Observer Variability of the Breast Imaging Reporting and Data System (BI-RADS) Lexicon for Mammography

Zehra H. Adibelli   Ruken Ergenc   Ozgur Oztekin   Suheyla Ecevit   Gokhan Unal   Yusuf Abalı

Radiology Department, Izmir Bozyaka Training and Research Hospital, Turkey

Summary

Aim: We aimed to determine the inter- and intra-observer variabilities between breast radiologists and a general radiologist in categorizing mammographic lesions using the Breast Imaging Reporting and Data System (BI-RADS), and to evaluate the effects of the histopathologic results on the variability. **Methods**: Mammograms from 142 women who underwent biopsy were evaluated. 3 breast radiologists (2 with >10 years experience and 1 with 1 year experience) and 1 general radiologist retrospectively reviewed mammograms twice within an 8-week interval. Inter- and intra-observer variabilities were assessed with Cohen's kappa statistic, and the positive predictive value for final assessments was calculated. **Results**: The intra-observer variability for mass and calcification assessments was moderate to almost perfect (kappa values: 0.41–1) for breast imagers and was fair to substantial for the general radiologist (kappa values: 0.21–0.8). Inter-observer agreement between the breast imagers was higher than between the breast and general radiologists. There was no apparent difference in agreement between observers for malignant and benign subgroups. **Conclusions**: The differences in intra- and inter-observer agreement between the breast imagers and the general radiologist affirm the utility of the BI-RADS lexicon. The histopathologic results of the lesions do not affect the agreement. BI-RADS is a simple and adequate tool for assessing mammograms, even after only 1 year of training.

Zusammenfassung

Ziel: Ziel der Studie war es, die Inter- und Intra-Observer-Variabilität zwischen mehreren Brustradiologen und einem Allgemeinradiologen bei der Einordnung von mammografisch festgestellten Läsionen mithilfe des Systems zur Erfassung und Datenverarbeitung von Bildgebungsmethoden der Brust (BI-RADS) zu bestimmen. Außerdem sollte der Einfluss der histopathologischen Ergebnisse auf die Variabilität festgestellt werden. **Methoden**: Die Mammogramme von 142 Frauen, die sich einer Biopsie unterzogen hatten, wurden evaluiert. Drei Brustradiologen (2 mit mehr als 10-jähriger Erfahrung und 1 mit 1-jähriger Erfahrung) und 1 Allgemeinradiologe führten in einem Abstand von 8 Wochen eine zweimalige retrospektive Begutachtung der Mammogramme durch. Die Inter- und Intra-Observer-Variabilität wurde mit Cohens Kappa-Statistik bestimmt. Der positive Vorhersagewert für die abschließende Beurteilung wurde berechnet. **Ergebnisse**: Die Intra-Observer-Variabilität war bei der Einschätzung von Tumoren und Kalzifizierungen für die Brustradiologen gut bis nahezu perfekt (Kappa-Werte: 0,41–1) und für den Allgemeinradiologen ausreichend bis gut (Kappa-Werte: 0,21–0,8). Die Inter-Observer-Übereinstimmung unter den Brustradiologen war höher als die zwischen den Brustradiologen und dem Allgemeinradiologen. Für die malignen und benignen Untergruppen gab es keinen deutlichen Unterschied in der Übereinstimmung zwischen den Observern. **Schlussfolgerung**: Die Unterschiede in der Intra- und Inter-Observer-Variabilität zwischen den Brustradiologen und dem Allgemeinradiologen bestätigen die Nützlichkeit des BI-RADS-Lexikons. Die histopathologischen Ergebnisse der Läsionen beeinflussen die Übereinstimmung nicht. BI-RADS ist ein einfaches und adäquates Werkzeug für die Einschätzung von Mammogrammen, sogar nach nur 1-jähriger Ausbildung.

Dr. Zehra Hilal Adibelli
Izmir Bozyaka Training and Research Hospital
Ugur Mumcu cad. No: 25 Buca, 35160 Izmir, Turkey
Tel. +90-532 7970472, Fax -232 2505050
adibellizh@mynet.com

Introduction

The American College of Radiology (ACR) has developed the Breast Imaging Reporting and Data System (BI-RADS®) lexicon in mammography for standardizing mammographic reporting [1]. Regarding the development of large mammographic databases, the ACR provides information for auditing mammography practices and facilities of research [2, 3]. According to the ACR guidelines, 5 main reporting categories are recommended for final assessment purposes: R1 (negative) and R2 (definitely benign) require no further assessment; R3 (probably benign, with a positive predictive value (PPV) < 2%) warrants early reassessment (3–6 months); R4 (suspicious, > 2–95% PPV) warrants histopathological confirmation; and R5 (malignant, > 95% PPV) warrants immediate treatment. BI-RADS has evolved over time, with the 4th edition issued in 2003 and extended to apply to both breast ultrasonography (US) and magnetic resonance imaging (MRI). One other major addition in the 4th edition involves the subcategorization of category 4 lesions. Dividing category 4 lesions into those with a low (category 4a), intermediate (category 4b) or moderate (category 4c) possibility of malignancy better informs the physicians and patients about the lesion and aids in making final decisions regarding biopsies or the potential need for follow-up.

The purpose of this study was to retrospectively evaluate inter-observer and intra-observer variability in describing breast density, lesion features and final assessments between breast and general radiologists, and to evaluate effects of the histopathologic results of lesions on the variability.

Methods

Institutional Review Board approval was obtained prior to commencement of the study. Individual-patient informed consent for this retrospective study was not required. Four radiologists were used in this study: 3 breast radiologists and 1 general radiologist. 2 of the breast radiologists (observers 1 and 2) have been interpreting breast images for more than 10 years, and observer 1 has been the chief of the mammography unit. The other breast radiologist (observer 3) was a last-year assistant who got her thesis on mammography, and has worked at the mammography unit for 1 year. These 3 breast radiologists practiced within the same group. The other radiologist (observer 4) has been a general radiologist for 7 years, and has less experience on breast imaging and more on MRI. Before the study, all observers received oral instructions and a summary of the BI-RADS lexicon, 4th edition.

Mammography was performed using a Senographe 700 T (General Electric Medical Systems, Milwaukee, USA). Screen-film mammograms of 142 women were evaluated over a 2-week period (approximately 15 patients' mammograms per day). To assess intra-observer variability, the observers reviewed all of the cases again 8 weeks later. During the study, we realized that 2 or 3 US images of a lesion were not enough to describe that lesion using the terminology of the 4th edition of the BI-RADS lexicon; therefore, only mammographic images were evaluated, and evaluations of US images were excluded from the study.

Each observer described each lesion by using the terminology of the 4th edition of the BI-RADS lexicon. Specifically, the following issues were evaluated: calcification description, distribution, and number; mass shape, margin, and density; architectural distortion, breast density, associ-

Table 1. Kappa values for intra-observer variability

Observer	Calcification description	Calcification distribution	Calcification number	Mass shape	Mass margin	Mass density	Architectural distortion	Associated findings & special cases	Breast density	Final BI-RADS assessment (4th edition)	Final BI-RADS assessment (3rd edition)
Observer 1											
Malignant cases	0.70	0.69	0.86	0.58	0.50	0.49	0.45	0.31	0.39	0.57	0.75
Benign cases	0.88	0.79	0.80	0.63	0.60	0.51	0.58	0.38	0.30	0.45	0.46
Total cases	0.81	0.76	0.83	0.64	0.60	0.51	0.51	0.42	0.23	0.56	0.61
Observer 2											
Malignant cases	0.70	0.67	0.80	0.72	0.61	0.69	0.35	0.02	0.60	0.35	0.38
Benign cases	0.74	0.71	0.84	0.64	0.57	0.71	0.55	0.47	0.58	0.26	0.41
Total cases	0.74	0.70	0.83	0.69	0.61	0.68	0.49	0.28	0.59	0.38	0.51
Observer 3											
Malignant cases	0.80	0.73	0.84	0.90	0.84	0.78	0.52	0.39	0.39	0.35	0.41
Benign cases	0.78	0.71	0.78	0.50	0.57	0.57	0.90	0.58	0.40	0.40	0.50
Total cases	0.80	0.72	0.81	0.66	0.70	0.66	0.69	0.53	0.42	0.48	0.54
Observer 4											
Malignant cases	0.56	0.50	0.71	0.66	0.49	0.50	0.40	0.02	0.24	0.05	0.05
Benign cases	0.57	0.66	0.64	0.48	0.32	0.35	0.55	0.26	0.22	0.12	0.11
Total cases	0.59	0.59	0.68	0.57	0.42	0.41	0.49	0.16	0.23	0.17	0.18

Kappa values are interpreted as follows: 0.81–1.00, almost perfect agreement; 0.61–0.80, substantial agreement; 0.41–0.60, moderate agreement; 0.21–0.40, fair agreement; 0.00–0.20, slight agreement.
Groups are as follows: malignant cases (n = 60); benign cases (n = 82); total cases (both benign and malignant, n = 142).
Observers 1 and 2 have been interpreting breast images for more than 10 years; observer 3 has been interpreting breast images for 1 year; observer 4 is a general radiologist.

Table 2. Kappa values for inter-observer variability

	Malignant cases	Benign cases	Total cases
	Observer 1	Observer 1	Observer 1
Calcification description			
Observer 2	0.75	0.62	0.83
Observer 3	0.70	0.56	0.77
Observer 4	0.65	0.50	0.73
Calcification distribution			
Observer 2	0.63	0.56	0.66
Observer 3	0.57	0.50	0.62
Observer 4	0.54	0.45	0.60
Calcification number			
Observer 2	0.82	0.74	0.85
Observer 3	0.73	0.73	0.70
Observer 4	0.67	0.63	0.68
Mass shape			
Observer 2	0.62	0.69	0.53
Observer 3	0.54	0.47	0.52
Observer 4	0.59	0.67	0.50
Mass margin			
Observer 2	0.65	0.63	0.62
Observer 3	0.51	0.47	0.47
Observer 4	0.54	0.51	0.50
Mass density			
Observer 2	0.56	0.58	0.53
Observer 3	0.33	0.31	0.34
Observer 4	0.50	0.49	0.50
Architectural distortion			
Observer 2	0.60	0.48	0.75
Observer 3	0.58	0.31	1.00
Observer 4	0.47	0.32	0.63
Associated findings and special cases			
Observer 2	0.28	0.18	0.37
Observer 3	0.21	0.02	0.29
Observer 4	0.24	0.04	0.42
Breast density			
Observer 2	0.50	0.36	0.58
Observer 3	0.41	0.14	0.53
Observer 4	0.40	0.19	0.50
BI-RADS (4th edition)			
Observer 2	0.27	0.15	0.23
Observer 3	0.17	0.00	0.11
Observer 4	0.18	0.14	0.08
BI-RADS (3rd edition)			
Observer 2	0.42	0.24	0.44
Observer 3	0.25	0.07	0.25
Observer 4	0.23	0.17	0.15

Kappa values are interpreted as follows: A = 0.81–1.00, almost perfect agreement; B = 0.61–0.80, substantial agreement;
C = 0.41–0.60, moderate agreement; D = 0.21–0.40, fair agreement; E = 0.00–0.20, slight agreement.
Groups are as follows: malignant cases (n = 60); benign cases (n = 82); total cases (both benign and malignant, n = 142).
Observers 1 and 2 have been interpreting breast images for more than 10 years; observer 3 has been interpreting breast images for 1 year; observer 4 is a general radiologist.

ated findings and special cases; and final BI-RADS assessments. Each observer assigned 2 final assessment BI-RADS categories. First, the observers assigned a final assessment by using the 4th edition of the BI-RADS lexicon including subcategories for category 4 and then they as-

signed another final assessment BI-RADS category by using the 3rd edition of the BI-RADS lexicon without including subcategories for category 4. The following BI-RADS categories for breast density were used: 1 (fatty, < 25% glandular), 2 (scattered fibroglandular densities, approxi-

Table 3. Number of malignancies, lesions, and PPVs according to the BI-RADS lexicon

BI-RADS categories	Malignancies/lesions			
	Observer 1	Observer 2	Observer 3	Observer 4
BI-RADS 2	0/3 (0%)	0/3 (0%)	0/6 (0%)	3/13 (23%)
BI-RADS 3	1/37 (3%)	1/25 (4%)	1/26 (4%)	4/30 (13%)
BI-RADS 4a	15/49 (31%)	6/33 (18%)	37/43 (86%)	8/32 (25%)
BI-RADS 4b	8/13 (62%)	2/5 (40%)	6/12 (50%)	1/4 (25%)
BI-RADS 4c	18/19 (95%)	12/32 (38%)	11/15 (73%)	4/12 (33%)
BI-RADS 5	18/21 (86%)	39/44 (89%)	36/40 (90%)	40/50 (80%)

Data are presented as number of malignancies/number of lesions, with positive predictive value (PPV) in parentheses.

mately 25–50% glandular), 3 (heterogeneously dense, approximately 51–75% glandular), and 4 (extremely dense, > 75% glandular) [1].

The lesions were evaluated on original mammograms without prior mammograms for comparison. Lesions were marked on both craniocaudal and mediolateral oblique views so that all observers would examine the same lesions. Spot and magnification views were also available to evaluate for some cases. The observers evaluated the mammograms alone and were instructed to select the most appropriate descriptors for each lesion and document their assessments on a sheet that was provided for each day's cases. After the evaluation, the observers gave this sheet to a supervisor who was independent of the study.

Pathologic diagnosis was available for all 142 lesions. Pathologic findings were evaluated with results from image-guided fine-needle aspiration (FNA) biopsy, core needle biopsy and excisional biopsy, in cases referred for needle biopsy because of atypia, positivity for malignancy or discordance.

In order to understand the effect of histopathological results on the variability, we divided the patients into subgroups based on the pathological diagnoses: (a) group 1: malignant cases (60 patients, 42%), (b) group 2: benign cases (82 patients, 58%), (c) group 3: total cases (142 patients, 100%). Kappa was calculated for each of these subgroups to understand whether malignant or benign lesions would decrease kappa or not.

Cohen's kappa statistic was used to assess inter- and intra-observer agreement for all descriptor variables, and kappa values were calculated using Stata software (Stata Press, College Station, TX, USA, Stata Release 10), to assess the proportion of inter-observer agreement between all pairs of observers. In the case of multiple observers and multiple categories, the guidelines of Landis and Koch were followed in interpreting kappa values: 0.81–1.00, almost perfect agreement; 0.61–0.80, substantial agreement; 0.41–0.60, moderate agreement; 0.21–0.40, fair agreement; 0.00–0.20, slight agreement [4].

Results

One-hundred-forty-two consecutive non-palpable lesions in 142 women (mean age = 52 years, range = 38–71 years) who underwent image-guided biopsy between January 1999 and December 2008 were evaluated in the study. The pathological diagnoses of this series were malignant in 60 cases (42%) (group 1) and benign in 82 cases (58%) (group 2). Group 3 represented the total patient population (142 cases). The distribution of histopathological results was as follows: invasive carcinoma (n = 50, 35%), in situ carcinoma (n = 10, 7%), fibrocystic changes and/or dysplasias (n = 42, 30%), fibroadenoma (n = 30, 21%), and other benign pathologies (including sclerosing adenosis, radial scar, intraductal papillamatosis, etc.) (n = 10, 7%).

When cases were reassessed by the same observer, substan-tial agreement was achieved in describing microcalcifications while substantial or moderate agreement was achieved in describing mass shape, margin and density. In describing breast density and associated findings, moderate or fair agreement was seen among the breast radiologists (observers 1–3), while fair or slight agreement was found for the general radiologist (observer 4). For the final BI-RADS assessements, fair to moderate agreement was found for the breast radiologists and slight agreement was found for the general radiologist. The intra-observer variability for each observer is summarized in table 1.

Inter-observer variability was examined by comparing observer 1, who was an experienced breast radiologist, to observer 2 (another experienced breast radiologist), to observer 3 (new breast radiologist), and to observer 4 (the general radiologist). The inter-observer variability for these comparisons are described below for the various lesion assessments based on the kappa values presented in table 2.

Pathologic correlations for all lesions were available. PPV of lesions categorized as BI-RADS 2, 3, 4 or 5 for all of the readers are presented in table 3. In general, the PPV increased accordingly as the BI-RADS category increased from 2 (definitely benign) to 5 (malignant).

Discussion

The ACR has developed BI-RADS to standardize mammographic reporting [1, 5]. However, even with the BI-RADS lexicon, variability in mammographic interpretation can still be observed and may be attributed both to differences in detection of lesions and variation in lesion characterization and subsequent management.

Perfect to moderate intra-observer agreement was achieved for the breast radiologists in describing microcalcifications and mass while moderate to fair agreement was achieved for the general radiologist (table 1). This suggests that each breast radiologists has a clear concept of the BI-RADS descriptors for mass and calcifications, and the high degree of intra-observer agreement in our study also affirms the utility of the BI-RADS lexicon for interpreting mammography. Furthermore, these results reveal an apparent difference between the breast imagers and the general radiologist. In contrast, the intra-observer variability of the

BI-RADS descriptors for architectural distortion, associated findings, special cases, and breast density was substantial to slight for all patients (group 3) for all of the observers (both breast and general radiologists). This suggests that the observers did not have a clear concept about these descriptors. Intra-observer variability of the general radiologist was slight for the final BI-RADS assessments, whereas intra-observer variability of the experienced and new breast imagers was fair to substantial. Intra-observer agreement of the breast radiologists was slightly lower when the BI-RADS 4th edition (categories 4a, 4b, 4c) was used compared to the 3rd edition (category 4). Furthermore, there was an apparent difference between the breast radiologists (even between experienced and new) and the general radiologist for the final BI-RADS assessment. This suggests that the general radiologist, even after receiving oral instructions and a summary of the BI-RADS lexicon, could not apply the BI-RADS lexicon as well as the breast imager who had 1 year experience.

Inter-observer agreements between experienced breast imagers and between experienced and new breast imagers were as high as in prior studies for calcification description, distribution and number [6–8]. In assessing the differences between malignant and benign cases (groups 1 and 2), the kappa values were slightly higher in the benign cases, which may be because it is easier to identify the typical benign calcifications (table 2). Inter-observer agreement between the experienced breast imagers for mass shape, margin and density was not as high as the agreement for calcifications, but was similar to previous studies [6–8], and there was no apparent difference between the malignant, benign, and total patients cases (table 2). The lower rate of agreement for mass density on mammograms suggests that the observers had difficulties in making up their minds about this categorization. Inter-observer agreement for interpreting mass and calcification among the breast radiologists (including 1 and 10 years experience) was apparently higher than the inter-observer agreement between one breast radiologist and one general radiologist. This shows that being experienced in breast imaging makes the agreement value higher with respect to interpreting mammograms.

Fair to moderate inter-observer variability was seen between breast imagers for breast density in our study. This was much different (and lower) from that reported by Ooms et al. [9] and Ciatto et al. [10]. In our practice, the use of BI-RADS breast density categorization like BI-RADS 3 or 5 led to some confusion for clinicians as to whether the numbers represented density-of-lesion categorization; as a result, we no longer use breast density categorization in our reports.

Fair to moderate inter-observer agreement between experienced breast imagers for overall BI-RADS 4th edition's categories in our study was very similar to the studies reported by Lazarus et al. [6] and Ciatto et al. [11]. However, the kappa values in our study for the BI-RADS 3rd edition were much higher than for the BI-RADS 4th edition, and this fact may be explained by a greater number of categories offered (i.e., inclusion of the 4a, 4b, and 4c subcategories in the 4th edition).

The PPV of BI-RADS 2 for the new and experienced breast imagers was 0%, but it was 23% for the general radiologist (table 3). Moreover, PPVs of BI-RADS 4a, 4b and 4c for breast imagers were similar to or slightly higher than previously reported values [6, 7, 12]. However, PPVs of final BI-RADS assessments from the general radiologist were apparently different and lower compared to those from the experienced breast imagers in our study and from prior studies [6, 7, 12, 13]. Sickles et al. [14] reported that specialist radiologists detected more cancer, recommended more biopsies and had lower recall rates than general radiologists. But in our study there was a higher rate of false-negative cases (BI-RADS 2 and 3) for the general radiologist (observer 4) than in previous studies and we discussed that this high rate was caused by the relatively short oral instructions and summary of the BI-RADS given before the study. Berg et al. [15] said that training in BI-RADS feature analysis and assessment resulted in improved consistency and, in their study, the benefit of such training were retained after 2–3 months in a subgroup of participants.

There are several limitations in our study. First, we studied non-palpable lesions that were referred for biopsy; so we had only few descriptors typically associated with benign disease. Because of this, there were few lesions that could be characterized as special cases. On the other hand, in trying to describe the lesions, sometimes the 'associated findings' were not described or noted. Also, as mentioned in the methods, it was stated that the 3 breast imagers all worked in the same practice. These imagers may have had more agreement in interpreting mammograms because of their training and processes within their practice; therefore, there is a possibility that such high inter-observer agreement might not have been obtained using independent breast imagers. And also the short oral instructions and summary of the BI-RADS given before the study was another limitation in our study that may have caused a high rate of false-negative cases for the general radiologist. Intensified training of the general radiologist with the BI-RADS lexicon might improve characterizations.

In conclusion, an apparent difference in intra-observer agreement and a moderate difference of inter-observer agreement between breast imagers and a general radiologist, even after receiving oral instructions and a summary of the BI-RADS lexicon, were observed. However, the nearly same intra- and inter-observer agreement of the breast imagers (regardless of the years of experience) affirms the utility of the BI-RADS lexicon for interpreting mammography. We also found that the histopathologic results of the lesions do not affect the agreement. Thus, BI-RADS appears to be a simple and adequate tool for assessing mammograms, even after only 1 year of experience in breast imaging.

Conflict of Interest

The authors indicate that they have no interest that may be perceived as posing a conflict or bias.

References

1 American College of Radiology (ACR): Breast Imaging Reporting and Data System: breast imaging atlas, 4th ed. Reston, American College of Radiology, 2003.

2 Geller BM, Barlow WE, Ballard-Barbash R, Ernster VL, Yankaskas BC, Sickles EA, Carney PA, Dignan MB, Rosenberg RD, Urban N, Zheng Y, Taplin SH: Use of the American College of Radiology BI-RADS to report on the mammographic evaluation of women with signs and symptoms of breast disease. Radiology 2002;222:536–542.

3 Taplin SH, Ichikawa LE, Kerlikowske K, Ernster VL, Rosenberg RD, Yankaskas BC, Carney PA, Geller BM, Urban N, Dignan MB, Barlow WE, Ballard-Barbash R, Sickles EA: Concordance of Breast Imaging Reporting and Data System assessments and management recommendations in screening mammography. Radiology 2002;222:529–535.

4 Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 1977; 33:159–174.

5 D'Orsi CJ: The American College of Radiology mammography lexicon: an attempt to standardize terminology. AJR Am J Roentgenol 1996;166:779–780.

6 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;239:385–391.

7 Berg WA, Campassi C, Langenberg P, Sexton MJ: Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. AJR Am J Roentgenol 2000;174: 1769–1777.

8 Baker JA, Kornguth PJ, Floyd CE Jr: Breast Imaging Reporting and Data System standardized mammography lexicon: observers variability in lesion description. AJR Am J Roentgenol 1996;166:773–778.

9 Ooms EA, Zonderland HM, Eijkemans MJ, Kriege M, Mahdavian Delavary B, Burger CW, Ansink AC: Mammography: interobserver variability in breast density assessment. Breast 2007;16:568–576.

10 Ciatto S, Houssami N, Apruzzese A, Bassetti E, Brancato B, Carozzi F, Catarzi S, Lamberini MP, Marcelli G, Pellizzoni R, Pesce B, Risso G, Russo F, Scorsolini A: Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. Breast 2005;14:269–275.

11 Ciatto S, Houssami N, Apruzzese A, Bassetti E, Brancato B, Carozzi F, Catarzi S, Lamberini MP, Marcelli G, Pellizzoni R, Pesce B, Risso G, Russo F, Scorsolini A: Reader variability in reporting breast imaging according to BI-RADS assessment categories (the Florence experience). Breast 2006; 15:44–51.

12 Cosar ZS, Cetin M, Tepe TK, Cetin R, Zarali AC: Concordance of mammographic classifications of microcalcifications in breast cancer diagnosis: Utility of the Breast Imaging Reporting and Data System (fourth edition). Clin Imaging 2005;29:389–395.

13 Lee HJ, Kim EK, Kim MJ, Youk JH, Lee JY, Kang DR, Oh KK: Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound. Eur J Radiol 2008;65:293–298.

14 Sickles EA, Wolverton DE, Dee KE: Performance parameters for screening and diagnostic mammography: specialist and general radiologists. Radiology 2002;224:861–869.

15 Berg WA, D'Orsi CJ, Jackson VP, Bassett LW, Beam CA, Lewis RS, Crewson PE: 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;224:871–880.