Effect of Training on Qualitative Mammographic Density Assessment

Sughra Raza, MD[a], Meaghan M. Mackesy, MD[a], Nicole S. Winkler, MD[a], Shelley Hurwitz, PhD[a], Robyn L. Birdwell, MD[a]

Abstract

Purpose: The aim of this study was to evaluate the accuracy of visual mammographic breast density assessment and determine if training can improve this assessment, to compare the accuracy of qualitative density assessment before and after training with a quantitative assessment tool, and to evaluate agreement between qualitative and quantitative density assessment methods.

Methods: Consecutive screening mammograms performed over a 4-month period were visually assessed by two study breast radiologists (the leads), who selected 200 cases equally distributed among the four BI-RADS density categories. These 200 cases were shown to 20 other breast radiologists (the readers) before and after viewing a training module on visual density assessment. Agreement between reader assessment and lead radiologist assessment was calculated for both reading sessions. Quantitative volumetric density of the 200 mammograms, determined using a commercially available tool, was compared with both sets of reader assessment and with lead radiologist assessment.

Results: Compared with lead radiologist assessment, reader accuracy of breast density assessment increased from 65% before training to 72% after training (odds ratio, 1.41; $P < .0001$). Training specifically improved assignment to BI-RADS categories 1 ($P < .0001$) and 4 ($P < .10$). Compared with quantitative assessment, reader accuracy showed statistically nonsignificant improvement with training (odds ratio, 1.1; $P = .26$). Substantial agreement between qualitative and quantitative breast density assessment was demonstrated ($\kappa = 0.78$).

Conclusions: Training may improve the accuracy of mammographic breast density assessment. Substantial agreement between qualitative and quantitative breast density assessment exists.

Key Words: Breast density, BI-RADS density, mammographic density, volumetric breast density, breast density legislation

J Am Coll Radiol 2016;13:310-315. Copyright © 2016 American College of Radiology

INTRODUCTION

Breast tissue density is a documented independent risk factor for breast cancer, irrespective of its ability to influence mammographic detection of malignancy [1-4]. Women with high (>75%) breast tissue density have an overall fourfold to sixfold increased risk for breast cancer compared with those with low (≤10%) breast tissue density [4]. Furthermore, the incidence of both benign and malignant pathology is increased in dense compared with fatty breast tissue [5,6], with ductal carcinoma in situ 10 times more likely to occur in dense tissue [6].

[a]Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts.

Corresponding author and reprints: Meaghan M. Mackesy, MD, Brigham and Women's Hospital, Department of Radiology, 75 Francis Street, Boston, MA 02115; e-mail: mmackesy@partners.org.

The authors have no conflicts of interest related to the material discussed in this article.

Mammographic breast density categorization has been based largely on breast imagers' visual assessment. This is a subjective assessment, influenced by individual radiologist, environmental, and software display factors. The fourth edition of the ACR's BI-RADS® manual established a four-tiered density categorization system: category 1 (almost entirely fatty), with less than 25% fibroglandular tissue; category 2 (scattered fibroglandular densities), with 25% to 49% fibroglandular tissue; category 3 (heterogeneously dense), with 50% to 74% fibroglandular tissue; and category 4 (extremely dense), with ≥75% fibroglandular tissue. Despite these guidelines, training, and the individual effort of interpreting radiologists, studies have shown variability in mammographic density categorization, with the reported interobserver agreement ranging from fair ($\kappa = 0.31$) to almost perfect ($\kappa > 0.80$) [7-10].

The recognition that breast tissue density is both an independent risk factor for breast cancer and a significant

1546-1440/15/$36.00 ■ http://dx.doi.org/10.1016/j.jacr.2015.10.025

influence on mammographic detection of malignancy has led to the stratification of women into those with "dense" and those with "not dense" breast tissue [5,11-14]. To date, 24 states in the United States have enacted legislation requiring that women with dense breast tissue be notified by radiologists so that they may consider alternative screening methods, such as whole-breast ultrasound or MRI [15-18]. There is concern that the subjectivity and variability of visual mammographic density assessment by radiologists makes this an unreliable and inexact method, yet with important and far-reaching implications [19].

The purposes of this study were to evaluate readers' accuracy of visual breast density categorization by comparing reader data with density categories established for this study as "truth" by two senior study radiologists and to determine if the accuracy of this assessment is affected by training. The accuracy of visual assessment before and after training was also compared with a quantitative breast density model. Additionally, we assessed agreement between lead radiologist assessment and the quantitative breast density model.

METHODS

For the sake of clarity, study participant roles are named and defined as follows, and are henceforth referred to simply by these names:

- Study leads: two senior breast imagers, each with more than 20 years of breast imaging experience;
- Study readers: radiologists who reviewed the 200 cases to determine density category;
- Study manager: the radiology resident who organized the project; and
- Study trainers: a breast imaging fellow and one of the leads who created the training module.

For this institutional review board-approved investigation, the leads visually assessed consecutive screening mammograms obtained at our academic institution from September 2012 through January 2013. With consensus, they selected 200 cases, equally distributed among the four BI-RADS density categories. The first 50 cases from each density category encountered without significant postsurgical changes or artifacts were chosen. An equal distribution among the density categories, rather than a distribution representative of the population, was used to aid in statistical analysis. Density categorization of these 200 mammograms by the leads provided the qualitative breast density truth (leads truth [LT]) against which subsequent readers' categorizations were compared.

All digital bilateral screening mammograms used in this study were obtained in the standard craniocaudal and mediolateral oblique projections, using GE digital mammography units (Senograph ES and Senograph DS; GE Healthcare, Milwaukee, Wisconsin) in accordance with ACR practice guidelines.

Study Design

Readers included 17 board-certified radiologists (13 staff radiologists, 4 breast imaging fellows) and 3 senior radiology residents; breast imaging experience ranged from 1 to 30 years. Each reader reviewed the 200 screening mammograms on a PACS workstation, assigning and recording BI-RADS breast density categories on the basis of their habitual assessment methods. The data were collected and tabulated by the manager, who did not participate as a reader. One month later, all readers viewed a custom-designed training module, which included a review of currently available density assessment methods; limitations of visual, semi-quantitative, and quantitative breast density measurements; and a systematic visual density assessment technique (Fig. 1). This custom-designed training module was created by the trainers on the basis of the guidelines of the fourth edition of the BI-RADS manual and was presented in PowerPoint format (Microsoft Corporation, Redmond, Washington), easily viewed individually by each reader. Within 1 week of viewing this training module, the readers reviewed the same 200 screening mammograms and once again assigned and recorded a BI-RADS density category to each mammogram. Data were again collected and tabulated by the manager. The pre- and posttraining reader density assessments were compared with LT.

Additionally, the 200 mammograms were processed using a commercially available computer-based volumetric density measurement tool (VolparaDensity; Volpara Solutions, Wellington, New Zealand), and quantitative breast density truth was established (quantitative truth [QT]). QT was compared with the pre- and posttraining reader density assessments and with LT.

Statistical Analysis

Data were summarized by frequencies and percentages for readings overall. Means, standard deviations, lowest frequencies, and highest frequencies were used to summarize accuracy before and after training for the cohort. The main analysis addressing the effect of training on accuracy was a logistic regression predicting a match with LT and QT, taking into account the hierarchical structure of the data with 200 mammograms evaluated by 20 readers before and after training. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated using the SAS version 9.2 genmod procedure with the logit link function to model