Reader Variability in Breast Density Estimation from Full-Field Digital Mammograms:

The Effect of Image Postprocessing on Relative and Absolute Measures

Brad M. Keller, PhD, Diane L. Nathan, MD, Sara C. Gavenonis, MD, Jinbo Chen, PhD, Emily F. Conant, MD, Despina Kontos, PhD

Rationale and Objectives: Mammographic breast density, a strong risk factor for breast cancer, may be measured as either a relative percentage of dense (ie, radiopaque) breast tissue or as an absolute area from either raw (ie, "for processing") or vendor postprocessed (ie, "for presentation") digital mammograms. Given the increasing interest in the incorporation of mammographic density in breast cancer risk assessment, the purpose of this study is to determine the inherent reader variability in breast density assessment from raw and vendor-processed digital mammograms, because inconsistent estimates could to lead to misclassification of an individual woman's risk for breast cancer.

Materials and Methods: Bilateral, mediolateral-oblique view, raw, and processed digital mammograms of 81 women were retrospectively collected for this study (N = 324 images). Mammographic percent density and absolute dense tissue area estimates for each image were obtained from two radiologists using a validated, interactive software tool.

Results: The variability of interreader agreement was not found to be affected by the image presentation style (ie, raw or processed, F-test: $P > .5$). Interreader estimates of relative and absolute breast density are strongly correlated (Pearson $r > 0.84$, $P < .001$) but systematically different (t-test, $P < .001$) between the two readers.

Conclusion: Our results show that mammographic density may be assessed with equal reliability from either raw or vendor postprocessed images. Furthermore, our results suggest that the primary source of density variability comes from the subjectivity of the individual reader in assessing the absolute amount of dense tissue present in the breast, indicating the need to use standardized tools to mitigate this effect.

Key Words: Digital mammography; breast density; reader variability; breast cancer risk.

Breast cancer is currently the most commonly diagnosed cancer in women and is projected to account for 29% of all new cancer cases in women in the United States this year (1). Although it is expected that one in eight women will develop breast cancer over the course of their life (2), previous studies have identified multiple demographic and lifestyle risk factors that are associated with an increased risk for developing breast cancer, such as age, weight, ethnicity, parity, and family history (3,4). Comprehensive assessment of an individual woman's risk for breast cancer could lead to personalized screening regimens using complementary or alternative imaging modalities to mammography such as ultrasound or magnetic resonance imaging (5).

In addition to demographic risk factors, several studies have also identified that mammographic breast density, commonly measured as the relative amount of radiopaque fibroglandular breast tissue, is a strong, independent risk factor for breast cancer (6–8). Clinically, breast density is most commonly estimated by radiologists via visual assessment as the amount of mammographically dense tissue, or "white areas," and then categorized using the American College of Radiology four-class breast-imaging reporting and data system (BIRADS) (9) or the Boyd six-category scale (10). In addition, continuous measures of breast percent density (PD%), acquired using interactive image thresholding software (8), have also been widely used, primarily in the research setting, as a more precise, quantitative measures in the effort to better estimate the risk for breast cancer associated with increasing amounts of fibroglandular tissue.

As the standard of practice moves toward the use of full-field digital mammography (DM) (11), a number of

Acad Radiol 2013; 20:560–568

From the Departments of Radiology (B.M.K., D.L.N., S.C.G., E.F.C., D.K.) and Biostatistics and Epidemiology (J.C.), University of Pennsylvania Perelman School of Medicine, 3600 Market St. Suite 360, Philadelphia, PA 19104. Received December 6, 2012; accepted January 9, 2013. **Address correspondence to:** B.M.K. e-mail: Brad.Keller@uphs.upenn.edu

http://dx.doi.org/10.1016/j.acra.2013.01.003

considerations arise that may affect breast density assessment. First, it has been suggested that the breast, and subsequently breast density, may have a different appearance between analog and digital mammograms (12), which can lead to differences in density assessment between the two modalities (13). In addition, DM in general produces two types of images, a raw (eg, "for processing") image with gray-level intensity values proportional to the x-ray attenuation through the breast and a vendor-processed (eg, "for presentation") image with increased tissue contrast and lesion conspicuity, which is used for radiological interpretation and diagnostic evaluation (Fig 1). It has been previously recommended that breast density should be assessed using the raw images (14) because they maintain a proportional relationship between image gray-level intensity and the underlying tissue x-ray attenuation properties. However, the majority of clinical density assessments performed by radiologists are primarily done on the vendor-processed images because these are the ones used for clinical interpretation and archived by most clinical centers (15). Thus, given the recent interest in the incorporation of breast density in breast cancer risk estimation (16), it becomes necessary to understand the variability in breast density assessment in raw and vendor postprocessed digital mammograms because inconsistent estimates could to lead to misclassification of an individual woman's risk for breast cancer (17).

Because the majority of reader variability studies to date have focused on the use of categorical estimates of breast density (12,18–21), the purpose of our study is to determine reader variability in estimating continuous, quantitative measures of mammographic breast density from raw versus processed DM images. Preliminary work by our group (22) analyzing intrareader assessment of PD% estimates made by a single reader showed that raw and processed PD% estimates are highly correlated ($r = 0.97$), yet have a small, statistically significant difference (approximately 1.5%) between them. In this work, we introduce a second reader to assess interreader variability, consider both quantitative and categorical estimates of breast density, and perform thorough statistical analysis (including Bland-Altman analysis) to analyze both intra- and interreader agreement and the expected ranges of reproducibility in density estimation. In addition, limited studies are currently available on the reproducibility of absolute versus relative (ie, percent) breast density measures, which are also suggested to be related to breast cancer risk (23). Thus, because absolute breast density may capture complementary information about breast cancer risk, it is also beneficial to characterize reader variability in the assessment of absolute dense tissue area in addition to the more commonly used relative (eg, percentage) breast density measures. Finally, this study aims to provide insight into the sources of reader disagreement. By investigating the reader variability of density estimation in raw versus postprocessed digital mammograms, our study offers understanding not only on the effect of the imaging format, but also on the potential biases introduced by inherent differences in the readers' assessment of the dense tissue. Identifying any such biases could be instrumental in guiding the appropriate use of DM images in breast density estimation and breast cancer risk assessment.

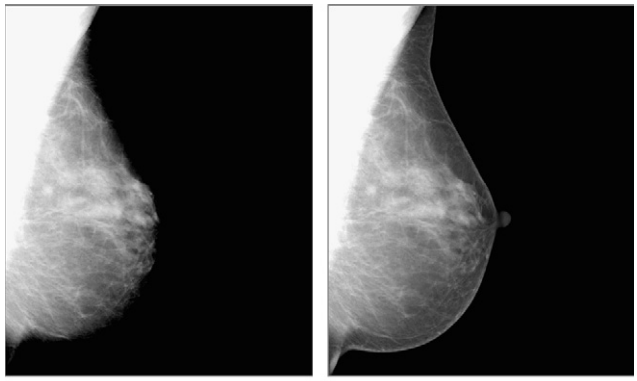

Figure 1. Examples of raw **(a)** and processed **(b)** mediolateral-oblique view mammograms of a breast-imaging reporting and data system category II breast with scattered densities from a 53-year-old woman. In general, improved tissue contrast and a more pronounced skin line can be seen in the processed image when compared to the raw digital mammogram.

MATERIALS AND METHODS

Study Population and DM Image Acquisition

We retrospectively analyzed, in compliance with the Health Insurance Portability and Accountability Act and University of Pennsylvania institutional review board approval #811761, DM images acquired as part of a separate multimodality imaging trial previously completed in our department (July 2007 to March 2008; trial sponsor, GE Healthcare, site principal investigator, E.F. Conant) as has been previously described (24). Trial participants were asymptomatic women who presented for annual screening mammography and had given written informed consent before their participation in the trial. Of the 83 women originally enrolled in the trial, two were excluded from this analysis: one because of a diagnosis of breast cancer and the other because of insufficient image quality. The remaining 81 women included in this study had a mean age of 52.9 years (standard deviation: 9.5 years) and an average Gail life-time risk for breast cancer of 11.22% ± 7.46%, which is considered the standard average risk for the general population. For these women, bilateral, mediolateral-oblique view mammograms with 100 μm isotropic resolution acquired using a full-field DM unit (Senographe DS; GE Healthcare, Chalfont St Giles, UK) were analyzed. The raw (eg, for processing) images were acquired at an original 14-bit gray-level depth. The raw mammograms were then processed using PremiumView (GE Healthcare), a vendor-specific algorithm, producing 12-bit gray level postprocessed (eg, for presentation) images for clinical interpretation. A total of 324 mediolateral-oblique (162 raw and