European Journal of Radiology Open 5 (2018) 67–72

Contents lists available at ScienceDirect

European Journal of Radiology Open

journal homepage: www.elsevier.com/locate/ejro

Comparison of inter- and intra-observer variability of breast density assessments using the fourth and fifth editions of Breast Imaging Reporting and Data System

Afsaneh Alikhassi[a,b,*], Hamed Esmaili Gourabi[a], Masoud Baikpour[c]

[a] Department of Radiology, Cancer Institute, Imam Khomeini Hospital, Tehran University of Medical Sciences, Tehran, Iran
[b] Cancer Research Center, Cancer Institute of Iran, Tehran University of Medical Sciences, Tehran, Iran
[c] Department of Medicine, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Keywords:
Inter-observer variability
Intra-observer variability
Breast density
Mammography

ABSTRACT

Background: Breast density is a well-known independent risk factor for breast cancer and can significantly affect the sensitivity of screening mammograms.
Objective: We aimed to evaluate the intra- and inter-observer consistencies of breast density assessments using methods outlined in the fourth and fifth editions of the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) guidelines to determine which method is more reliable.
Materials and methods: Three radiologists with subspecialties in breast imaging defined breast density in 72 mammograms four times each: twice using the fourth edition of the ACR BI-RADS guidelines and twice using the fifth edition. The intra- and inter-observer agreements were calculated and compared for each method.
Results: The weighted kappa values for the overall intra-observer agreement were 0.955 (95% confidence interval [CI]: 0.931–0.980) and 0.938 (95% CI: 0.907–0.968) when breast densities were assessed according to criteria outlined in the fourth and fifth ACR BI-RADS editions, respectively. The difference between these values was not statistically significant ($p = .4$). The overall Fleiss-Cohen (quadratic) weighted kappa for inter-observer agreement were 0.623 (95% CI: 0.517–0.729) and 0.702 (95% CI: 0.589–0.815) when breast densities were assessed according to criteria outlined in the fourth and fifth ACR BI-RADS editions, respectively. The difference between these values was not statistically significant ($p = .32$). Similarly, there were no significant differences in the evaluation of breast density (overall) when comparing breast density assignment using criteria outlined in the fourth and fifth ACR BI-RADS edition ($p = .582$).
Conclusion: The ACR BI-RADS guideline is an acceptable method to classify breast density, resulting in substantial inter-observer agreements using criteria outlined in both the fourth and fifth editions. The intra-observer agreement was nearly perfect for radiologists using criteria outlined in both sets of guidelines. Moreover, although the percentage of women who were classified as having dense breasts was higher when radiologists used the fifth edition of ACR BI-RADS guidelines than when they used the fourth edition, this difference was not statistically significant.

1. Introduction

Breast density refers to the amount of radiographically dense tissue, comprising glandular, stromal, and connective tissue, in a woman's breast. Mammographic breast density (MBD) is an important feature evaluated during a mammogram for several reasons: 1) Breast density is among the few known independent risk factors for breast cancer [1]. 2) Dense breast tissue decreases the mammography's sensitivity to identifying breast cancers [2]. 3) Women with high MBD may be at increased risk for local recurrence compared to those with low MBD [3]. 4) Some reports have shown that women with high MBD have more widespread cancers of higher grades with more frequent lymph node positivity [4]. 5) Interval cancers have worse prognoses and are more common in patients with high MBD [5].

For these reasons, women with dense breast tissue should be identified and offered additional screening modalities to properly assess their disease risk. As such, radiologists must use an accurate, consistent, and reproducible method of assessing breast density [6,7]. Different methods for measuring breast density have been proposed, some qualitative and some quantitative, including classification systems

* Corresponding author at: Department of Radiology, Cancer Institute, Imam Khomein Hospital, Tehran University of Medical Sciences, Tehran, 3314114197, Iran.
E-mail addresses: a-alikhasi@sina.tums.ac.ir (A. Alikhassi), esmaili.hamed@yahoo.com (H. Esmaili Gourabi), m-beikpour@alumnus.tums.ac.ir (M. Baikpour).

https://doi.org/10.1016/j.ejro.2018.04.002
Received 15 September 2017; Received in revised form 2 April 2018; Accepted 9 April 2018
Available online 20 April 2018

described by Wolfe [8], Boyd et al. [9], and Tabar [10], and an automated computer-based density measurement system described by Jeffreys et al. [11].

Presently, qualitative assessments for breast density are more common than quantitative techniques in clinical setting, which are not widely available or easy to use.

The American College of Radiology (ACR) developed the Breast Imaging Reporting and Data System (BI-RADS) lexicon [12,13]. This system aimed to standardize the description of breast density and other aspects of breast imaging reports and provides information for auditing mammography practices. Currently 33 states in United States regarding breast density reporting legislation in the U.S. require some level of breast density notification in mammography reports. BI-RADS has become widely used outside America as well, and currently is a valuable method of standardization worldwide. In 2013, the fifth edition of the ACR BI-RADS was released [13]. The goal of this study was to compare the consistency of the fifth BI-RADS edition with the previous edition, which was released in 2003 [12].

2. Methods and materials

2.1. Study design

The literature suggests that a minimum of 3 radiologists evaluating 30 mammograms are necessary to calculate statistical accuracy of inter-observer agreement [14,15]. To optimize the study, we used a sample of 72 consecutive full digital mammograms of Persian women performed during opportunistic screenings over a period of 3 days in the first half of 2017 in Tehran University of medical science. Patient who had previous cancer surgery, cosmetic breast surgery, breast implants, or chemotherapy were excluded from this study to decrease any intervening factor in breast density determination.

The three radiologists involved in this study worked in different imaging centers in diverse cities and were present in our ward for a breast imaging fellowship. They participated in this study in the second half of their fellowship period. To standardize the criteria by which the mammography data was reported, an oral training session was held for all mammogram readers before the start of the study. The training session focused on methods for reporting breast density outlined by each edition of the ACR guidelines and included ACR atlas images. The radiologists were blinded to the mammogram interpretations of the other radiologists and had no knowledge of the patient's history, including the original mammogram interpretation.

Each woman was evaluated using two-view mammography (mediolateral–oblique and craniocaudal) in a full-field digital mammography unit (Selenia, Hologic). At this university hospital, all mammograms are routinely interpreted by one radiologist specialized in breast imaging, according to the BI-RADS classification. A woman was considered to have higher breast density when the density of one breast was different from that of the other breast.

All mammograms were read four times by each participating radiologist who had access to all four complete views of each mammogram; each reading was separated by a 1 month interval and the reading order was changed. The radiologists provided an ACR density (BI-RADS fourth edition guidelines) and breast composition (BI-RADS fifth edition guidelines) for each mammography.

2.2. ACR BI-RADS density

The fourth edition of the ACR BI-RADS guidelines described a subjective method for visually assessing the percentage of fibroglandular tissue within the total breast using mediolateral oblique and craniocaudal images. Breasts with less than 25% glandular density, 25%–50% glandular density, 50%–75% glandular density, and above 75% glandular density were assigned BI-RADS density values of 1, 2, 3, and 4, respectively [12] (Fig. 1).

In the fifth edition of the ACR BI-RAS guidelines, the percentage system was omitted, and emphasis was placed on the decreasing sensitivity of a mammography to detect dense tissues. Four categories of breast density were defined based on visual estimation. The categories were defined as A, B, C, and D so as not to be confused with the numbering system utilized by the fourth edition. A) The breasts are almost entirely fatty; B) There are scattered areas of fibroglandular density; C) The breasts are heterogeneously dense, which may obscure small masses; D) The breasts are extremely dense, which lowers the sensitivity of mammography [13] (Fig. 2).

2.3. Ethical considerations

Informed written consent was obtained from all participants in this study. Gathered information was considered confidential and used anonymously and was only accessible to the authors of the survey. This study was conducted according to the principles of the Declaration of Helsinki, and the participating researchers declare no conflicts of interests.

2.4. Statistical analyses

The intra-observer agreement was calculated for each radiologist assessing breast density using criteria outlined in each ACR BI-RADS edition and reported as weighted kappa values and 95% confidence intervals (CIs). Considering the number of radiologists and the ordinal scale of breast density (from 1 to 4 in the fourth edition and from A to D in the fifth edition), Fleiss-Cohen weighted kappa coefficients and their corresponding 95% CIs were calculated to determine the inter-observer agreement. Using this method, the cells closer in agreement were assigned larger weights than those further from agreement [16,17].

To examine the differences between the criteria outlined in the fourth and fifth ACR BI-RADS editions, kappa values were compared between the three observers using the z-scores calculated according to the differences in kappa values and their corresponding asymptotic standard errors. A multinomial logistic regression model accounting for clustering on each examination was also used to compare the distribution of density categories assigned using criteria outlined in the two editions of the BI-RADS guidelines.

Levels of agreement were classified in the following ways: a kappa value of 1.0 was considered perfect agreement; a kappa value of 0 was considered no agreement; a kappa value below 0.20 was considered slight agreement; a kappa value of 0.21–0.41 was considered fair agreement; a kappa value of 0.41–0.60 was considered moderate agreement, a kappa value of 0.61–0.80 was considered substantial agreement, and a kappa value of 0.81–0.99 was considered almost perfect agreement.

Finally, breast densities were further categorized into two groups: non-dense (density categories 1 and 2 in the fourth edition BI-RADS guidelines and categories A and B in the fifth edition ACR BI-RADS guidelines) and dense (density categories 3 and 4 in the fourth edition ACR BI-RADS guidelines and categories C and D in the fifth edition BI-RADS guidelines). Assignment of cases to these groups was compared between radiologists using criteria outlined in the two ACR BI-RADS editions using a logistic regression model accounting for clustering on each examination. Inter-observer agreements and comparison of density assignment distributions were performed according the radiologists' first reports. All analyses were performed using SPSS software for Windows v.22 (IBM corp., Armonk, NY, USA).

3. Results

Three radiologists were asked to review mammograms obtained from a total of 72 subjects (mean age = 50.4 ± 10.7 years) and determine breast density for each subject according to criteria outlined in the fourth and fifth editions of the ACR BI-RADS guidelines.

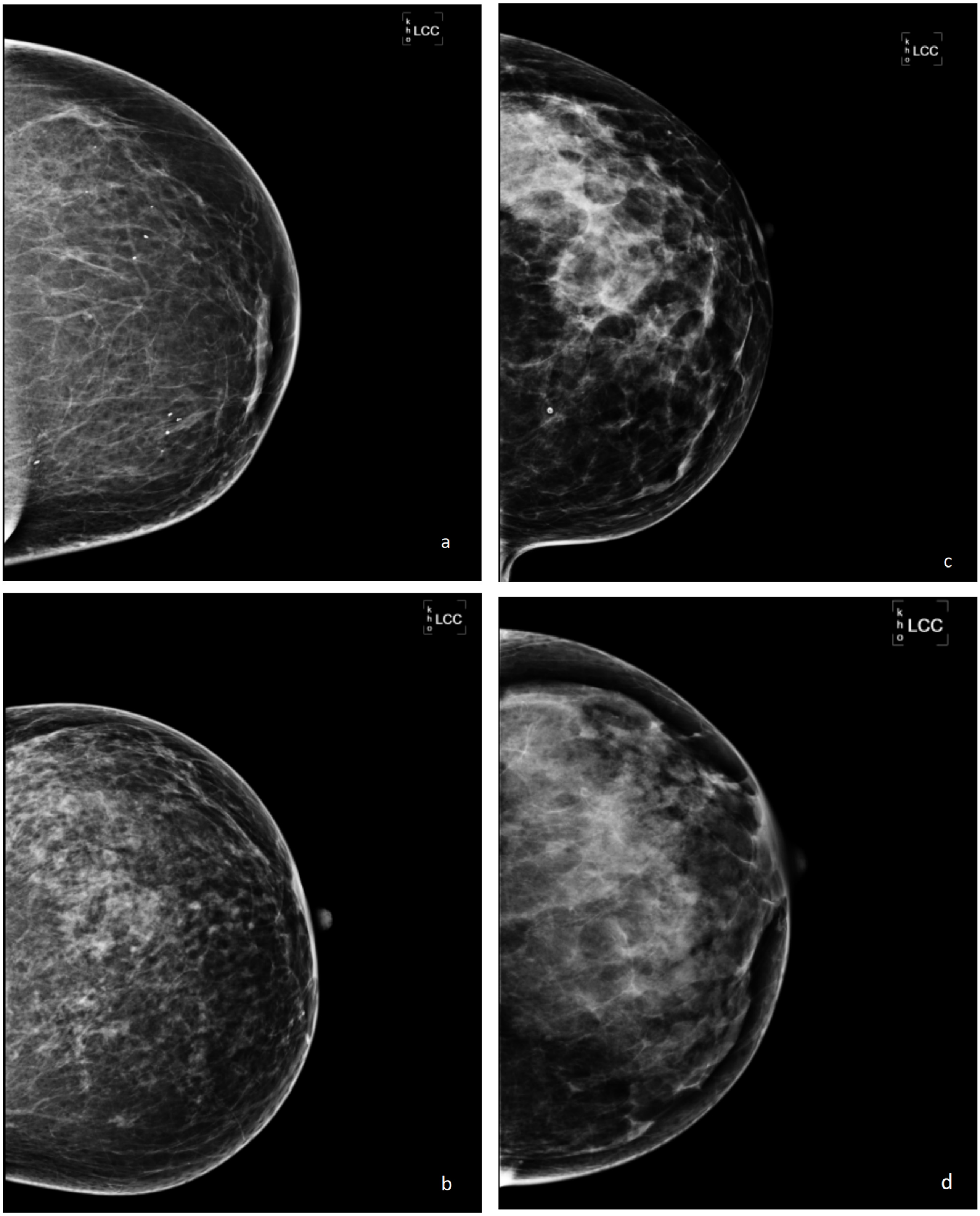

Fig. 1. CC views mammographies of four different patients: breasts with less than 25% glandular density (a; BI-RADS density 1), 25%–50% glandular density (b; BI-RADS density 2), 50%–75% density (c; BI-RADS density 3), and 50%–75% glandular density (d; BI-RADS density 4) are shown; values are based on the fourth edition BI-RADS guidelines (2003).

3.1. Intra-observer agreement

The weighted kappa values for the overall intra-observer agreement were 0.955 (95% CI: 0.931–0.980) and 0.938 (95% CI: 0.907–0.968) when the breast densities were assessed according to the criteria outlined in the fourth and fifth ACR BI-RADS editions, respectively. The difference between these two values was found to be statistically insignificant (p = .4). Table 1 presents the individual intra-observer

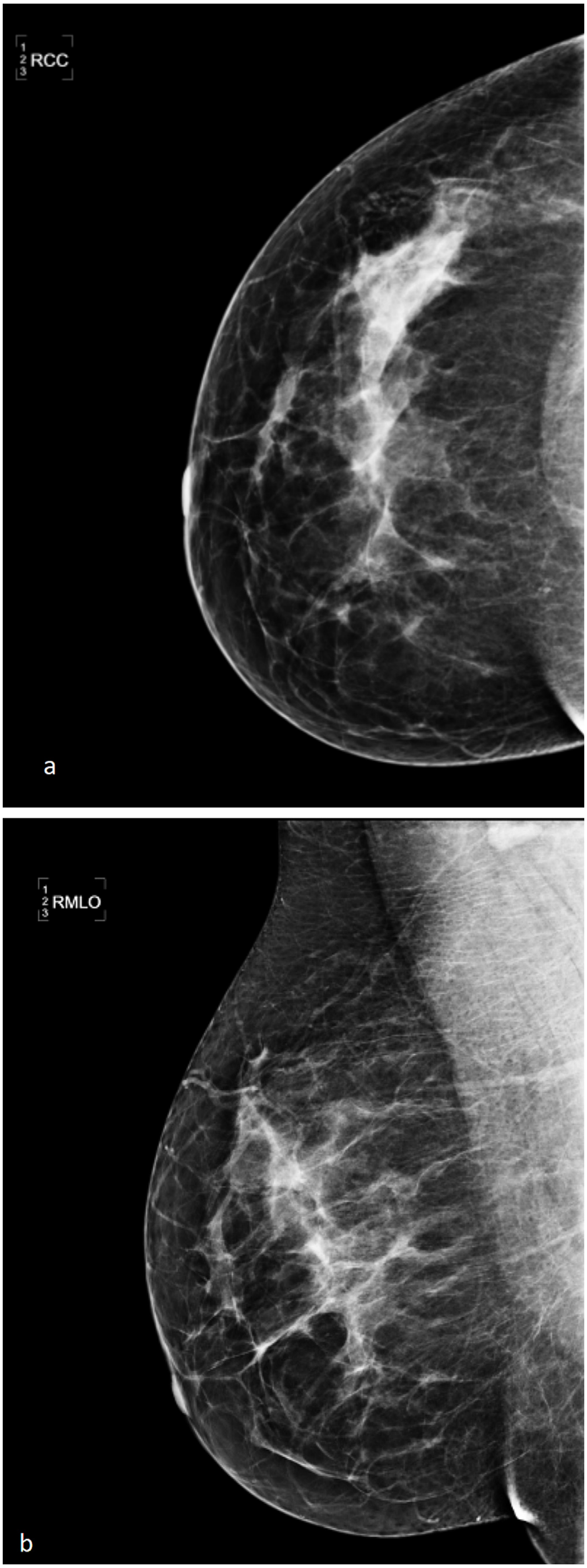

Fig. 2. Right breast mammography of a 45-year-old woman who was referred for a screening mammogram. a) Craniocaudal, CC; b) Mediolateral oblique, MLO. Her full digital mammogram was reported and her breast density was categorized as BI-RADS density 2 based on criteria outlined in the fourth ACR BI-RADS edition. Although the amount of fibroglandular tissue in her breast was below 50 percent, her breast density was categorized as BI-RADS density C based on criteria outlined in the fifth ACR BI-RADS edition because the tissue in the upper outer region of the breast was capable of obscuring a small mass.

Table 1
Intra-observer agreements for the two editions of BI-RADS criteria presented in Weighted Kappa (95% CI).

Readers	Fourth edition BI-RADS	P value*	Fifth edition BI-RADS	P value*
Reader 1	0.975 (0.946–1.000)	< 0.001	0.908 (0.849–0.967)	< 0.001
Reader 2	0.945 (0.891–0.998)	0.011	0.946 (0.894–1.000)	0.037
Reader 3	0.936 (0.885–0.987)	< 0.001	0.965 (0.924–1.000)	< 0.001
Overall	0.955 (0.931–0.980)	–	0.938 (0.907–0.968)	–

* p value calculated for the difference between the reader's kappa and the overall kappa.

Table 2
Inter-observer Agreement for Breast Density Assessment presented in Fleiss-Cohen (Quadratic) Weighted Kappa (95% CI).

Comparisons	Fourth edition BI-RADS	P value*	Fifth edition BI-RADS	P value*
Readers 1 and 2	0.805 (0.705–0.905)	< 0.001	0.885 (0.811–0.958)	< 0.001
Readers 1 and 3	0.879 (0.805–0.954)	< 0.001	0.815 (0.694–0.937)	< 0.001
Readers 2 and 3	0.777 (0.649–0.904)	< 0.001	0.828 (0.705–0.951)	< 0.001
All Readers	0.623 (0.517–0.729)	–	0.702 (0.589–0.815)	–

* p value calculated for the difference between the pairwise kappa and the overall kappa.

agreement values for each of the three radiologists when using criteria from each of the two ACR BI-RADS editions (Table 1).

3.2. Inter-observer agreement

Table 2 presents the inter-observer agreement amongst the three radiologists when using criteria from the fourth and fifth ACR BI-RADS editions, based on the Fleiss-Cohen (quadratic) weighted kappa. The overall kappa values were calculated to be 0.623 (95% CI: 0.517–0.729) and 0.702 (95% CI: 0.589–0.815) when the assessments were in accordance with criteria from the fourth and fifth ACR BI-RADS editions, respectively. The difference between these two values was not statistically significant (p = .32).

3.3. Density distribution

As presented in Table 3, the distribution of breast density

Table 3
Distribution of Density Assignment to Two Categories of Non-dense and Dense, According to the Two BI-RADS Editions.

	Fourth edition BI-RADS		Fifth edition BI-RADS		P value
	Non-dense	Dense	Non-dense	Dense	
Reader 1 (n = 72)	50	22	44	28	0.294
Reader 2 (n = 53)	35	18	34	19	0.839
Reader 3 (n = 52)	29	23	31	21	0.691
Overall (n = 177)	114	63	109	68	0.582

assignments was not significantly different when radiologists used criteria outlined in the fourth versus the fifth ACR BI-RADS editions (p = .582). This was true when evaluating overall distribution and when comparing each radiologist individually. Among the 216 assessments, the assigned density categories of 168 cases (77.7%) were comparable when using criteria outlined the fourth and fifth ACR BI-RADS editions; Additionally, when using criteria outlined in the fifth versus fourth the ACR BI-RADS editions, 34 cases (15.7%) increased in assigned density category, whereas 14 cases (6.4%) decreased in assigned density category: 14 cases (41.2%) changed from category 1 to category B, 14 cases (41.2%) changed from category 2 to category C, and 6 cases (17.6%) changed from category 3 to category D. Among the 14 cases that decreased density categories, 4 cases (28.6%) decreased from category 4 to category C, 9 cases (64.3%) decreased from category 3 to category B, and 1 case (7.1%) decreased from 2 category to category A.

When patients were categorized as having non-dense (1, 2, A and B) and dense (3, 4, C and D) breasts, use of criteria from the fourth versus fifth ACR BI-RADS editions resulted in 14 cases (6.5%) of non-dense breasts being reclassified as dense breasts and 9 cases (4.1%) of dense breasts being reclassified as non-dense breasts. Overall, the percentage of cases assigned to the dense category based on the criteria described in the fourth BI-RADS edition was 35.6%, which increased to 38.4% when assignments were made based on the criteria described in the fifth BI-RADS edition; however, this change was not statistically significant (p = .550).

4. Discussion

Different methods of measuring MBD have been proposed, including subjective qualitative and objective quantitative techniques [8–13]. Radiologists understand the importance of defining MBD correctly and therefore strive to be as steadfast as possible, particularly when using subjective methods [6]. There are a limited number of studies that have assessed variability in reporting breast density according to ACR BI-RADS criteria; this is in contrast to the large number of studies that have evaluated the effectiveness of ACR BI-RADS features and BI-RADS final assessment in mammography [18–21].

Previous studies that evaluated variability in reporting breast density suffered from several limitations. 1) They were based on a small number of radiologists or mammograms; 2) They were often conducted using film-based mammography methods; 3) In studies involving more than two radiologists, confidence intervals (CIs) for agreement estimates were not provided; 4) They failed to control for important variables, such as using different mammographic settings or failing to account for intra-observer variation [18–22].

To our knowledge, this is just the second study performed that compares methods described in the fourth and fifth ACR BI-RADS editions and their effect on observer concordance in breast density measurement. The number of radiologists and the number of mammograms in our study are both acceptable to calculate statistical accuracy of inter-observer agreement. Enough time was allowed between readings so as to avoid memory bias. The mammograms were full-field and taken digitally using two views in a screening setting.

In this study, the overall Fleiss-Cohen (quadratic) weighted kappas for inter-observer agreement were 0.623 (95% CI: 0.517–0.729) and 0.702 (95% CI: 0.589–0.815) when breast density was assessed using criteria from the fourth and fifth ACR BI-RADS editions, respectively. These values are sufficiently large to suggest inter-observer agreement with no statistically significant differences between the two figures (p = .32).

Previous studies, such as Berg et al. [19] and Ciatto et al. [20] reported moderate variability in inter-observer mammogram interpretation using criteria outlined in the fourth ACR BI-RADS edition. In a multi-center study, Ooms et al. determined the interobserver variability of breast density assessment according to fourth edition BI-RADS criteria and evaluated the relationship between breast density and risk for breast cancer. Four instructed and experienced breast radiologists assessed 57 film-based mammograms and categorized each case into ACR BI-RADS density categories 1–4. Overall, inter-observer agreement was substantial [21]. However, the authors did not measure the intra-observer variability and data were based on assessments of a mixture of screening and diagnostic mammograms, which could have biased the results.

Gard et al. [22] enrolled 19 experienced radiologists from a single center to assess breast density using 341 screening mammograms at two different time points 6 months apart. Inter-observer agreement varied considerably across radiologist pairs from slight to substantial, with kappa values ranging from 0.02 to 0.72 (mean = 0.46, 95% CI: 0.36–0.55).

In the current study, we obtained near perfect overall intra-observer agreement using weighted kappa values and 95% CIs when breast density was measured according to criteria outlined in both the fourth and fifth ACR BI-RADS editions. The weighted kappa values for the overall intra-observer agreement were 0.955 (95% CI: 0.931-0.980) and 0.938 (95% CI: 0.907–0.968) when criteria outlined in the fourth and fifth BI-RADS editions, respectively, were used, and the difference between the two figures was found to be statistically insignificant (p = .4).

Redondo et al. [23] also showed substantial intra-observer agreement in measurements of breast density ($\kappa = 0.69$, 95% CI: 0.68–0.70). Similarly, Ciatto et al. [20] reported a substantial intra-observer agreement ($\kappa = 0.71$, range 0.32–0.88) on a four-grade scale and near perfect intra-observer agreement ($\kappa = 0.81$, range 0.62–1.00) on a two-grade scale. Gard et al. [22] showed that intra-observer agreement was higher for radiologists with ≥ 10 years' experience interpreting mammograms (difference in mean kappa values = 0.10, 95% CI: 0.01–0.24).

Ekpo et al. [24] reported results similar to that of the current study, showing intra-observer agreement ranging from $\kappa = 0.86$ (95% CI: 0.77–0.93) to $\kappa = 0.89$ (95% CI: 0.81–0.95) between five readers. Inter-observer agreement ranged from substantial ($\kappa = 0.76$, 95% CI: 0.73–0.78) to almost perfect ($\kappa = 0.87$, 95% CI: 0.86–0.89).

A reason for the high levels of intra- and inter-observer agreement in the current study may be use of the training session, which used proper ACR BI-RADS atlas images to educate radiologists participating in the study.

The only other study to compare MBD analyses using criteria outlined in the fourth and fifth ACR BIRADS editions was a study by Irshad et al. [25], who reported the analysis of breast density by five radiologists of 104 mammograms four times each: twice using criteria outlined in the fourth ACR BI-RADS edition and twice using criteria outlined in the fifth edition. They found that the intra-observer agreement for MBD assessment was lower when radiologists used criteria outlined in fifth edition than when they used criteria outlined in the fourth edition (p = .0179). The overall intra-observer agreement (weighted kappa) using criteria outlined in the fourth ACR BI-RADS edition was 0.84 (95% CI: 0.80–0.87) and the overall intra-observer agreement using criteria outlined in the fifth ACR BI-RADS edition was 0.77 (95% CI: 0.73–0.81). The inter-observer agreement was higher when radiologists used the criteria outlined in the fourth ACR BI-RADS edition than that obtained when radiologists used criteria outlined in the fifth ACR BI-RADS edition (p = .006). The overall inter-observer agreement was 0.65 (95% CI: 0.61–0.69) when radiologists used criteria outlined in the fourth ACR BI-RADS edition, whereas the overall inter-observer agreement using criteria outlined in the fifth ACR BI-RADS edition was 0.57 (95% CI: 0.53–0.61) [25]. Their limitations were an outlier reader that might have affected their results and retrospective case selection from a screening mammography database

In the current study, no significant differences were found between intra-observer (p = .4) or inter-observer agreements (p = .32) calculated from MBD assessments made using criteria from the fourth or fifth ACR BI-RADS edition. It may be in some part because of the fact that three radiologists who participated in this study had a few years of experience in using BI-RADS fifth edition and in addition, prior to the start of the study they were administered an informational training session that highlighted the importance of breast cancer density and reviewed the ACR BI-RADS methods of density measurement with proper imaging demonstrated using the ACR atlas.

It is important to accurately measure breast density because of the potential for decreased mammogram sensitivity and increased cancer risk in patients with high breast density. Patients with high breast density may require supplementary screening modalities, such as sonography or MRI [6].

The changes implemented by the fifth edition release of the ACR BI-RADS guidelines will likely result in an increase in the number of patients classified as having dense breasts because the new guidelines consider even a small amount of dense tissue as heterogeneous dense fibroglandular tissue capable of hiding an underlying mass [13].

Irshad et al. [25] reported that overall, less patients were classified as having dense breasts when radiologists used the criteria outlined in the fourth edition than when they used those outlined in the fifth edition of the ACR BI-RADS guidelines ($p < 0.0001$). By contrast, the current study found no significant differences ($p = .582$) in classification of breasts as dense (3, 4, C, and D) according to the criteria outlined in the fourth edition (35.6% of cases) and criteria outlined in the fifth edition of the ACR BI-RADS guidelines (38.4% of cases). The difference between two studies might be due limitation in number of cases and different kinds of breast density in various races.

The data outlined by the current study should be interpreted with caution for two reasons. First, this study was performed at a university hospital by radiologists who were specialist in breast imaging and who had been provided a training course prior to the initiation of the study. It is not clear if these results would extend to general radiologists that work in private hospitals. Second, the high inter- and intra-observer agreements were reached when asking radiologists to specifically focus on breast density. It is unclear if this finding would be relevant in routine practice, where density is an accessory part of report.

5. Conclusion

Although overall inter- and intra-observer variability exists, ACR BI-RADS density classification is an acceptable method with substantial agreement. This study showed a substantial inter-observer level of agreement when radiologists used criteria outlined in the fourth and fifth editions of the ACR BI-RADS guidelines. The intra-observer agreement was near perfect when using criteria outlined in both the fourth and fifth editions, with no significant statistical differences calculated between the two. Although the percentage of women who were categorized as having dense breasts was lower when radiologists used criteria outlined in the fourth edition of the BI-RADS guidelines, this difference was not statistically significant.

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] V. Tesic, B. Kolaric, A. Znaor, S.K. Kuna, B. Brkljacic, Mammographic density and estimation of breast cancer risk in intermediate risk population, Breast J. 19 (1) (2013) 71–78, http://dx.doi.org/10.1111/tbj.12051.

[2] I. Saarenmaa, T. Salminen, U. Geiger, P. Heikkinen, S. Hyvärinen, J. Isola, The effect of age and density of the breast on the sensitivity of breast cancer diagnostic by mammography and ultrasonography, Breast Cancer Res. Treat. 67 (2) (2001) 117–123.

[3] C.C. Park, J. Rembert, K. Chew, D. Moore, K. Kerlikowske, High mammographic breast density is independent predictor of local but not distant recurrence for invasive breast cancer, Int. J. Radiat. Oncol. Biol. Phys. 73 (2009) 75, http://dx.doi.org/10.1016/j.ijrobp.2008.04.007.

[4] V.A. McCormack, I. dos Santos Silva, Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis, Cancer Epidemiol. Biomark. Prev. 15 (2006) 1159–1169, http://dx.doi.org/10.1158/1055-9965.EPI-06-0034.

[5] L. Yaghjyan, G.A. Colditz, L.C. Collins, S.J. Schnitt, B. Rosner, C. Vachon, et al., Mammographic breast density and subsequent risk of breast cancer in postmenopausal women according to tumor characteristics, J. Natl. Cancer Inst. 103 (15) (2011) 1179, http://dx.doi.org/10.1093/jnci/djr225.

[6] J.A. Tice, S.R. Cummings, R. Smith-Bindman, L. Ichikawa, W.E. Barlow, K. Kerlikowske, Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model, Ann. Intern. Med. 148 (2008) 337–347.

[7] W.A. Berg, Z. Zhang, D. Lehrer, et al., Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk, JAMA 307 (13) (2012) 1394–1404, http://dx.doi.org/10.1001/jama.2012.388.

[8] J.N. Wolfe, Breast patterns as an index of risk for developing breast cancer, AJR Am. J. Roentgenol. 126 (6) (1976) 1130–1137, http://dx.doi.org/10.2214/ajr.126.6.1130.

[9] N.F. Boyd, J.W. Byng, R.A. Jong, E.K. Fishell, E. Little, A.B. Miller, et al., Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian national breast screening study, J. Natl. Cancer Inst. 87 (9) (1995) 670–675.

[10] I.T. Gram, E. Funkhouser, L. Tabar, The Tabar classification of mammographic parenchymal patterns, Eur. J. Radiol. 24 (1997) 131–136.

[11] M. Jeffreys, R. Warren, R. Highnam, G.D. Smith, Initial experiences of using an automated volumetric measure of breast density: the standard mammogram form, Br. J. Radiol. 79 (941) (2006) 378–382, http://dx.doi.org/10.1259/bjr/24769358.

[12] C.J. D'Orsi, E.B. Mendelson, D.M. Ikeda, et al., Breast imaging reporting and data system, ACR BI-RADS, 4th ed., American College of Radiology, Reston, VA, 2003.

[13] C.J. D'Orsi, E.B. Mendelson, D.M. Ikeda, et al., Breast imaging reporting and data system, ACR BI-RADS, 5th ed., American College of Radiology, Reston, VA, 2013.

[14] B. Giraudeau, J.Y. Mary, Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient, Stat. Med. 20 (2001) 3205–3214.

[15] H.L. Kundel, M. Polansky, Measurement of observer agreement, Radiology 228 (2003) 303–308, http://dx.doi.org/10.1148/radiol.2282011860.

[16] J.L. Fleiss, Measuring nominal scale agreement among many raters, Psychol. Bull. 76 (5) (1971) 378.

[17] J.L. Fleiss, J. Cohen, The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, Educ. Psychol. Meas. 33 (3) (1973) 613–619.

[18] K. Kerlikowske, D. Grady, J. Barclay, S.D. Frankel, S.H. Ominsky, E.A. Sickles, et al., Variability and accuracy in mammographic interpretation using the American college of radiology breast imaging reporting and data system, J. Nat. Cancer Inst. 90 (1998) 1801–1809.

[19] W.A. Berg, C. Campassi, P. Langenberg, M.J. Sexton, Breast imaging reporting and data system: inter- and intraobserver variability in feature analysis and final assessment, Am. J. Roentgenol. 174 (6) (2000) 1769–1777, http://dx.doi.org/10.2214/ajr.174.6.1741769.

[20] S. Ciatto, N. Houssami, A. Apruzzese, E. Bassetti, B. Brancato, F. Carozzi, et al., Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories, Breast 14 (4) (2005) 269–275, http://dx.doi.org/10.1016/j.breast.2004.12.004.

[21] E.A. Ooms, H.M. Zonderland, M.J. Eijkemans, M. Kriege, D.B. Mahdavian, C.W. Burger, et al., Mammography: interobserver variability in breast density assessment, Breast 16 (6) (2007) 568–576, http://dx.doi.org/10.1016/j.breast.2007.04.007.

[22] C.C. Gard, E.J. Aiello Bowles, D.L. Miglioretti, S.H. Taplin, C.M. Rutter, Misclassification of breast imaging reporting and data system (bi-rads) mammographic density and implications for breast density reporting legislation, Breast J. 21 (5) (2015) 481–489, http://dx.doi.org/10.1111/tbj.12443.

[23] A. Redondo, M. Comas, F. Macia, F. Ferrer, C. Murta-nascimento, M.T. Maristany, et al., Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms, Br. J. Radiol. 85 (1019) (2012) 1465–1470, http://dx.doi.org/10.1259/bjr/21256379.

[24] E. Ekpo, U. Ujong, C. Mello-Thoms, M. McEntee, Assessment of interradiologist agreement regarding mammographic breast density classification using the fifth edition of the BI-RADS Atlas, Am. J. Roentgenol. 206 (2016) 1119–1123, http://dx.doi.org/10.2214/AJR.15.15049.

[25] A. Irshad, R. Leddy, S. Ackerman, A. Cluver, D. Pavic, A. Abid, et al., Effects of changes in bi-rads density assessment guidelines (fourth versus fifth edition) on breast density assessment: intra- and interreader agreements and density distribution, Am. J. Roentgenol. 207 (6) (2016) 1366–1371.