Radiology

ORIGINAL RESEARCH • BREAST IMAGING

Comparison of Mammography and Mammography with Supplemental Whole-Breast US Tomography for Cancer Detection in Patients with Dense Breasts

Mary W. Yamashita, MD • Linda H. Larsen, MD • Jeremiah Perez, PhD • Alexandra V. Edwards, MA • John Papaioannou, MS • Yulei Jiang, PhD

From the Department of Radiology, University of Southern California, Keck School of Medicine, Keck Hospital, 1500 San Pablo St, 2nd Floor, Suite 2250, Los Angeles, CA 90033 (M.W.Y., L.H.L.); Department of Biostatistics, Avania U.S., Marlborough, Mass (J. Perez); and Department of Radiology, The University of Chicago, Chicago, Ill (A.V.E., J. Papaioannou, Y.J.). Received July 6, 2023; revision requested August 1; final revision received April 1, 2024; accepted April 8. **Address correspondence to** M.W.Y. (email: *mary.yamashita@med.usc.edu*).

Supported by Delphinus Medical Technologies.

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

See also the editorial by Mann in this issue.

Radiology 2024; 311(3):e231680 • https://doi.org/10.1148/radiol.231680 • Content codes: BR US

Background: Women with dense breasts benefit from supplemental cancer screening with US, but US has low specificity.

Purpose: To evaluate the performance of breast US tomography (UST) combined with full-field digital mammography (FFDM) compared with FFDM alone for breast cancer screening in women with dense breasts.

Materials and Methods: This retrospective multireader multicase study included women with dense breasts who underwent FFDM and UST at 10 centers between August 2017 and October 2019 as part of a prospective case collection registry. All patients in the registry with cancer were included; patients with benign biopsy or negative follow-up imaging findings were randomly selected for inclusion. Thirty-two Mammography Quality Standards Act–qualified radiologists independently evaluated FFDM followed immediately by FFDM plus UST for suspicious findings and assigned a Breast Imaging Reporting and Data System (BI-RADS) category. The superiority of FFDM plus UST versus FFDM alone for cancer detection (assessed with area under the receiver operating characteristic curve [AUC]), BI-RADS 4 sensitivity, and BI-RADS 3 sensitivity and specificity were evaluated using the two-sided significance level of $\alpha = .05$. Noninferiority of BI-RADS 4 specificity was evaluated at the one-sided significance level of $\alpha = .025$ with a –10% margin.

Results: Among 140 women (mean age, 56 years ±10 [SD]; 36 with cancer, 104 without), FFDM plus UST achieved superior performance compared with FFDM alone (AUC, 0.60 [95% CI: 0.51, 0.69] vs 0.54 [95% CI: 0.45, 0.64]; $P = .03$). For FFDM plus UST versus FFDM alone, BI-RADS 4 mean sensitivity was superior (37% [428 of 1152] vs 30% [343 of 1152]; $P = .03$) and BI-RADS 4 mean specificity was noninferior (82% [2741 of 3328] vs 88% [2916 of 3328]; $P = .004$). For FFDM plus UST versus FFDM, no difference in BI-RADS 3 mean sensitivity was observed (40% [461 of 1152] vs 33% [385 of 1152]; $P = .08$), but BI-RADS 3 mean specificity was superior (75% [2491 of 3328] vs 69% [2299 of 3328]; $P = .04$).

Conclusion: In women with dense breasts, FFDM plus UST improved cancer detection by radiologists versus FFDM alone.

Clinical trial registration nos. NCT03257839 and NCT04260620

Supplemental material is available for this article.

Breast cancer screening with mammography reduces mortality by 41% in women with nondense breasts, but reduces mortality by only 13% in women with dense breasts (1). In addition, women with scattered fibroglandular, heterogeneously dense, and extremely dense breasts have a 2.1-, 2.9-, and 4.6-fold increase in the risk for developing breast cancer and a 6-, 16-, and 31-fold increase in the likelihood for interval cancer diagnosis, respectively, compared with women with fatty breasts (2–4). Underscoring the need for supplemental screening, the U.S. government has mandated federal breast density notifications effective September 2024 (5,6).

Several studies have shown that mammography followed by supplemental US screening of dense breasts increases cancer detection rate—detecting two to five additional cancers per 1000 women—and increases detection of lower-stage node-negative invasive cancers and reduces interval cancers (7–9). In the landmark American College of Radiology Imaging Network, or ACRIN, 6666 clinical trial in women with dense breasts and an elevated risk of breast cancer, mammography plus US depicted four additional cancers per 1000 women screened compared with mammography alone (10). The Japan Strategic Anticancer Randomized Trial, or J-START, demonstrated similar results with mammography plus US screening, showing improved sensitivity and cancer detection rate compared with mammography alone (9). Furthermore, the Adjunct Screening with Tomosynthesis or Ultrasound in Women with Mammography-Negative Dense Breasts, or ASTOUND, trial in Italy found four additional cancers per 1000 by adding tomosynthesis and seven additional cancers per 1000 by adding physician-performed handheld

Abbreviations

AUC = area under the receiver operating characteristic curve, BI-RADS = Breast Imaging Reporting and Data System, FFDM = full-field digital mammography, MRMC = multireader multicase, UST = US tomography

Summary

In women with dense breasts, automated breast US tomography added to screening mammography led to increased breast cancer detection and increased specificity for Breast Imaging Reporting and Data System category 3 lesions compared with mammography alone.

Key Results

- In this retrospective study of 140 patients (36 with cancer) and 32 readers, mammography plus breast US tomography (UST) improved breast cancer detection versus mammography alone (area under the receiver operating characteristic curve, 0.60 vs 0.54; $P = .03$).
- For mammography plus UST versus mammography alone, Breast Imaging Reporting and Data System (BI-RADS) 4 mean sensitivity was superior (37% [428 of 1152] vs 30% [343 of 1152], $P = .03$), BI-RADS 4 mean specificity was noninferior (82% [2741 of 3328] vs 88% [2916 of 3328], $P = .004$), and BI-RADS 3 mean specificity was superior (75% [2491 of 3328] vs 69% [2299 of 3328], $P = .04$).

US to mammography (11). However, both handheld US and automated US reduce specificity when added to mammography. In a review of 35 studies, Berg and Vourtsis (8) found that mammography with automated US or handheld US increased the recall rate by 10.6% and 7.5%, respectively. While automated US reduces operator dependence, the long interpretation time burdens clinical workflow (12–14).

Unlike US with B mode only, breast US tomography (UST) system provides four coronal volumetric image sequences of B mode and quantitative tissue characterization information, including tissue stiffness, speed of sound, and attenuation (15). By merging reflection images with these parameters, UST can potentially differentiate cancer from normal tissue or benign disease in women with dense breasts. The aim of this study was to evaluate the performance of UST combined with full-field digital mammography (FFDM) compared with FFDM alone for breast cancer screening in women with dense breasts.

Materials and Methods

In this retrospective study, images obtained in the Delphinus SoftVue Prospective Case Collection-Arm 1 study were evaluated (ClinicalTrials.gov: NCT03257839). The current study was approved by the institutional review board, compliant with the Health Insurance Portability and Accountability Act, and registered as the Delphinus SoftVue™ ROC Reader Study (ClinicalTrials.gov: NCT04260620). The original study obtained written informed consent from patients and the current study obtained written informed consent from radiologist readers. Delphinus Medical Technologies funded the study. The University of Chicago and Avania investigators had sole control of study data. All authors had control of reporting of data.

Study Sample

The Delphinus SoftVue Prospective Case Collection-Arm 1 study enrolled asymptomatic women scheduled for a screening mammogram across 10 sites in the United States between August 2017 and October 2019. Female patients aged 18 years or older of any race or ethnicity and with heterogeneous or extremely dense breasts were eligible to enroll. Women with prior breast intervention, including breast-conserving surgery, and implants were also eligible. Patients were not eligible if they *(a)* weighed more than 350 lbs (159 kg), *(b)* were pregnant or lactating, *(c)* were diagnosed with breast cancer in the prior year, *(d)* had a benign biopsy less than 7 days prior, *(e)* had an open wound, or *(f)* were unable to lie prone for 15 minutes. Patients who were lost to follow-up, withdrew, had images collected on older versions of the SoftVue system or underwent unilateral imaging, had Breast Imaging Reporting and Data System (BI-RADS) density A or B, or were

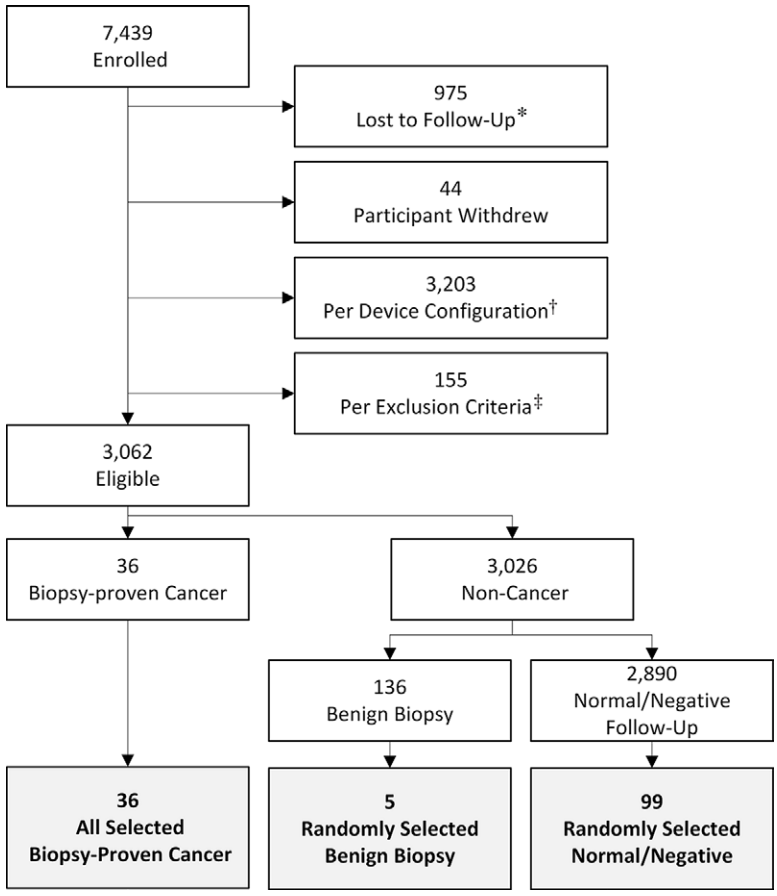


Figure 1: Flowchart shows selection of patients from the prospective case collection registry. *Patients did not return for follow-up mammogram or biopsy within 456 days. †Images acquired on early-device iterations that did not match the commercially released configuration were excluded. ‡Exclusion criteria included unilateral imaging, Breast Imaging Reporting and Data System density A or B, and simultaneous enrollment in another study arm.

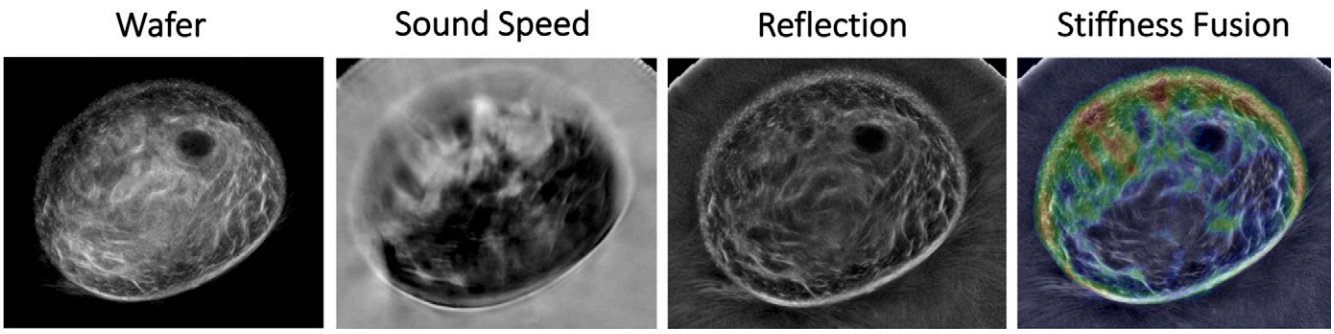


Figure 2: Images from the automated breast US tomography system (SoftVue; Delphinus Medical Technologies). Four coronal volumetric image-stack sequences are produced: wafer (waveform-enhanced reflection, which enhances fat so that dark masses are better visualized); sound speed (a direct output of image acquisition showing the change in the speed of sound moving through breast tissue); reflection (equivalent to B mode); and stiffness fusion (transmission properties of sound speed and attenuation overlaid on coregistered reflection images to highlight relative differences in tissue stiffness). A single slice from each sequence is shown.

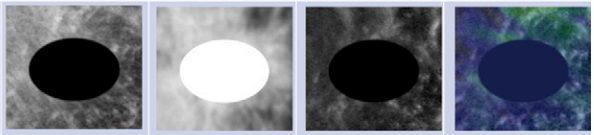
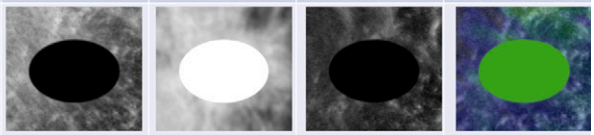
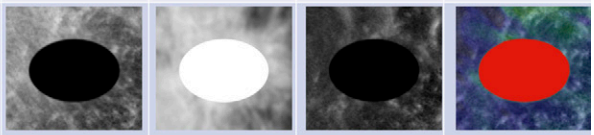
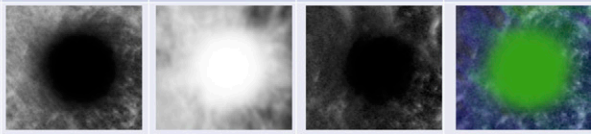
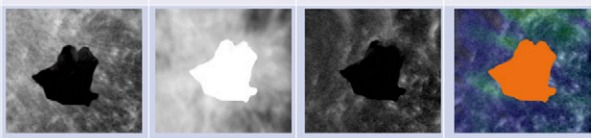
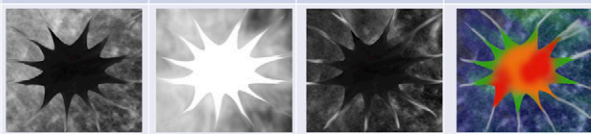
BI-RADS® Assessment Category		SoftVue Findings	Schematic of image sequences			
			Wafer	Sound Speed	Reflection	Stiffness Fusion
1	Negative	Negative				
2	Benign	Circumscribed/ Indistinct + Blue				
3	Probably Benign	Circumscribed + Green				
4a	Low suspicion for malignancy	Circumscribed + Orange/Red				
4b	Moderate suspicion for malignancy	Indistinct + Green				
4c	High suspicion for malignancy	Indistinct/ Irregular + Orange/ Red				
5	Highly suggestive of malignancy	Spiculations/AD + Green/Orange/Red				

Figure 3: Customized Breast Imaging Reporting and Data System (BI-RADS) assessment categories to enable consistent ratings when breast US tomographic (UST) image was interpreted together with full-field digital mammogram in the reader study. The schematics of UST image sequences are added as examples of one of several ways that each BI-RADS category can present on UST image. AD = architectural distortion.

simultaneously enrolled in another arm of the original study were excluded in the current study (Fig 1).

Cancers were diagnosed with image-guided or excisional biopsy, and noncancers were diagnosed with either biopsy or negative findings at imaging follow-up, all within 456 days of the study entry screening (Appendix S1). All patients diagnosed with breast cancer in the original study were included in the current study. Imaging findings proven to be benign at biopsy (five of 136, 3.7%) or negative at follow-up imaging (99 of 2890, 3.4%) were randomly selected for inclusion in

Table 1: Patient Characteristics

Characteristic	Cancer Cases (*n* = 36)	Noncancer Cases (*n* = 104)	Total (*n* = 140)
Age*	58.1 ± 10.8	54.7 ± 9.8	55.6 ± 10.2
History of breast cancer			
Yes	5.6 (2/36)	0.0 (0/104)	1.4 (2/140)
No	94.4 (34/36)	100.0 (104/104)	98.6 (138/140)
History of prior breast procedures			
Biopsy, not otherwise specified†	33.3 (12/36)	19.2 (20/104)	22.9 (32/140)
Needle biopsy	25.0 (9/36)	13.5 (14/104)	16.4 (23/140)
Cyst aspiration	11.1 (4/36)	9.6 (10/104)	10.0 (14/140)
Excisional biopsy	11.1 (4/36)	7.7 (8/104)	8.6 (12/140)
Current breast implants	2.8 (1/36)	9.6 (10/104)	7.9 (11/140)
Lumpectomy‡	13.9 (5/36)	0.0 (0/104)	3.6 (5/140)
Prior breast implant removal	2.8 (1/36)	1.0 (1/104)	1.4 (2/140)
None	55.6 (20/36)	71.2 (74/104)	67.1 (94/140)
Family history of breast cancer§			
Yes	55.6 (20/36)	51.0 (53/104)	52.1 (73/140)
No	44.4 (16/36)	48.1 (50/104)	47.1 (66/140)
Unknown	0.0 (0/36)	1.0 (1/104)	0.7 (1/140)
Body mass index*	26.7 ± 4.6	24.7 ± 4.8	25.2 ± 4.8
Currently menstruating			
No	47.2 (17/36)	45.2 (47/104)	45.7 (64/140)
Yes	30.6 (11/36)	41.3 (43/104)	38.6 (54/140)
Unknown	22.2 (8/36)	13.5 (14/104)	15.7 (22/140)
BI-RADS breast density			
C	94.4 (34/36)	79.8 (83/104)	83.6 (117/140)
D	5.6 (2/36)	20.2 (21/104)	16.4 (23/140)

Note.—Unless otherwise indicated, data are percentages and data in parentheses are the numbers used to calculate the percentages. BI-RADS = Breast Imaging Reporting and Data System. Body mass index was calculated as weight in kilograms divided by height in meters squared.

* Data are means ± SDs.

† Self-reported by patient and biopsy method was not specified.

‡ Out of five patients, two had lumpectomy for breast cancer and three for benign lesions.

§ Family history includes all blood relatives.

this study. The biopsy yield of cancer in this cohort was 21% (36/[36 + 136]).

The primary end point was the difference in performance between FFDM plus UST and FFDM alone. The secondary end points were patient-level sensitivity and specificity based on a BI-RADS category 4 threshold (main analysis) and a BI-RADS 3 threshold (supportive analysis).

Definitions

The sensitivity of BI-RADS 4 was defined as the fraction of cancer cases that a reader reported as BI-RADS 4 or higher, and the specificity of BI-RADS 4 was defined as the fraction of noncancer cases that a reader reported as BI-RADS 3 or lower. The BI-RADS 3 sensitivity and specificity were defined analogously.

Breast UST Procedure

UST (SoftVue; Delphinus Medical Technologies) was performed with a ring transducer while the participant lay prone, with the breast stabilized in a water-filled imaging chamber. The water provided coupling of sound energy between the transducer and the breast. The transducer scanned the entire breast from the nipple to the chest wall in 2–4 minutes. Reflection, sound speed, and attenuation data were processed to produce four coronal volumetric image sequences: wafer, sound speed, reflection, and stiffness fusion (15) (Fig 2, Appendix S2).

Analysis of Breast UST

Thirty-two Mammography Quality Standards Act–qualified general or breast radiologists from academic, private, and community-based practices in diverse areas participated in the study. They had a wide range of breast imaging experience (range, 2–37 years) and annual mammography volume (*n* = 500–25 000) (Table S1). The readers had no prior UST experience, were not associated with the sites where imaging was performed, and had not previously reviewed any of the images included in the study. Each reader completed 12 hours of UST training, including video modules and hands-on training by a radiologist with 22 years of experience in breast imaging.

Each reader independently interpreted sets of patient images in a unique random order in an environment that simulated clinical reading. Readers were provided a lexicon of lesion characteristics corresponding to BI-RADS assessment

Table 2: Characteristics of Cancer Lesions

Characteristic	Percentage of Total Cancer Cases
Type of cancer	
Invasive ductal carcinoma, not otherwise specified	50.0 (18/36)
Ductal carcinoma in situ	30.6 (11/36)
Invasive lobular carcinoma	11.1 (4/36)
Other malignancy	8.3 (3/36)
Lesion type	
Mass	61.1 (22/36)
Calcifications	30.6 (11/36)
Mass plus calcifications	5.6 (2/36)
Asymmetry	2.8 (1/36)
Laterality	
Left	55.6 (20/36)
Right	44.4 (16/36)
Location	
Upper outer quadrant	33.3 (12/36)
Upper inner quadrant	30.6 (11/36)
Lower outer quadrant	19.4 (7/36)
Lower inner quadrant	8.3 (3/36)
Retroareolar/periareolar/central	8.3 (3/36)
Tumor size stage*	
T0	2.8 (1/36)
Tis	27.8 (10/36)
T1	50.0 (18/36)
T2	16.7 (6/36)
T3	2.8 (1/36)
Node invasion stage†	
N0	66.7 (24/36)
N1mi	2.8 (1/36)
N1	8.3 (3/36)
NX	22.2 (8/36)

Note.—Data are percentages and data in parentheses are the numbers used to calculate the percentages.

* T0 = No evidence of primary tumor, Tis = carcinoma in situ, T1 = tumor ≤ 20 mm in greatest dimension, T2 = tumor > 20 mm but ≤ 50 mm, T3 = tumor > 50 mm.

† N0 = No regional lymph node metastases, N1mi = micrometastases (> 0.2 mm and ≤ 2 mm) to axillary lymph node(s), N1 = metastases to one to three axillary lymph nodes(s) and/or the internal mammary lymph nodes, NX = regional lymph nodes cannot be assessed.

categories (Fig 3) and were blinded to the ground truth (cancer vs noncancer) and prior images. Readers were informed literally that "cancer prevalence was greater than typically seen in clinical practice of asymptomatic screening populations" without any other specific detail. For each patient, FFDM with four standard views (mediolateral oblique and craniocaudal) and UST with four coronal volumetric sequences were reviewed. The reader interpreted FFDM first and marked no more than three suspicious findings. The reader also assigned a BI-RADS assessment category (1 through 5; BI-RADS 0 was not permitted) and a malignancy score (between 0 and 100) for each lesion and for the patient overall. This use of the BI-RADS categories followed other receiver operating characteristic studies (16–19). The FFDM report was locked, and then the reader interpreted both modalities together of the same patient and reported suspicious finding(s), BI-RADS assessment(s), and malignancy score(s) as previously described. There was no washout period between readings. The sequential design was chosen as it more closely mimics the clinical use and little evidence suggests the design would influence results (20,21). The amount of time it took for each reader to assess each patient's FFDM and FFDM plus UST images was recorded. Readings of 30 minutes or longer, which likely resulted from intentional breaks, were not included in analysis.

Statistical Analysis

Patient characteristics, lesion characteristics, and reading time were summarized with fractions and/or percentages for categorical variables and means with SDs or medians with IQRs for continuous variables. The receiver operating characteristic analysis was based on the reader-reported malignancy score for each patient (not those for individual findings). The nonparametric area under the receiver operating characteristic curve (AUC) value was estimated by using multireader multicase (MRMC) analysis of variance method of Obuchowski and Rockette (22). Sensitivity and specificity were assessed by using MRMC analysis of variance model for binary data (23,24). All analyses required a reader to correctly localize at least one malignant lesion in a cancer case to be credited as a true-positive finding (Appendix S1).

For AUC, BI-RADS 4 sensitivity, and BI-RADS 3 sensitivity and specificity, the superiority of FFDM plus UST compared with FFDM alone was evaluated at the two-sided significance level of $\alpha = .05$. Per the prespecified statistical analysis plan, noninferiority of BI-RADS 4 specificity was evaluated at the one-sided significance level of $\alpha = .025$ for noninferiority margin of –10%. As a supplemental analysis, post hoc, a partial AUC value for cancer detection was analyzed within an operating range based on the readers' sensitivity and specificity operating points, evaluated at the two-sided significance level of $\alpha = .05$ (25,26). Average positive predictive value was calculated. Statistical analyses were performed by an author (J. Perez) using SAS (version 9.4; SAS Institute), OR-DBM MRMC (version 2.51; University of Iowa), and R (version 3.4.1; The R Foundation) software.

Sample size calculations were performed prior to the conduct of the study. Thirty-two readers and 140 cases (36 cancer, 104 noncancer) were required for 80% statistical power for superiority in AUC of FFDM plus UST compared with FFDM alone at two-sided α of .05 (Appendix S3).

Results

Patient and Lesion Characteristics

The prospective case collection enrolled 7439 patients. After exclusion of 975 patients who were lost to follow-up, 44 who withdrew from the original study, 3203 who had images collected using older versions of the SoftVue system, and 155 who had unilateral imaging, BI-RADS density A or B, or were simultaneously enrolled in another arm of the original study, a total

Table 3: Diagnostic Performance for Cancer Detection and Mean Diagnostic Accuracy Metrics for BI-RADS Category 3 and 4 Assessments at Mammography and Mammography Plus UST

Variable	FFDM*	FFDM Plus UST*	Change from FFDM to FFDM Plus UST		
			Relative Change (%)*†	Absolute Change*‡	*P* Value§
AUC	0.54 ± 0.05 (0.45, 0.64)	0.60 ± 0.05 (0.51, 0.69)	10.08 [(0.600 – .54)/0.54]	0.05 ± 0.02 (0.01, 0.10)	.03‖
Partial AUC#	0.21 ± 0.03 (0.15, 0.27)	0.25 ± 0.03 (0.20, 0.31)	19.26 [(0.25 – 0.21)/0.21]	0.04 ± 0.02 (0.01, 0.07)	.02‖
BI-RADS 4					
Sensitivity (%)	30 ± 6 [343/1152] (17, 42)	37 ± 6 [428/1152] (25, 50)	25 [(428 – 343)/343]	7 ± 3 [85/1152] (1, 14)	.03‖
Specificity (%)	88 ± 2 [2916/3328] (83, 92)	82 ± 3 [2741/3328] (77, 87)	–6 [(2741 – 2916)/2916]	–5 ± 2 [–175/3328] (–9, –2)	.004**
BI-RADS 3					
Sensitivity (%)	33 ± 7 [385/1152] (20, 46)	40 ± 6 [461/1152] (27, 53)	20 [(461 – 385)/385]	7 ± 4 [76/1152] (–1, 14)	.08‖
Specificity (%)	69 ± 3 [2299/3328] (63, 76)	75 ± 3 [2491/3328] (69, 81)	8 [(2491 – 2299)/2299]	6 ± 3 [192/3328] (0.3, 11)	.04‖

Note.—Unless otherwise indicated, data are means ± standard errors; data in parentheses are 95% CIs and data in brackets are the numbers used to calculate the percentages. FFDM and FFDM plus UST from 140 patients were reviewed by 32 readers and correct lesion localization was required for this analysis. Among these patients, 36 had breast cancer and 104 had no cancer. AUC was estimated using the analysis of variance method of Obuchowski and Rockette (22). Sensitivity and specificity were assessed using the analysis of variance model for binary data. The numerical values may not add up exactly due to rounding. AUC = area under the receiver operating characteristic curve, BI-RADS = Breast Imaging Reporting and Data System, FFDM = full-field digital mammography, MRMC = multireader multicase, UST = US tomography.

* For sensitivity and specificity, the numbers in parentheses are the total number of correct cases over all readers divided by the number of cases times the number of readers.

† Calculated as (FFDM + UST – FFDM)/FFDM.

‡ Calculated as (FFDM + UST – FFDM).

§ All *P* values were calculated based on absolute change (FFDM + UST – FFDM).

‖ Two-sided test of superiority (α = .05) based on *t* test from analysis of variance model.

Based on area under the receiver operating characteristic curve and "1 – specificity" values between 0.1 and 0.6 were determined to be the relevant operating range.

** One-sided test of noninferiority (α = .025) based on *t* test from analysis of variance model with prespecified noninferiority margin of –10% absolute change.

of 3062 patients were eligible for inclusion in this study (Fig 1). Ultimately 140 women (mean age, 56 years ± 10 [SD]) were included in this study, of whom 36 had breast cancer and 104 did not (Table 1). Of the 36 patients with breast cancer, 18 had invasive ductal carcinoma, 11 had ductal carcinoma in situ, four had invasive lobular carcinoma, and three had other malignancies, manifesting as mass, calcifications (visible only on mammography), mass and calcifications, and asymmetry (Table 2). Two patients had a prior history of cancer with breast-conserving surgery, and their postsurgical changes did not result in false-positive findings with either modality (Table 1). The benign biopsy-proven lesions (*n* = 5) were fibrocystic change, fibroadenoma, papilloma, duct ectasia, and apocrine metaplasia.

Overall Performance

The overall performance of the 32 readers for cancer detection was higher with FFDM combined with UST (AUC = 0.60; 95% CI: 0.51, 0.69) compared with FFDM alone (AUC = 0.54; 95% CI: 0.45, 0.64), with FFDM plus UST demonstrating superiority to FFDM alone (AUC increase = 0.05; 95% CI: 0.01, 0.10; *P* = .03) (Table 3). Pooled receiver operating characteristic curves across the readers are shown in Figure 4A.

AUC in a Relevant Operating Range

The relevant operating range for this study was determined to be 1 minus specificity values between 0.1 and 0.6 (Fig 4B, 4C). The partial AUC value for cancer detection for FFDM plus UST demonstrated superiority compared with FFDM alone (partial AUC = 0.25 [95% CI: 0.20, 0.31] vs 0.21 [95% CI: 0.15, 0.27]; increase = 0.04 [95% CI: 0.01, 0.07], *P* = .02) (Table 3, Fig 4A).

BI-RADS 4 Sensitivity and Specificity

For BI-RADS 4 assessment, FFDM plus UST achieved a relative increase in sensitivity of 25% (85 of 343) and was found to be superior to FFDM alone (37% [428 of 1152] vs 30% [343 of 1152]; *P* = .03) (Tables 3, S2; Fig 4B). The mean specificity of FFDM plus UST was noninferior to that of FFDM alone (82% [2741 of 3328] versus 88% [2916 of 3328]; one-sided *P* = .004), with an absolute decrease of 5% (–175 of 3328). Figure

5 shows example patient images demonstrating an increase in sensitivity for UST images compared with FFDM alone. UST led to a mean of 3.4 ± 1.6 more true-positive findings per reader (Fig S1) and a mean of 0.8 ± 0.8 more false-negative findings (Fig S2), resulting in a net gain of 2.7 more true-positive findings. UST also led to a mean of 9.0 ± 6.1 more false-positive findings per reader and a mean of 3.5 ± 3.7 more true-negative findings, resulting in a net increase of 5.5 more false-positive findings. The mean positive predictive value across readers was 45% (343 of 755) for FFDM alone, equivalent to 2.2 biopsies per cancer detected. For FFDM plus UST, the positive predictive value was 42% (428 of 1015), or 2.4 biopsies per cancer detected.

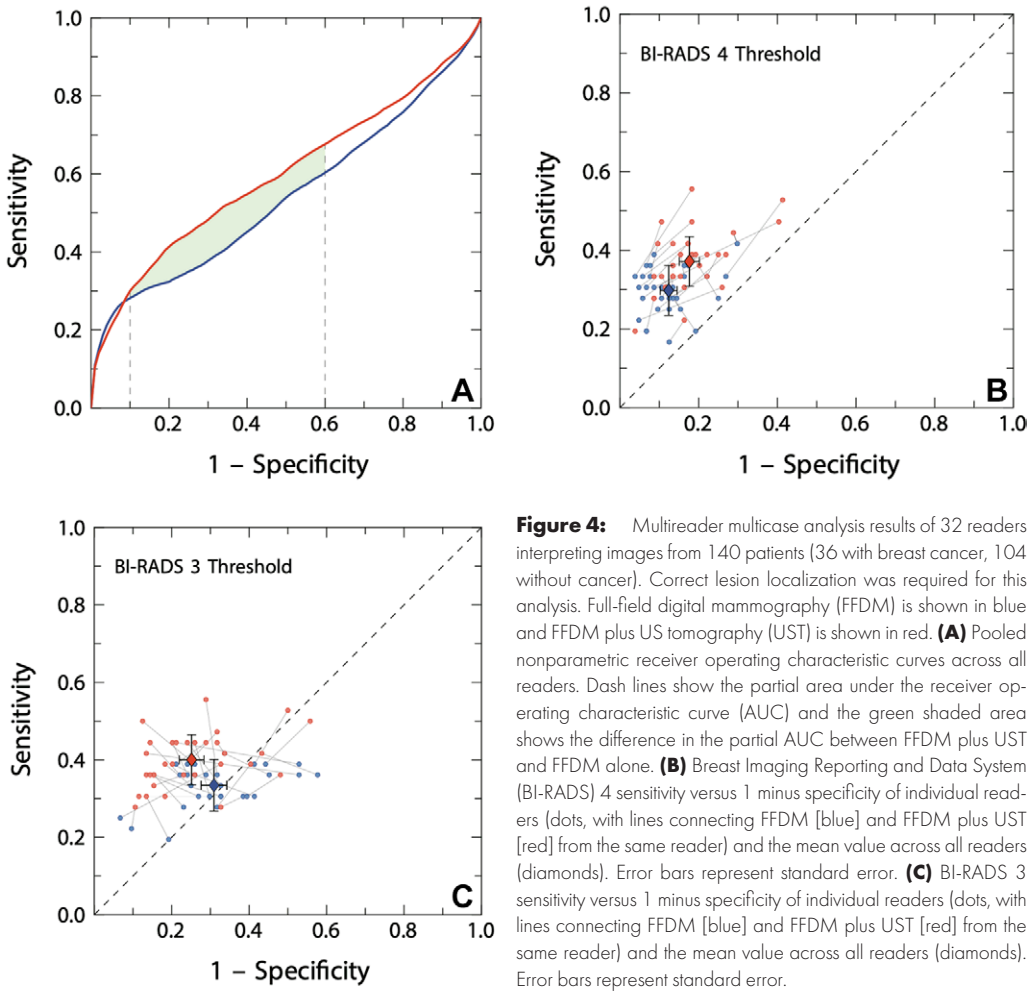


Figure 4: Multireader multicase analysis results of 32 readers interpreting images from 140 patients (36 with breast cancer, 104 without cancer). Correct lesion localization was required for this analysis. Full-field digital mammography (FFDM) is shown in blue and FFDM plus US tomography (UST) is shown in red. **(A)** Pooled nonparametric receiver operating characteristic curves across all readers. Dash lines show the partial area under the receiver operating characteristic curve (AUC) and the green shaded area shows the difference in the partial AUC between FFDM plus UST and FFDM alone. **(B)** Breast Imaging Reporting and Data System (BI-RADS) 4 sensitivity versus 1 minus specificity of individual readers (dots, with lines connecting FFDM [blue] and FFDM plus UST [red] from the same reader) and the mean value across all readers (diamonds). Error bars represent standard error. **(C)** BI-RADS 3 sensitivity versus 1 minus specificity of individual readers (dots, with lines connecting FFDM [blue] and FFDM plus UST [red] from the same reader) and the mean value across all readers (diamonds). Error bars represent standard error.

BI-RADS 3 Sensitivity and Specificity

For BI-RADS 3 assessment, FFDM plus UST showed no evidence of an increase in mean sensitivity relative to FFDM alone (20% [76 of 385]; 40% [461 of 1152] vs 33% [385 of 1152]; P = .08) (Tables 3, S2; Fig 4C). The FFDM plus UST mean specificity had a relative increase of 8% (192 of 2299), which was superior to that of FFDM alone (75% [2491 of 3328] vs 69% [2299 of 3328]; P = .04). Figure 6 shows example patient images that demonstrate an increase in specificity for FFDM plus UST compared with FFDM alone. UST led to a mean of 3.3 ± 1.4 more true-positive findings per reader and a mean of 0.9 ± 0.8 more false-negative findings, resulting in a net gain of 2.4 more true-positive findings. UST also led to a mean of 13.2 ± 9.6 more true-negative findings per reader and a mean of 7.3 ± 6.1 more false-positive findings, resulting in a net decrease of 6.0 fewer false-positive findings.

Reading Time

There were a total of 4480 readings (32 readers assessing 140 sets of patient images) for FFDM plus UST and for FFDM alone. Of the 4480 readings, 4475 FFDM plus UST readings and 4332 FFDM-alone readings were less than 30 minutes. Among images read in 30 minutes or less, the median reading time was 3.4 minutes (IQR: 2.3, 4.7) for FFDM plus UST and 1.2 minutes (IQR: 0.8, 1.7) for FFDM alone. For FFDM plus UST, images with normal findings appeared to have a faster reading time than images where cancer was detected (median, 3.2 minutes [IQR: 2.2, 4.4] vs 3.9 minutes [IQR: 2.7, 5.4]) (Table 4).

Discussion

Women with dense breasts benefit from supplemental screening with US (9–11). The purpose of this multireader multicase study was to evaluate the performance of US tomography (UST) combined with full-field digital mammography (FFDM) compared with FFDM alone for breast cancer screening in women with dense breasts. We found that FFDM plus UST improved breast cancer detection among 32 readers compared with FFDM alone (area under the receiver operating characteristic curve, 0.60 [95% CI: 0.51, 0.69] vs 0.54 [95% CI: 0.45, 0.64]; P = .03). Evaluation of FFDM with UST led to superior Breast Imaging Reporting and Data System (BI-RADS) 4 sensitivity (relative change in mean sensitivity, 25% [85 of 343]; 37% [428 of 1152] vs 30% [343 of 1152]; P = .03) and noninferior BI-RADS 4 specificity (relative change in mean specificity, −6% [−175 of 2916]; 82% [2741 of 3328] vs 88% [2916 of 3328]; margin, −10%; P = .004) compared with FFDM alone, respectively. For BI-RADS 3 assessment, mean sensitivity did not show evidence

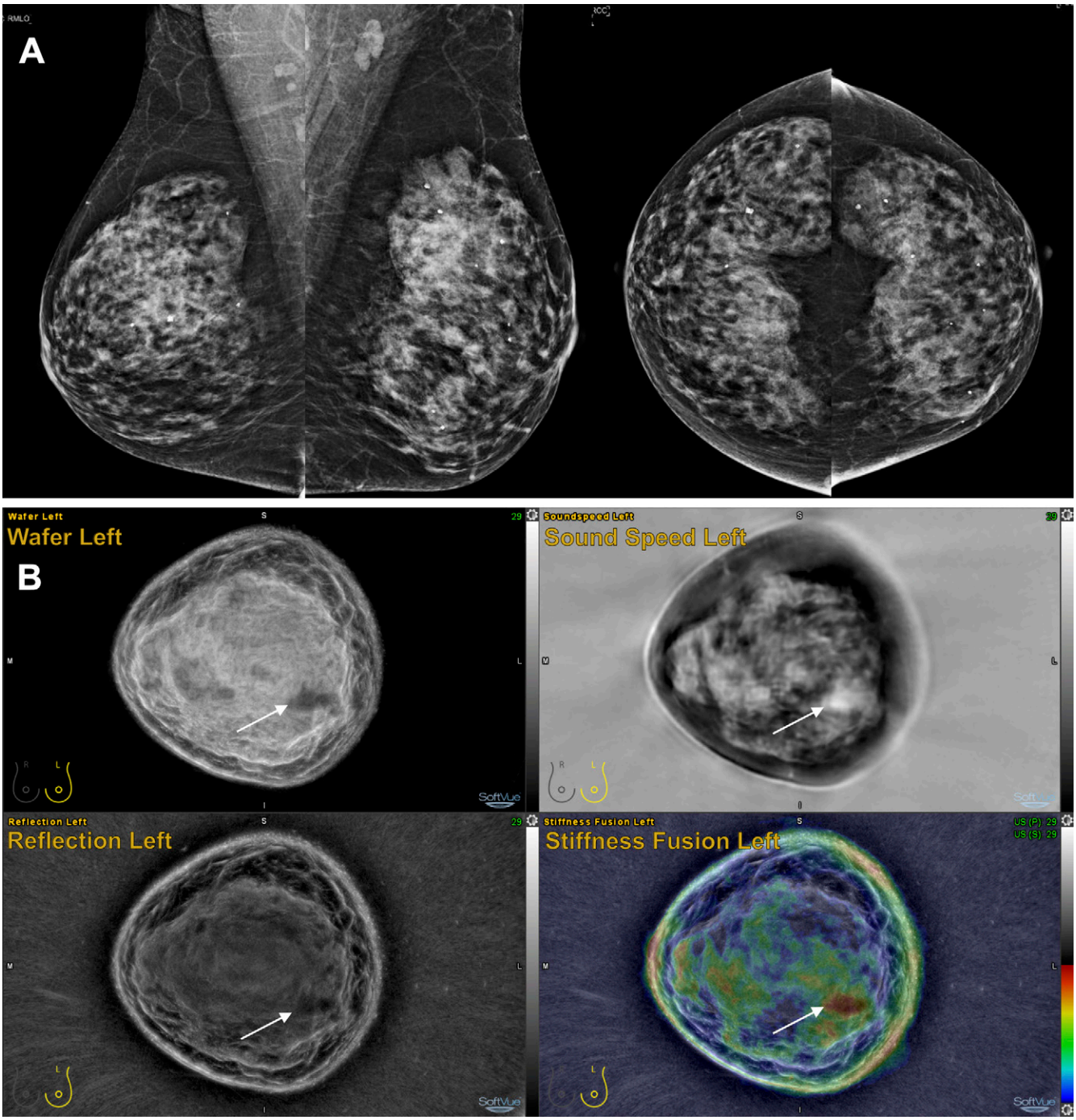


Figure 5: Example images in a 60-year-old woman with a biopsy-proven triple-negative invasive ductal carcinoma. **(A)** Negative bilateral mediolateral oblique (left) and cranial caudal (right) mammographic views demonstrate heterogeneously dense breasts. **(B)** Image slices from US tomography (UST). Coronal volumetric image-stack sequences demonstrate an irregular mass (arrows) with focal stiffness in the lower outer quadrant of the left breast, not seen at full-field digital mammography (FFDM). The four image sequences show the mass, which is dark on wafer, bright on sound speed, persistent on reflection, and focally stiff (orange/red) on stiffness fusion sequences. All 32 readers rated this case as Breast Imaging Reporting and Data System (BI-RADS) 3 or lower (median malignancy score, 23.5 [IQR: 16.0, 35.0]) at FFDM alone, and 17 rated it as BI-RADS 4a or higher (median malignancy score, 54.0 [IQR: 34.0, 74.0]) at FFDM plus UST.

of an increase with FFDM plus UST relative to FFDM alone (change, 20% [76 of 385]; 40% [461 of 1152] vs 33% [385 of 1152]; $P = .08$) but mean specificity increased 8% (192 of 2299; 75% [2491 of 3328] vs 69% [2299 of 3328]; $P = .04$). The median case reading time was 1.2 minutes (IQR: 0.8, 1.7) for FFDM alone and 3.4 minutes (IQR: 2.3, 4.7) for FFDM plus UST.

Handheld US and automated breast US as a supplement to FFDM increase cancer detection in women with dense breasts (8,10). Our study suggests UST can be another supplemental screening tool. The increase in sensitivity of UST (relative change in mean BI-RADS-4 sensitivity, 25% [85 of 343]; 37% [428 of 1152] vs 30% [343 of 1152]; $P = .03$) is within the reported 22%–41% improvement from a review

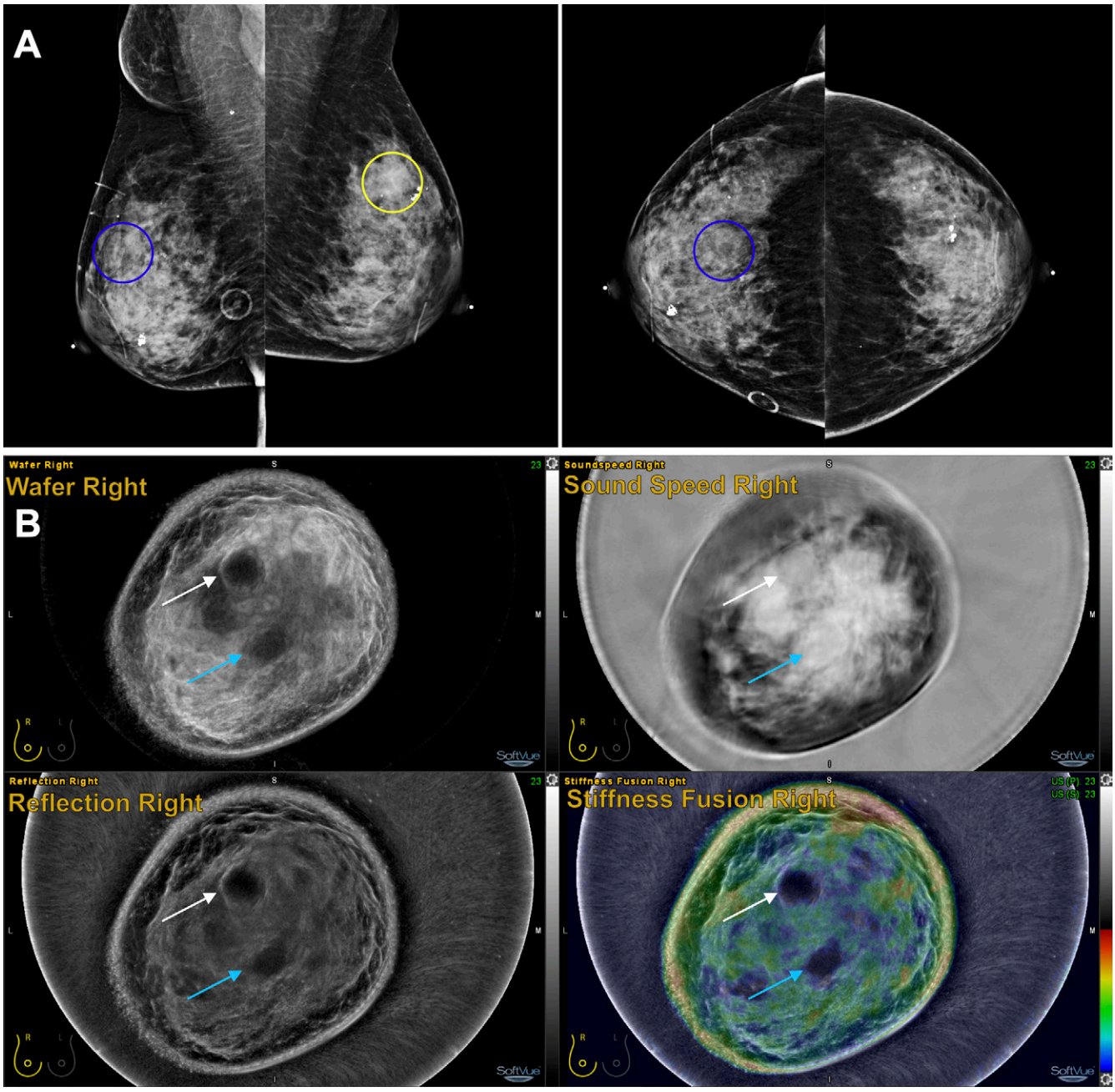


Figure 6: Example images in a 52-year-old asymptomatic woman with a history of multiple bilateral excisional biopsies and bilateral benign calcifications. **(A)** Bilateral mediolateral oblique (left) and cranial caudal (right) mammographic views demonstrate extremely dense breasts with a round mass at 11 o'clock in the right breast (blue circle) and asymmetry in the superior left breast seen only on the mediolateral oblique view (yellow circle). **(B)** Image slices from US tomography (UST). Coronal volumetric image-stack sequences of the right breast demonstrate a cyst (white arrows) at 11 o' clock which corresponds to the mass seen at full-field digital mammography (FFDM). Another cyst is seen incidentally in the central right breast (blue arrows). The cysts are dark on wafer, gray on sound speed, persistent on reflection, and blue (soft) on stiffness fusion sequences *(Fig 6 continues)*.

of automated breast US studies (27). UST provides two specific features that aid in cancer detection. First, coronal volumetric image sequences are useful to identify and characterize lesions. They provide a better view of the fat-glandular interface where most breast cancers are located (28). Second, the stiffness fusion sequence provides tissue stiffness information, which can help differentiate cancer from benign masses and is not readily available with handheld US or automated breast US.

We found that for BI-RADS 4 assessment, addition of UST to FFDM would lead to additional benign biopsies. For screening US, reduced positive predictive value is a known limitation, with reductions ranging from 4.2% to 15.8% (8,27). In comparison, addition of UST to FFDM resulted in only a slight reduction in positive predictive value compared with FFDM alone (42.2% vs 45.4%, respectively). Clinical studies are needed to determine whether supplemental UST can improve cancer detection without substantially increasing the number of benign biopsies.

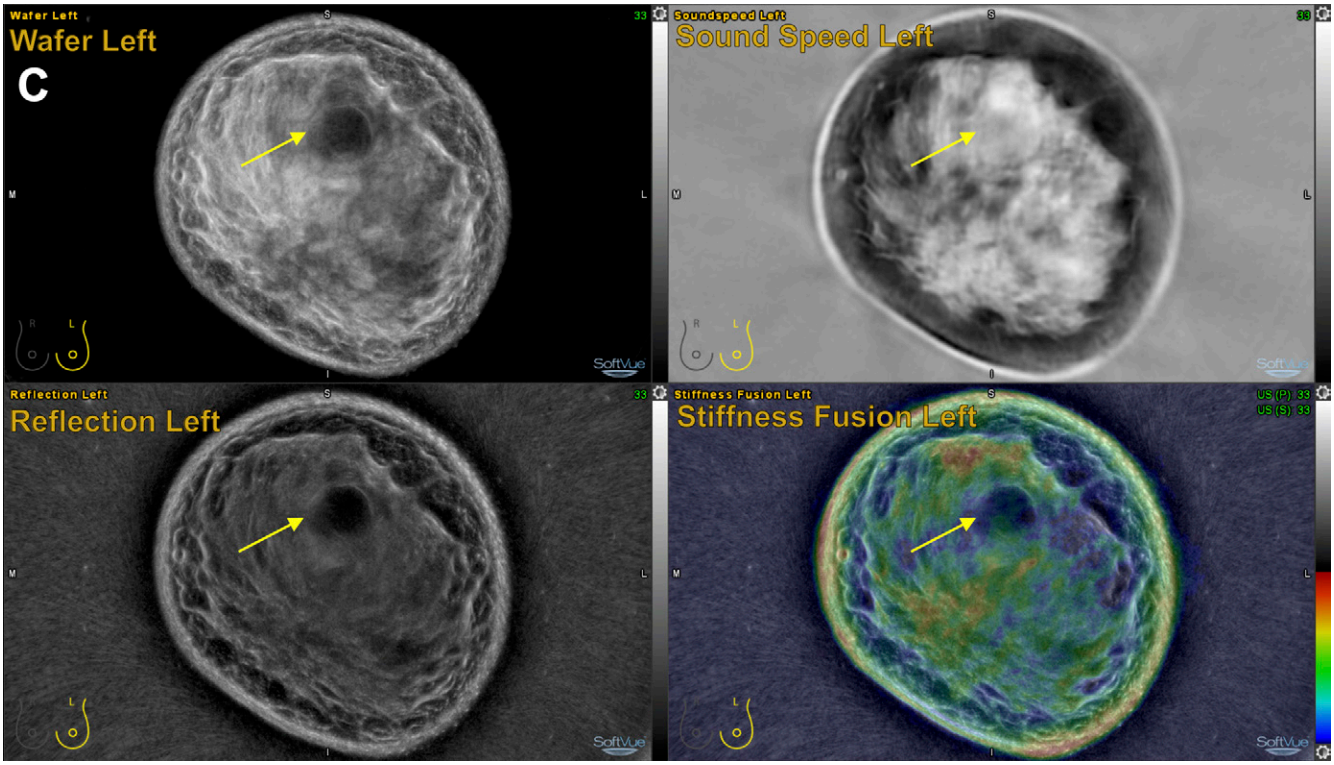


Figure 6 *(continued)*: **(C)** Image slices from UST coronal volumetric image-stack sequences of the left breast demonstrate a cyst at 12 o'clock (yellow arrows), which corresponds to the asymmetry seen at FFDM. Of the 32 readers, 22 rated this case as Breast Imaging Reporting and Data System (BI-RADS) 2 or lower (median malignancy score, 39.0 [IQR: 30.0, 41.5]) at FFDM alone, and 28 rated it as BI-RADS 2 or lower (median malignancy score, 30.0 [IQR: 22.0, 39.0]) at FFDM plus UST.

Table 4: Reading Time for FFDM Plus UST and FFDM Alone Images

	FFDM Plus UST				FFDM Alone
Reading Time	Normal (n = 3164)	Benign (n = 160)	Cancer (n = 1151)	All (n = 4475)	All (n = 4332)
Median (min)*	3.2 (2.2, 4.4)	3.8 (2.7, 4.8)	3.9 (2.7, 5.4)	3.4 (2.3, 4.7)	1.2 (0.8, 1.7)
Range (min)	0.6–26.3	0.3–10.0	0.6–25.3	0.3–26.3	0.3–28.3

Note.—FFDM = full-field digital mammography, UST = US tomography.
* Values in parentheses are the 1st and 3rd quartiles, respectively.

The specificity improvement with FFDM plus UST at BI-RADS 3 assessment is encouraging. A BI-RADS 3 assessment is the most common source of false-positive findings in handheld US and automated breast US screening (8,29). With improved BI-RADS 3 specificity, radiologists can more accurately characterize masses as benign (BI-RADS 2), leading to fewer BI-RADS 3 assessments. The combined interpretation time of FFDM plus UST is also promising. While our experience is limited to a reader study, the 3–4 minutes per case reading time compares favorably with that of automated breast US, reported as 9 minutes by Skaane et al (14) and as 3–7 minutes by Berg and Vourtsis (8). This should allow successful integration of UST into the clinical workflow.

Similar to other reader studies (16–18), the patient samples included in this study differ in some ways from those in a prospective clinical trial. All cancers and only 4% of noncancers were selected. One reason for this is that readers completed this reader study over a weekend, whereas a clinical trial typically spans many months. Another is the extremely low breast cancer prevalence in screening. Combined, these two reasons necessitate to raise the cancer prevalence many folds and to include only a small fraction of noncancer cases. Therefore, not all clinical cases are represented equally. The five biopsied noncancers probably did not cover the full spectrum of benign diseases that require biopsy. Evidence indicates that raised prevalence likely does not affect the AUC value, thus supporting reader studies as a useful means to measure clinical performance (30,31).

Our study had limitations. First, our study was conducted in a laboratory setting and subject to possible "laboratory effect," where readers behave differently from clinical practice (32). The reading environment in our study also differed from that in clinical practice in that readers had an abbreviated clinical history without prior images and the cancer prevalence was greater than seen in typical clinical practice. Second, BI-RADS

category 0 was not used, and readers made biopsy decisions without diagnostic mammography or targeted US. This deviated from clinical practice but allowed analysis of sensitivity and specificity specifically pertaining to UST and not confounded by other additional imaging modalities, similar to other reader studies (16–19). Third, we had only five biopsied noncancer cases, which could have caused underestimation of the impact of these cases on specificity for both FFDM alone and FFDM plus UST. Fourth, our study did not have uncommon cancers such as mucinous, medullary, tubular, or intracystic papillary carcinomas. Fifth, none of the readers had prior experience reading images from SoftVue systems. As this new technology becomes more widespread, radiologists will gain more experience with it and their performance for detecting cancer on UST images is likely to improve.

In summary, this multireader multicase observer performance study showed that in women with dense breasts, full-field digital mammography (FFDM) with US tomography (UST) led to improved cancer detection by radiologists compared with FFDM alone. Large clinical studies and greater clinical experience are required to understand the clinical benefits of UST.

Deputy Editor: Linda Moy
Scientific Editor: Linda Moy

Acknowledgments: We thank the many radiologists who participated as investigators in the SoftVue Prospective Case Collection and the many radiologist readers who participated in the MRMC observer performance study.

Author contributions: Guarantor of integrity of entire study, **M.W.Y.**; 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, **M.W.Y., L.H.L., Y.J.**; clinical studies, **M.W.Y., L.H.L., A.V.E., J. Papaioannou, Y.J**; statistical analysis, **J. Perez, Y.J.**; and manuscript editing, **M.W.Y., L.H.L., J. Perez, Y.J.**

Disclosures of conflicts of interest: M.W.Y. Institution received a grant from Delphinus Medical Technologies; author is a consultant for and received payments for presentations, training of SoftVue interpretation, and SoftVue image reviews for radiologists from Delphinus Medical Technologies; received support for attending certain meetings and travel, if related to SoftVue, from Delphinus Medical Technologies. **L.H.L.** Received consulting fees from Delphinus Medical Technologies; payment for expert testimony from Burkwitz law firm; Society of Breast Imaging, social media group member. **J.Perez** Delphinus Medical Technologies paid Avania, a contract research organization, through a statement of work outlined in a clinical trial agreement, to provide statistical consulting and statistical analysis services. The author, an employee of Avania, was the statistical lead of the study protocol and executed the analyses for this study and manuscript. **A.V.E.** Delphinus Medical Technologies provided funding through a research agreement paid to the University of Chicago, Y. Jiang, PI; received consulting fees related to research studies not reported in this manuscript; study travel reimbursement. **J.Papaioannou** Delphinus Medical Technologies provided funding through a research agreement paid to the University of Chicago, Y. Jiang, PI; received consulting fees related to research studies not reported in this manuscript; study travel reimbursement. **Y.J.** Delphinus Medical Technologies provided funding through a research agreement paid to the University of Chicago, Y. Jiang, PI; received consulting fees related to research studies not reported in this manuscript; study travel reimbursement; travel reimbursement for presenting this study at SBI 2023; consulting fees from Quantitative Transmission Ultrasound Imaging.

References

1. van der Waal D, Ripping TM, Verbeek ALM, Broeders MJM. Breast cancer screening effect across breast density strata: A case-control study. Int J Cancer 2017;140(1):41–49.
2. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2006;15(6):1159–1169.
3. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med 2007;356(3):227–236.
4. Strand F, Azavedo E, Hellgren R, et al. Localized mammographic density is associated with interval cancer and large breast cancer: a nested case-control study. Breast Cancer Res 2019;21(1):8.
5. Draft Recommendation: Breast Cancer: Screening | United States Preventive Services Taskforce. https://www.uspreventiveservicestaskforce.org/uspstf/draft-recommendation/breast-cancer-screening-adults#bcei-recommendation-title-area. Accessed June 1, 2023.
6. Mammography Quality Standards Act,21 CFR Part 900 – Mammography. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-I/part-900. Accessed June 1, 2023.
7. Vourtsis A, Berg WA. Breast density implications and supplemental screening. Eur Radiol 2019;29(4):1762–1777.
8. Berg WA, Vourtsis A. Screening breast ultrasound using handheld or automated technique in women with dense breasts. J Breast Imaging 2019;1(4):283–296.
9. Ohuchi N, Suzuki A, Sobue T, et al. Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): a randomised controlled trial. Lancet 2016;387(10016):341–348.
10. Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. JAMA 2008;299(18):2151–2163. [Published correction appears in JAMA 2010;303(15):1482.]
11. Tagliafico AS, Calabrese M, Mariscotti G, et al. Adjunct Screening With Tomosynthesis or Ultrasound in Women With Mammography-Negative Dense Breasts: Interim Report of a Prospective Comparative Trial. J Clin Oncol 2016;34(16):1882–1888.
12. Berg WA, Rafferty EA, Friedewald SM, Hruska CB, Rahbar H. Screening algorithms in dense breasts: AJR expert panel narrative review. AJR Am J Roentgenol 2021;216(2):275–294.
13. Arleo EK, Saleh M, Ionescu D, Drotman M, Min RJ, Hentel K. Recall rate of screening ultrasound with automated breast volumetric scanning (ABVS) in women with dense breasts: a first quarter experience. Clin Imaging 2014;38(4):439–444.
14. Skaane P, Gullien R, Eben EB, Sandhaug M, Schulz-Wendtland R, Stoeblen F. Interpretation of automated breast ultrasound (ABUS) with and without knowledge of mammography: a reader performance study. Acta Radiol 2015;56(4):404–412.
15. Littrup PJ, Duric N, Sak M, et al. Multicenter study of whole breast stiffness imaging by ultrasound tomography (SoftVue) for characterization of breast tissues and masses. J Clin Med 2021;10(23):5528.
16. Jiang Y, Edwards AV, Newstead GM. Artificial Intelligence Applied to Breast MRI for Improved Diagnosis. Radiology 2021;298(1):38–46.
17. Jiang Y, Inciardi MF, Edwards AV, Papaioannou J. Interpretation time using a concurrent-read computer-aided detection system for automated breast ultrasound in breast cancer screening of women with dense breast tissue. AJR Am J Roentgenol 2018;211(2):452–461.
18. Giger ML, Inciardi MF, Edwards A, et al. Automated breast ultrasound in breast cancer screening of women with dense breasts: Reader study of mammography-negative and mammography-positive cancers. AJR Am J Roentgenol 2016;206(6):1341–1350.
19. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005;353(17):1773–1783. [Published correction appears in N Engl J Med 2006;355(17):1840.]
20. Kobayashi T, Xu XW, MacMahon H, Metz CE, Doi K. Effect of a computer-aided diagnosis scheme on radiologists' performance in detection of lung nodules on radiographs. Radiology 1996;199(3):843–848.
21. Beiden SV, Wagner RF, Doi K, et al. Independent versus sequential reading in ROC studies of computer-assist modalities: analysis of components of variance. Acad Radiol 2002;9(9):1036–1043.
22. Obuchowski NA, Rockette HE. Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests an anova approach with dependent observations. Commun Stat Simul Comput 1995;24(2):285–308.
23. Gallas BD, Pennello GA, Myers KJ. Multireader multicase variance analysis for binary data. J Opt Soc Am A Opt Image Sci Vis 2007;24(12):B70–B80.
24. Smith BJ, Hillis SL. Multi-reader multi-case analysis of variance software for diagnostic performance comparison of imaging modalities. In: Proceedings Volume 11316, Medical Imaging 2020: Image Perception, Observer Performance, and Technology Assessment. SPIE, 2020; 113160K.

25. McClish DK. Analyzing a portion of the ROC curve. Med Decis Making 1989;9(3):190–195.
26. Jiang Y, Metz CE, Nishikawa RM. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. Radiology 1996;201(3):745–750.
27. Kim SH, Kim HH, Moon WK. Automated Breast Ultrasound Screening for Dense Breasts. Korean J Radiol 2020;21(1):15–24.
28. Littrup PJ, Duric N, Sak M, et al. The fat-glandular interface and breast tumor locations: Appearances on ultrasound tomography are supported by quantitative peritumoral analyses. J Breast Imaging 2021;3(4):455–464.
29. Berg WA. Current status of supplemental screening in dense breasts. J Clin Oncol 2016;34(16):1840–1843.
30. Gur D, Rockette HE, Armfield DR, et al. Prevalence effect in a laboratory environment. Radiology 2003;228(1):10–14.
31. Swets JA. Measuring the accuracy of diagnostic systems. Science 1988;240(4857):1285–1293.
32. Gur D, Bandos AI, Cohen CS, et al. The "laboratory" effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. Radiology 2008;249(1):47–53.