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

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.