

Reader Trial  
*European Radiology*  
DOI 10.1007/s00330-010-1844-1

Breast cancer detection:  
radiologists' performance using  
mammography with and without  
automated whole-breast ultrasound

by:  
Kevin M. Kelly  
Judy Dean  
Sung-Jae Lee  
W. Scott Comulada

Published on-line: 15 July 2010

# Automated Whole-Breast Ultrasound Screening (AWBUS)

## Reader Trial Summary

- Examine reproducibility and trainability of the technique
- 12 radiologists
- 102 cases
- All dense-breasted patients
- Mammograms read first
- AWBUS read second
- More and smaller cancers found with AWBUS
  - ✓ Radiologists found approximately 50% of all cancers in dense-breasted patients with mammogram.
  - ✓ Radiologists found approximately 80% of all cancers by adding AWBUS to the mammogram.
  - ✓ Radiologists found approximately 25% of sub-centimeter cancers in dense-breasted patients with mammogram.
  - ✓ Radiologists found approximately 66% of sub-centimeter cancers by adding AWBUS to the mammogram.

Kevin M. Kelly  
Judy Dean  
Sung-Jae Lee  
W. Scott Comulada

## Breast cancer detection: radiologists' performance using mammography with and without automated whole-breast ultrasound

Received: 2 March 2010  
Revised: 23 April 2010  
Accepted: 19 May 2010  
© The Author(s) 2010.  
This article is published with open access at Springerlink.com

The authors dedicate this article to the memory of Dr. Robert F. Wagner, who died May 2008 after a short illness. He was a pioneer in the development of a variety of the methods for assessment of the accuracy, safety, and effectiveness of diagnostic medical devices, particularly in the application of receiver operating characteristic (ROC) analysis.

K. M. Kelly (✉)  
The Hall Center, Santa Monica, CA, USA  
e-mail: drkelly@ca.rr.com  
Tel.: +1-310-3443800  
Fax: +1-775-2561845

J. Dean  
1525 State St. Suite 102, Santa Barbara,  
CA 93101, USA  
e-mail: judy@judydeanmd.net

S.-J. Lee · W. S. Comulada  
Semel Institute Center for Community Health,  
University of California, Los Angeles,  
CA, USA  
e-mail: sjlee@mednet.ucla.edu  
e-mail: scomulad@ucla.edu

### Abstract

**Objective** Radiologist reader performance for breast cancer detection using mammography plus automated whole-breast ultrasound (AWBU) was compared with mammography alone. **Methods** Screenings for non-palpable breast malignancies in women with radiographically dense breasts with contemporaneous mammograms and AWBU were reviewed by 12 radiologists blinded to the diagnoses; half the studies were abnormal. Readers first reviewed the 102 mammograms. The American College of Radiology (ACR) Breast Imaging Reporting and Data System (BIRADS) and Digital Mammographic Imaging Screening Trial (DMIST) likelihood ratings were recorded with location information for identified abnormalities. Readers then reviewed the mammograms and AWBU with knowledge of previous mammogram-only evaluation. We compared reader performance across

screening techniques using absolute callback, areas under the curve (AUC), and figure of merit (FOM). **Results** True positivity of cancer detection increased 63%, with only a 4% decrease in true negativity. Reader-averaged AUC was higher for mammography plus AWBU compared with mammography alone by BIRADS (0.808 versus 0.701) and likelihood scores (0.810 versus 0.703). Similarly, FOM was higher for mammography plus AWBU compared with mammography alone by BIRADS (0.786 versus 0.613) and likelihood scores (0.791 versus 0.614). **Conclusion** Adding AWBU to mammography improved callback rates, accuracy of breast cancer detection, and confidence in callbacks for dense-breasted women.

**Keywords** Multi-reader multi-case (MRMC) study · Diagnostic accuracy · JAFROC analysis · Breast screening · Breast ultrasound

### Introduction

Screening with mammography has been shown to reduce mortality from breast cancer [1, 2]. However, the sensitivity to non-palpable cancer of screening mammography in radiographically dense-breasted women is as low as 30–48% [3]. Extremely dense-breasted women have an 18-fold increase in interval cancer found between annual mammograms, compared with fatty-breasted women [4].

Magnetic resonance imaging (MRI) has been demonstrated to be, and recommended as, an efficacious adjunct to mammography for very high-risk, dense-breasted women [5, 6]. It has not been recommended for all dense-breasted women. Three limitations to MRI screening for breast cancer are cost, intravenous injection of gadolinium-containing contrast medium, and lower specificity of MRI compared with mammography [7] with increased false positive callbacks and biopsies.

For radiographically dense-breasted women, whole-breast ultrasound as an adjunct to screening mammography has shown promise. Berg et al. increased cancer discovery 42% by adding handheld whole-breast ultrasound performed by radiologists [8]. Kelly et al. used an automated whole-breast ultrasound (AWBU) device capturing a ciné loop of 2D breast images [9]. This blinded study of mostly dense-breasted women showed a 100% increase in cancer detection, and a 200% increase in discovery of invasive cancers 1 cm or less, compared with mammograms alone. These ciné loops were recorded and are available for reader trials similar to those performed for comparison of screening mammography with and without computer-aided detection (CAD) [10].

For AWBU to be a useful adjunct to screening mammography for dense-breasted women, interpretation of examinations must be shown as beneficial, when performed by community radiologists. This paper evaluates the performance of such radiologists in detection before and after AWBU is added to a test set of screening mammograms of radiographically dense-breasted women.

## Materials and methods

### Imaging studies

**Mammograms** Standard cranio-caudal (CC) and medio-lateral oblique (MLO) views of each breast were available for all cases. If implants were present, displacement views were included. Original analog films (66 cases) or prints of digital films (36 cases) were used for review. All cases used in the study provided informed consent, and the protocol was approved by the Institutional Review Boards at each hospital, or The Western Institutional Review Board [9].

**AWBUs** Automated whole-breast ultrasound (AWBU) is a computer-based system for performing, recording, and reading whole-breast ultrasound examinations similar in appearance to 2D freehand imaging (SonoCine, Reno, NV). Images were collected with 7- to 12-MHz multi-frequency transducers. The transducer is attached to a computer-guided mechanical arm that acquires images in CC rows overlapping 7 to 10 mm insuring complete coverage of both breasts. Images are collected approximately 0.8 mm apart. The AWBU software creates a ciné loop for review of approximately 3,000 images, simulating real-time imaging. The Windows®-based reading station uses a high-definition 1,600×1,200 monitor and special software to increase cancers' conspicuity. The AWBU procedure was described more fully in a previous publication [9].

### Readers

Twelve board-certified breast radiologists who use breast ultrasound in their practices were recruited as readers for

the trial. Remuneration for 3.5 days was at the prevailing US rate. Eleven readers were from the USA and one from Great Britain. Eleven had no experience with AWBU. One had reviewed limited AWBUs 8 years earlier during the developmental phase of the technology. No reader had foreknowledge of the positivity rate of the test set.

Each reader had a 4-h tutorial with one author (KK) explaining the AWBU reading station operation. The readers reviewed and discussed approximately 12 AWBUs with known cancers, not part of the test set. They were not in the test set because either palpable findings were present or there were no concurrent mammograms. Nothing concerning the study was discussed, other than the use of the data form (Appendix A) and the number of cases to be reviewed.

### Procedure

A set of 51 malignant cases (3 cases with bilateral cancers), including invasive and in situ cancer were collected for the trial (Table 1). Screening mammography and AWBU were performed within 2 months of each other. No cancers were associated with prospective palpable findings or symptoms suggestive of cancer. The mammograms were heterogeneously dense or extremely dense breast tissue (BIRADS 3 or 4) on the original reports. All imaging was performed from 2003 to 2008. The data set included all cases meeting the above criteria in the AWBU archives. Twelve cancers were included that were not prospectively reported on either imaging technique, but are visible in retrospect. Four of these became palpable within 1 year, three in more than 1 year; five were discovered in a subsequent screening round, three by AWBU only, and two by both AWBU and mammography.

Fifty-one normal cases performed from 2003 to 2008 were matched with each of the positive cases for the following factors:

1. Facility
2. Digital or analog mammogram
3. Ultrasound machine model
4. American breast cup size (A–DD)
5. ACR BIRADS breast density
6. Implant (saline or silicone) and location (pre- or retropectoral)
7. Breast cancer history
8. Age

**Table 1** Pathological diagnosis of 51 positive cases (54 cancers)

	≤1 cm	>1 to ≤2 cm	>2 cm	Total
DCIS	2	0	4	6
IDC	17	19	5	41
ILC	3	2	1	6
Mixed IDC and ILC	0	1	0	1
Total	22	22	10	54

*DCIS* ductal carcinoma in situ, *IDC* invasive ductal carcinoma, *ILC* invasive lobular carcinoma

---

The normal case matching factors 1 to 7 closest to the age of the positive case was matched as the normal partner case. The mean difference in age between the positive case and its matched normal was 31 days.

Testing occurred on a subsequent date at each reader's own site with only the reader and a research assistant (monitor) present. The same monitor was present for all readers. She had no knowledge of the test set makeup, had no mammography or ultrasound training, reviewed the test data forms in real-time for completeness, and transferred the data to the study database.

At each test site 102 mammograms were placed on a film alternator in random order, generated once, and used for all readers. Excluding breaks, the test subject's time for review was recorded. The upper half of a data form (Appendix A) was completed for each case, checked by the monitor, and entered into the database.

Four questions were asked:

1. Would you request further evaluation based on this mammogram, or recommend routine screening?
2. Where is/are the most suspicious (up to 3) lesions, identifying their location by breast and clock face position?
3. What would be your prediction of the final ACR BIRADS after any needed diagnostic workup was completed?
4. What is the reader's confidence level that the woman has or does not have cancer (DMIST likelihood scale)?

The American College of Radiology Breast Imaging Reporting and Data System (BIRADS) is a seven-point scale (0 = incomplete, needs additional assessment; 1 = normal; 2 = benign; 3 = probably benign; 4a = possible malignancy; 4b = probable malignancy, or 5 = highly suggestive of malignancy) designed to categorize the results of mammography and other imaging studies [3, 11]. Scores from 1 to 5 were allowed. Similar to the DMIST [12], readers were asked to predict a BIRADS score before any diagnostic workup.

The DMIST likelihood rating is a seven-point scale to express the confidence of the diagnosis, and ranges from definitely not cancer to definitely cancer [3, 11, 12].

A correct location response was recorded for an hour position marked within the half of the breast centered at the middle of the cancer.

A true positive (TP) was recorded for mammography for any malignant case if 'callback' was marked for mammography and any correct tumor location was identified. A TP was recorded for mammography plus AWBU if 'callback' was marked on either or both halves of the form in the malignant cases, with at least one correct location identified. Thus, a correctly identified TP found with mammography would remain TP even were it not identified again on AWBU. AWBU findings could change the outcome to TP if a cancer was correctly identified with AWBU, but missed with mammography. We evaluated readings on a per-case (i.e., per-patient) basis rather than a per-score basis because screening

serves as a "go no-go" gatekeeper for subsequent workup [13].

A true negative (TN) was recorded for mammography for any normal case if 'callback' was not marked for mammography. A TN was recorded for mammography plus AWBU for any normal case if 'callback' was not marked on the second half of the form. This allowed the reader to reverse a callback for an asymmetric density seen mammographically but cleared by the AWBU as no suspicion. To validate TN cases, all cases were followed for at least 1 year or more.

A false positive (FP) was recorded for mammography in two situations:

1. Callback was marked for mammography in a normal case.
2. Callback was marked for mammography in a cancer case, but none of the marked locations corresponded to the cancer.

An FP was recorded for mammography plus AWBU in the same two situations as above when callback was marked for AWBU. A false negative (FN) was recorded for mammography when callback was not marked in a cancer case in the mammography portion of the form. Similarly, an FN was recorded for mammography plus AWBU when callback was not marked in a cancer case in either portion of the form.

The 102 ABWUs were reviewed by readers on a review station brought by the research assistant acting as a monitor. They worked approximately 8 h daily for 3 days, with breaks at the readers' choosing. The readers were given the corresponding mammograms with each AWBU and completed the second half of the data sheet with the knowledge from the mammogram-only evaluation available. The same questions were answered for AWBU and the reading time of each AWBU recorded.

#### Statistical analysis

Analyses were conducted in a multi-reader multi-case (MRMC) framework where each reader screened all cases and each case contained both screening techniques. The MRMC design efficiently reduces the number of readers and cases needed to detect improvements across techniques [14]. Analyses appropriate for an MRMC design were chosen both to correctly model correlations between readings on the same case across readers and to estimate correctly standard errors. Unless specified otherwise, analyses were conducted in SAS software version 9.2 (SAS Institute Inc., Cary, NC, USA). We present  $F$  statistics, shown as  $F(\text{numerator degrees of freedom}, \text{denominator degrees of freedom})$ , and  $p$  values for comparisons between mammography plus AWBU and mammography alone.

*Cases identified for further imaging* were assessed by four binary measures: sensitivity = number of TP/number

**Table 2** Reader performance categorized by imaging technique ( $n=102$ , 51 positive cases)

Reader # <sup>a</sup>	True positives		True negatives		False positives		False negatives	
	M	M+A	M	M+A	M	M+A	M	M+A
1	28	45	32	21	27	30	15	6
2	28	45	25	21	33	30	16	6
3	25	44	30	20	32	31	15	7
4	26	43	20	28	43	23	13	8
5	26	43	32	30	28	21	16	8
6	32	43	20	37	43	14	7	8
7	26	41	25	27	33	24	18	10
8	23	40	35	31	17	20	27	11
9	16	40	43	25	21	26	22	11
10	27	39	34	36	27	15	14	12
11	26	37	35	41	24	10	17	14
12	22	37	37	34	20	17	23	14
Mean # of cases	25.4	41.4	30.7	29.3	29.0	21.8	16.9	9.6
% of 51 cases	49.8%	81.2%	60.2%	57.5%	56.9%	42.7%	33.1%	18.8%
Mean # of added cases	16.0		-1.4		-7.2		-7.3	
Mean % of 51 cases added	31.4%		-2.7%		-14.1%		-14.3%	
% improvement compared with M alone	63%		-4%		-25%		-43%	

M mammography, M+A mammography plus automated whole-breast ultrasound (AWBU)

<sup>a</sup>Reader # presented by best to worst performance based on sensitivity on M+A

of cancer cases; specificity = number of TN/number of non-cancer cases; positive predictive value (PPV) = number of cancer cases/(number of TP + FP cases); and negative predictive value (NPV) = number of non-cancer cases/(number of FN + TN). Random-effect logistic regression models were used to test whether each binary measure differed significantly between mammography plus AWBU versus mammography alone. To account for the MRMC framework, we included random effects for readers and cases similar to the DBM model [15].

*Accuracy* was assessed through BIRADS ratings and DMIST likelihood scores, comparing two commonly used indicators of accuracy between mammography plus AWBU versus mammography alone: areas under the curve (AUC) and figures of merit (FOM). The FOM incorporates information from each reader on the region of suspected malignancy, as well as their confidence level in the finding, incorporated in an AUC. Because it includes both confidence level and location accuracy, the FOM is more powerful than AUC in detecting differences between techniques. We include both analyses, as described below:

Areas under the curve (AUC) were estimated in DBM MRMC 2.1 [15] (available from <http://perception.radiology.uiowa.edu>) using the trapezoidal/Wilcoxon method.

Readers and patients were treated as random factors. We also present reader-averaged receiver operating characteristic (ROC) curves; average values were calculated from separate ROC analyses conducted on each reader in the PROC LOGISTIC procedure.

Figures of merit (FOM) were estimated by using jackknife alternative free-response receiver operating characteristic methodology as implemented in JAFROC Version 1.0 [16] (available from <http://www.devchakra.com>). The FOM is defined as the probability that a cancer on an abnormal image is scored higher than a falsely marked location on a normal image and is analogous to the ROC curve; a higher FOM indicates improvement in reader performance.

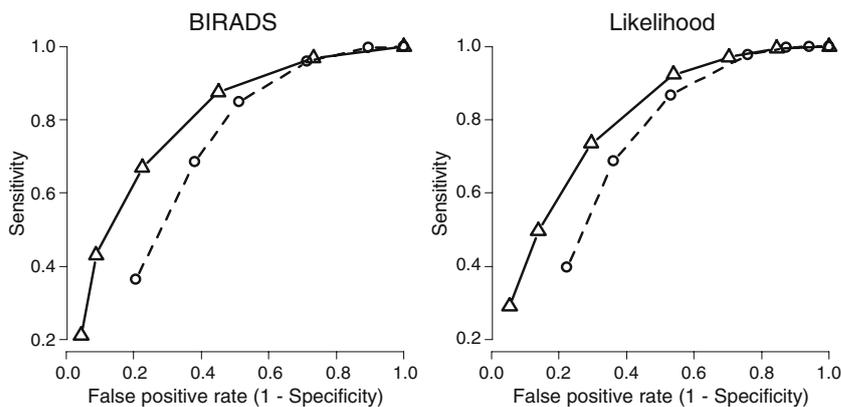
*Confidence in identification of cases for further imaging* We used linear regression, comparing BIRADS ratings and DMIST likelihood scores across the two screening techniques among TP cases; mean ratings and scores are estimated by the regression for each screening technique. To account for the MRMC framework, we included random effects similar to the DBM model [15]; the model included a fixed effect for technique, classified as mammography plus AWBU or mammography alone, and random effects for readers and cases.

**Table 3** Reader performance with 45 invasive cases

	≤1 cm		>1 to ≤2 cm		>2 cm		Total	
	#	%	#	%	#	%	#	%
# of cancers	17	100	22	100	6	100	45	100
Mean cancers by mammography	4.4	26	13.5	61	3.0	50	20.9	46
Mean additional cancers by AWBU	6.7	39	6.6	30	2.0	33	15.3	34
Mean total cases detected	11.1	65	20.1	91	5.0	83	36.2	80
% improvement compared to mammography alone	151%		49%		67%		73%	

For cases with more than one invasive tumor, the larger of the two was used. For interval cancers after imaging, size is the greatest diameter of the tumor seen retrospectively on the AWBU or mammogram, otherwise the diameter is that reported by pathological diagnosis

**Fig. 1** Receiver operating characteristic curves averaged across 12 readers for mammography alone (circles and dashed line) and mammography plus AWBU (triangles and solid line)



## Results

*Sample* Subjects averaged 59.4 years of age (SD = 10.2; range = 41–83). The 51 cancer patients and 51 normal subjects were well-matched with an insignificant mean difference of 31.0 days in age between abnormal and normal cases ( $t$  test=1.47,  $df=50$ ,  $p=0.15$ ). Table 1 lists the types and size of cancers in the test set.

*Identification of cases for further imaging* Table 2 details individual performance in the identification of cancer cases for further imaging. Mean sensitivity increased from 50% to 81%, an improvement of 63% in the number of cancer cases identified (25.4 vs. 41.4,  $F(1, 1,161)=165.95$ ,  $p<0.001$ ). Specificity (60–58%; 30.7 vs. 29.1,  $F(1, 1,161)=1.11$ ,  $p=0.29$ ), PPV (mean=47–67%;  $F(1, 1,297)=0.02$ ,  $p=0.89$ ), and NPV (mean=65–75%;  $F(1, 933)=0.61$ ,  $p=0.44$ ) did not change significantly with the addition of AWBU.

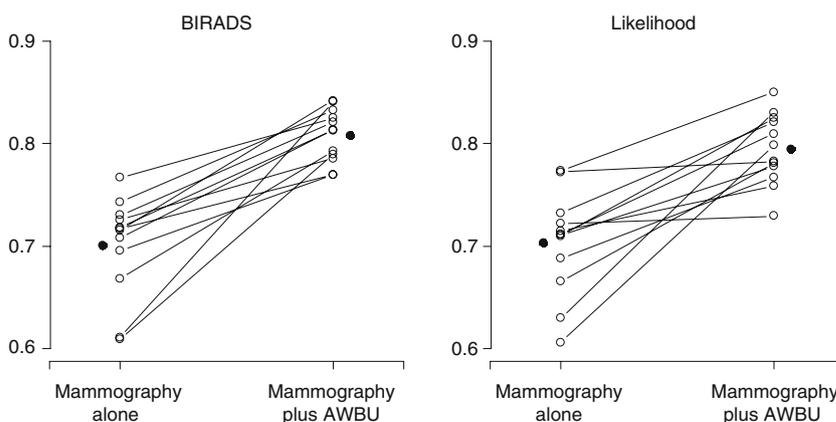
Individual success varied from 11 to 24 more cancer cases detected by AWBU. As a percentage of the cancers detected with mammography the range in improvement was 42–150%. Not only did all readers find more cancers individually, but all found 16–29% more cancers than the best mammography reader did with mammography alone. For the best performing mammography reader the cancer detections added by AWBU was predictably lower, as

more cancers had already been identified with mammography. For the poorest performer on mammography, the addition of AWBU resulted in a 150% improvement, bringing his overall cancer detection rate near the average for the group.

Table 3 shows the average reader performance by tumor size for the 45 image sets of patients with invasive cancer. The greatest percentage increase was for cancers 1 cm and under. This is due largely to the relatively poor performance at detecting these cancers with mammography, where only 26% of cases were correctly identified. By adding AWBU, the detection of these small cancers was increased to 65%.

*Accuracy* The ROC area was greater for mammography plus AWBU for both BIRADS (0.808 versus 0.701;  $F(1, 123) = 14.79$ ,  $p<0.001$ ) and likelihood scores (0.810 versus 0.703;  $F(1, 85) = 17.88$ ,  $p<0.001$ ) as estimated by multi-reader multi-case analyses. This is highlighted in Fig. 1 by ROC curves that are generated by averaging the results of separate ROC analyses for each reader. The BIRADS and likelihood AUC curves for mammography and mammography plus AWBU in both cases almost superimpose when confidence in malignancy by mammography is high, but when confidence in malignancy by mammography is low, as in the lower portions of the

**Fig. 2** Changes in areas under the receiver operating characteristic curve(s) for each reader (hollow circles) and averaged across 12 readers (solid circles)



---

graphs, the curves in both cases diverge significantly. In both cases the mammography plus AWBU approaches the y-axis indicating better cancer recognition.

Figure 2 shows the areas under the ROC curves for each reader and for the average of all readers as estimated by multi-reader multi-case analyses. These individual line graphs mirror the improvement in reader performance shown in Table 2.

Similar to ROC areas, the figures of merit (FOM) were higher for mammography plus AWBU across all readers, compared with mammography alone using both the BIRADS scores (0.786 versus 0.613;  $F(1, 270)=34.1$ ,  $p<0.001$ ) and DMIST likelihood scores (0.791 versus 0.614;  $F(1, 238) = 37.9$ ,  $p<0.001$ ) as accuracy indices.

*Confidence in identification of cases for further imaging* Readers reviewing cancer cases were more confident in correctly identifying cases for further imaging, i.e., TP reading, using mammography plus AWBU compared with mammography alone. On average, both BIRADS scores (mean = 4.8 versus 4.2,  $F(1, 740) = 81.91$ ,  $p<0.001$ ) and DMIST likelihood scores (mean = 4.8 versus 4.1,  $F(1, 740) = 82.21$ ,  $p<0.001$ ) were higher.

*Interpretation times* Average reading time per study for the 102 AWBUs was 7 min 58 s (7:58) varying from 5:54 to 12:51. The difference in review time was unrelated to the number of cancers identified by each reader (correlation = 0.02,  $p=0.96$ ).

---

## Discussion

Significant improvement in identification of asymptomatic cancers occurred for all readers in this study. This is shown by a 63% increase in callbacks of cancer cases with only a 4% decrease in correct identification of the true negative cases. The confidence of the diagnoses of the 102 cases with predictive BIRADS and DMIST likelihood scales was confirmed by using AUC and FOM methodology.

With a short training period experienced radiologists using 2D AWBU significantly improve their ability to diagnose cancer in dense-breasted women. This type of AWBU is similar in appearance to real-time ultrasound images. The slower transducer speed enforced by the AWBU decreases inter-image distance, allowing the reader more time to identify small masses. At a review speed of 10 images per second the observer has 0.5 s to identify a 5-mm mass. A high-resolution computer screen, along with a post-processing technique to expand the grayscale at the black end of the spectrum, results in visually sharper margins and more contrast of masses against the background tissue. These factors are designed to make recognition of invasive cancers easier and more reliable. This automated process for breast ultrasound

eliminates operator variability, provides greater consistency, and ensures reproducibility of quality images. Study radiologists increased discovery of T1a and T1b invasive cancers 150% over mammography alone (Table 3).

The average review time per AWBU study was about 11 min shorter than the 19 min for radiologists in the ACRIN 6666 trial of handheld screening ultrasound [8]. As half of our test set subjects had cancers, it would be expected that the average review time we observed for AWBU would be significantly longer than in a typical screening population with mostly normal studies.

Our study had a number of inherent weaknesses. Although the test set was confidential, the readers probably quickly realized that it was enriched. They may have been extraordinarily vigilant resulting in increases in both TPs and FPs.

A false increase in TPs would occur if all the correctly identified cancers were not subsequently confirmed with biopsy. Also, analysis was performed on a case basis in the three patients in whom cancers were present in both breasts; it was assumed if one of the cancers was identified, the cancer in the other breast would be found by the subsequent workup. This assumption might have falsely raised the TPs and reduced the FNs. In addition, we did not have a comparison with hand-held ultrasound.

Any of the following factors could have decreased the readers' accuracy with AWBU (decreased true positives and negatives, and increased false positives and negatives) compared with a normal screening situation.

1. Fatigue—Readers reviewed an average of 34 AWBUs daily.
2. Inexperience with ultrasound screening—Some readers do not perform screening ultrasound.
3. Limited experience with AWBU—This was the first exposure to AWBU for 11 of the readers.
4. Unfamiliarity with some ultrasound formats—Images from many different manufacturers were used.

In spite of these hindrances our observations clearly show that radiologists improve detection of cancers, especially small invasive ones, by adding AWBU to mammography findings.

---

## Conclusion

This article demonstrates that experienced breast radiologists can learn to interpret 2D AWBU quickly. Radiologists will significantly improve their cancer detection rates in dense-breasted women by adding AWBU to mammography. This procedure has the potential for both standardizing the performance of whole-breast ultrasound and shortening the time required for radiologists.

**Acknowledgments** We thank the radiologists participating in this trial: Catherine Babcock, Debra Copit, Ruth English, Jon Fish, Thalia B. Forte, Michael N. Linver, Laleh Lourie, Carrie Morrisson, Susan Roux, Thomas Stavros, Lakshmi Tegulapalle, and Richard Vanesian.

consultants for the study and have no conflict of interest relevant to Sonocine, Inc. The study was funded by Sonocine, Inc.

**Financial disclosures** Dr. Kelly is the Majority Stockholder of Sonocine, Inc. Dr. Dean owns stock in Sonocine, Inc. Neither author has received any form of payment from the company. Drs. Lee and Comulada served as statistical

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Appendix A

**Screening Mammogram – SonoCiné Data Form**

Reader Number: \_\_\_\_\_ Case Number: \_\_\_\_\_ Date: \_\_\_/\_\_\_/\_\_\_

Mammogram: Usual Follow-up Short-term Follow-up Callback

BIRADS: 1 2 3 4a 4b 4c 5

Likelihood of Cancer: \_\_\_\_\_ 1 - not ca, 2 - almost def not ca, 3- prob not ca, 4 – may be ca, 5 - prob ca, 6 – almost def ca, 7 - def ca

Mark only lesions with requiring follow up or callback to rule in or out malignancy:

Lesion #1: Breast: R L Location: \_\_\_\_\_ o'clock Axillary Diffuse/Multiple

Lesion #2: Breast: R L Location: \_\_\_\_\_ o'clock Axillary Diffuse/Multiple

Lesion #3: Breast: R L Location: \_\_\_\_\_ o'clock Axillary Diffuse/Multiple

Time in Seconds: Mammography \_\_\_\_\_ SonoCiné \_\_\_\_\_

Mammogram & SonoCiné: Change \_\_\_\_\_ No Change \_\_\_\_\_ from Mammogram reading

SonoCiné: Usual Follow-up Short-term Follow-up Callback

BIRADS: 1 2 3 4a 4b 4c 5

Likelihood of Cancer: \_\_\_\_\_ 1 - not ca, 2 - almost def not ca, 3- prob not ca, 4 – may be ca, 5 - prob ca, 6 – almost def ca, 7 - def ca

Mark only lesions with requiring follow up or callback to rule in or out malignancy:

Lesion #1: Breast: R L Location: \_\_\_\_\_ o'clock Axillary Diffuse/Multiple

Lesion #2: Breast: R L Location: \_\_\_\_\_ o'clock Axillary Diffuse/Multiple

Lesion #3: Breast: R L Location: \_\_\_\_\_ o'clock Axillary Diffuse/Multiple

Note: \_\_\_\_\_

---

## References

1. Tabár L, Fagerberg CJ, Gad A et al (1985) Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet* 325:829–832
2. Shapiro S, Venet W, Strax P, Venet L, Roeser R (1982) Ten to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst* 69:349–355
3. Mendelson EB, Baum JK, Berg WA, Merritt CRB, Rubin E (2003) Breast imaging reporting and data system, BI-RADS: ultrasound, 1st edn. American College of Radiology, Reston, VA
4. Boyd NF, Guo H, Martin LJ et al (2007) Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 356:227–236
5. Kriege M, Brekelmans CT, Boetes C et al (2004) Efficacy of MRI and mammography for breast cancer screening in women with a familial or genetic predisposition. *N Engl J Med* 351:427–500
6. Saslow D, Boetes C, Burke W et al (2007) American cancer society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 57:75–89
7. Leach MO, Boggis CR, Dixon AK et al (2005) Screening with magnetic imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). *Lancet* 365:1769–1778
8. Berg WA, Blume JD, Cormack JB et al (2008) Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA* 299:2151–2163
9. Kelly KM, Dean J, Comulada WS, Lee SJ (2010) Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *Eur Radiol* 20:734–742
10. Ciatto S, Roselli Del Turco M, Risso G et al (2003) Comparison of standard reading and computer aided detection (CAD) on a national proficiency test of screening mammography. *Eur J Radiol* 45:135–138
11. D'Orsi CJ, Bassett LW, Berg WA et al (2003) Breast imaging reporting and data system, BI-RADS: mammography, 4th edn. American College of Radiology, Reston, VA
12. Pisano ED, Gastonis C, Hendrick E et al (2005) Diagnostic performance of digital versus film mammography for breast cancer screening. *N Engl J Med* 353:1–11
13. Wagner RF, Metz CE, Campbell G (2007) Assessment of medical imaging systems and computer aids: a tutorial review. *Acad Radiol* 14:723–748
14. Obuchowski NA (1995) Multi-reader ROC studies: a comparison of study designs. *Acad Radiol* 2:709–716
15. Dorfman DD, Berbaum KS, Metz CE (1992) Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. *Invest Radiol* 27:723–731
16. Chakraborty DP, Berbaum KS (2004) Observer studies involving detection and localization: modeling, analysis and validation. *Med Phys* 31:2313–2330