Radiology

Performance and Reading Time of Automated Breast US with or without Computer-aided Detection

Shanling Yang, MD • Xican Gao, MD • Liwen Liu, PhD, MD • Rui Shu, MD • Jingru Yan, MD • Ge Zhang, MD • Yao Xiao, MD • Yan Ju, MS • Ni Zhao, MD • Hongping Song, PhD, MD

From the Department of Ultrasonic Medicine, Xijing Hospital of the Fourth Military Medical University, No. 127 Changle West Road, Xi'an, Shaanxi, China 710032. Received August 2, 2018; revision requested September 11; revision received April 22, 2019; accepted April 30. **Address correspondence to** H.S. (e-mail: *song.hp@foxmail.com*).

Study supported by National Natural Science Foundation of China (81401416), Shaanxi International Scientific and Technological Cooperation and Exchange Program (2017KW-057), 2016 Key Project of Clinical High and New Technology of Xijing Hospital, and the 2017 Shaanxi Scientific Research Projects in Higher Education (XGH17281).

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

See also the editorial by Slanetz in this issue.

Radiology 2019; 292:540–549 • https://doi.org/10.1148/radiol.2019181816 • Content code: BR

Background: Computer-aided detection (CAD) systems may be used to help radiologists interpret automated breast (AB) US images. However, the optimal use of CAD with AB US has, to the knowledge of the authors, not been determined.

Purpose: To compare the performance and reading time of different readers by using AB US CAD system to detect breast cancer in different reading modes.

Materials and Methods: In this retrospective study, 1485 AB US images (282 with malignant lesions, 695 with benign lesions, and 508 healthy) in 1452 women (mean age, 43.7 years; age range, 19–82 years) including 529 (36.4%) women who were asymptomatic were collected between 2016 and 2017. A CAD system was used to interpret the images. Three novice readers with 1–3 years of US experience and three experienced readers with 5–10 years of US experience were assigned to read AB US images without CAD, at a second reading (after the reader completed a full unaided interpretation), and at concurrent reading (use of CAD at the start of the assessment). Diagnostic performances and reading times were compared by using analysis of variance.

Results: For all readers, the mean area under the receiver operating characteristic curve improved from 0.88 (95% confidence interval [CI]: 0.85, 0.91) at without-CAD mode to 0.91 (95% CI: 0.89, 0.92; $P < .001$) at the second-reading mode and 0.90 (95% CI: 0.89, 0.92; $P = .002$) at the concurrent-reading mode. The mean sensitivity of novice readers in women who were asymptomatic improved from 67% (95% CI: 63%, 74%) at without-CAD mode to 88% (95% CI: 84%, 89%) at both the second-reading mode and the concurrent-reading mode ($P = .003$). Compared with the without-CAD and second-reading modes, the mean reading time per volume of concurrent reading was 16 seconds (95% CI: 11, 22; $P < .001$) and 27 seconds (95% CI: 21, 32; $P < .001$) shorter, respectively.

Conclusion: Computer-aided detection (CAD) was helpful for novice readers to improve cancer detection at automated breast US in women who were asymptomatic. CAD was more efficient when used concurrently for all readers.

Online supplemental material is available for this article.

Use of mammography has been shown in randomized controlled trials (1,2) to reduce breast cancer mortality. However, the sensitivity of mammography is poor in women with dense fibroglandular tissue (3,4). As a supplement to mammography, handheld US and automated breast (AB) US can help to improve detection of cancer (5–8). Handheld US is operator dependent, and therefore the experience of the operator is important. At AB US, full coverage of each breast consists of two to five acquisitions (8,9). Each acquisition contains more than 300 transverse images and reconstructions in coronal and sagittal planes, forming a volume of US images. The high-resolution AB US volume acquires up to 3000 two-dimensional images per woman, and interpretation is longer than that of a conventional US image (8). In addition, the probability of overlooking subtle lesions may be substantial in women who are asymptomatic (9). Thus, a commercially developed computer-aided detection (CAD) system may be used to help radiologists interpret AB US images (10).

A CAD system can be used either as a second or concurrent reader (11–13). The classic CAD implementation is second-reading mode at which CAD is applied after the reader has completed a full, unaided assessment (14). However, such implementation increases interpretation time. A potentially more efficient paradigm is concurrent-reading mode, in which CAD is applied at the start of the assessment. However, concurrent application of CAD may reduce reader's vigilance, reducing sensitivity (12). It is also contrary to the recommendations for CAD at mammography, at which the CAD display of prompts should be displayed only after the radiologist has completed their initial assessment. In addition, the usefulness of applying CAD to

Abbreviations

AB = automated breast, AUC = area under the receiver operating characteristic curve, BI-RADS = Breast Imaging Reporting and Data System, CAD = computer-aided detection, CI = confidence interval

Summary

Both the second-reading mode and concurrent-reading mode have the potential to improve novice readers' performance for breast cancer screening on automated breast US images.

Key Points

- By using computer-aided detection (CAD) for interpretation of automated breast US images, the mean sensitivity of novice readers in women who are asymptomatic improved from 67% at without-CAD mode to 88% at both the second-reading and the concurrent-reading modes (P = .003).
- By using CAD at the concurrent-reading mode, all readers could save 32% (16 seconds per 50 seconds per volume) of the reading time with higher area under the receiver operating characteristic curve values compared with without-CAD mode.

handheld breast US depends on the operator's experience with breast imaging. CAD was more useful for less experienced readers at handheld US (15,16). It is unclear how CAD for AB US will be incorporated into the clinical workflow. Therefore, the purpose of our study was to compare performance and reading time of different readers by using AB US CAD to detect breast cancer in different reading modes.

Materials and Methods

Our retrospective study was approved by the Xijing Hospital institutional review board and the requirement for informed consent was waived. QView Medical (Los Altos, Calif) supported this research by providing a CAD workstation.

Patient Selection

We performed a retrospective search for all consecutive women with AB US examinations in a tertiary care center in northwest China between February 2016 and February 2017. An initial search from our breast US database found 1809 women (672 women who underwent screening and 1137 women who underwent diagnostic examinations). AB US has been routinely used at our institution for screening and diagnostic examinations since 2015. For screening, AB US was performed first. If a lesion was suspicious for malignancy, handheld US was performed for further confirmation. For diagnostic examinations, AB US and handheld US of the ipsilateral breast and contralateral breast were performed simultaneously. Of the 1809 women, we included 1452 women (mean age, 43.7 years; age range, 19–82 years) in our study (Fig 1, Appendix E1 [online]).

AB US Image Acquisition

Images were obtained with AB US (Invenia; GE Healthcare, Chicago, Ill). This system automatically images the breast by using a large (15.4-cm), 6- to 15-MHz bandwidth linear-array transducer at 10-MHz center frequency and produces a three-dimensional B-scan US image volume of 15.4 × 16.9 × 5.0 cm. All AB US examinations were performed by professionally trained technicians with at least 1 year of experience. To ensure full coverage, two to four US volumes per breast were performed at predefined locations (8,9). In our study, 80.2% (1165 of 1452) of the women had six volumes, 14.8% (215 of 1452) had four volumes, and 5.0% (72 of 1452) had eight volumes. The images were anonymized and transmitted via the network to the CAD workstation.

CAD System

Our CAD system extracts features by using an ensemble of artificial neural-network classifiers. It is designed to aid radiologists in their search for areas suspicious for cancer on the AB US images (Fig 2). In addition, the CAD system can enhance areas possibly suspicious for malignancy by providing a minimum intensity projection of the breast tissue on a three-dimensional AB US volume acquisition that can be used for rapid navigation through AB US scans (10). The sensitivity of the instrument can be adjusted, with greater sensitivity yielding a higher potential rate of false-positive findings. We chose the default setting of one false-positive CAD mark per AB US volume. To test the reproducibility of CAD marks with the same image, we randomly selected 5.0% (72 of 1452) of examinations and sent them through the CAD system three times, and the results showed that marking was consistent in all images.

Readers, Reading Modes, and Training

Six radiologists who had performed more than 500 AB US examinations in the last year participated as readers in our study. Experience with the general US for readers one to six was 1, 2, 3, 5, 6, and 10 years and specifically with breast imaging (performing and interpreting at least 2500 examinations of breast US per year) was 1, 1, 2, 3, 3, and 4 years, respectively. Because 2 or 3 years of fellowship training is usually required in China, we divided the six readers into two groups: the novice group (G.Z., J.Y., and N.Z., with 1–3 years of US experience and ≤2 years of experience with breast US) and the experienced group (S.Y., R.S., and H.S., with 5–10 years of US experience and ≥3 years of experience with breast US). All readers were trained on the reading procedures with 30 AB US images that were not part of the study set, of which 10 were read by using without-CAD mode, 10 by using second-reading mode, and 10 by using concurrent-reading mode. In second-reading mode, readers read AB US images without CAD first, then combined the indications of CAD marks to make the final decision. In concurrent-reading mode, readers identified CAD marks first, then quickly browsed the entire AB US examination.

Study Design and Data Analysis

The task was performed in two reading sessions at an interval of 4 weeks apart. In the first session, half of the data sets were read without CAD, then with CAD in the second-reading mode; the remaining data sets were read in the concurrent-reading mode. All data sets were presented to readers in randomized order, and orders were different for every reader. Four weeks later, in the second session, data sets were given to readers with orders different than those used in the first session (Fig 1).

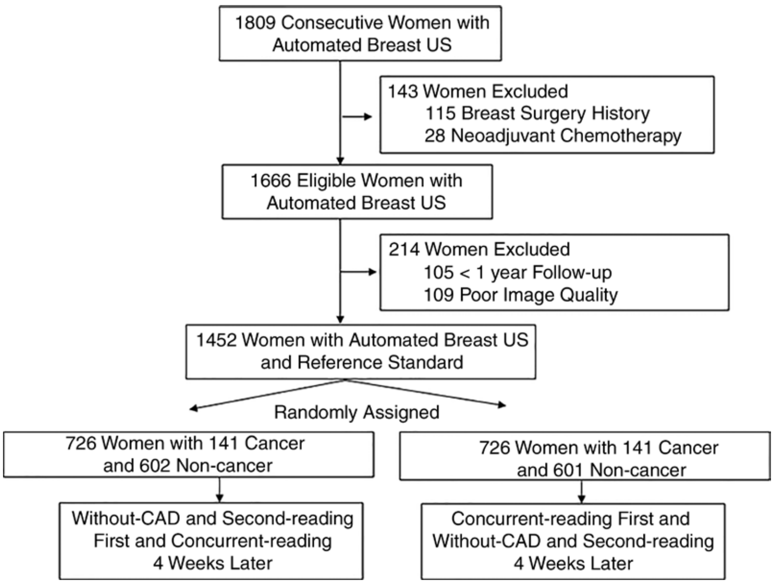


Figure 1: Flowchart of patient selection and randomization for reader study. CAD = computer-aided detection.

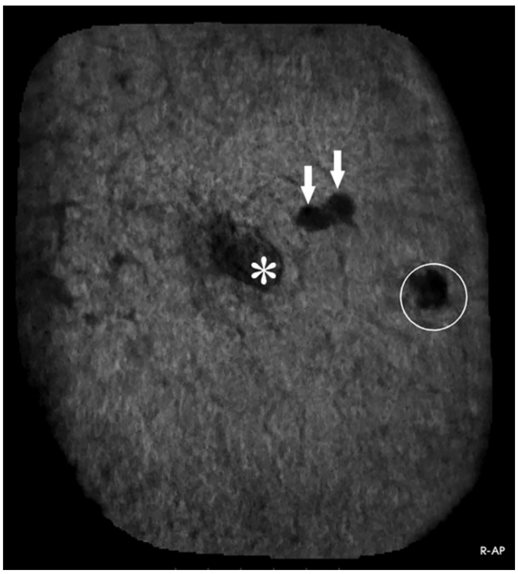


Figure 2: Screen-capture image shows computer monitor display of computer-aided detection (CAD) system output in a 41-year-old woman. Minimum intensity projection image in right anteroposterior view shows a CAD mark (circle) on the mass, indicating a lesion suspicious for cancer that was proved to be an invasive carcinoma at pathologic analysis. However, two cysts (arrows) near the nipple (*) were not marked by the CAD system. R-AP = right anteroposterior.

All readers reviewed every examination at each reading mode and were blinded to any information about the women, including the age, manifestation of symptoms, and previous radiology and histopathologic reports. The readers were asked to read at least 3 hours per day to simulate batch reading of examinations.

At each reading mode, the readers were instructed to mark lesions and subsequently determine a final Breast Imaging Reporting and Data System (BI-RADS) (17) assessment category. The decision regarding whether each CAD marking highlighted a previously overlooked lesion or only a false-positive finding depended mainly on the doctor's experience and the image characteristics. In general, we allowed BI-RADS category 1 (negative), 2 (benign), or 3 (probably benign) assessment by readers to be upgraded to 4 (suspicious), and BI-RADS category 4 or 3 assessment could be downgraded to 3 or 2 by addition of CAD. However, BI-RADS category 5 (highly suggestive of malignancy) assessment could not be downgraded. Because a quasicontinuous linear scale is required to perform receiver operating characteristic analysis (18), readers were also asked to provide a level-of-suspiciousness score between 0 and 100. When analyzing the management decision changes of readers, we defined BI-RADS category 1, 2, and 3 as negative for cancer and recommended for follow-up and BI-RADS category 4 and 5 as positive for cancer and recommended for biopsy. The reference standard for each AB US image was established by a radiologist with access to the case report and histopathologic reports.

Reading time, measured by a software plug-in (QView Medical) integrated in the CAD system, started when the reader opened images and ended when the reader closed the images. For second-reading mode, we counted the reading time from the beginning of the reading without CAD to the end of the reading with CAD. Reading time only included the time to read images, and the timing device was not visible to the reader. Because volumes of each patient varied from four to eight, we recorded reading time per volume in each reading modes for comparison.

Statistical Analysis

Multireader multicase analysis was performed by using the alternative free-response receiver operating characteristic method (19). This method uses jackknifing and two-way analysis of variance to compare the mean area under the receiver operating characteristic curve (AUC), which was used for the three reading modes in our study. Sensitivity and specificity in the three reading modes for each reader were calculated on the basis of the final BI-RADS assessment data with BI-RADS category 3 and BI-RADS category 4 as two different thresholds (20). For reading time analysis, outliers, defined as a distance greater than 1.5 times the interquartile range below the first quartile or above the third quartile, were removed. These were considered unreliable because readers might have been interrupted (21). The mean AUC, sensitivity, specificity, and reading time with 95% confidence intervals (CIs) were calculated. Management decision changes by using CAD were classified into no change, downgraded, and upgraded, and their mean number was calculated. Detailed statistical methods are in Appendix E1 (online).

The receiver operating characteristic analyses were performed by using software (Jackknife Alternative Free-Response Receiver-Operating Characteristic, version 4.2.1; *http://www.devchakraborty.com*). All other analyses were performed with statistical software (SPSS, version 19.0; IBM Statistics,

Table 1: Patient and Lesion Characteristics

Characteristic	All Women (n = 1452)	Asymptomatic Women (n = 529)	Symptomatic Women (n = 923)	P Value
Patient age (y)				
Mean	43.7 ± 18.8	45.9 ± 17.0	42.1 ± 19.2	.07
Median*	44 (19–82)	46 (28–69)	41 (19–82)	
Menopausal status				.19
Premenopause	690 (48)	239 (45)	451 (49)	
Postmenopause	762 (52)	290 (55)	472 (51)	
Biopsy or surgery				<.001
Yes	590 (41)	104 (20)	486 (53)	
No	862 (59)	425 (80)	437 (47)	
Lesion size at AB US (cm)				
All				<.001
Mean	2.5 ± 1.9	1.1 ± 0.5	2.6 ± 1.7	
Median*	2.5 (0.4–6.4)	1.0 (0.4–1.8)	2.6 (0.8–6.4)	
Malignant				<.001
Mean	2.7 ± 1.8	1.2 ± 0.5	2.7 ± 1.7	
Median*	2.6 (0.4–6.4)	1.0 (0.4–1.8)	2.6 (0.8–6.4)	
Benign				<.001
Mean	2.2 ± 1.8	1.3 ± 0.4	2.5 ± 1.6	
Median*	2.1 (0.5–5.8)	1.2 (0.5–1.7)	2.4 (0.8–5.8)	
Histologic type†				
Malignant	282	19	263	.81
Invasive carcinoma	223	16	207	
Invasive carcinoma of no specific type	209 (74)	16 (84)	193 (73)	
Invasive lobular carcinoma	8 (3)	0 (0)	8 (3)	
Mucinous carcinoma	6 (2)	0 (0)	6 (2)	
Ductal carcinoma in situ	58 (21)	3 (16)	55 (21)	
Malignant phyllodes tumor	1 (0)	0 (0)	1 (0)	
Benign	320	35	285	.73
Fibroadenoma	115 (36)	16 (46)	99 (35)	
Intraductal papilloma	45 (14)	5 (14)	40 (14)	
Inflammation	37 (12)	1 (3)	36 (13)	
Usual ductal hyperplasia	29 (9)	4 (11)	25 (9)	
Adenosis	25 (8)	3 (9)	22 (8)	
Cysts	16 (5)	1 (3)	15 (5)	
Benign phyllodes tumor	6 (2)	0 (0)	6 (2)	
Lipoma	2 (1)	0 (0)	2 (1)	
Other	45 (14)	5 (14)	40 (14)	

Note.—Unless otherwise specified, data are numbers of patients, with percentages in parentheses. Mean data are ± standard deviation. P values are a comparison between women with and without symptoms. They were calculated by using Student t test or the χ^2 test. AB = automated breast.

* Data in parentheses are range.

† Data are numbers of lesions, with percentages in parentheses. Results of 590 women with malignant (n = 282) or benign (n = 320) lesions at biopsy or surgery. Twelve women had two malignant lesions.

Armonk, NY). In all analyses, a P value less than .05 was considered to indicate a statistically significant difference.

Results

Patient Characteristics

Of the 1452 women, 529 (36.4%) were asymptomatic and 923 (63.6%) had symptoms in the breast such as palpable mass, pain, or nipple discharge. In the asymptomatic group, there were 19 women with 19 malignant lesions, 85 women with 85 benign lesions, and 425 healthy women. In the symptomatic group, there were 251 women with 263 malignant lesions, 589 women with 610 benign lesions, and 83 healthy women. Mean size for all lesions, malignant lesions, and benign lesions at AB US was 2.5 cm ± 1.9 (standard deviation; range, 0.4–6.4 cm), 2.7 cm ± 1.8 (range, 0.4–6.4 cm; n = 282), and 2.2 cm ± 1.8 (range, 0.5–5.8 cm; n = 695), respectively. Patient and lesion characteristics on the basis of manifestation of symptoms in women are summarized in Table 1.

Table 2: Diagnostic Performance of Readers in Three Reading Modes with a BI-RADS Category 3 Threshold

	Sensitivity (%)			Specificity (%)			AUC		
Parameter	Without-CAD Mode	Second-reading Mode	Concurrent-reading Mode	Without-CAD Mode	Second-reading Mode	Concurrent-reading Mode	Without-CAD Mode	Second-reading Mode	Concurrent-reading Mode
Asymptomatic women									
Novice readers	67 (63, 74)	88 (84, 89)	88 (84, 89)	92 (91, 92)	91 (91, 92)	91 (91, 92)	0.84 (0.79, 0.87)	0.95 (0.95, 0.95)	0.93 (0.92, 0.95)
P value	.003	>.99	.003	>.99	>.99	>.99	.014	>.99	.04
Experienced readers	89 (84, 95)	93 (89, 95)	93 (89, 95)	92 (92, 93)	92 (92, 93)	92 (92, 93)	0.95 (0.93, 0.97)	0.97 (0.95, 0.99)	0.97 (0.95, 0.98)
P value	.95	>.99	.95	>.99	>.99	>.99	.66	>.99	.9
All readers	78 (67, 89)	90 (87, 93)	90 (87, 93)	92 (91, 92)	92 (91, 92)	92 (91, 92)	0.90 (0.84, 0.95)	0.96 (0.95, 0.98)	0.95 (0.92, 0.97)
P value	.001	>.99	.001	>.99	>.99	>.99	.003	>.99	.011
Symptomatic women									
Novice readers	88 (87, 89)	94 (93, 95)	94 (92, 95)	33 (30, 35)	33 (31, 35)	33 (30, 35)	0.79 (0.78, 0.81)	0.85 (0.84, 0.86)	0.84 (0.83, 0.84)
P value	.002	>.99	.004	>.99	>.99	>.99	.003	.95	.008
Experienced readers	93 (92, 95)	94 (93, 95)	94 (93, 95)	42 (38, 47)	42 (38, 47)	42 (38, 47)	0.89 (0.87, 0.90)	0.90 (0.89, 0.91)	0.90 (0.89, 0.91)
P value	>.99	>.99	>.99	>.99	>.99	>.99	>.99	>.99	>.99
All readers	91 (89, 93)	94 (93, 95)	94 (93, 94)	37 (33, 42)	37 (33, 43)	38 (33, 43)	0.84 (0.79, 0.88)	0.87 (0.85, 0.90)	0.87 (0.84, 0.89)
P value	.001	>.99	.002	>.99	>.99	>.99	.005	>.99	.017
All women									
Novice readers	87 (85, 88)	94 (92, 94)	93 (92, 94)	58 (56, 59)	57 (56, 59)	58 (56, 59)	0.85 (0.84, 0.85)	0.89 (0.89, 0.90)	0.88 (0.88, 0.89)
P value	.002	>.99	.002	>.99	>.99	>.99	<.001	.48	.001
Experienced readers	93 (92, 94)	94 (93, 95)	94 (93, 94)	63 (61, 66)	63 (61, 66)	63 (61, 66)	0.91 (0.90, 0.93)	0.92 (0.92, 0.94)	0.92 (0.92, 0.93)
P value	.51	>.99	.77	>.99	>.99	>.99	.84	>.99	.96
All readers	90 (87, 93)	94 (93, 94)	93 (93, 94)	60 (58, 63)	60 (58, 64)	61 (58, 64)	0.88 (0.85, 0.91)	0.91 (0.89, 0.92)	0.90 (0.89, 0.92)
P value	<.001	>.99	<.001	>.99	>.99	>.99	<.001	>.99	.002

Note.—Unless otherwise indicated, data are mean and data in parentheses are 95% confidence intervals. Breast Imaging Reporting and Data System (BI-RADS) assessment category 3, 4, and 5 were considered positive for cancer for calculation of sensitivity and specificity. P values in the without-CAD mode column were calculated by comparing with second-reading mode. P values in the second-reading mode column were calculated by comparing with concurrent-reading mode. P values in the concurrent-reading mode column were calculated by comparing with without-CAD mode. AUC = area under the receiver operating characteristic curve, CAD = computer-aided detection.

Reader Performances

The mean AUC of the reading was improved from 0.88 (95% CI: 0.85, 0.91) without CAD to 0.91 (95% CI: 0.89, 0.92; $P < .001$) at the second-reading mode and 0.90 (95% CI: 0.89, 0.92; $P = .002$) at the concurrent-reading mode for all readers (Table 2). For novice readers, the improvements of mean AUCs were significant at both the second-reading mode (mean AUC, without CAD vs second-reading mode, 0.85 [95% CI: 0.84, 0.85] vs 0.89 [95% CI: 0.89, 0.90], respectively; $P < .001$) and concurrent-reading mode (mean AUC, without CAD vs concurrent-reading mode, 0.85 [95% CI: 0.84, 0.85] vs 0.88 [95% CI: 0.88, 0.89], respectively; $P = .001$) (Fig 3). For experienced readers, there were no differences in AUC among the three reading modes.

When a BI-RADS category 3 threshold was used, compared with the without-CAD mode, the improvements of mean sensitivity were significant at both the second-reading mode (87% [95% CI: 85%, 88%] vs 94% [95% CI: 92%, 94%], respectively; $P = .002$) and concurrent-reading mode (87% [95% CI: 85%, 88%] vs 93% [95% CI: 92%, 94%], respectively; $P = .002$) for novice readers, but they were not significant for experienced readers (Table 2). When the second-reading mode and the concurrent-reading mode were compared, the mean sensitivity, specificity, and AUC were similar for both novice and experienced readers, and all readers ($P > .05$). The performances of each reader are in Table E1 (online). When subgroup analysis was performed, for novice and all readers, both second-reading mode and concurrent-reading mode had significantly greater

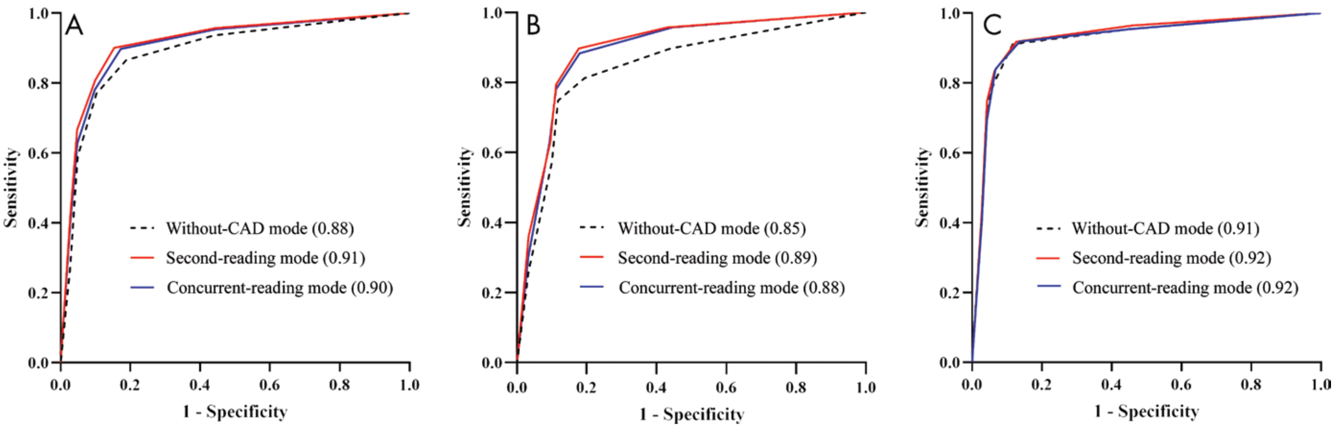


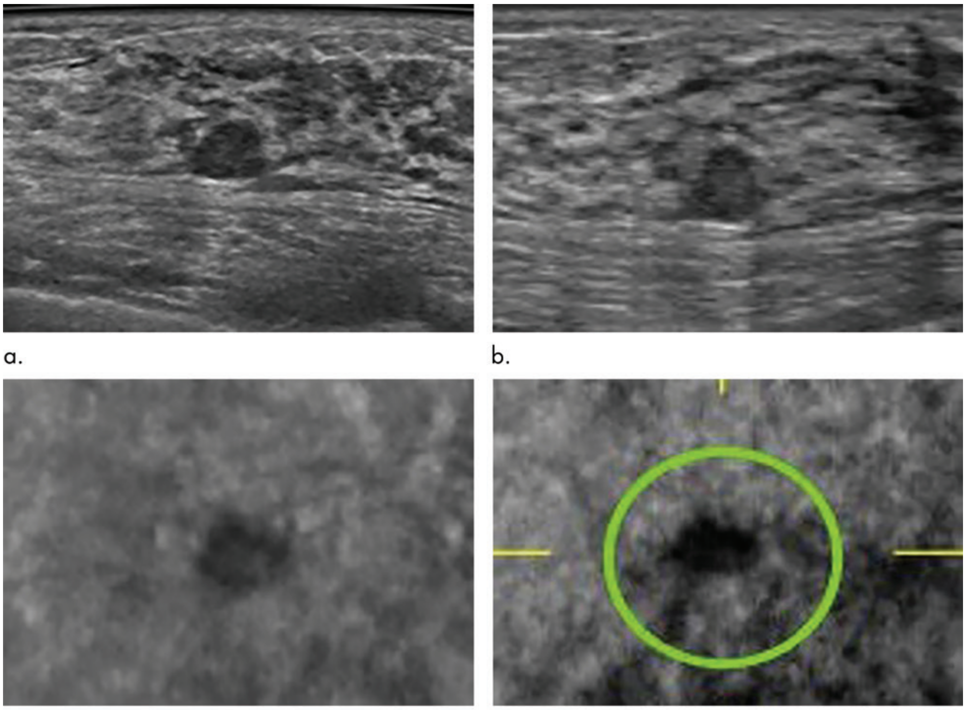
Figure 3: Receiver operating characteristic curves of three reading modes. Average is computed across, *A*, all readers, *B*, three novice readers, and, *C*, three experienced readers. Data in parentheses are areas under the receiver operating characteristic curve. CAD = computer-aided detection.

Table 3: Management Decision Changes of Readers with Computer-aided Detection System

Parameter	Second-reading Mode			Concurrent-reading Mode		
	No Change	Downgraded	Upgraded	No Change	Downgraded	Upgraded
Malignant lesions						
Asymptomatic women (*n* = 19)						
Novice readers	13.0 ± 1.0	0.0 ± 0.0	6.0 ± 1.0	12.3 ± 0.6	0.7 ± 0.6	6.0 ± 1.0
Experienced readers	17.0 ± 1.0	0.0 ± 0.0	2.0 ± 1.0	16.7 ± 1.2	0.3 ± 0.6	2.0 ± 1.0
All readers	15.0 ± 2.4	0.0 ± 0.0	4.0 ± 2.4	14.5 ± 2.5	0.5 ± 0.6	4.0 ± 2.4
Symptomatic women (*n* = 263)						
Novice readers	239.0 ± 5.0	1.3 ± 1.5	22.7 ± 3.5	238.3 ± 4.0	3.3 ± 2.3	21.3 ± 2.1
Experienced readers	259.3 ± 1.2	0.3 ± 0.6	3.3 ± 0.6	258.0 ± 2.7	1.7 ± 1.5	3.3 ± 1.2
All readers	249.2 ± 11.6	0.8 ± 1.2	13.0 ± 10.8	248.2 ± 11.2	2.5 ± 2.0	12.3 ± 10.0
All women (*n* = 282)						
Novice readers	252.0 ± 4.0	1.3 ± 1.5	28.7 ± 2.5	250.7 ± 3.5	4.0 ± 2.7	27.3 ± 1.2
Experienced readers	276.0 ± 0.6	0.3 ± 0.6	5.3 ± 0.6	274.7 ± 2.1	2.0 ± 1.0	5.3 ± 1.5
All readers	264.2 ± 13.6	0.8 ± 1.2	17.0 ± 12.9	262.7 ± 13.4	3.0 ± 2.1	16.3 ± 12.1
Benign lesions and healthy US*						
Asymptomatic women (*n* = 510)						
Novice readers	498.0 ± 2.0	10.0 ± 1.7	2.0 ± 1.0	498.7 ± 1.2	9.7 ± 1.5	1.7 ± 0.6
Experienced readers	496.3 ± 0.6	13.0 ± 1.0	0.7 ± 1.2	496.7 ± 1.2	12.3 ± 1.5	1.0 ± 1.0
All readers	497.2 ± 1.6	11.5 ± 2.1	1.3 ± 1.2	497.7 ± 1.5	11.0 ± 2.0	1.3 ± 0.8
Symptomatic women (*n* = 693)						
Novice readers	675.3 ± 3.2	15.3 ± 2.1	2.3 ± 1.2	675.7 ± 1.5	15.3 ± 2.1	2.0 ± 1.0
Experienced readers	672.3 ± 6.4	18.3 ± 7.5	2.3 ± 1.5	671.3 ± 8.3	19.7 ± 8.3	2.0 ± 0.0
All readers	673.8 ± 4.8	16.8 ± 5.2	2.3 ± 1.2	673.5 ± 5.9	17.5 ± 5.9	2.0 ± 0.6
All women (*n* = 1203)						
Novice readers	1173.3 ± 4.9	25.3 ± 3.8	4.3 ± 1.5	1174.3 ± 2.5	25.0 ± 3.6	3.7 ± 1.2
Experienced readers	1168.7 ± 5.9	31.3 ± 7.6	3.0 ± 1.7	1168.0 ± 7.2	32.0 ± 7.0	3.0 ± 1.0
All readers	1171.0 ± 5.5	28.3 ± 6.3	3.7 ± 1.6	1171.2 ± 6.0	28.5 ± 6.3	3.3 ± 1.0

Note.—Data are mean ± standard deviation.
* Healthy US indicates normal automated breast US images.

AUC than without-CAD mode in both asymptomatic and symptomatic women (Table 2). For experienced readers, however, the difference was not significant in both asymptomatic and symptomatic women. In women who were asymptomatic, the mean sensitivity of novice readers at both second-reading mode and concurrent-reading mode was improved by 21% (67% [95% CI: 63%, 74%] vs 88% [95% CI: 84%, 89%], respectively; *P* = .003).



a. b. c. d.

Figure 4: Automated breast US images in 51-year-old woman with invasive breast cancer. **(a)** Transverse, **(b)** sagittal, and **(c)** coronal plane images show an 8-mm ovoid hypoechoic mass with mostly circumscribed margin in the left breast. **(d)** Computer-aided detection (CAD) system output display shows a lesion suspicious for cancer (circle). Without CAD support, the mass was categorized as probably benign (Breast Imaging Reporting and Database System [BI-RADS] category 3) by three novice readers and one experienced reader, and as suspicious (BI-RADS category 4) by two experienced readers. With CAD support, two novice readers and one experienced reader who initially categorized this as BI-RADS category 3 upgraded the lesion to BI-RADS category 4, whereas the other three readers did not change their final assessment.

When a BI-RADS category 4 threshold was used, all sensitivities decreased, and all specificities improved at different reading modes and in subgroups of women (Table E2 [online]). However, there were significant differences of sensitivity between without-CAD and with-CAD reading modes for novice and all readers ($P < .001$). In women who were asymptomatic, the mean specificity for experienced readers at both the second- and concurrent-reading modes was improved by 2% (95.0% [95% CI: 94.7%, 95.3%] vs 97.4% [95% CI: 97.2%, 97.4%] and 97.2% [95% CI: 96.9%, 97.4%], respectively; $P < .001$). The performances of each reader at three reading modes with BI-RADS category 4 considered to be test-positive are shown in detail in Table E3 (online).

Management Decision Change

At the second-reading mode and concurrent-reading mode, the management decision changes compared with the without-CAD mode were mean of 3.4% (50 of 1485), ranging from 2.4% (36 of 1485) to 4.4% (66 of 1485) of AB US images for all readers (Tables 3, E4 [online]). For novice readers, with CAD support, 26–31 of 282 malignant lesions (9.2%–11.0%) were correctly changed from follow-up to biopsy (Fig 4), whereas none to seven (0%–2.5%) malignant lesions were incorrectly changed from biopsy to follow-up. For experienced readers, 26–40 of 1203 benign or healthy images (2.2%–3.3%) were correctly changed from biopsy to follow-up (Fig 5), whereas one to four benign or healthy images (0.1%–0.3%) were incorrectly changed from follow-up to biopsy.

When we analyzed women who were asymptomatic, for novice readers, five to seven of 19 malignant lesions (26%–37%) were correctly changed from follow-up to biopsy, whereas zero to one malignant lesion (0%–5%) was incorrectly changed from biopsy to follow-up. For experienced readers, 11–14 of 510 benign lesions and healthy images (2.2%–2.7%) were correctly changed from biopsy to follow-up, whereas none to two benign lesions (0%–0.4%) were incorrectly changed from follow-up to biopsy. The management decision changes of each reader are provided in Table E4 (online).

When we analyzed the management decision changes according to histologic types, novice readers with CAD support detected more invasive carcinomas (11.2%; 25.0 of 223) than ductal carcinoma in situ (6.4%; 3.7 of 58). For experienced readers, more inflammation (24.3%; 9.0 of 37) and adenosis (20.0%; 5.0 of 25) were downgraded compared with other biopsy-proved benign lesions. Management decision changes according to the histologic type are summarized in Table E5 (online).

Reading Time

For all readers, the mean reading time per volume was 50 seconds ± 9 (range, 40–63 seconds) at without-CAD mode, 61 seconds ± 8 (range, 52–74 seconds) at second-reading mode, and 34 seconds ± 9 (range, 25–52 seconds) at concurrent-reading mode (Table 4). The reading time of all readers at concurrent-reading mode was shorter than without-CAD mode and second-reading mode (Fig 6), with a mean difference of 16 seconds (95% CI: 11, 22; $P < .001$) and 27 seconds (95% CI: 21, 32; $P < .001$) per volume, respectively. Similar results were found for both novice and experienced readers when we compared the results in both asymptomatic and symptomatic women. For all readers, the average reading time per examination at without-CAD, second-reading, and concurrent-reading modes were 4:50 minutes, 5:54 minutes, and 3:17 minutes, respectively.

Discussion

We found the mean area under the receiver operating characteristic curve (AUC) for all readers improved in both the second-reading mode (from 0.88 to 0.91; $P < .001$) and concurrent reading mode (from 0.88 to 0.90; $P = .002$), and oversight errors in malignancies occurred less often with sensitivity improvement (from 90% to 94% in the second-reading mode and 93% in the concurrent-reading mode; $P < .001$ for each). For subgroup analysis with use of a Breast Imaging Reporting and Database System (BI-RADS) category 3 threshold, the mean sensitivity of novice readers in women who were asymptomatic improved from 67% without computer-aided detection (CAD) to 88% ($P = .003$) at both the second-reading and the concurrent-reading modes. The concurrent-reading mode, however, could save 32% (16 seconds of 50 seconds per volume) time with higher AUC compared with without-CAD mode, and 44% (27 seconds of 61 seconds per volume) time with similar AUC compared with the second-reading mode. Thus, our results suggest that the concurrent-reading mode with shorter reading time is preferable for workflow efficiency.

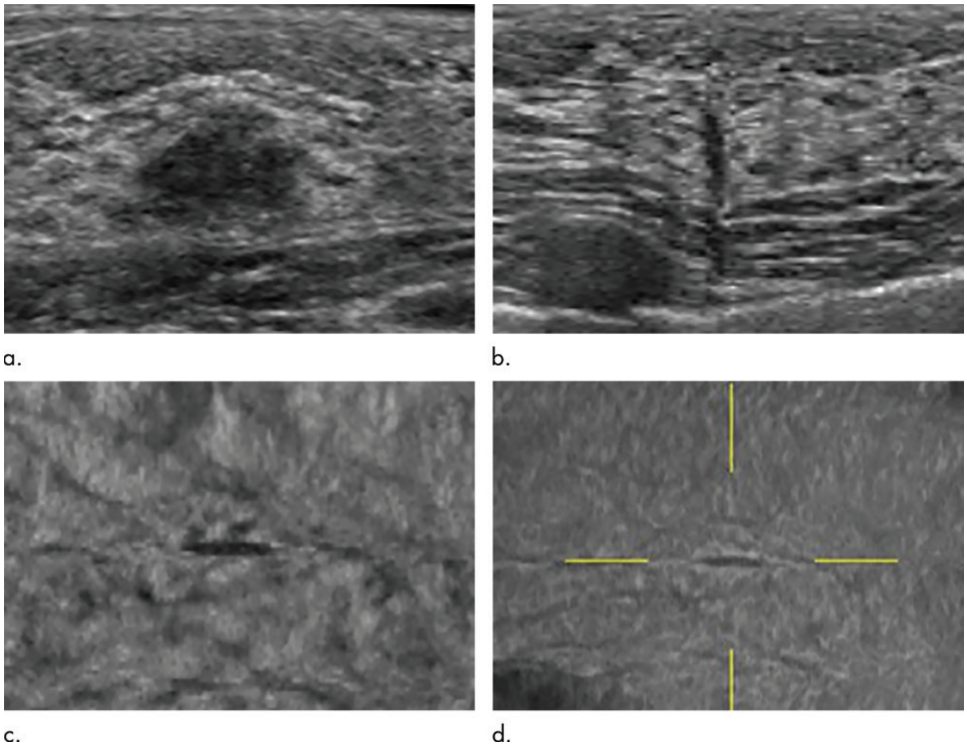


Figure 5: Automated breast US scans obtained in a 37-year-old woman with sclerosing adenosis. **(a)** Transverse, **(b)** sagittal, and **(c)** coronal plane images show a 13-mm irregular hypoechoic mass with ill-defined margin in the left breast. **(d)** Computer-aided detection (CAD) system output display shows no mark on the lesion. Without CAD support, the mass was categorized as Breast Imaging Reporting and Data System (BI-RADS) category 4 by three novice and two experienced readers, and BI-RADS category 3 by one experienced reader. With CAD support, two novice and two experienced readers who initially categorized this as BI-RADS category 4 downgraded the lesion to BI-RADS category 3, whereas other two readers did not change their final assessments.

Our results showed that CAD was more helpful for novice readers than for experienced readers, and the mean AUC difference between the two groups changed from 0.06 (0.91 vs 0.85, respectively) at without-CAD mode to 0.03 (0.92 vs 0.89, respectively) at second-reading mode. Furthermore, for upgraded malignancies, novice readers who used CAD support detected more invasive carcinomas (11.2%; 25.0 of 223) than ductal carcinoma in situ (6.4%; 3.7 of 58). Similar results were obtained by Xu et al (22), who reported that the CAD system improved the performance of less experienced readers for detection of breast cancer with AB US. For experienced readers, CAD allowed for benign lesions to be downgraded from biopsy to follow-up recommendation. This result is encouraging because the low specificity is an important downside of supplemental US screening. A significant improvement in specificity with CAD was also found in other studies (10,22).

Currently, most CAD systems are used at the second-reading mode for breast cancer screening (23–25). However, this mode prolongs the reading time, especially for large volume imaging (26–28). Our study showed the use of concurrent-reading mode reduced the time required to read AB US images with similar sensitivity and specificity compared with second-reading mode. Similar CAD results were reported in studies of AB US and other imaging modalities (20,29–32). In our study, the average reading time per examination at without-CAD and concurrent-reading modes were 4:50 minutes and 3:17 minutes, respectively, whereas those in the study by Jiang et al (20) were 3:33 minutes and 2:24 minutes, respectively. AB US image interpretation times probably depend on the examination type (ie, screening vs diagnostic), lesion difficulty, and reader experience (20,32). According to our study and other reports (20,32), a shorter reading time without decreasing performance for interpretation of AB US studies can be expected with CAD when used concurrently versus when used as a second reader in screening.

There are several reader studies (10,20,22,32) that used the commercial AB US CAD system and evaluated effect of CAD system on the accuracy and efficiency of interpretation of AB US images by radiologists, and they showed the benefit of CAD in both novice and experienced readers. However, the incremental benefit of CAD when used as either a second or concurrent reader of AB US images, to our knowledge, has not been compared. Our study dealt with the CAD application at AB US image reading for both novice and experienced readers and it evaluated how the CAD system was best implemented in conjunction with AB US.

Table 4: Reading Time at Three Reading Modes

Parameter	Without-CAD Mode	Second-reading Mode	Concurrent-reading Mode	Difference between Without-CAD Mode and Second-reading Mode	Difference between Second-reading Mode and Concurrent-reading Mode	Difference between Without-CAD Mode and Concurrent-reading Mode
Asymptomatic women						
Novice readers	55 ± 6	66 ± 6	38 ± 9	11 (9, 13) [20]	−28 (−21, −36) [−42]	−17 (−9, −26) [−31]
Experienced readers	42 ± 4	53 ± 2	26 ± 4	11 (6, 17) [26]	−27 (−22, −32) [−51]	−16 (−12, −18) [−38]
All readers	48 ± 8	59 ± 9	32 ± 9	11 (10, 13) [23]	−27 (−25, −30) [−46]	−16 (−14, −19) [−33]
Symptomatic women						
Novice readers	60 ± 6	69 ± 6	43 ± 11	9 (3, 14) [15]	−26 (−13, −39) [−38]	−17 (−3, −32) [−28]
Experienced readers	46 ± 6	57 ± 2	31 ± 3	11 (1, 21) [24]	−26 (−23, −29) [−46]	−15 (−7, −23) [−33]
All readers	54 ± 9	63 ± 8	37 ± 10	9 (6, 13) [17]	−26 (−22, −30) [−41]	−17 (−12, −21) [−31]
All women						
Novice readers	58 ± 5	67 ± 6	40 ± 10	9 (5, 13) [16]	−27 (−16, −37) [−40]	−18 (−3, −32) [−31]
Experienced readers	43 ± 4	55 ± 2	28 ± 3	12 (5, 18) [28]	−27 (−21, −32) [−49]	−15 (−13, −16) [−35]
All readers	50 ± 9	61 ± 8	34 ± 9	11 (8, 13) [22]	−27 (−21, −32) [−44]	−16 (−11, −22) [−32]

Note.—Data are seconds per volume acquisition in automated breast US scanning. Mean data are ± standard deviation; 95% confidence intervals are in parentheses. Data in brackets are percentage change. CAD = computer-aided detection.

Our study had limitations. First, it was a retrospective study and there could be bias caused by memorization of previous images. We recorded the reading time and diagnostic results of the second-reading mode after the without-CAD mode. The results may be different if we separate the second-reading mode from the without-CAD mode. If a review for the second-reading mode was performed another day, the interpretation time might be shorter than the time reported in our study. Second, all of the readers were less-experienced breast imagers (with breast specialty experience ranging from 1 to 4 years). In China, the specialty of breast imaging is newer, and staff and faculty in breast imaging are young compared with other imaging specialties. Third, we did not assess the association between lesion difficulty or lesion types and decision changes in different experienced readers. Certain types of lesions such as circumscribed or not circumscribed masses may be more likely associated with decision changes than others. Finally, we did not analyze the false-positive CAD marks. Further study is needed to track the outcome of all CAD marks at AB US.

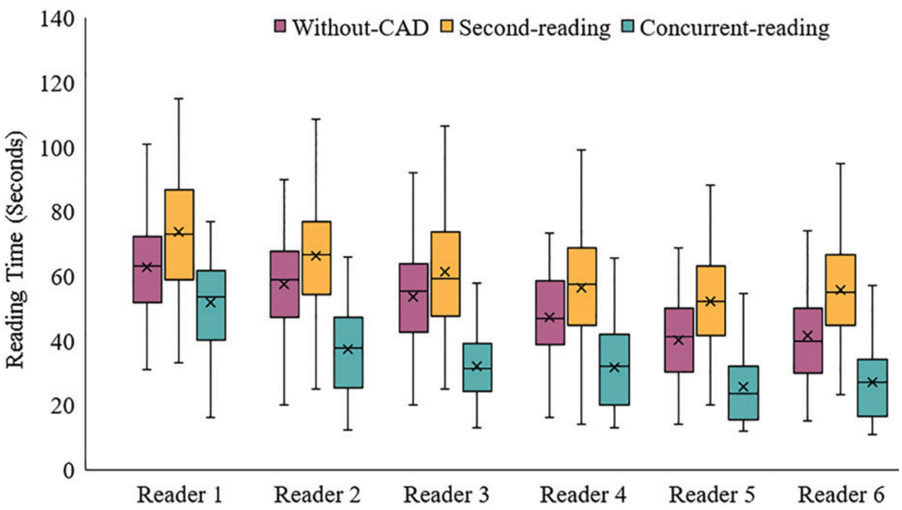


Figure 6: Box-and-whisker plot of the reading time per volume of each reader with three reading modes. Readers 1, 2, and 3 are novice readers with 1–3 years of US experience and readers 4, 5, and 6 are experienced readers with 5–10 years of US experience. Boxes indicate interquartile range, crosses indicate mean, center lines indicate median, and whiskers indicate range. CAD = computer-aided detection.

In conclusion, our study showed computer-aided detection (CAD) at both the second-reading mode and concurrent-reading mode may improve novice readers' performance of breast cancer detection at automated breast US, especially in women who are asymptomatic. At the concurrent-reading mode, a significantly shorter reading time can be expected without decreasing performance for both novice and experienced readers. Further research is warranted in a multicenter prospective study to investigate the usefulness of CAD at breast cancer screening with automated breast US.

Author contributions: Guarantors of integrity of entire study, S.Y., L.L., G.Z., H.S.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, S.Y., X.G., L.L., H.S.; clinical studies, S.Y., L.L., R.S., J.Y., G.Z., Y.X., N.Z., H.S.; experimental studies, S.Y., X.G., L.L., R.S., J.Y., Y.X., Y.J., N.Z., H.S.; statistical analysis, S.Y., X.G., L.L., H.S.; and manuscript editing, S.Y., X.G., L.L., R.S., J.Y., Y.X., N.Z., H.S.

Disclosures of Conflicts of Interest: **S.Y.** disclosed no relevant relationships. **X.G.** disclosed no relevant relationships. **L.L.** disclosed no relevant relationships. **R.S.** disclosed no relevant relationships. **J.Y.** disclosed no relevant relationships. **G.Z.** disclosed no relevant relationships. **Y.X.** disclosed no relevant relationships. **Y.J.** disclosed no relevant relationships. **N.Z.** disclosed no relevant relationships. **H.S.** disclosed no relevant relationships.

References

1. Tabár L, Vitak B, Chen TH, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011;260(3):658–663.
2. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet 2012;380(9855):1778–1786.
3. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst 2000;92(13):1081–1087.
4. 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.
5. 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.
6. 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.
7. Weigert J, Steenbergen S. The connecticut experiments second year: ultrasound in the screening of women with dense breasts. Breast J 2015;21(2):175–180.
8. Brem RF, Tabár L, Duffy SW, et al. Assessing improvement in detection of breast cancer with three-dimensional automated breast US in women with dense breast tissue: the SomoInsight Study. Radiology 2015;274(3):663–673.
9. Chang JM, Moon WK, Cho N, Park JS, Kim SJ. Breast cancers initially detected by hand-held ultrasound: detection performance of radiologists using automated breast ultrasound data. Acta Radiol 2011;52(1):8–14.
10. van Zelst JCM, Tan T, Clauser P, et al. Dedicated computer-aided detection software for automated 3D breast ultrasound; an efficient tool for the radiologist in supplemental screening of women with dense breasts. Eur Radiol 2018;28(7):2996–3006.
11. Halligan S, Mallett S, Altman DG, et al. Incremental benefit of computer-aided detection when used as a second and concurrent reader of CT colonographic data: multiobserver study. Radiology 2011;258(2):469–476.
12. Zheng B, Swensson RG, Golla S, et al. Detection and classification performance levels of mammographic masses under different computer-aided detection cueing environments. Acad Radiol 2004;11(4):398–406.
13. Matsumoto S, Ohno Y, Aoki T, et al. Computer-aided detection of lung nodules on multidetector CT in concurrent-reader and second-reader modes: a comparative study. Eur J Radiol 2013;82(8):1332–1337.
14. Warren Burhenne LJ, Wood SA, D'Orsi CJ, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology 2000;215(2):554–562.
15. Choi JH, Kang BJ, Baek JE, Lee HS, Kim SH. Application of computer-aided diagnosis in breast ultrasound interpretation: improvements in diagnostic performance according to reader experience. Ultrasonography 2018;37(3):217–225.
16. Chabi ML, Borget I, Ardiles R, et al. Evaluation of the accuracy of a computer-aided diagnosis (CAD) system in breast ultrasound according to the radiologist's experience. Acad Radiol 2012;19(3):311–319.
17. Mendelson EB, Böhm-Vélez M, Berg WA, et al. ACR BI-RADS Ultrasound. In: ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Reston, Va: American College of Radiology, 2013.
18. Jiang Y, Metz CE. BI-RADS data should not be used to estimate ROC curves. Radiology 2010;256(1):29–31.
19. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. Invest Radiol 1992;27(9):723–731.
20. 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.
21. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. Radiology 2019;290(2):305–314.
22. Xu X, Bao L, Tan Y, Zhu L, Kong F, Wang W. 1000-case reader study of radiologists' performance in interpretation of automated breast volume scanner images with a computer-aided detection system. Ultrasound Med Biol 2018;44(8):1694–1702.
23. Azavedo E, Zackrisson S, Mejàre I, Heibert Arnlind M. Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. BMC Med Imaging 2012;12(1):22.
24. Gilbert FJ, Astley SM, McGee MA, et al. Single reading with computer-aided detection and double reading of screening mammograms in the United Kingdom National Breast Screening Program. Radiology 2006;241(1):47–53.
25. Lehman CD, Blume JD, DeMartini WB, Hylton NM, Herman B, Schnall MD. Accuracy and interpretation time of computer-aided detection among novice and experienced breast MRI readers. AJR Am J Roentgenol 2013;200(6):W683–W689.
26. Wittenberg R, Berger FH, Peters JF, et al. Acute pulmonary embolism: effect of a computer-assisted detection prototype on diagnosis--an observer study. Radiology 2012;262(1):305–313.
27. Tchou PM, Haygood TM, Atkinson EN, et al. Interpretation time of computer-aided detection at screening mammography. Radiology 2010;257(1):40–46.
28. Petrick N, Haider M, Summers RM, et al. CT colonography with computer-aided detection as a second reader: observer performance study. Radiology 2008;246(1):148–156.
29. Balleyguier C, Arfi-Rouche J, Levy L, et al. Improving digital breast tomosynthesis reading time: A pilot multi-reader, multi-case study using concurrent computer-aided detection (CAD). Eur J Radiol 2017;97:83–89.
30. Benedikt RA, Boatsman JE, Swann CA, Kirkpatrick AD, Toledano AY. Concurrent computer-aided detection improves reading time of digital breast tomosynthesis and maintains interpretation performance in a multireader multicase study. AJR Am J Roentgenol 2018;210(3):685–694.
31. Beyer F, Zierott L, Fallenberg EM, et al. Comparison of sensitivity and reading time for the use of computer-aided detection (CAD) of pulmonary nodules at MDCT as concurrent or second reader. Eur Radiol 2007;17(11):2941–2947.
32. van Zelst JCM, Tan T, Platel B, et al. Improved cancer detection in automated breast ultrasound by radiologists using computer aided detection. Eur J Radiol 2017;89:54–59.