# Variability in Observer Performance Between Faculty Members and Residents Using Breast Imaging Reporting and Data System (BI-RADS)-Ultrasound, Fifth Edition (2013)

## Youn Joo Lee,[1] So Young Choi,[2] Kyu Sun Kim,[2] and Po Song Yang[1,*]

[1]Department of Radiology, Daejeon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Daejeon, Republic of Korea
[2]Department of Radiology, Eulji University Hospital, Daejeon, Republic of Korea

[*] *Corresponding author*: Po Song Yang, Department of Radiology, Daejeon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Daejeon, Republic of Korea. Tel: +82-422209700, Fax: +82-422209087, E-mail: psyrad@naver.com

## Abstract

**Background:** Ultrasonography (US) is a useful tool for breast imaging, yet is highly operator-dependent.

**Objectives:** We evaluated inter-observer variability and performance discrepancies between faculty members and radiology residents when describing breast lesions, by the fifth edition of breast imaging reporting and data system (BI-RADS)-US lexicon, and then attempted to identify whether inter-observer variability could be improved after one education session.

**Patients and Methods:** In total, 50 malignant lesions and 70 benign lesions were considered in our retrospective study. Two faculty members, two senior residents, and two junior residents separately assessed the US images. After the first assessment, the readers received one education session, and then reassessed the images in a random order. Inter-observer variability was measured using the kappa coefficient ($\kappa$). Performance discrepancy was evaluated by receiver operating characteristic (ROC) curves.

**Results:** For the faculty members, fair-to-good agreement was obtained in all descriptors and final assessment, while for residents, poor-to-moderate agreement was obtained. The areas under the ROC curves were 0.78 for the faculty members, 0.59 for the senior residents, and 0.52 for the junior residents, respectively. Diagnostic performance was significantly higher in the faculty members than the senior and junior residents (P = 0.0001 and < 0.0001, respectively). After one education session, the agreement in the final assessment was one level higher in the faculty members and senior residents, yet in the senior residents, the degree of agreement was still only fair. Moreover, in the junior residents, there was no improvement.

**Conclusion:** Investigative assessment of breast US by residents is inadvisable. We recommend continued professional resident training to improve the degree of agreement and performance.

*Keywords:* Breast, Ultrasonography, Neoplasms, Faculty, Residency

## 1. Background

Ultrasonography (US) is an indispensable tool in breast imaging, and complements both mammography and magnetic resonance imaging of the breast (1-4). It increases the specificity of mammography; in particular, both the number of false-negative lesions in dense breasts and false-positive lesions that could lead to unnecessary biopsies are reduced (5, 6). However, lack of reproducibility concerning lesion assessment and operator-dependency of breast US interpretation still remains problematic (6-12). The breast imaging, reporting and data system (BI-RADS) was published by the American college of radiology (ACR) to offer a common description for the characterization of breast abnormalities, and to communicate effectively with the physician (5, 13-16).

At some hospitals, US images are examined by skilled faculty members. At resident-training hospitals, however, the examinations may be conducted by residents. Part of the lack of reproducibility may be the product of inherent operator skill, which represents a small but unavoidable limitation for US as a modality. More commonly, however, the inconsistencies result from disparities in lesion assessment. Misunderstanding about the BI-RADS US lexicon or the lack of clinical practice may disturb accurate assessment.

Many reports (17-21) have published the rates of discrepancies between radiology residents and experienced radiologists in emergency diagnostic imaging examinations of the head and abdomen. However, to the best of our knowledge, a few studies have reported the difference in the interpretation of breast US images between residents and faculty radiologists. Thus, we evaluated the discrepancy between faculty members and radiology residents in breast US interpretations.

## 2. Objectives

The aim of our study was to evaluate inter-observer variability and performance discrepancies between faculty members and radiology residents, when assessing breast lesions, using the fifth edition (2013) of the ACR BI-RADS US lexicon. We also examined whether inter-observer agreement would improve in a second assessment after one education session.

## 3. Patients and Methods

### 3.1. Patients and Imaging

This retrospective study was approved by the institutional review board of our institution. We selected cases from female patients, who had undergone bilateral whole breast US and US guided core needle biopsies or surgical excisions between January 2008 and August 2014 (Figure 1). All US imaging was scanned using high-resolution US equipment (IU22; Philips, Bothell, WA, USA) by a 12 MHz linear transducer. We recruited only histologically proven malignant and benign lesions. Benign lesions that underwent biopsies, had been radiographically stable for two years. We excluded lesions at risk of developing cancer, such as atypical lobular hyperplasia, atypical ductal hyperplasia, radial scar, papilloma, and phyllodes tumors. Ultimately, 120 lesions from 108 patients were selected for our study.

### 3.2. Study Protocol

The training program of residents at our institution includes six months of training in breast radiology imaging (three months for second-year residents, two additional months for third-year residents, and one additional month for fourth-year residents), scanning breast US and reporting using the ACR BI-RADS lexicon.

To measure performance discrepancy, we divided faculty members, senior residents and junior residents to three subgroups. We then analyzed inter-observer agreement and receiver operating characteristic (ROC) curves for each subgroup (22, 23). To estimate observer variability and performance discrepancies depending on years of practice, two faculty members with four and > ten years of breast imaging experience respectively, two junior residents who had completed the three months of breast imaging training, and two senior residents who had completed the full six months of training, assessed the images without a special education session.

To identify whether inter-observer agreement could be enhanced by one education session, the participant gathered for a day, after the first assessment. A total of six participants checked the fifth edition of BI-RADS-US lexicon in the education session, and then keenly discussed thirty cases that had been excluded in the present study. At four weeks after this education session, the original US images were rearranged randomly by the investigator, and then participants reassessed the images separately.

### 3.3. Image Assessment

We reviewed pre-existing breast US images retrospectively to compare the performance of the faculty members and the residents. For assessments of lesions, the participants examined at least two static US images including radial and anti-radial images or transverse and longitudinal images. Each participant chose the most proper terms to assess breast abnormalities, using the fifth edition of BI-RADS-US. The BI-RADS lexicon for US is presented in Box 1. Following the description of the lesion, the final category was divided to five groups, based on the BI-RADS-US: category 3 (probably benign), 4a (low suspicion for malignancy), 4b (moderate suspicion for malignancy), 4c (high suspicion for malignancy), and 5 (highly suggestive of malignancy). In the present study, agreement among observers for the presence of associated findings and special cases was not assessed, and we concentrated on US features of masses, such as shape, orientation, margin, echo pattern, and posterior features. For the designation of tissue composition, at least four static images including each quadrant of the ipsilateral breast were also provided. Mammographic images and clinical background were not given to eliminate the probability of bias in the assessment of the US images.

### 3.4. Statistical Analysis

The statistical analysis was performed using the SPSS software (version 11.5 for Windows; SPSS Inc. Chicago, IL, USA). The kappa coefficient ($\kappa$) was calculated to determine inter-observer variability for BI-RADS terminology and the final category between the two readers. The $\kappa$ values were interpreted, using the guidelines of Landis and Koch (24, 25): poor agreement, $\kappa = 0.00$-$0.20$; fair agreement, $\kappa = 0.21$ - $0.40$; moderate agreement, $\kappa = 0.41$ - $0.60$; good agreement $\kappa = 0.61$ - $0.80$; and excellent agreement, $\kappa = 0.81$-$1.00$. To determine the inter-observer reliability between faculty radiologists, senior residents, and junior residents, intraclass correlation coefficients were calculated with 95% confidence intervals.

The area under the receiver operating characteristic (ROC) curves (AUC) was analysed with 95% confidence Intervals (CI), to compare the diagnostic performance of faculty radiologists, senior residents, and junior residents, using MedCalc (version 10.1.6.0; MedCalc software, Ghent, Belgium). The sensitivity and specificity for each category
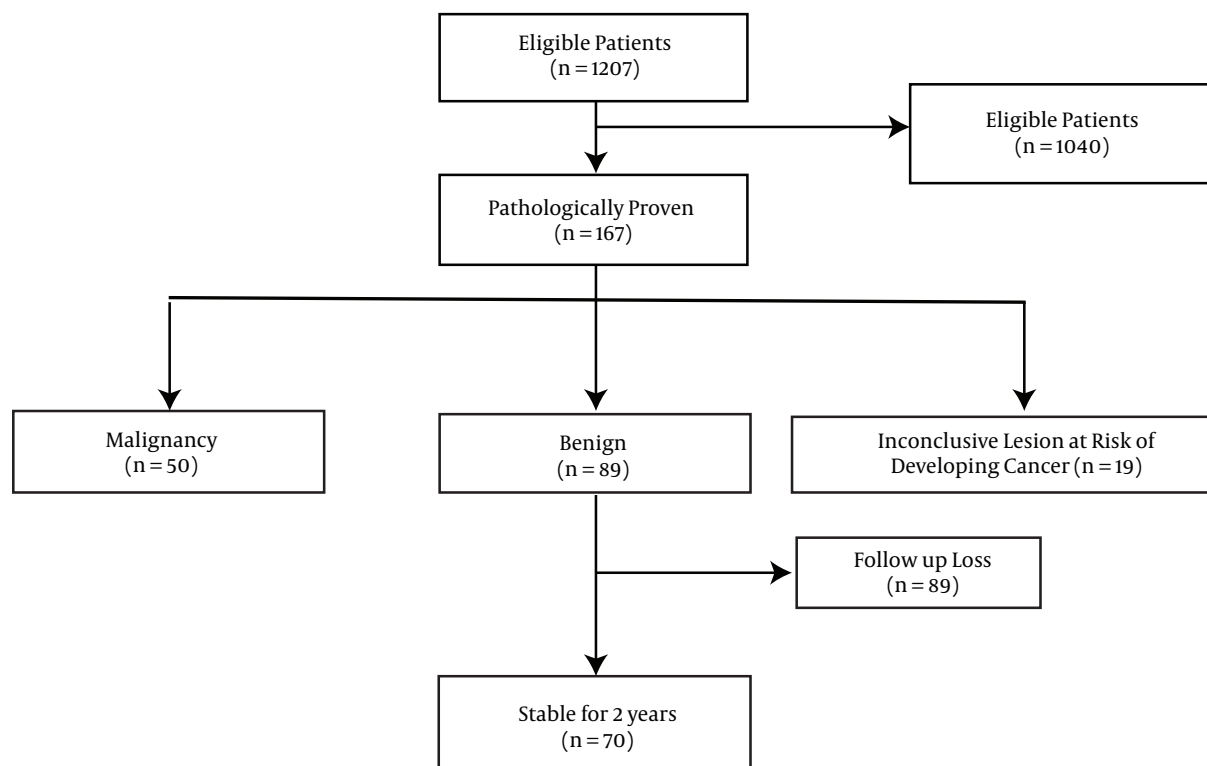
**Figure 1.** Study Flow Diagram

were also evaluated for a subgroup of the observers. A P value of < 0.05 was considered to indicate statistical significance.

## 4. Results

In total, 120 lesions in 108 patients met the selection criteria and were included in this study. The 108 patients ranged in age from 23 to 76 years (mean, 51.7 years). The size range of the lesions was 3 - 37 mm (mean, 17.4 mm) in the maximal dimension. Of the lesions, 50 were found to be malignant: invasive ductal carcinoma in 39 lesions, ductal carcinoma in situ in eight, invasive lobular carcinoma in two, and medullary carcinoma in one. The other 70 lesions were benign: fibroadenoma in 21 lesions, fibrocystic change in 19, stromal fibrosis in 12, ductal epithelial hyperplasia in 11, fibroadenomatous hyperplasia in five, fat necrosis in one, and adenosis in one.

### 4.1. Observer Agreement

Agreement between the faculty members was fair-to-good for all criteria; however, between residents, agreement was poor-to-moderate. The inter-observer variability for each subgroup (faculty, senior resident, junior resident) is presented in Table 1. For the two faculty members, inter-observer agreement was moderate for tissue composition, shape, posterior features and calcifications ($\kappa =$ 0.47, 0.49, 0.53 and 0.48, respectively) (95% CI: 0.35 - 59; 0.36 - 61; 0.42 - 0.65; 0.36 - 0.60), good for orientation ($\kappa = 0.66$) (95% CI: 0.59 - 75), and fair for margin, echo pattern, and final assessment ($\kappa =$ 0.32, 0.36 and 0.40, respectively) (95% CI: 0.15 - 47; 0.20 - 51; 0.25 - 0.52). For the two senior residents, inter-observer agreement was fair for tissue composition, shape, posterior features and calcifications ($\kappa =$ 0.36, 0.33, 0.40 and 0.39, respectively) (95% CI: 0.17 - 50; 0.20 - 47; 0.25 - 0.51; 0.23 - 55), moderate for orientation ($\kappa = 0.59$) (95% CI: 0.45 - 73), and poor for margin, echo pattern and final assessment ($\kappa =$ 0.11, 0.15 and 0.19, respectively) (95% CI: 0.02 - 0.21; 0.04 - 0.25; 0.07 - 0.30). For the two junior residents, inter-observer agreement was fair for shape, orientation and calcifications ($\kappa =$ 0.22, 0.30 and 0.23, respectively) (95% CI: 0.12 - 0.33; 0.20 - 0.41; 0.13 - 0.33). For all the remaining criteria such as tissue composition, margin, echo pattern, posterior features and final assessment, poor agreement was observed ($\kappa =$ 0.15, 0.18, 0.12 and 0.14, respectively) (95% CIs: 0.04 - 0.25; 0.07 - 0.28; 0.02 - 0.22 and

**Box 1.** The Contents of the New Breast Imaging, Reporting and Data System for Breast Sonography

**Tissue Composition (Screening Only)**

A. Homogeneous background echotexture-fat [a]

B. Homogeneous background echotexture-fibroglandular [a]

C. Heterogeneous background echotexture [a]

**Shape**

Oval

Round

Irregular

**Orientation**

Parallel

Nonparallel

**Margin**

Circumscribed

Noncircumscribed: indistinct, angular

Microlobulated, spiculated

**Echo pattern**

Anechoic

Hyperechoic

Complex cystic and solid [a]

Hypoechoic

Isoechoic

Heterogeneous [a]

**Posterior features**

No posterior features

Enhancement

Shadowing

Combined pattern

**Calcifications**

Calcifications in a mass

Calcifications outside of a mass

Intraductal calcifications [a]

**Final assessment**

Category 3

Category 4 (a, b, c)

Category 5

[a] New in the fifth BI-RADS lexicon for breast sonography.

0.03 - 0.24, respectively).

The observers then received an education session on the US criteria, and then performed a second assessment. With regards to the changes seen in each group, the agree-ment was one level higher in the faculty radiologists and junior residents for two criteria, and in the senior residents for seven criteria, after the education session. In the faculty radiologists, the degree of agreement was higher than moderate for all criteria, excluding the margin. In the senior residents, the agreement was one level higher for all of the criteria except orientation, following the training. However, in the junior residents, there was no increase in degree of agreement for any of the criteria except for tissue composition and posterior features, and the degree of agreement was lower than fair for all criteria. At second assessment, the respective 95% CI of the faculty radiologist, senior resident, and junior resident were 0.39 - 0.62, 0.25 - 0.59 and 0.13 - 0.47 for tissue composition, 0.35 - 0.60, 0.25 - 0.55 and 0.17 - 0.53 for shape, 0.54 - 0.77, 0.46 - 0.70 and 0.25 - 0.55 for orientation, 0.21 - 0.52, 0.06 - 0.37 and 0.05-0.30 for margin, 0.27 - 0.55, 0.06 - 0.52 and 0.05-0.32 for echo pattern, 0.44-0.64, 0.40-0.60 and 0.33-51 for posterior features, 0.39 - 0.62, 0.32 - 0.55 and 0.25 - 0.45 for calcifications, and 0.35 - 0.59, 0.13 - 0.34 and 0.04 - 0.29 for final assessment.

The intra-class correlation coefficients for final assessment between the faculty radiologist, senior resident, and junior resident were 0.55 (95% CI, 0.44 - 0.66) at primary assessment, and 0.57 (95% CI, 0.42 - 0.69) at the secondary assessment. The agreement was considered to be poor.

*4.2. Diagnostic Performance*

The sensitivity and specificity acquired from the faculty members, senior residents, and junior residents are shown in Table 2. In the ROC curve analysis, the respective AUC values of the faculty radiologists, senior residents, and junior residents were 0.78 (95% CI, 0.70 - 0.85), 0.59 ((95% CI, 0.50 - 0.68), and 0.52 (95% CI, 0.43 - 0.61) (Figure 2). There was a significantly higher diagnostic accuracy for the faculty radiologists than the senior and junior residents (P = 0.0001 and < 0.0001, respectively). However, there was no significant difference in diagnostic performance between the senior and junior residents.

## 5. Discussion

In our study, agreement between the faculty radiologists was fair-to-good for all criteria; however, between residents, agreement was poor-to-moderate. Therefore, investigative reporting of breast US by residents is inadvisable. At the academic medical centres under study, US of the breast, as well as the abdominal organs, are examined by radiology residents and experienced faculty members. The residents actually operate the ultrasonic equipment themselves and then interpret the images that they, themselves, have produced, in a preliminary form. At the end, the final report is confirmed by the attending faculty member.

**Table 1.** Inter-Observer Agreement

| Final Assessments | Faculty Member [a] | | Senior Resident [a] | | Junior Resident [a] | |
|---|---|---|---|---|---|---|
| | Primary Assessment | Secondary Assessment | Primary Assessment | Secondary Assessment | Primary Assessment | Secondary Assessment |
| **Tissue composition** | 0.47 | 0.51 | 0.36 | 0.42 [b] | 0.15 | 0.30 [b] |
| **Shape** | 0.49 | 0.48 | 0.33 | 0.40 [b] | 0.22 | 0.35 |
| **Orientation** | 0.66 | 0.66 | 0.59 | 0.60 | 0.30 | 0.40 |
| **Margin** | 0.32 | 0.37 | 0.11 | 0.21 [b] | 0.18 | 0.17 |
| **Echo pattern** | 0.36 | 0.41 [b] | 0.15 | 0.29 [b] | 0.12 | 0.19 |
| **Posterior features** | 0.53 | 0.54 | 0.40 | 0.49 [b] | 0.20 | 0.38 [b] |
| **Calcifications** | 0.48 | 0.51 | 0.39 | 0.45 | 0.23 | 0.35 |
| **Final assessment** | 0.40 | 0.47 [b] | 0.19 | 0.23 [b] | 0.14 | 0.16 |

[a] n = 2.
[b] The degree of agreement was higher at the second assessment, which was performed following the education session.

**Table 2.** Sensitivity and Specificity of Ultrasonography Interpretations by the Three Observer Groups[a]

| | Faculty Member | Senior Resident | Junior Resident |
|---|---|---|---|
| **Sensitivity** | 98.0 | 66.0 | 58.0 |
| **Specificity** | 58.6 | 52.9 | 45.7 |

[a] Values unit is %.
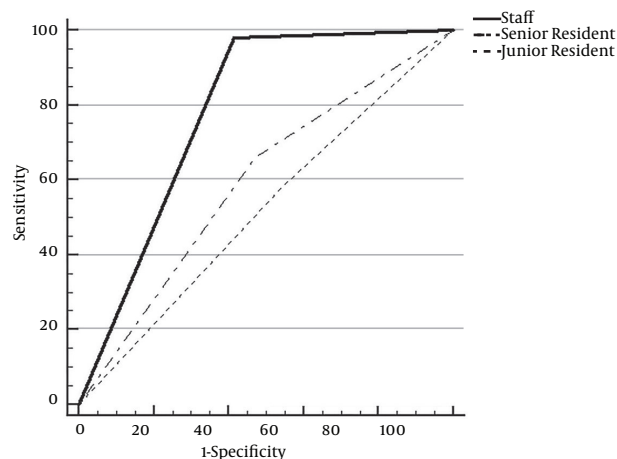


**Figure 2.** Receiver operating characteristic curves for faculty members, senior and junior residents; The area under the curve was 0.78 for faculty members, 0.59 for senior residents, and 0.52 for junior residents. There was a significant difference between area under the receiver operating characteristic curves of faculty members and residents.

The variability in agreement regarding image assessment between radiology residents and skilled faculty members has been previously studied, and these reports (19, 26-28)

showed that the degree of agreement was greater than 90% between radiology residents and faculty members in the assessment of head and pulmonary angiography computerized tomography (CT). However, to date, the discrepancies between faculty radiologists and residents in the assessment of breast US images have not been fully established; this was the motivation for our research. We found that, overall, the agreement between the two faculty radiologists was greater than for the residents. The comparison of diagnostic performance was also significantly higher for the faculty radiologists, yet there was no significant difference between junior and senior residents.

Previous reports have demonstrated that certain breast US features are reliable for differentiation of benign and malignant breast lesions (29-33). However, reader discrepancy in mass assessment by US is responsible for differences in lesion detection and variation in lesion description and subsequent management. Several researches have reported inter-observer variability in the assessment of breast masses with the use of the BI-RADS US lexicon, fourth edition (7, 11). The fifth edition of BI-RADS for US was published in 2013 (13). However, to the best of our knowledge, observer variability, using the BI-RADS new lexicon for US, fifth edition (2013), has not been widely studied.

Our results using the BI-RADS fifth edition are similar to previous studies using the fourth edition. In the new BI-RADS lexicon for US, the boundary term was removed, yet in our data, this seemed to have little effect in determining a final category. The agreement for the final category of faculty radiologists was fair-to-moderate in our study. This result was similar to that in studies by Elverici et al. (7), Lee et al. (10), and Berg et al. (8), using the BI-RADS

fourth edition. According to a study by Elverici et al. (7), the observer agreement was good for orientation, moderate for shape, echo pattern and posterior feature, and was fair for margin and final category for two experienced radiologists. Regarding our results, data of the faculty members was similar to that of Elverici et al. (7), yet our results for residents were poor. Also, as in other studies (34), the agreement of all three subgroups in the current study was lowest for margin and highest for orientation. We assume the reason is that the terminology for characterization of the mass margin is multiple, and even overlapped. Our study showed that the disagreement among readers, when labelling mass margin and echo pattern, was more with a heterogeneous background echo-texture than homogeneous background. This is probably because the observers were uncertain and confused when trying to detect and classify abnormalities in heterogeneous tissue composition with posterior shadowing. This confusion is not minor, because the designation of circumscribed margin can encourage observers to decide that the lesion is benign and thus produce a false negative. In contrast, the designation of a not-circumscribed margin may contribute to a false positive lesion and an unnecessary biopsy.

Interestingly, following the education session, the agreement was one level higher for all of the criteria except orientation in the senior residents; however, there was no improvement in the degree of agreement for the six criteria in the junior residents. The agreement for the final category was one level higher in the senior residents after the education session, yet the degree of agreement was still only fair. These data imply that a single education session was not adequate to improve the agreement level and performance of the residents. Therefore, we suggest that attending radiologists need a more careful review and confirmation of the preliminary interpretation. The successful training of residents would appear to require clinical experience through the practice of breast US, continuous consensus reading, and steady feedback and correlation between US findings and the pathology results.

Our study had several limitations. First, this study included only benign masses that underwent biopsies. Thus, typical 'probably benign' or 'benign' lesions on US findings were not included in our study; this exclusion may have led to lower specificity. In real clinics, the specificity may be higher. Second, readers did not actually operate the ultrasonic equipment themselves, and were instead provided by at least two static US images. We think it is very difficult to measure performance discrepancy. Thus, we divided faculty members, senior residents and junior residents into three subgroups, and then we analyzed interobserver agreement and ROC curves for each subgroups. Finally, there was selection bias because only images par-

ticularly chosen by an investigator were evaluated. Our results are limited due to our small sample size. Additional studies are required with a larger series of patients.

Our study showed that the reader agreement for sonographic BIRADS lexicon was higher among faculty radiologists than among residents. In addition, there was significantly higher diagnostic accuracy for the faculty members when compared to the senior and junior residents. Therefore, we recommend continued professional resident training to improve the degree of agreement and performance for breast US.

## Footnotes

## References

1. Mendelson EB, Berg WA, Merritt CR. Toward a standardized breast ultrasound lexicon, BI-RADS: ultrasound. *Semin Roentgenol.* 2001;**36**(3):217–25. [PubMed: 11475068].

2. Arger PH, Sehgal CM, Conant EF, Zuckerman J, Rowling SE, Patton JA. Interreader variability and predictive value of US descriptions of solid breast masses: pilot study. *Acad Radiol.* 2001;**8**(4):335–42. doi: 10.1016/s1076-6332(03)80503-2. [PubMed: 11293782].

3. Heinig J, Witteler R, Schmitz R, Kiesel L, Steinhard J. Accuracy of classification of breast ultrasound findings based on criteria used for BI-RADS. *Ultrasound Obstet Gynecol.* 2008;**32**(4):573–8. doi: 10.1002/uog.5191. [PubMed: 18421795].

4. Kim EK, Ko KH, Oh KK, Kwak JY, You JK, Kim MJ, et al. Clinical application of the BI-RADS final assessment to breast sonography in conjunction with mammography. *AJR Am J Roentgenol.* 2008;**190**(5):1209–15. doi: 10.2214/AJR.07.3259. [PubMed: 18430833].

5. American College of Radiology. . Breast imaging reporting and data system, Breast imaging atlas. fourth ed. Reston, VA: American College of Radiology; 2003.

6. 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):1260–5. doi: 10.2214/ajr.184.4.01841260. [PubMed: 15788607].

7. Elverici E, Zengin B, Nurdan Barca A, Didem Yilmaz P, Alimli A, Araz L. Interobserver and Intraobserver Agreement of Sonographic BIRADS Lexicon in the Assessment of Breast Masses. *Iran J Radiol.* 2013;**10**(3):122–7. doi: 10.5812/iranjradiol.10708. [PubMed: 24348596].

8. Berg WA, Blume JD, Cormack JB, Mendelson EB. Operator dependence of physician-performed whole-breast US: lesion detection and characterization. *Radiology.* 2006;**241**(2):355–65. doi: 10.1148/radiol.2412051710. [PubMed: 17057064].

9. Baker JA, Kornguth PJ, Soo MS, Walsh R, Mengoni P. Sonography of solid breast lesions: observer variability of lesion description and assessment. *AJR Am J Roentgenol.* 1999;**172**(6):1621–5. doi: 10.2214/ajr.172.6.10350302. [PubMed: 10350302].

10. Lee HJ, Kim EK, Kim MJ, Youk JH, Lee JY, Kang DR, et al. Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound. *Eur J Radiol.* 2008;**65**(2):293–8. doi: 10.1016/j.ejrad.2007.04.008. [PubMed: 17531417].

11. Park CS, Lee JH, Yim HW, Kang BJ, Kim HS, Jung JI, 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. [PubMed: 17923782].

12. Calas MJ, Almeida RM, Gutfilen B, Pereira WC. Interobserver concordance in the BI-RADS classification of breast ultrasound exams. *Clinics (Sao Paulo).* 2012;**67**(2):185–9. [PubMed: 22358246].

13. American College of Radiology. . Breast imaging reporting and data system, Breast imaging atlas. fifth ed. Reston, VA: American College of Radiology; 2013.

14. Raza S, Goldkamp AL, Chikarmane SA, Birdwell RL. US of breast masses categorized as BI-RADS 3, 4, and 5: pictorial review of factors influencing clinical management. *Radiographics.* 2010;**30**(5):1199–213. doi: 10.1148/rg.305095144. [PubMed: 20833845].

15. Sedgwick E. The breast ultrasound lexicon: breast imaging reporting and data system (BI-RADS). *Semin Roentgenol.* 2011;**46**(4):245–51. doi: 10.1053/j.ro.2011.04.001. [PubMed: 22035666].

16. Costantini M, Belli P, Lombardi R, Franceschini G, Mule A, Bonomo L. Characterization of solid breast masses: use of the sonographic breast imaging reporting and data system lexicon. *J Ultrasound Med.* 2006;**25**(5):649–59. [PubMed: 16632790].

17. Carney E, Kempf J, DeCarvalho V, Yudd A, Nosher J. Preliminary interpretations of after-hours CT and sonography by radiology residents versus final interpretations by body imaging radiologists at a level 1 trauma center. *AJR Am J Roentgenol.* 2003;**181**(2):367–73. doi: 10.2214/ajr.181.2.1810367. [PubMed: 12876012].

18. Bechtold RE, Chen MY, Ott DJ, Zagoria RJ, Scharling ES, Wolfman NT, et al. Interpretation of abdominal CT: analysis of errors and their causes. *J Comput Assist Tomogr.* 1997;**21**(5):681–5. [PubMed: 9294552].

19. Erly WK, Berger WG, Krupinski E, Seeger JF, Guisto JA. Radiology resident evaluation of head CT scan orders in the emergency department. *AJNR Am J Neuroradiol.* 2002;**23**(1):103–7. [PubMed: 11827881].

20. Yoon LS, Haims AH, Brink JA, Rabinovici R, Forman HP. Evaluation of an emergency radiology quality assurance program at a level I trauma center: abdominal and pelvic CT studies. *Radiology.* 2002;**224**(1):42–6. doi: 10.1148/radiol.2241011470. [PubMed: 12091660].

21. Albano MC, Ross GW, Ditchek JJ, Duke GL, Teeger S, Sostman HD, et al. Resident interpretation of emergency CT scans in the evaluation of acute appendicitis. *Acad Radiol.* 2001;**8**(9):915–8. [PubMed: 11724048].

22. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;**143**(1):29–36. doi: 10.1148/radiology.143.1.7063747. [PubMed: 7063747].

23. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean J Radiol.* 2004;**5**(1):11–8. [PubMed: 15064554].

24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;**33**(1):159–74. [PubMed: 843571].

25. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med.* 2005;**37**(5):360–3. [PubMed: 15883903].

26. Filippi CG, Schneider B, Burbank HN, Alsofrom GF, Linnell G, Ratkovits B. Discrepancy rates of radiology resident interpretations of on-call neuroradiology MR imaging studies. *Radiology.* 2008;**249**(3):972–9. doi: 10.1148/radiol.2493071543. [PubMed: 19011191].

27. Wysoki MG, Nassar CJ, Koenigsberg RA, Novelline RA, Faro SH, Faerber EN. Head trauma: CT scan interpretation by radiology residents versus staff radiologists. *Radiology.* 1998;**208**(1):125–8. doi: 10.1148/radiology.208.1.9646802. [PubMed: 9646802].

28. Yavas US, Calisir C, Ozkan IR. The interobserver agreement between residents and experienced radiologists for detecting pulmonary embolism and DVT with using CT pulmonary angiography and indirect CT venography. *Korean J Radiol.* 2008;**9**(6):498–502. doi: 10.3348/kjr.2008.9.6.498. [PubMed: 19039265].

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):665–72. doi: 10.1148/radiol.2523080670. [PubMed: 19567644].

30. Rahbar G, Sie AC, Hansen GC, Prince JS, Melany ML, Reynolds HE, et al. Benign versus malignant solid breast masses: US differentiation. *Radiology.* 1999;**213**(3):889–94. doi: 10.1148/radiology.213.3.r99dc20889. [PubMed: 10580971].

31. 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):123–34. doi: 10.1148/radiology.196.1.7784555. [PubMed: 7784555].

32. Harper AP, Kelly-Fry E, Noe JS, Bies JR, Jackson VP. Ultrasound in the evaluation of solid breast masses. *Radiology.* 1983;**146**(3):731–6. doi: 10.1148/radiology.146.3.6298858. [PubMed: 6298858].

33. Heywang SH, Lipsit ER, Glassman LM, Thomas MA. Specificity of ultrasonography in the diagnosis of benign breast masses. *J Ultrasound Med.* 1984;**3**(10):453–61. [PubMed: 6092659].

34. 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):385–91. doi: 10.1148/radiol.2392042127. [PubMed: 16569780].