Full Field Digital Mammography and Breast Density:

Comparison of Calibrated and Noncalibrated Measurements

John J. Heine, PhD, Erin E. E. Fowler, BS, Chris I. Flowers, MD

Rationale and Objectives: Mammographic breast density is an important and widely accepted risk factor for breast cancer. A statement about breast density in the mammographic report is becoming a requirement in many States. However, there is significant inter-observer variation between radiologists in their interpretation of breast density. A properly designed automated system could provide benefits in maintaining consistency and reproducibility. We have developed a new automated and calibrated measure of breast density using full field digital mammography (FFDM). This new measure assesses spatial variation within a mammogram and produced significant associations with breast cancer in a small study. The costs of this automation are delays from advanced image and data analyses before the study can be processed. We evaluated this new calibrated variation measure using a larger dataset than previously. We also explored the possibility of developing an automated measure from unprocessed (raw data) mammograms as an approximation for this calibrated breast density measure.

Materials and Methods: A case-control study comprised of 160 cases and 160 controls matched by age, screening history, and hormone replacement therapy was used to compare the calibrated variation measure of breast density with three variants of a noncalibrated measure of spatial variation. The operator-assisted percentage of breast density measure (PD) was used as a standard reference for comparison. Odds ratio (OR) quartile analysis was used to compare these measures. Linear regression analysis was applied to assess the calibration's impact on the raw pixel distribution.

Results: All breast density measures showed significant breast cancer associations. The calibrated spatial variation measure produced the strongest associations (OR: 1.0 [ref.], 4.6, 4.3, 7.4). The associations for PD were diminished in comparison (OR: 1.0 [ref.], 2.7, 2.9, 5.2). Two additional non-calibrated measures restricted in region size also showed significant associations (OR: 1.0 [ref.], 2.9, 4.4, 5.4), and (OR: 1.0 [ref.], 3.5, 3.1, 4.9). Regression analyses indicated the raw image mean is influenced by the calibration more so than its standard deviation.

Conclusion: Breast density measures can be automated. The associated calibration produced risk information not retrievable from the raw data representation. Although the calibrated measure produced the stronger association, the non-calibrated measures may offer an alternative to PD and other operator based methods after further evaluation, because they can be implemented automatically with a simple processing algorithm.

Key Words: Breast density; calibration; automation; breast cancer risk.

Mammographic breast density is a significant factor for breast cancer risk that has been studied for many years (1,2). A statement about breast density is a part of the radiological report according to the fourth edition of the Breast Imaging Reporting and Data System (BI-RADS) (3). BI-RADS breaks down the estimate of breast density into quartiles: from almost entirely fat (0%–25% glandular tissue) to extremely dense (75%–100% glandular tissue). The breast density part of this report is meant to guide referring physicians to the risk of a cancer being obscured by the background tissue. The downside is that there is significant variation in the way breast density is reported from the 2D examination read by the radiologist (4). BI-RADS breast density is also used as a measure of risk in research (2) as a coarse approximation for the percentage of breast density measure. Breast density as a breast cancer risk factor is not currently used in clinical practice due to the lack of standardization and automation of its measurement (5). The attributes of an automated breast density measure for clinical applications should have a high degree of replication and translate across imaging platforms without extensive modification.

There are various methods used to assess breast density, as reviewed previously (6). For the most part, the breast density and breast cancer associations have been developed with measurements that did not consider the inter-image

Acad Radiol 2011; 18:1430–1436

From the Cancer Prevention & Control Division (J.J.H., E.E.E.F.) and Diagnostic Imaging (C.I.F.), H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, Tampa, FL 33612. Received May 27, 2011; accepted July 27, 2011. Supported by NIH grant #R01 CA114491. The authors thank Gail Tiffenberg for her outstanding recruiting efforts. **Address correspondence to:** J.J.H. e-mail: john.heine@moffitt.org

doi:10.1016/j.acra.2011.07.011

acquisition technique differences. In particular, the operator-assisted percentage of breast density approach (or PD) has shown repeatedly to correlate well with breast cancer (2) without considering the acquisition technique. Methods for automating PD are not widely used (6). An alternative method of assessing breast density is to calibrate, or adjust, for the acquisition technique differences (7–11).

Calibration should reduce unwanted measurement variation and produce a measure of mammographic density that shows stronger associations with breast cancer than non-calibrated methods such as PD. However, measurements based on calibration with digitized film mammography have produced mixed findings. Some work shows that calibration does not produce anything beyond PD (12,13). Other researchers found that calibration strengthens the breast density associations with film mammography (14). Using full field digital mammography (FFDM), we have shown that calibration can be used to both describe PD (15) and to develop a new measure of breast density (16). This new measure is calculated as the standard deviation (SD) of the calibrated pixels within the breast area, which captures spatial variation. This measure provided stronger associations with breast cancer than PD in a small study (16).

Our calibration methodology was described in detail previously (17–20) and is briefly discussed here to put the various measurements in context. The calibration produces image data normalized for the inter-image acquisition technique differences at the pixel level (or more coarse scales) referred to as the percent glandular representation, which is a normalized effective x-ray attenuation coefficient metric. Differences in the compressed breast thickness, target-filter combination, x-ray tube voltage, and exposure are rectified by the calibration process. There are many technical problems (15,18) that if not addressed will introduce considerable error into the calibration output.

The calibration may not influence the moments of the raw pixel distribution uniformly. If the calibration primarily operates on the central (or mean) value of the pixel distribution for a given image, the standard deviation of the raw pixel values (derived from non-calibrated images) may also be a measure of breast cancer risk. The objectives of this work were 1) evaluate the new calibrated standard deviation measure (or PG_{SD}) with a larger dataset than used previously, 2) explore the possibility of developing a breast density measure without calibration that shows a similar association with breast cancer as PG_{SD}, and 3) characterize the calibration influence on the raw pixel distributions. To meet these objectives, we performed a case-control study to evaluate PG_{SD} and explored the standard deviation from the raw FFDM images as the breast density metric. For one measure, the standard division was calculated from the raw data using the same region as for PG_{SD}. Two additional standard deviation measures were considered from the raw data by restricting the region sizes. These noncalibrated measures were compared with PG_{SD} using their association with breast cancer as the endpoint. PD was applied to raw mammograms and used as a common reference for comparison. Regression analysis was used to compare calibrated and noncalibration pixel distribution characteristics to understand the quantities most influenced by calibration. Our previous work was performed at a more coarse calibration scale (15,16). In contrast, the calibration was performed at the pixel level for this report.

METHODS

Study Population

The patient accrual was part of an ongoing case-control study. The study population, selection methods, and matching particulars have been discussed previously (15) and are not discussed here in detail. In brief, the study accrual has been updated in this report to include more participants. In this IRB approved study, women diagnosed with a primary breast cancer (September 2007-March 2011) were included as cases ($n = 160$) identified from those attending the breast clinics at the H. Lee Moffitt Cancer Center. For the controls, three groups of cases were considered based on their screening history. Group 1 was comprised of women that had a negative screening mammogram within 30 months prior to their breast cancer diagnosis ($n_1 = 141$). Group 2 was comprised of women who had a negative screening history that fell outside of the group 1 parameters, such as a woman who had a screening in 2007 but not again until 2010 at which time she was diagnosed with cancer ($n_2 = 14$). Group 3 was comprised of women who were just starting screening and were diagnosed at their baseline mammogram ($n_3 = 5$). Case data and images were either located by retrospective records review ($n = 52$) for those women with images archived on the study FFDM unit or recruited, consented, and imaged for the study ($n = 108$). Controls ($n = 160$) were identified retrospectively from the pool of women undergoing breast cancer screening mammography at the H. Lee Moffitt Cancer Center with archived images acquired with the study FFDM unit and were matched (individual) to their cases by age (±2 years) and hormone replacement therapy usage and duration (±1 year).

Spatial Variation Breast Density

Various breast density measures and their association with breast cancer were compared using a matched case-control design. To reduce anomalous spatial variation, the analysis was contained to the portion of the image that was in contact with the compression paddle during imaging. Using methods described previously (15,19), the breast image area was eroded by 25% along a radial direction. This area defined the effective breast area. The degree of breast area reduction is an approximation that eliminates anomalous region where the compressed breast thickness is not well defined. Both PG_{SD} and the standard deviation calculated from the raw data (or R_{SD}) were derived from this modified breast area. The measures R_{SDL} and raw image