 5.3% (670 of 12 621), 4.6% (582 of 12 621), 4.6% (560 of 12 270), and 4.5% (555 of 12 270), respectively (Tables 1, 2). False-positive rates differed significantly between FFDM plus DBT and initial synthesized 2D images plus DBT in period 1 (ratio = 0.87; $P = .012$), but there was no significant difference (ratio = 0.99; 95% confidence interval [CI]: 0.88, 1.11; $P = .85$) between FFDM plus DBT and current synthesized 2D images plus DBT during period 2, primarily owing to the improvement in recall rates for FFDM plus DBT over time. The results remained virtually the same after exclusion of either data corresponding to the low-volume radiologist (reader 1 in Table 2) or women imaged twice in period 2. False-positive rates after the

Table 1

True-Positive and False-Positive Interpretations, Recalls, and Cancers Detected during Each Period in Each Study Arm

	Study Period 1 (n = 12 621)		Study Period 2 (n = 12 270)	
Parameter	FFDM plus DBT	Initial Synthesized 2D Images plus DBT	FFDM plus DBT	Current Synthesized 2D Images plus DBT
No. of false-positive scores	670	582	560	555
No. of true-positive scores*	101	94	96	94
No. of cases recalled at arbitration	351	310	296	269
No. of women with detected cancer	100	94	95	94
PPV (%)	28.5	30.3	32.1	34.9

* In each time period, the FFDM-plus-DBT reader detected a bilateral cancer; therefore, the true-positive scores were 101 and 96, while the numbers of women with cancer were 100 and 95.

exclusion of reader 1 were 5.2% (624 of 11 996) for FFDM plus DBT in period 1, 4.5% (541 of 12 099) for initial synthesized 2D images plus DBT in period 1, 4.5% (555 of 12 232) for FFDM plus DBT in period 2, and 4.5% (550 of 12 228) for current synthesized 2D images plus DBT in period 2. Among the 992 women who were imaged with FFDM plus DBT twice during the study, there were 32 false-positives with FFDM plus DBT and 24 false-positives with FFDM plus synthesized 2D images during trial period 2. Also, with the exception of a few additional false-positive ratings of benign microcalcifications when synthesized 2D images were used during period 2, no substantial differences were seen in breast density or radiologic sign of false-positive ratings of suspected abnormalities when interpreted by using the two methods.

Cancer Detection

The number of cancers detected with FFDM plus DBT and initial synthesized 2D images plus DBT in the first study period and with FFDM plus DBT and current synthesized 2D images plus DBT in the second study period were 101 (8.0 per 1000 screening examinations), 94 (7.4 per 1000 screening examinations), 96 (7.8 per 1000 screening examinations), and 94 (7.7 per 1000 screening examinations), respectively. Paired comparisons for period 1 showed a statistically nonsignificant 7% decrease in cancer detection rate for the synthesized mode (ratio = 0.93; 95% CI: 0.70, 1.23; P = .62). In period 2, the cancer detection rate was substantially more comparable (2% difference) between current synthesized 2D images plus DBT and FFDM plus DBT (ratio = 0.98; 95% CI: 0.74, 1.30; P = .89). In period 1, there were 25 cancers that were detected only by the FFDM-plus-DBT readers and 18 cancers that were detected only by the initial-synthesized-2D-images-plus-DBT readers. In period 2, there were 14 cancers that were detected only by the FFDM-plus-DBT readers and 12 cancers that were detected only by the current-synthesized-2D-images-plus-DBT readers. Therefore, the total number of discordant referral recommendations for verified cancers decreased substantially, from 43 (25 + 18) to 26 (14 + 12) (P = .05). During the first and second periods, 111 and 77 cases, respectively, were rated as negative by using FFDM plus DBT while being rated as positive by using synthesized 2D images plus DBT. Thirty cancers (18 in period 1 plus 12 in period 2) were confirmed as a result of the diagnostic work-ups that followed. During the same two periods, 154 and 106 cases, respectively, were rated as positive by using FFDM plus DBT and as negative by using synthesized 2D images plus DBT, resulting in 39 confirmed cancers (25 in period 1 plus 14 in period 2). No substantial differences were seen in terms of breast density, grade, size, or radiologic signs for cancers given discordant scores with the two interpretation methods. In the 992 women who were imaged with FFDM plus DBT twice during the study, nine cancers were detected in eight women (one with bilateral cancer) with all four modes during period 2 (expected rate, 8.96 [110 of 12 270]). Cancer detection rates were the same with each of the two modes in question. Hence, exclusion of these women resulted in very similar results and did not change any of the study conclusions.

For women with detected cancer, the PPVs of recalls that were attributed to FFDM plus DBT and initial synthesized 2D images plus DBT in period 1 were 28.5% (100 of 351) and 30.3% (94 of 310) (P = .61), and the PPVs of recalls that were attributed to FFDM plus DBT and current synthesized 2D images plus DBT in period 2 were 32.1% (95 of 296) and 34.9% (94 of 269) (P = .47). The exclusion of data from reader 1 or from the 992 repeat examinations had only a minor effect on our numeric results and resulted in no change in our conclusions. Figures 1 and 2 illustrate two screening-detected cancers.

Discussion

If DBT is to be widely used in screening mammography, dose-related issues will have to be carefully addressed (18). To date, to our knowledge, the use of synthesized 2D images has been investigated only in retrospective observer performance studies that used early versions of synthesized 2D images. These early versions of synthesized 2D images showed relatively high performance levels, although not as high as the performance achieved by using the original radiation dose–requiring FFDM images (13). These results are largely in concordance with our results for period 1 of our trial. In period 2, however, with use of an

Table 2

Overall Performance Levels of Each of the Participating Radiologists in Each of the Study Arms during the Two Study Periods

A: Study Period 1

	FFDM plus DBT							Initial Synthesized 2D Images plus DBT						
Reader No.	No. of Patients	No. of False-Positive Scores	False-Positive Rate*	No. of Known Cancers†	No. of True-Positive Scores	Percentage Detected Cancers‡	Cancer Detection Rate*	No. of Patients	No. of False-Positive Scores	False-Positive Rate*	No. of Known Cancers†	No. of True-Positive Scores	Percentage Detected Cancers‡	Cancer Detection Rate*
1	625	46	73.6	8	7	87.5	11.2	522	41	78.5	9	8	88.9	15.3
2	1743	119	68.3	15	13	86.7	7.5	2051	143	69.7	23	14	60.9	6.8
3	1483	82	55.3	9	7	77.8	4.7	1771	60	33.9	19	16	84.2	9.0
4	1758	78	44.4	10	9	90.0	5.1	1828	76	41.6	20	16	80.0	8.8
5	2790	147	52.7	43	37	86.0	13.3	2757	102	37.0	11	9	81.8	3.3
6	1402	71	50.6	14	8	57.1	5.7	1277	74	57.9	13	9	69.2	7.0
7	1355	71	52.4	14	12	85.7	8.9	1250	36	28.8	12	8	66.7	6.4
8	1465	56	38.2	8	8	100.0	5.5	1165	50	42.9	14	14	100.0	12.0
All	12621	670	53.1	121	101	83.5	8.0	12621	582	46.1	121	94	77.7	7.4

B: Study Period 2

	FFDM plus DBT							Current Synthesized 2D Images plus DBT						
Reader No.	No. of Patients	No. of False-Positive Scores	False-Positive Rate*	No. of Known Cancers†	No. of True-Positive Scores	Percentage Detected Cancers‡	Cancer Detection Rate*	No. of Patients	No. of False-Positive Scores	False-Positive Rate*	No. of Known Cancers†	No. of True-Positive Scores	Percentage Detected Cancers‡	Cancer Detection Rate*
1	38	5	131.6	0	0	...	...	42	5	119.0	0	0	...	...
2	2030	87	42.9	15	12	80.0	5.9	1812	77	42.5	14	9	64.3	5.0
3	1830	79	43.2	17	15	88.2	8.2	1724	67	38.9	15	12	80.0	7.0
4	2661	103	38.7	26	24	92.3	9.0	2335	99	42.4	25	24	96.0	10.3
5	2595	118	45.5	27	23	85.2	8.9	2729	135	49.5	29	25	86.2	9.2
6	572	41	71.7	3	2	66.7	3.5	814	55	67.6	4	2	50.0	2.5
7	1392	68	48.9	10	10	100.0	7.2	1354	62	45.8	15	14	93.3	10.3
8	1152	59	51.2	12	10	83.3	8.7	1460	55	37.7	8	8	100.0	5.5
All	12270	560	45.6	110	96	87.3	7.8	12270	555	45.2	110	94	85.5	7.7

* Rates are given as numbers of cases per 1000 screening examinations.

† Number of screening-detected cancers in any of the four study arms (detected by any of the four radiologists who interpreted the examinations).

‡ Percentage detected cancers refers to the fraction of all screening-detected cancers found by any of the four radiologists independently reading each examination that were detected by the radiologist using the specific mode of interest.

Figure 1

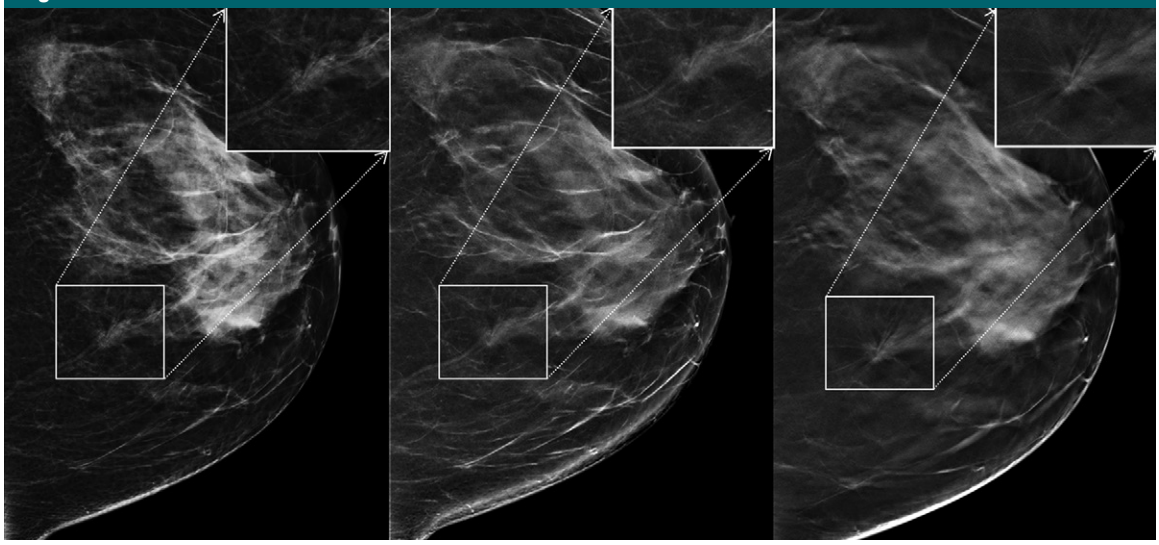

Figure 1: Craniocaudal images in 57-year-old woman with 8-mm invasive ductal carcinoma in the left breast. Left: FFDM image. Middle: Initial synthesized 2D image. Right: DBT image. The region of interest is magnified for each image. Reader scores were 1 for FFDM, 3 for FFDM plus DBT, and 4 for synthesized 2D images plus DBT. This case was from study period 1.

Figure 2

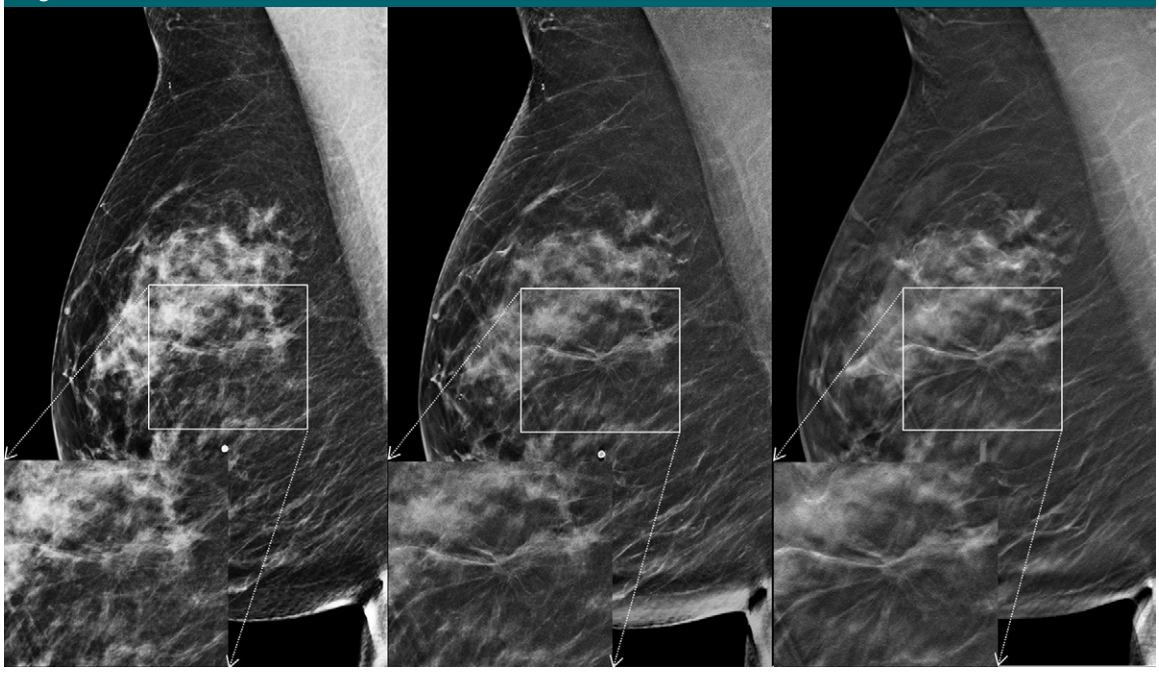

Figure 2: Mediolateral oblique images in 61-year-old woman with 12-mm invasive lobular carcinoma in the right breast. Left: FFDM image. Middle: Current synthesized 2D image. Right: DBT image. The region of interest is magnified for each image. Reader scores for the study were 1 for FFDM, 3 for FFDM plus DBT, and 4 for synthesized 2D images plus DBT. This case was from study period 2.

improved version of the reconstructed synthesized image processing software (C-View) that was recently approved for clinical use in the United States in combination with DBT, we showed that the performance of the synthesized 2D images was reasonably comparable to that of FFDM in combination with DBT. The difference in performance levels decreased, as well as the total number of discordant interpretations in the cancers detected by only one

of the readers with one mode or the other. On the basis of these results and the consistency among studies in demonstrating gradual and consistent improvement in performance with improvement of the synthesized images (13,14), we believe that the use of a high-quality synthesized view plus DBT should be viewed as acceptable for replacing radiation dose–requiring FFDM projection images when interpreting screening mammography examinations. The use of computer-aided detection for both synthesized 2D views and three-dimensional DBT may further improve performance. We also note that, surprisingly, even in period 1 of the study and despite the use of an early version of the synthesized 2D images, a single reader of synthesized 2D images plus DBT detected more cancers than double reading of FFDM, as previously reported (94 cancers detected by a single reader using synthesized 2D images plus DBT compared with 90 cancers detected by using double reading of FFDM) (15). Hence, the performance improvement noted in these studies suggest that one can use this improvement to find additional cancers, to eliminate the need for double readings, and/or for some combination of both (eg, double read only specific types of examinations, such as those in dense breasts).

Regarding the goal of reducing radiation dose, there have been recent reports on the use of one-view DBT with or without FFDM (2,19,20). Because the primary objective of screening is improved cancer detection and we and others have previously observed a substantially larger improvement in cancer detection with the use of two-view DBT (6,9,15), we believe that the current lower-dose approach of synthesized 2D images plus DBT is more appropriate for screening purposes.

Our study had several limitations. First, the consensus/arbitration step could have preferentially decreased the actual recall rates of women referred to arbitration (rating > 1) on the basis of only one of the two reading modes being compared who were later dismissed during arbitration. Second, the reference standard for positive cases was a cancer detected in any of the study arms. Absolute sensitivity and specificity cannot be calculated until 2-year follow-up is complete. Third, while overall our trial is reasonably well balanced in terms of reader-by-mode interpretations, we could not completely balance the interpretations, namely, by having each reader interpret the same number of examinations in each of the reading modes during each of the periods being compared. This is a very difficult task in the clinical environment, and we had to account for this imbalance by adjusting for reader-specific differences in modality-specific performance levels. Fourth, we note that more than 90% of the women participating in this trial underwent the DBT procedure for the first time. Hence, data on performance during repeat DBT-based studies are not presented in this analysis. Fifth, it is possible that the increasing familiarity with the synthesized 2D images (the learning curve effect) affected our results, especially in period 2 of the study. Last, this was a single-institution, single-vendor study with vendor-specific software, and our results should be validated by other practices with different workflows, decision processes, or case presentations.

In conclusion, synthesized 2D images combined with DBT performed comparably to FFDM plus DBT when interpreting screening mammograms in terms of cancer detection rates and false-positive scores. The two-view combination of a synthesized 2D view plus DBT should be considered acceptable for routine use in mammographic screening.

Disclosures of Conflicts of Interest: P.S. No relevant conflicts of interest to disclose. **A.I.B.** Financial activities related to the present article: the Department of Biostatistics at the University of Pittsburgh was contracted to independently perform all statistical analyses. Financial activities not related to the present article: none to disclose. Other relationships: none to disclose. **E.B.E.** No relevant conflicts of interest to disclose. **I.N.J.** No relevant conflicts of interest to disclose. **M.K.** No relevant conflicts of interest to disclose. **U.H.** No relevant conflicts of interest to disclose. **U.E.** No relevant conflicts of interest to disclose. **M.I.** No relevant conflicts of interest to disclose. **S.H.** No relevant conflicts of interest to disclose. **R.G.** Financial activities related to the present article: none to disclose. Financial activities not related to the present article: has received travel costs and a fee for one lecture (a meeting in 2012) from Hologic. Other relationships: none to disclose.

References

1. Bernardi D, Ciatto S, Pellegrini M, et al. Prospective study of breast tomosynthesis as a triage to assessment in screening. Breast Cancer Res Treat 2012;133(1):267–271.
2. Gennaro G, Hendrick RE, Ruppel P, et al. Performance comparison of single-view digital breast tomosynthesis plus single-view digital mammography with two-view digital mammography. Eur Radiol 2013;23(3):664–672.
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. 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.
5. 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.
6. 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.
7. Spangler ML, Zuley ML, Sumkin JH, et al. Detection and classification of calcifications on digital breast tomosynthesis and 2D digital mammography: a comparison. AJR Am J Roentgenol 2011;196(2):320–324.
8. Zuley ML, Bandos AI, Abrams GS, et al. Time to diagnosis and performance levels during repeat interpretations of digital breast tomosynthesis: preliminary observations. Acad Radiol 2010;17(4):450–455.
9. 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.
10. 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.

11. Olgar T, Kahn T, Gosch D. Average glandular dose in digital mammography and breast tomosynthesis. Rofo 2012;184(10):911–918.

12. Sechopoulos I. A review of breast tomosynthesis. I. The image acquisition process. Med Phys 2013;40(1):014301.

13. Gur D, Zuley ML, Anello MI, et al. Dose reduction in digital breast tomosynthesis (DBT) screening using synthetically reconstructed projection images: an observer performance study. Acad Radiol 2012;19(2):166–171.

14. U.S. Food and Drug Administration. Meeting of the radiological devices advisory panel. P080003/S001 Hologic Selenia Dimensions 3D System, FDA executive summary. http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/RadiologicalDevicesPanel/UCM324861.pdf. Published October 24, 2012. Accessed February 12, 2013.

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

16. Skaane P, Young K, Skjennald A. Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading—Oslo I study. Radiology 2003;229(3):877–884.

17. Skaane P, Skjennald A. Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program—the Oslo II Study. Radiology 2004;232(1):197–204.

18. Baker JA, Lo JY. Breast tomosynthesis: state-of-the-art and review of the literature. Acad Radiol 2011;18(10):1298–1310.

19. Svahn T, Andersson I, Chakraborty D, et al. The diagnostic accuracy of dual-view digital mammography, single-view breast tomosynthesis and a dual-view combination of breast tomosynthesis and digital mammography in a free-response observer performance study. Radiat Prot Dosimetry 2010;139(1-3):113–117.

20. Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution X-ray imaging observer study. Radiology 2012;262(3):788–796.