ORIGINAL RESEARCH

A National Chinese Survey on Ultrasound Feature Interpretation and Risk Assessment of Breast Masses Under ACR BI-RADS

Wen Wen [1]
Jingyan Liu[1]
Junren Wang[2]
Heng Jiang[3]
Yulan Peng[1]

[1]Department of Ultrasound, West China Hospital, Sichuan University, Chengdu, Sichuan, People's Republic of China; [2]Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, Sichuan, People's Republic of China; [3]College of Medicine, University of Nebraska Medical Center, Omaha, NE, USA

Purpose: Through this nationwide survey on ACR BI-RADS including ultrasound images of 10 selected breast lesions, we aimed to learn about consistency in feature interpretation and assessment categories and to identify factors that might contribute to inconsistencies, thereby promoting the application of BI-RADS in China.

Materials and Methods: The survey was delivered through a self-developed website about blinded image interpretation and was released to the public through online platforms and social media. A total of 10 representative lesions were selected by an experienced radiologist to gather information about the general practice of BI-RADS lexicons and categories. The Kappa statistic, the chi-squared test, and descriptive statistics were used for data analysis.

Results: Nine hundred ultrasound workers completed the questionnaire, coming from all provinces and major cities in China. They had different positions, grades of work organization, and seniority. The interrater agreement of BI-RADS features was fair to substantial (kappa value: 0.37–0.66). For BI-RADS categories, the highest agreement was observed in the typical benign group (average constituent rate = 74.78%), and generally lower agreement was observed in the typical malignant (average constituent rate = 36.03%) and suspicious groups (average constituent rate = 39.02%).

Conclusion: We found inconsistencies in BI-RADS applications, providing direction for image feature research using big data. Therefore, we call for more efforts to improve the consistency of BI-RADS application and provide an evidence-based basis for identifying benign and malignant lesions by sonographic features.

Keywords: interrater agreement, BI-RADS lexicons, breast ultrasound, national survey

Introduction

Breast cancer, with approximately 2.26 million new cases in 2020, has become the most common malignancy in humans and the leading cause of cancer death in women worldwide.[1] Breast cancer is a heterogeneous group of tumors with over 20 histological types and various biological features, making it challenging to accurately diagnose. Imaging techniques play a critical role in beast screening and diagnosis, especially mammography and ultrasonography. While mammography detects up to 98% of carcinomas in fatty breasts, its sensitivity declines to as low as 48% in extremely dense breasts.[2] Higher-density breast tissue has been associated with a higher risk of breast cancer,[3] and the incidence of dense breast cancer in Asians is higher than that in women from other ethnic groups.[4] Ultrasonography is a low-cost imaging modality that increases the sensitivity and detection rates of

Correspondence: Yulan Peng
Department of Ultrasound, West China Hospital, Sichuan University, No. 37, Guo Xue Road, Chengdu, 610041, Sichuan, People's Republic of China
Tel +86 189-8060-1606
Email yulanpeng@163.com

cancers in dense breast populations. China has a relatively higher proportion of dense breast lesions than other countries,[5] which explains the popularity and importance of ultrasound for Chinese breast screening.

The Breast Imaging Reporting and Data System (BI-RADS) was developed by the American College of Radiology to standardize the sonographic feature interpretation and comprehensive risk assessment of breast lesions identified on images and provide corresponding management recommendations for different BI-RADS categories. The emergence of BI-RADS significantly improved the management of breast abnormalities and made both lesion reporting and clinical recommendations more consistent. The first edition of BI-RADS, created in 1992,[6] mainly introduced the practice of standardized reporting in mammography, and the fourth (2003)[7] edition introduced ultrasound and MRI standardized reporting. The latest, 5th edition of BI-RADS,[8] revised in 2013 and published in 2014, harmonized the terminology across imaging modalities and has been widely recognized in China and all over the world. The ultrasound-related content in BI-RADS has also been interpreted by Chinese ultrasound experts through various academic conferences or articles. However, the popularity of BI-RADS is still restricted in China by the inconsistencies in sonographic feature interpretation by hospitals and doctors of different levels.

To promote the application of BI-RADS, a breast ultrasound training system and a related website for blind ultrasound image interpretation were developed with the support of big data and artificial intelligence technology. A national survey based on the system was released to the public online. The aim of this nationwide survey was to learn about the consistency of BI-RADS application, including feature interpretation and risk assessment, and to identify factors that account for the inconsistencies.

Materials and Methods
Survey Implementation
This prospective cross-sectional study was approved by the institutional review board (West China Hospital, Sichuan University). Informed consent was obtained from all the survey participants when they completed the online questionnaire. The survey was filled out through a self-developed website about blinded ultrasound image interpretation and was released to the public through online platforms and social media (WeChat etc.) from December 17, 2019, to December 20, 2020

(Supplementary Figure 1). The survey collected information about the agreement and disagreement on breast ultrasound according to ACR BI-RADS 5th Edition,[9] including shape (round, oral and irregular), orientation (parallel or not), margin (circumscribed or not), echo pattern (anechoic, hyperechoic, complex cystic and solid, hypoechoic, isoechoic and heterogeneous), posterior features (no posterior features, enhancement, shadowing, or combined pattern), and calcification (yes or no), and the final assessment based on sonographic features. All respondents were instructed to choose the most appropriate term to describe each lesion, without a clinical history or histopathologic report. Pathology diagnoses were made based on image-guided core-needle biopsy and surgical specimens. Associated features and surrounding tissue were not included because only static images were provided.

The lesions included in this survey were selected by an experienced radiologist specializing in breast ultrasound to ensure representativeness. A total of 10 breast lesions with pathology results were selected (Supplementary Figures 2–11), consisting of 3 typical benign tumors, 3 typical malignant tumors, 4 suspicious lesions according to the BI-RADS risk assessment based on expert consensus. Typical benign tumors were originally diagnosed as BI-RADS 3 or 2 masses with benign pathology results; typical malignant tumors were diagnosed as BI-RADS 5 masses with malignant pathology results; and suspicious tumors were diagnosed as BI-RADS 4 masses with benign or malignant pathology results. After localizing the target breast mass by using ultrasound, representative B-mode scans of the mass were obtained, and color Doppler imaging was performed in all cases. The names of the respondents were deidentified when analyzing the data.

Data Analysis
All statistical and graphical analyses were conducted with SPSS system (Version 21, IBM SPSS Statistics) and R (Version 4.0.3). Kappa statistics were calculated to assess interrater reliability in BI-RADS descriptors. The method for estimating the kappa value in the case of multiple observers and multiple categories is based on the work of Fleiss.[10] A value of $\kappa = 1.0$ corresponds to complete agreement, 0 to no agreement, and less than 0 to disagreement. Landis and Koch[11] suggested that a kappa value (κ) of equal to or less than 0.20 indicates 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.

Additionally, the agreement of each lesion and BI-RADS lexicons was analyzed by descriptive statistics. To measure the agreement of each BI-RADS lexicon in each lesion, we defined a new statistical term, "constituent rate", as the percentage of most selected options in every question when doing statistical analysis. The chi-squared test was used to compare differences in agreement between sonographers and ultrasound doctors.

Results
Respondent Characteristics
In this survey, a total of 930 ultrasound workers completed the online questionnaire. After excluding the respondents without complete basic information, 900 answers were included in our analysis of agreement and disagreement on BI-RADS descriptors and classifications. The respondents came from 31 provinces and 219 cities, covering all the provinces and major cities of mainland China (Figure 1A and B). The basic characteristics of the respondents are listed in Table 1. Some 225 (25%) were male and 675 (75%) were female, which is probably consistent with the sex ratio of ultrasound workers in China. According to the three-tier system of hospitals in China, the organizations the respondents worked for covered all grades of medical centers in China (511 from tertiary hospitals, 253 from secondary hospitals, and 136 from primary hospitals). As for working seniority, the respondents had three levels: 120 senior workers (experience in ultrasound >10 years), 334 middle-seniority workers (experience in ultrasound >5 years), and 446 junior workers (experience in ultrasound >2 years). Ultrasound doctors (736) and sonographers (164) answered the questionnaire based on their understanding of BI-RADS when the survey was available. Therefore, the results

from the survey were representative of the overall application of BI-RADS in China.

BI-RADS Features
The interrater reliability of the individual BI-RADS lexicons was evaluated. We saw substantial agreement about the margins of tumor masses (Fleiss's kappa = 0.66). Moderate agreement among the respondents was observed for the description of calcification (Fleiss's kappa = 0.60), shape (Fleiss's kappa = 0.58), and posterior acoustic changes (Fleiss's kappa = 0.43), whereas fair agreement was obtained for orientation (Fleiss's kappa = 0.38) and echo pattern (Fleiss's kappa = 0.37).

The descriptor "margin" proved to be the most consistent feature among all relevant sonographic features in this survey, with the highest average constituent rate (89.10%, Figure 2). Lower constituent rates were observed for lesion 5 and lesion 10 (constituent rate = 72.11% and 89.10%, respectively, Figure 2). The percentage of respondents believing that lesion 5 (Figure 3A and B) was not circumscribed was 72% (n = 649). In addition, 63% of respondents considered lesion 10 (Figure 3C and D) to have circumscribed margins, while 37% did not.

Sonographers and ultrasound doctors did not show significant differences when assessing the margins of lesions. The survey showed that the constituent rates of most lesions were higher than 60% (average 84.42%) when describing the shape of the breast mass, except lesion in 10 (constituent rate = 46.44%). For the shape of lesion 10 (Figure 3C and D), 46% of respondents (n = 418) answered that it was a round mass, whereas 38% of respondents (n = 346) thought it was oval and 15% of respondents (n = 136) thought it was irregular. Regarding
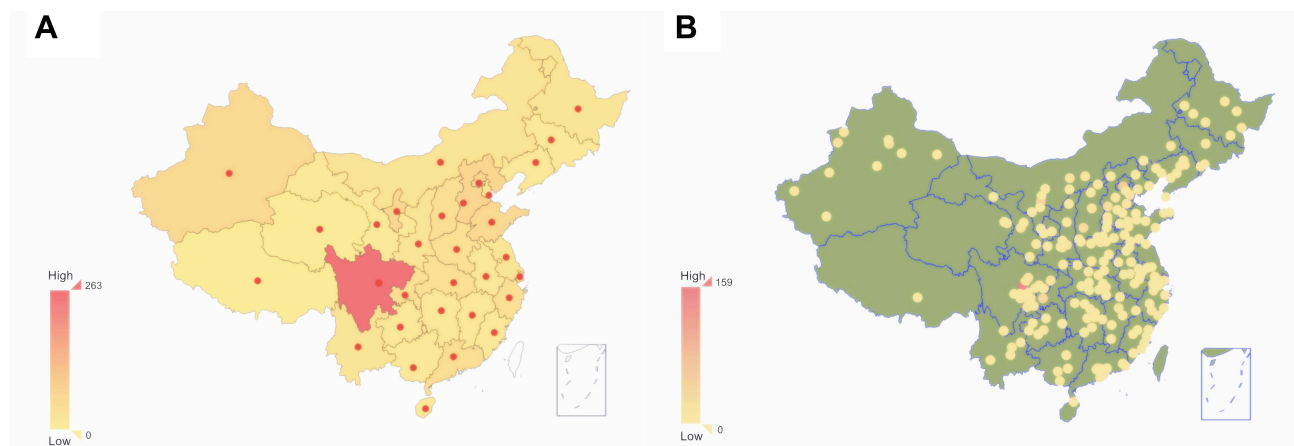
Figure 1 (**A**) Response rate (provinces); (**B**) response rate (cities).

Table 1 Respondents Characteristics

Characteristics		No. of Respondents	Count (%)
Gender	Male	225	25.00%
	Female	675	75.00%
Position	Doctor	736	81.78%
	Sonographer	164	18.22%
Grades of Hospital	Tertiary	511	56.78%
	Secondary	253	28.11%
	Primary	136	15.11%
Seniority	Senior (>10 years)	120	13.33%
	Middle (>5 years)	334	37.11%
	Junior (>2 years)	446	49.56%

Abbreviations: ACR, American College of Radiology; BI-RADS, Breast Imaging Reporting and Data System.

the comparison of sonographers and ultrasound doctors, no significant difference was found in these BI-RADS descriptors.

For calcification, the constituent rates of lesions 3 and 10 were lower than 75% when evaluating calcification (constituent rate = 51.44% and 65.56%, respectively). In detail, 51% of respondents (n = 463) did not find calcification in lesion 3, whereas the others found macrocalcification and microcalcification (n = 57, 6% and n = 384, 43%, respectively). For lesion 10, 523 respondents (58%) recognized microcalcification, 310 respondents (34%) did not, and 96 respondents (11%) found macrocalcification inside. Divergence between the sonographer and ultrasound worker groups was seen in the image interpretation of lesions 2 and 3 ($p < 0.05$).

The interpretation of posterior acoustic changes had the lowest mean constituent rate (73.30%), lesions 1 and 9 showing extremely low agreement. Similar numbers of respondents believed that there were no acoustic changes and that there was enhancement in the posterior area (n = 415, 46% and n = 428, 48%, respectively) of lesion 1. Assessments of lesion 9 were divergent because only 46% of respondents did not find acoustic changes, while the others did, especially posterior shadowing (n = 319, 35%).

Fair agreement was observed for the orientation and echo pattern of breast lesions in interrater reliability analysis. Regarding the assessment of orientation, the lowest
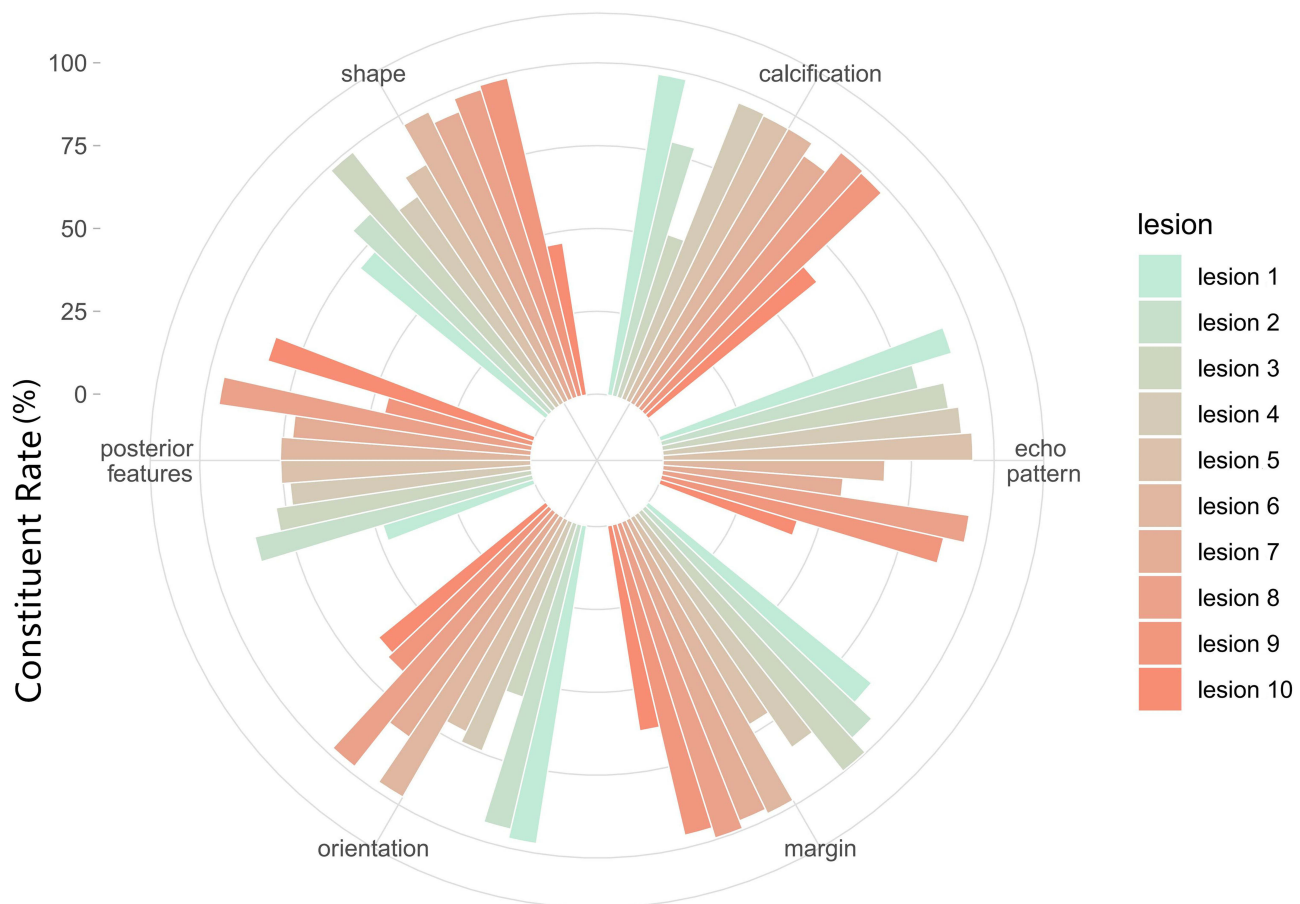
Figure 2 Constituent rate of each lesion and BI-RADS lexicon.
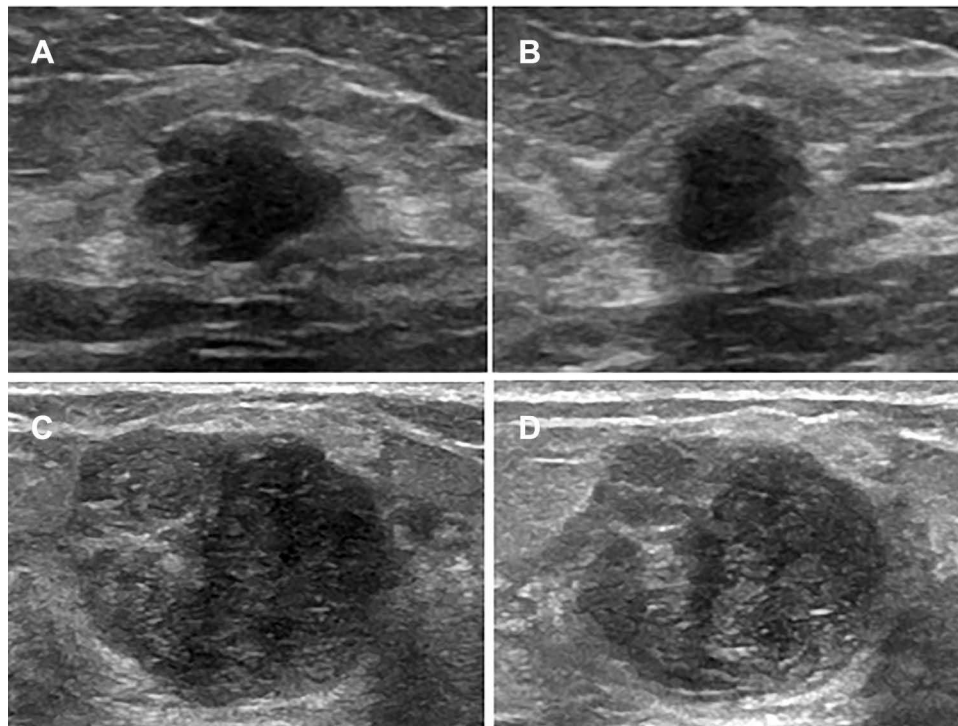
Figure 3 (**A**) Long-axis plane of lesion 5; (**B**) plane perpendicular to the long-axis plane of lesion 5; (**C**) long-axis plane of lesion 10; (**D**) plane perpendicular to the long-axis plane of lesion 10.

agreement was obtained in lesion 3 (constituent rate = 54.89%), 494 respondents (55%) thinking it was parallel and the others (n = 406, 45%) not. Of the echo pattern estimates, "heterogeneous" was the most controversial selection, accounting for low constituent rates in lesion 7 and lesion 10 (n = 352, 39% and n = 389, 43%, respectively).
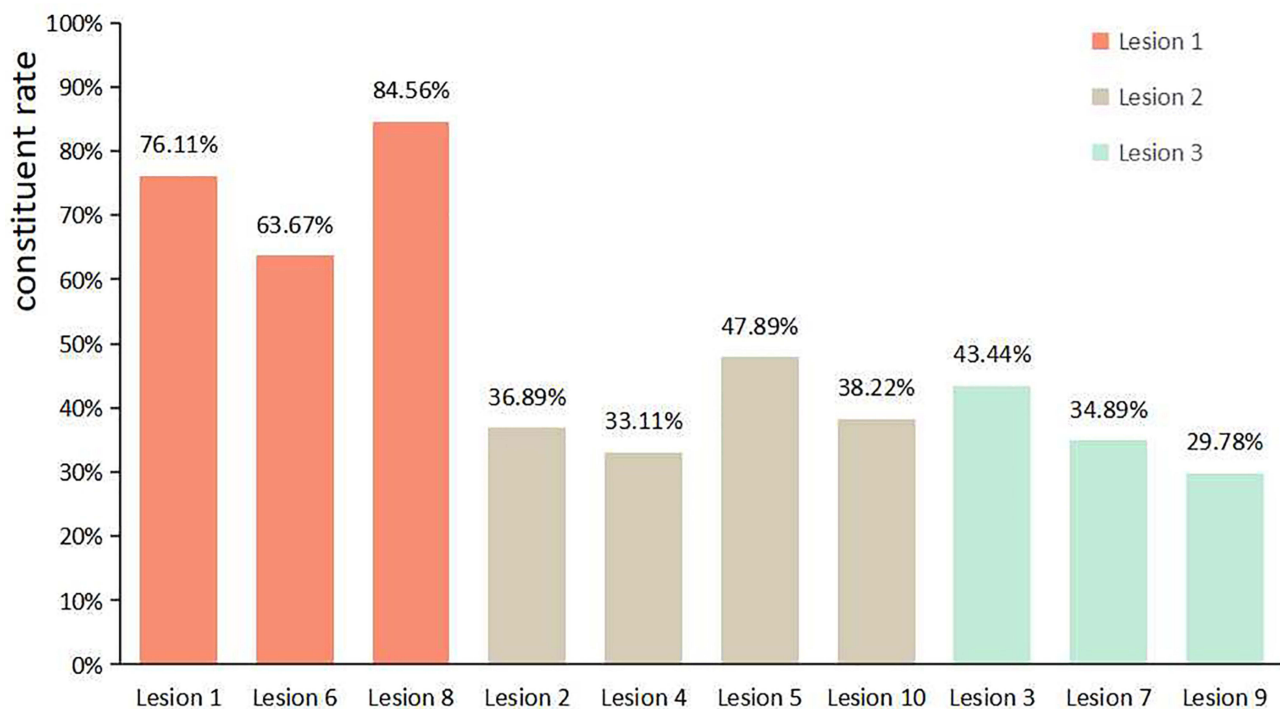
BI-RADS Assessment

The agreement of the final BI-RADS assessment is shown in Figure 4. Of the 10 lesions, 5 were malignant and 5 were benign according to the pathology diagnoses (5 invasive ductal carcinomas, 2 fibroadenomas, 1 fibroadenoma with adenosis, 1 fibroepithelial tumor, and 1 cyst). The lesions were divided into three groups according to their original BI-RADS assessment: typical benign (lesions 1, 6, and 8), typical malignant (lesions 3, 7, and 9), and suspicious (lesions 2, 4, 5, and 10). The typical benign group had higher agreement than other groups (constituent rates = 76.11%, 63.67%, and 84.56%, respectively), but a few respondents had problems evaluating category 3 or 4a lesions, especially in lesion 6 (n = 260, 29%, for category 4a). However, a generally lower constituent rate was observed in the evaluation of the typical malignant and suspicious groups. For suspicious lesions, controversy

mainly arose in differentiating the subcategorization of BI-RADS 4, the majority of respondents evaluating lesion 2 as BI-RADS 4a (n = 332, 37%), lesion 4 as BI-RADS 4b (n = 298, 33%), lesion 5 as BI-RADS 4a (n = 431, 48%), and lesion 10 as 4a (n = 344, 38%). The final risk assessment of lesion 9 obtained the lowest constituent rate (29.78%) in the typical malignant group, the numbers of respondents classifying it as BI-RADS 4a, 4b, and 4c were similar (n = 267, 30%; n = 268, 30%; and n = 202, 22%, respectively).

Discussion

As a structural reporting platform and data tracking system, BI-RADS provides standard sonographic feature descriptions, assessment categories, and management recommendations. A lexicon is a list of standardization terms used to describe imaging findings concisely and reproducibly,[12] and harmonization of terminology was reached across imaging modalities in the 5th Edition of BI-RADS. Consistent use of the BI-RADS lexicon terminology and assessment categories allows result communication, accurate risk evaluation of malignancies, and appropriate recommendations to patients and clinicians. Several studies have reported the intrarater and interrater reliability between two observers with similar ultrasound

Figure 4 Constituent rate of the final BI-RADS assessment categories of each lesion.

experience and education backgrounds, in which the kappa value was relatively higher in all descriptors.[13–18] However, the interrater reliability of most important features and assessment categories was at most moderate in a recent study.[19] Perfect agreement was not seen in our study (Kappa value ranges from 0.37 to 0.66) when we assessed the interrater reliability of BI-RADS lexicons. In addition, distinguishing between BI-RADS 3 and 4a, between BI-RADS 4c and 5, and between subcategories of BI-RADS 4 were the main points that confused ultrasound workers, clinicians, and patients in the survey. Malignancy descriptors are more useful for predicting malignancy when combined than when used alone.[20–22] Some rules have been suggested to make consistent assessments according to sonographic features,[19,23] and they have excellent diagnostic performance and variability. Therefore, the standardization of BI-RADS lexicons seems necessary because it can not only reduce operator dependency and enhance the consistency of image interpretation but also improve diagnostic accuracy of breast cancer. In the current survey, several disagreements that may contribute to the divergence were revealed and should be further discussed and researched.

A mass is a three-dimensional object, and BI-RADS stipulates that it should be seen in two different planes on two-dimensional ultrasound.[9] It is hard to make a decision when the observer cannot see the same features are seen in different planes, especially when evaluating the shape, margin, and orientation of breast lesions. This could explain the lower agreement for these features of lesions 5 and 10 (Figure 3). In BI-RADS, the long axis of the lesion that is parallel to the skin surface is defined as the parallel orientation.[9] We assumed that the insufficient understanding of BI-RADS by the respondents might lead to the misidentification of orientation, which periodic BI-RADS training of Chinese ultrasound workers with different seniority could overcome. Berg et al[24] demonstrated that BI-RADS training improved agreement even among experienced breast observers. To raise consistency, we suggest evaluating orientation in the long-axis planes, but studies on nonparallel planes should be conducted to determine whether they are associated with malignancy. Unlike for orientation, there is no suggestion about which plane is standard for evaluating and reporting sonographic features. Some descriptors — nonparallel orientation of the mass, spiculated margins, angular margins, microcalcifications, or posterior shadowing — have been associated with malignancy.[19,25–27] To address the disagreements about these features, we suggested that malignant descriptors might probably be considered first

when more than one descriptor of BI-RADS lexicons simultaneously exists, whereas the recommendation is not evidence-based until the diagnostic performance of these adjustments is further studied by ultrasound scholars.

Slight local changes may exist in certain planes when evaluating some lexicons, which leads to different features existing in one plane. We suppose that partly irregular changes and indistinct margins may account for the lower agreement of shape and margin in lesion 10 (Figure 3C and D), and partly posterior shadowing may explain the lower agreement of posterior acoustic changes in lesion 9 (Supplementary Figure 10). Indistinct margins are defined as no clear demarcation between a mass and the surrounding tissue anywhere on the margin.[9] Thus, we suggest that lesions with partial changes in margin be interpreted as uncircumscribed, following their original definition. Regarding other lexicons that might give rise to partial changes in interpretation, such as shape and posterior acoustic changes, we think that importance should be attached to probably malignant sonographic features. More clinical and imaging studies are needed to confirm the diagnostic performance and effectiveness of this principle.

Orientation is an important descriptor in differentiating benign from malignant lesions, as parallel orientation is highly correlated with malignancy,[19,25–27] but only qualitative criteria for judgment are regulated in BI-RADS. Compared to the agreement obtained in studies with kappa values of 0.56–0.70,[13–17,27] fair agreement on orientation was found in this study. In addition to the interference of different planes, the long axis of the lesion was neither parallel nor perpendicular to the skin line in some lesions, and there was a certain angle between the long axis and the surface of the skin. Thus, quantitative indices such as angles or aspect ratios would seem helpful for improving the consistency of orientation. Certainly, further research should be conducted to identify the diagnostic performance and reproducibility of these indices. A relevant project on artificial intelligence and computer-aided evaluation of orientation is underway by our team.

Heterogeneous echogenicity of a solid mass was added to the ACR BI-RADS 5th Edition, and although it has had little diagnostic value in differentiating between benign and malignant lesions, it may elevate the suspicion for malignancy, especially when seen with noncircumscribed margins and irregular shapes.[9] With several studies reporting low variability for echogenicity (kappa 0.26–0.58),[18,28–30] the echo pattern was the most inconsistent descriptor in this study.

However, we found that "heterogeneous" was the selection that mostly interfered with the choices of observers when analyzing all selections in echo patterns. Heterogeneous echogenicity is actually not mutually exclusive with other options in a single-choice matter such as the echo pattern. Yang et al found no significant difference between homogenous and heterogeneous echo patterns in malignant lesions.[31] Thus, we believe that eliminating "heterogeneous" echo patterns and adding "internal echogenic homogeneity of solid mass" to the BI-RADS lexicon would make for a good new descriptor of heterogeneous masses that would let observers describe masses more reproducibly.

The respondents differed from the expert consensus when describing the margin of lesion 5 (Figure 3A and B). That is, 649 (72%) respondents believed it was uncircumscribed, a majority selecting microlobular, whereas experts in the research team thought it was a macrolobulated mass. The misidentification between micro- and macrolobules might lead to different interpretations of margin and shape and even different evaluations of malignant risk. Microlobules are defined as margins characterized by short-cycle undulations, and gently lobulated or macrolobulated margins are not clearly defined in BI-RADS. Confusion in distinguishing them might result in difficulty in interpreting shape and margin: lesions with microlobular shapes are uncircumscribed, and lesions with macrolobular shapes are circumscribed with oval or irregular shapes. We hope this question will be discussed more to enhance the consistency of interpretation and diagnosis in the future.

This study has several limitations. First, the number of lesions in the survey was small, which may affect the interrater reliability analysis. Second, as the products of a real-time scanning method, only static images were included. This will affect the confidence of interpretation and diagnosis and indirectly affect the results of BI-RADS lexicons and assessment categories. To maintain a high response rate, 10 representative lesions were selected by experts in breast imaging, and the total number of questions was kept to be less than 100, which is why only a small number of lesions and static images were included in the questionnaire. Additionally, we intended to reveal some disagreements deserving discussion found through this survey, rather than finding all the reasons for all inconsistencies. Third, we offer some suggestions on how to improve the consistency of image interpretation, without data supporting them. The prospective study design and online questionnaire

modality did not allow retests in the same population to control for all variables that could influence the outcomes. We call for larger prospective studies on feature interpretation and risk assessment to improve the consistency of BI-RADS application, including studies that cover the disagreements we mentioned above.

Conclusion

Although BI-RADS has been highly accepted worldwide, the authority of a popular guideline comes largely from its stability. More than an application guideline, it should become a model for this field. This survey reflects some disagreement deserving further attention in BI-RADS lexicon descriptions, and we call for more participation in scientific studies and education to overcome the challenges of reducing operator dependency during ultrasound examinations and raising the integral consistency of breast ultrasound evaluations in the future, including discussion and research on the inconsistencies mentioned above. These disagreements provide some directions for image feature research on big data, and we hope that further feature-related big data research can provide an empirical basis for identifying both benign and malignant features.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Wild C, Weiderpass E, Stewart BW. World cancer report: cancer research for cancer prevention. Lyon: International Agency for Research on Cancer. 2020.

2. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology*. 2002;225(1):165–175. doi:10.1148/radiol.2251011667

3. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2006;15(6):1159–1169. doi:10.1158/1055-9965.EPI-06-0034

4. Del Carmen MG, Halpern EF, Kopans DB, et al. Mammographic breast density and race. *AJR Am J Roentgenol*. 2007;188(4):1147–1150. doi:10.2214/AJR.06.0619

5. Yang Y, Liu J, Gu R, et al. Influence of factors on mammographic density in premenopausal Chinese women. *Eur J Cancer Prev*. 2016;25(4):306–311. doi:10.1097/CEJ.0000000000000177

6. American College of Radiology. Breast Imaging Reporting and Data System (BIRADS). Reston, VA; 1993.

7. Mendelson E, Baum J, Berg W, Merritt C, Rubin E. Breast imaging reporting and data system (BI-RADS®). American College of Radiology; 2003.

8. D'Orsi C, Morris E, Mendelson E. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. American College of Radiology; 2013.

9. Mendelson EB, Böhm-Vélez M, Berg WA. ACR BI-RADS® Ultrasound. In: *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. American College of Radiology; 2013.

10. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76(5):378–382. doi:10.1037/h0031619

11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. doi:10.2307/2529310

12. Eghtedari M, Chong A, Rakow-Penner R, Ojeda-Fournier H. Current status and future of BI-RADS in multimodality imaging, from the AJR special series on radiology reporting and data systems. *AJR Am J Roentgenol*. 2021;216(4):860–873. doi:10.2214/AJR.20.24894

13. Schwab F, Redling K, Siebert M, Schötzau A, Schoenenberger CA, Zanetti-Dällenbach R. Inter- and intra-observer agreement in ultrasound BI-RADS classification and real-time elastography Tsukuba score assessment of breast lesions. *Ultrasound Med Biol*. 2016;42(11):2622–2629. doi:10.1016/j.ultrasmedbio.2016.06.017

14. Lee YJ, Choi SY, Kim KS, Yang PS. Variability in observer performance between faculty members and residents using Breast Imaging Reporting and Data System (BI-RADS)-ultrasound, fifth edition (2013). *Iran J Radiol*. 2016;13(3):e28281. doi:10.5812/iranjradiol.28281

15. Park CS, Kim SH, Jung NY, Choi JJ, Kang BJ, Jung HS. Interobserver variability of ultrasound elastography and the ultrasound BI-RADS lexicon of breast lesions. *Breast Cancer*. 2015;22(2):153–160. doi:10.1007/s12282-013-0465-3

16. Lee HJ, Kim EK, Kim MJ, et al. Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound. *Eur J Radiol*. 2008;65(2):293–298. doi:10.1016/j.ejrad.2007.04.008

17. Park CS, Lee JH, Yim HW, et al. Observer agreement using the ACR Breast Imaging Reporting and Data System (BI-RADS)-ultrasound, first edition (2003). *Korean J Radiol*. 2007;8(5):397–402. doi:10.3348/kjr.2007.8.5.397

18. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. *Radiology*. 2006;239(2):385–391. doi:10.1148/radiol.2392042127

19. Pfob A, Barr RG, Duda V, et al. A new practical decision rule to better differentiate BI-RADS © 3 or 4 breast masses on breast ultrasound. *J Ultrasound Med*. 2021. doi:10.1002/jum.15722

20. Liberman L, Abramson AF, Squires FB, Glassman JR, Morris EA, Dershaw DD. The breast imaging reporting and data system: positive predictive value of mammographic features and final assessment categories. *AJR Am J Roentgenol*. 1998;171(1):35–40. doi:10.2214/ajr.171.1.9648759

21. Bent CK, Bassett LW, D'Orsi CJ, Sayre JW. The positive predictive value of BI-RADS microcalcification descriptors and final assessment categories. *AJR Am J Roentgenol*. 2010;194(5):1378–1383. doi:10.2214/AJR.09.3423

22. Burnside ES, Ochsner JE, Fowler KJ, et al. Use of microcalcification descriptors in BI-RADS 4th edition to stratify risk of malignancy. *Radiology*. 2007;242(2):388–395. doi:10.1148/radiol.2422052130

23. Yao MM, Joe BN, Sickles EA, Lee CS. BI-RADS category 5 assessments at diagnostic breast imaging: outcomes analysis based on lesion descriptors. *Acad Radiol*. 2019;26(8):1048–1052. doi:10.1016/j.acra.2018.07.018

24. Berg WA, D'Orsi CJ, Jackson VP, et al. Does training in the Breast Imaging Reporting and Data System (BI-RADS) improve biopsy recommendations or feature analysis agreement with experienced breast imagers at mammography? *Radiology*. 2002;224(3):871–880. doi:10.1148/radiol.2243011626

25. 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

26. Lawson BT, Vinnicombe S, Whelehan P, Macaskill EJ, Sim YT, Evans A. Associations between the ultrasound features of invasive breast cancer and breast cancer specific survival. *Clin Radiol*. 2020;75(11):879.e813–879.e821. doi:10.1016/j.crad.2020.07.012

27. Korpraphong P, Tritanon O, Tangcharoensathien W, Angsusinha T, Chuthapisith S. Ultrasonographic characteristics of mammographically occult small breast cancer. *J Breast Cancer*. 2012;15 (3):344–349. doi:10.4048/jbc.2012.15.3.344

28. Santana Montesdeoca JM, Gómez Arnáiz A, Fuentes Pavón R, Alemán Flores P, Limiñana Cañal JM, Jorrín Moreno A. [Diagnostic accuracy and interobserver variability in the BI-RADS ultrasound system]. *Radiologia*. 2009;51(5):477–486. Spanish. doi:10.1016/j.rx.2009.03.001

29. Abdullah N, Mesurolle B, El-Khoury M, Kao E. Breast imaging reporting and data system lexicon for US: interobserver agreement for assessment of breast masses. *Radiology*. 2009;252(3):665–672. doi:10.1148/radiol.2523080670

30. Hong AS, Rosen EL, Soo MS, Baker JA. BI-RADS for sonography: positive and negative predictive values of sonographic features. *AJR Am J Roentgenol*. 2005;184(4):1260–1265. doi:10.2214/ajr.184.4.01841260

31. Yang Y, Hu Y, Shen S, et al. A new nomogram for predicting the malignant diagnosis of Breast Imaging Reporting and Data System (BI-RADS) ultrasonography category 4A lesions in women with dense breast tissue in the diagnostic setting. *Quant Imaging Med Surg*. 2021;11(7):3005–3017. doi:10.21037/qims-20-1203