ORIGINAL RESEARCH

Does the radiologist need to rescan the breast lesion to validate the final BI-RADS US assessment made on the static images in the diagnostic setting?

Yue Hu[1,2,*]
Jingsi Mei[1,2,*]
Xiaofang Jiang[1,2]
Ran Gu[1,2]
Fengtao Liu[1,2]
Yaping Yang[1,2]
Hongli Wang[1,2]
Shiyu Shen[1,2]
Haixia Jia[3]
Qiang Liu[1,2]
Chang Gong[1,2]

[1]Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation; [2]Breast Tumor Center, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University; [3]Department of Breast Surgery, Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, People's Republic of China

*These authors contributed equally to this work

Correspondence: Chang Gong; Qiang Liu
Breast Tumor Center, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Yingfeng Road No. 33, HaiZhu District, 510260 Guangzhou, Guangdong, People's Republic of China
Tel/Fax +860 203 407 1512
Email changgong282@163.com;
victorlq@hotmail.com

Purpose: To assess whether radiologist needs to rescan the breast lesion to validate the final American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) ultrasonography (US) assessment made on the static images in the diagnostic setting.

Patients and methods: Image data on 1,070 patients with 1,070 category 3–5 breast lesions with a pathological diagnosis scanned between January and June 2016 were included. Both real-time and static image assessments were acquired for each lesion. The diagnostic performance was evaluated by receiver operating characteristic (ROC) curves. The positive predictive values (PPVs) of each category in the two groups were calculated according to the ACR BI-RADS manual and compared. Kappas were determined for agreement on two assessment approaches.

Results: The sensitivity, specificity, PPV, and negative predictive value for real-time US were 98.9%, 58.2%, 44.8% and 99.4%, and for static images were 98.9%, 57.1%, 44.1% and 99.3%, respectively. The performance of the two groups was not significantly different (areas under ROCs: 0.786 vs 0.780, P=0.566) if the final assessment was only dichotomized as negative (category 3) and positive (categories 4 and 5). All PPVs of each category for each assessment were within the reference range provided by the ACR in 2013 except subcategory 4B (reference range: >10% and ≤50%) of static image evaluation, which was also significantly higher than that of real-time assessment (54.8% vs 40.7%, P=0.037). The overall agreement of the two approaches was moderate (κ=0.43–0.56 according to different detailed assessment).

Conclusion: Both static image and real-time assessment had similar diagnostic performance if only the treatment recommendations were considered, that is, follow-up or biopsy. However, as for subcategory 4B lesions without obviously benign or malignant US features, real-time scanning by the interpreter is recommended to obtain a more accurate BI-RADS assessment after assessing static images.

Keywords: Breast Imaging Reporting and Data System, ultrasonography, diagnosis, real-time scanning

Introduction

Ultrasonography (US) is playing an increasingly important role in both breast disease screening and diagnosis,[1–3] especially in Asian women with higher breast density and younger age at diagnosis of breast cancer (BC).[4–6] The use of US in breast diagnostic imaging has expanded from primary differentiation of cystic from solid lesions to accurate characterization of benign and malignant lesions.[7–9]

As with other breast imaging modalities like mammography (MG) and magnetic resonance imaging, proper US diagnosis of breast lesions requires both successful identification and accurate feature analysis. In some countries, technologists perform the diagnostic breast US examination and acquire static images for interpreting physicians.[10] The ability of technologists to detect lesions have been proved, several studies have reported that representative static images of breast lesions can be obtained successfully by well-trained technologists in BC screening.[11–13] However, the accuracy of interpreting static images by radiologists is unknown comparing with the real-time assessment. Because breast lesions are three-dimensional, even representative orthogonal static images may miss some detailed diagnostic features, especially for lesions with vague US features. From this point of view, real-time scanning by the interpreter may give extra information for accurate assessment of breast lesions.

Thus, the purpose of our study was to retrospectively assess whether the radiologist needs to rescan the breast lesion to validate the final American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) assessment made on the static images in the diagnostic setting.

Material and methods
Patients and biopsy

The institutional ethics committee of Sun Yat-Sen Memorial hospital approved this retrospective study. The requirement for informed consent was waived by the ethics committee due to the nature of the study, and no personal information was disclosed. This study strictly abides the principles of the World Medical Association Declaration of Helsinki. Using the diagnostic breast US database of our institution from January 2016 to June 2016, we reviewed female patients who underwent breast US examination because of palpable abnormalities, or for symptoms such as breast pain and nipple discharge. For patients with more than one breast lesion, only the one with the highest score was analyzed to guarantee statistical independence of each observation. In total, 1,070 breast lesions in 1,070 subjects scored as BI-RADS category 3–5 were biopsied and diagnosed histologically.

Although an interventional procedure was usually recommended for lesions of subcategories ≥4A, some patients with lesions of category 3 underwent a biopsy in consideration of other factors after clinical consultation with referring physicians. Pathological results were obtained by US-guided core needle biopsy, vacuum-assisted biopsy, or excisional biopsy.

Real-time ultrasonography examination and image documentation

Examinations were performed with a high-resolution US unit (S2000 or S1000; Siemens Medical Solutions, Erlangen, Germany; or MyLab 30; Esaote, Genoa, Italy) with a high-frequency linear transducer. Three radiologists who specialized in breast US diagnosis with 5–10 years of experience performed a bilateral whole-breast US examination. A final US score (category 0–6, including subcategory 4A–4C) was assigned by the same radiologist immediately after scanning according to the second edition of BI-RADS US.[9] Our radiologists assessed the US findings without referring to MG, because US and MG are usually in different departments in China and few radiologists routinely interpret both of them.[14,15]

Static images of lesions were saved by the radiologists during real-time scanning according to the ACR guidelines.[10] For each lesion, at least two representative B-mode images with and without calipers were obtained in orthogonal planes. Additionally, at least one image of color/power Doppler was acquired to assess the vascularity.

Static image evaluation

The 1,070 sets of static US images were retrospectively evaluated by the same three radiologists who performed the examinations at least 1 year after data collection. Images were anonymized and in random order. During the review, the radiologists were blinded to the previous real-time assessments and pathological results, but they were informed of the relevant clinical information including patient age, chief complaint, breast disease history, and physical examination results, which was also known to radiologists during the real-time assessment. They were asked to record a final BI-RADS category (category 3, 4A–C, 5) for each case.

Statistical analysis

According to biopsy recommendations described in BI-RADS, we set a cutoff point between category 3 and 4A/4. The sensitivity, specificity, positive predictive values (PPV) and negative predictive value (NPV) of the real-time and static image groups were calculated, respectively. Receiver operating characteristic (ROC) curves were constructed, and areas under the ROC curves (AUC) were calculated.

We calculated the PPV for each category of the two groups according to the ACR BI-RADS manual. PPV was defined as the percentage of BC cases within the total number of biopsies. Then, we compared the differences in PPVs of each category between the two groups, of which Pearson's chi-squared statistic was used for subcategories 4A–4C, and Fisher's exact test was used for categories 3 and 5.

Kappa statistics were used to calculate the degree of agreement between the final assessment of the two groups. A kappa value (κ) 0–0.20 corresponds to slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement.[16]

SPSS 22 (IBM Corp., Armonk NY, USA) and Medcalc software version 11.4 (Medcalc Software, Mariakerke, Belgium) were used for the statistical analyses. All P-values were two-tailed, and $P<0.05$ was considered statistically significant.

Results

Of the 1,070 patients (age range 12–80 years, mean age 39.8 ±12.3 years) with 1,070 lesions 273 (273/1,070, 25.5%) lesions were malignant. The most common histological type was invasive ductal carcinoma (IDC) (220/273, 80.6%). A total of 797 (797/1,070, 74.5%) lesions were benign. Fibroadenoma (501/797, 62.9%) was the most common diagnosis. Table 1 describes the pathological results.

The sensitivity, specificity, PPV, and NPV for real-time group were 98.9%, 58.2%, 44.8% and 99.4%, respectively, and for the static image group were 98.9%, 57.1%, 44.1% and 99.3% (Table S1), respectively. Discordant treatment decisions (either follow-up or biopsy) appeared in 231 cases. Among them, 111 lesions were assessed as positive in real-time scanning but negative in static image evaluation, and one of these lesions was IDC. A total of 120 lesions were assessed as negative in real-time scanning but positive in static image evaluation, and one of them was ductal carcinoma in situ (Figure 1). The other 229 lesions were all proven to be benign.

ROC curves showed that the diagnostic performance of real-time analysis and static image evaluation was not statistically different (0.786 vs 0.780, $P=0.566$) if the final assessment was only dichotomized as negative (category 3) and positive (categories 4 and 5). The performance of the two groups was statistically different if calculated as category 3–5 no matter with or without

Table 1 Final pathological results of 1,070 breast lesions

Pathological results	No. of lesions (%)
Benign	
Fibroadenoma	501 (62.9)
Fibrocystic change	70 (8.8)
Intraductal papilloma	66 (8.3)
Sclerosing adenosis	39 (4.9)
Inflammation	36 (4.5)
Phyllodes tumor	35 (4.4)
Radial scar	20 (2.5)
Hamartoma	4 (0.5)
Postoperative change	3 (0.4)
Fibromatosis	3 (0.4)
Others	20 (2.5)
Total	797
Malignant	
IDC	220 (80.6)
ILC	14 (5.1)
DCIS	13 (4.8)
Papillary carcinoma	10 (3.7)
Mucinous carcinoma	3 (1.1)
Invasive micropapillary carcinoma	2 (0.7)
Mixed invasive carcinoma	2 (0.7)
Metaplastic carcinoma	2 (0.7)
Neuroendocrine carcinoma	2 (0.7)
Others	5 (1.8)
Total	273

Abbreviations: IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; DCIS, ductal carcinoma in situ.

subcategory 4A–4C (0.969 vs 0.955, $P=0.011$; 0.915 vs 0.855, $P<0.001$, respectively) (Figure 2).

The distribution of BI-RADS US categories in the two groups is shown in Table 2. In the real-time group, the PPVs for categories 3–5 were 0.6% (3/467), 23.1% (99/429) and 98.3% (171/174), respectively. PPVs for subcategories 4A–4C were 5.9% (17/288), 40.7% (37/91) and 90.0% (45/50), respectively. In static image group, the PPVs for categories 3–5 were 0.7% (3/458), 33.8% (174/515) and 99.0% (96/97), respectively; that for subcategories 4A–4C were 8.3% (25/301), 54.8% (74/135) and 94.9% (75/79), respectively. The PPVs of each BI-RADS US category/subcategory were all within the reference range provided by the ACR, except subcategory 4B (54.8%, 74/135) in the static image group, which exceeded the reference range (10–50%). The PPV of subcategory 4B in the static image group was significantly higher than that in the real-time group (54.8% vs 40.7%, $P=0.037$). There were no significant differences in the PPVs of the other categories/subcategories (3, 4A, 4C, 5) between the two groups.
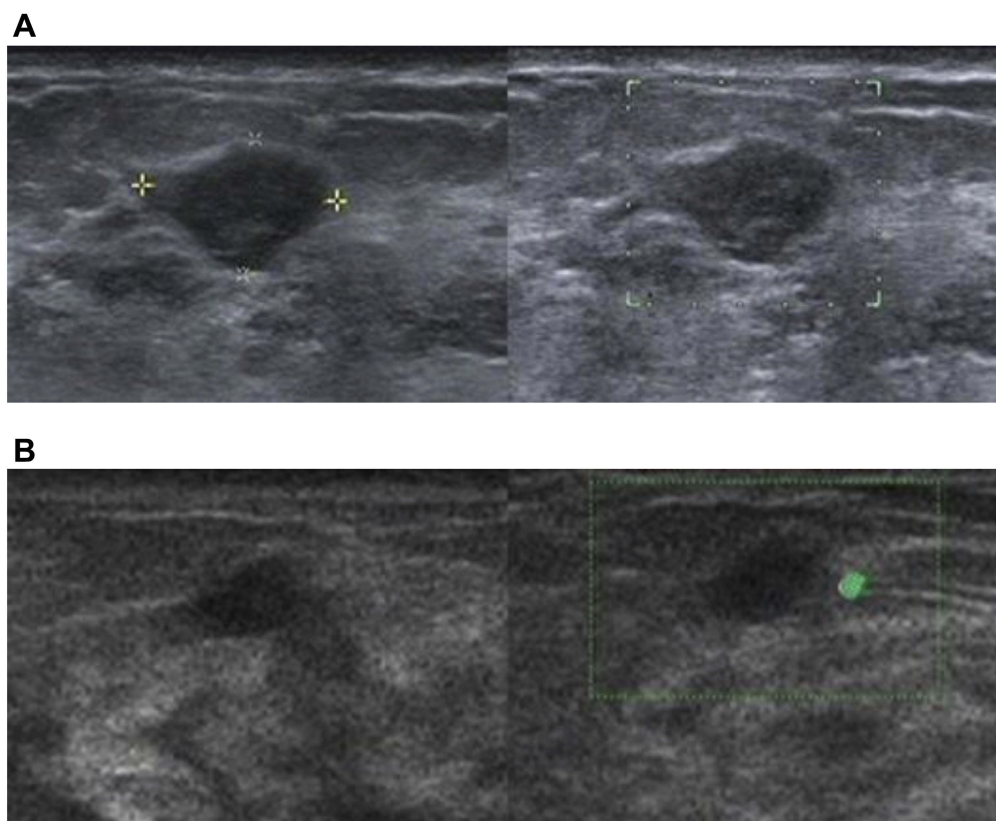
A

B

Figure 1 (A) A solid hypo-echoic breast lesion categorized as BI-RADS-US 4A in real-time scanning and BI-RADS-US 3 in static image evaluation. The pathologic diagnosis was IDC. (B) A solid hypo-echoic breast lesion categorized as BI-RADS-US 3 in real-time scanning and BI-RADS-US 4A in static image evaluation. The pathologic diagnosis was DCIS. **Abbreviations:** IDC, invasive ductal carcinoma; DCIS, ductal carcinoma in situ.
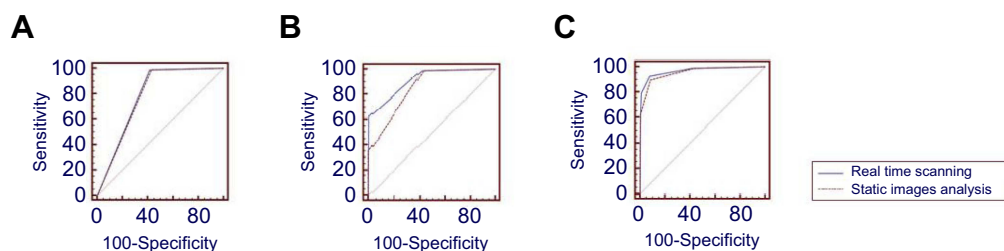
Figure 2 Receiver operating characteristic curves. (A) Lesions were assigned to negative (category 3) or positive (category 4 and 5) in each group. The AUC for real-time and static image group was 0.786 vs 0.780, respectively. (B) Lesions were assigned to 3, 4 or 5 categories in each group. The AUC for real-time and static image group was 0.915 vs 0.855, respectively. (C) Lesions were assigned to 3, 4A–4C or 5 subcategories/categories in each group. The AUC for real-time and static image group was 0.969 vs 0.955, respectively. **Abbreviation:** AUC, Area under the ROC curve.

According to Cohen's κ statistic, the overall degree of agreement between the final assessments of the two groups was always moderate according to different detailed assessment (category 3–5 with or without subcategory 4A–4C: κ=0.43 and 0.48; assessment dichotomized as negative or positive: κ=0.56) (Table S2). In addition, the agreement of the two groups was fair for category 4 (κ=0.37), but moderate for categories 3 and 5 (both κ=0.56).

Discussion

US is indispensable in the diagnostic evaluation of breast lesions. Physician-performed real-time scanning is the optimal model for breast US examination because of its high operator dependence.[7,17,18] However, breast US scanning is time-consuming; the reported average time to complete a bilateral whole-breast US examination is 10–31 mins.[11,17,19]

Table 2 BI-RADS US of real-time and static image evaluation correlate with pathology

BI-RADS US category/subcategory	PPV provided by ACR		Real-time US n (%)		Static image US n (%)		P
3	>0 & ≤2%		467		458		>0.999
Benign			464	(99.4%)	455	(99.3%)	
Malignant			3	(0.6%)	3	(0.7%)	
4A	>2% & ≤10%		288		301		0.257
Benign			271	(94.1%)	276	(91.7%)	
Malignant			17	(5.9%)	25	(8.3%)	
4B	>10% & ≤50%		91		135		0.037
Benign			54	(59.3%)	61	(45.2%)	
Malignant			37	(40.7%)	74	(54.8%)	
4C	>50% & <95%		50		79		0.473
Benign			5	(10.0%)	4	(5.1%)	
Malignant			45	(90.0%)	75	(94.9%)	
5	≥95%		174		97		>0.999
Benign			3	(1.7%)	1	(1.0%)	
Malignant			171	(98.3%)	96	(99.0%)	

Abbreviations: BI-RADS US, Breast Imaging Reporting and Data System for Ultrasonography; PPV, positive predictive value; ACR, American College of Radiology.

In many countries, especially in high-volume clinical practices, radiologists usually render a final interpretation directly by reviewing static images obtained by trained technologists, although real-time scanning by the interpreter is encouraged.[10] This division of labor may make up the shortage of physicians to perform the examination. However, this workflow may lead to some unavailability of significant dynamic US information for the interpreters and thus influences their interpretive performance. To our knowledge, our study is the first to compare the diagnostic efficiency of real-time analysis and retrospective static image evaluation of breast lesions, based on the second edition of ACR BI-RADS US published in 2013.

In this study, both methods performed well in the diagnosis of breast lesions no matter how many classifications of assessment were given, although real-time analysis performed much better than static images except for the dichotomized assessment. Youk et al and Foldi et al did not find overall significant differences in diagnostic performance between video and static image evaluation (assessment based on the second edition of ACR BI-RADS US,[20,21] categories 2–5 with subcategory, AUC =0.830 vs 0.800, P=0.08; assessment based on the first edition of ACR BI-RADS US, categories 2–5 without subcategory, AUC =0.719 vs 0.762, respectively). Our study did not show a significantly better diagnostic performance of real-time US when lesions were assessed as negative or positive (P=0.566), which implies that treatment recommendation (follow-up or biopsy) would not be influenced by interpretations of static vs real-time images.

The PPVs of each category in the real-time group were all within the reference ranges provided by the ACR in 2013. The PPVs of subcategory 4B in the static image group exceeded the reference value, and also was significantly higher than that of the real-time group. In addition, most malignant lesions of subcategory 4B in the static image group had been previously assigned to a higher category (4C or 5) based on real-time US findings, while only 21.6% (8/37) malignant lesions of subcategory 4B assessed by the real-time US were assigned to a higher category by retrospective static image evaluation.

Unlike lesions of other categories/subcategories, category 4B has the vaguest US features, which may cause a lack of confidence for the likelihood of benign or malignant diagnosis. Zou et al found that when malignancy was not obvious, most radiologists chose a relatively lower category to ease patients' concern.[22] The PPV of subcategory 4B by static image evaluation in their study was 70.8% (74/135), which is significantly higher than the reference value. In dynamic real-time scanning, the examiner can scrutinize every part of the lesion and acquire supplementary information for more accurate diagnosis, which accounts for the difference from static image evaluation in subcategory 4B. However, for lesions with more obvious benign or malignant US features, final assessment can be made by static images as accurately as real-time analysis. Similar findings were reported by Van Holsbeke et al in the assessment of adnexal masses by real-time versus static image evaluation.[23]

The overall degree of agreement between the final assessments of real-time and static images remained moderate, whether or not subcategories were assigned, and even when the assessment was only dichotomized as negative and positive. Category 4 was divided into three subcategories by the ACR because of its wide-ranging likelihood of malignancy (>2 to <95%).[9] However, these subcategories were assigned subjectively by radiologists in clinical practice. Previous studies have reported some unsatisfactory results in subdividing category 4, but they focused on the interobserver agreement of static image evaluation.[14,22] In our study, category 4 (fair agreement, κ=0.37) had the lowest agreement compared with categories 3 and 5 (moderate agreement, both κ=0.56) between real-time and static images.

The present study has several limitations. First, the ACR management recommendation for category 3 was not strictly followed in our institution, with a high biopsy rate of 34.3% (472/1,376 of real-time assessment cases). Other authors have also reported high biopsy rates for category 3 lesions. The biopsy rates reported by Raza et al and Berg et al were 22.5% and 71.0%, respectively.[24,25] Second, elastography was not involved in the study, which has been confirmed to improve the specificity of breast US interpretation and was added to the second edition of ACR BI-RADS US.[9,25] Third, there were some biases and limitations because of the real situation of breast imaging in China. For example, few radiologists routinely interpret both breast US and MG because they are in different departments; there are no breast US

technologists, radiologists need to scan and interpret all patients simultaneously, therefore static images of this study were saved by the radiologists instead of technologists. Finally, this was a single-institution retrospective study; multi-institutional prospective studies with greater numbers of patients are needed to confirm our findings.

Conclusion

In conclusion, the diagnostic performance of static image evaluation was similar to real-time assessment if only the treatment recommendations were considered, that is, follow-up or biopsy. However, real-time scanning by the interpreter is recommended for lesions of subcategory 4B assessed through static images to obtain a more accurate BI-RADS assessment. Considering the limitations, further verification should be needed before extrapolating this study to everyday breast US diagnosis.

Abbreviation list

US, ultrasonography; BC, breast cancer; MG, mammography; ACR, American College of Radiology; BI-RADS, Breast Imaging Reporting and Data System; CNB, core needle biopsy; VAB, vacuum-assisted biopsy; PPV, positive predictive value; NPV, negative predictive value; ROC, receiver operating characteristic; AUC, area under the ROC curve; IDC, invasive ductal carcinoma.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. 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 (Lond, Engl)*. 2016;387(10016):341–348. doi:10.1016/S0140-6736(15)00774-6

2. Berg WA, Bandos AI, Mendelson EB, Lehrer D, Jong RA, Pisano ED. Ultrasound as the primary screening test for breast cancer: analysis from ACRIN 6666. *J Natl Cancer Inst*. 2016;108(4):djv367. doi:10.1093/jnci/djv367

3. Shen S, Zhou Y, Xu Y, et al. A multi-centre randomised trial comparing ultrasound vs mammography for screening breast cancer in high-risk Chinese women. *Br J Cancer*. 2015;112(6):998–1004. doi:10.1038/bjc.2015.33

4. Fan L, Strasser-Weippl K, Li J-J, et al. Breast cancer in China. *Lancet Oncol*. 2014;15(7):e279–e289. doi:10.1016/S1470-2045(13)70567-9

5. Dai H, Yan Y, Wang P, et al. Distribution of mammographic density and its influential factors among Chinese women. *Int J Epidemiol*. 2014;43(4):1240–1251. doi:10.1093/ije/dyu042

6. Nelson HD, Pappas M, Cantor A, Griffin J, Daeges M, Humphrey L. Harms of breast cancer screening: systematic review to update the 2009 U.S. Preventive Services Task Force recommendation. *Ann Intern Med*. 2016;164(4):256–267. doi:10.7326/M15-0970

7. Hooley RJ, Scoutt LM, Philpotts LE. Breast ultrasonography: state of the art. *Radiology*. 2013;268(3):642–659. doi:10.1148/radiol.13121606

8. Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiology*. 1995;196(1):123–134. doi:10.1148/radiology.196.1.7784555

9. Sickles E, D'Orsi C, Bassett L, Appleton C, Berg W, Burnside E. *ACR BI-RADS® Atlas Breast Imaging Reporting and Data System*. Reston (VA): American College of Radiology; 2013.

10. American College of Radiology. ACR practice parameter for the performance of a breast ultrasound examination. Available at: https://www.acr.org/-/media/ACR/Files/Practice-Parameters/US-Breast.pdf?la=en. Res.38-2016. Accessed Oct 20, 2018.

11. Kaplan SS. Clinical utility of bilateral whole-breast US in the evaluation of women with dense breast tissue. *Radiology*. 2001;221(3):641–649. doi:10.1148/radiol.2213010364

12. Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE. Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. *Radiology*. 2012;265(1):59–69. doi:10.1148/radiol.12120621

13. Tohno E, Takahashi H, Tamada T, Fujimoto Y, Yasuda H, Ohuchi N. Educational program and testing using images for the standardization of breast cancer screening by ultrasonography. *Breast Cancer*. 2012;19(2):138–146. doi:10.1007/s12282-010-0221-x

14. Berg WA, Blume JD, Cormack JB, Mendelson EB. Training the ACRIN 6666 Investigators and effects of feedback on breast ultrasound interpretive performance and agreement in BI-RADS ultrasound feature analysis. *AJR Am J Roentgenol*. 2012;199(1):224–235. doi:10.2214/AJR.11.7324

15. Hu Y, Yang Y, Gu R, et al. Does patient age affect the PPV3 of ACR BI-RADS ultrasound categories 4 and 5 in the diagnostic setting? *Eur Radiol*. 2018;28(6):2492–2498. doi:10.1007/s00330-017-5203-3

16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.

17. Berg WA, Mendelson EB. Technologist-performed handheld screening breast US imaging: how is it performed and what are the outcomes to date? *Radiology*. 2014;272(1):12–27. doi:10.1148/radiol.14132628

18. Sung JS. High-quality breast ultrasonography. *Radiol Clin North Am*. 2014;52(3):519–526. doi:10.1016/j.rcl.2014.02.012

19. Berg WA, Blume JD, Cormack JB, Mendelson EB. Operator dependence of physician-performed whole-breast US: lesion detection and characterization. *Radiology*. 2006;241(2):355–365. doi:10.1148/radiol.2412051710

20. Youk JH, Jung I, Yoon JH, et al. Comparison of inter-observer variability and diagnostic performance of the fifth edition of BI-RADS for breast ultrasound of static versus video images. *Ultrasound Med Biol*. 2016;42(9):2083–2088. doi:10.1016/j.ultrasmedbio.2016.05.006

21. Foldi M, Hanjalic-Beck A, Klar M, et al. Video sequence compared to conventional freeze image documentation: a way to improve the sonographic assessment of breast lesions? *Ultraschall Med.* 2011;32 (5):497–503. doi:10.1055/s-0029-1245797

22. Zou X, Wang J, Lan X, et al. Assessment of diagnostic accuracy and efficiency of categories 4 and 5 of the second edition of the BI-RADS ultrasound lexicon in diagnosing breast lesions. *Ultrasound Med Biol.* 2016;42(9):2065–2071. doi:10.1016/j.ultrasmedbio.2016.04.020

23. Van Holsbeke C, Yazbek J, Holland TK, et al. Real-time ultrasound vs. evaluation of static images in the preoperative assessment of adnexal masses. *Ultrasound Obstet Gynecol.* 2008;32(6):828–831. doi:10.1002/uog.6214

24. Raza S, Chikarmane SA, Neilsen SS, Zorn LM, Birdwell RL. BI-RADS 3, 4, and 5 lesions: value of US in management–follow-up and outcome. *Radiology.* 2008;248(3):773–781. doi:10.1148/radiol.2483071786

25. Berg WA, Cosgrove DO, Dore CJ, et al. Shear-wave elastography improves the specificity of breast US: the BE1 multinational study of 939 masses. *Radiology.* 2012;262(2):435–449. doi:10.1148/radiol.11110640

Supplementary materials

Table S1 Sensitivity, specificity, PPV and NPV of real-time and static image ultrasonography

Group	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Real-time US	98.9	58.2	44.8	99.4
Static image US	98.9	57.1	44.1	99.3

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; US, ultrasonography.

Table S2 Agreement for category/subcategory of real-time and static image ultrasonography

BI-RADS	ASE	κ value	95% confidence limits	
			Upper	Lower
Subcategory (3, 4A-4C and 5)	0.020	0.43	0.39	0.47
Category (3, 4 and 5)	0.023	0.48	0.44	0.53
Dichotomy (3 and 4/5)	0.026	0.56	0.51	0.61

Abbreviation: ASE, asymptotic standard error.