Published in final edited form as:

Acad Radiol. 2013 May ; 20(5): 560–568. doi:10.1016/j.acra.2013.01.003.

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, and Despina Kontos, PhD

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

Abstract

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 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.

Keywords

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 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 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.

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 $11.22\% \pm 7.46\%$, 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 162 processed) digital images were thus available for density analysis in this study. For this retrospective analysis, the requirement of informed consent was waived under institutional review board approval.

Radiologist Estimation of PD%

For the 324 images analyzed in this study, PD% and absolute dense tissue area were estimated by two fellowship-trained, board certified breast-imaging radiologists (with 2.5 years' and 5 years' breast imaging experience, respectively) using a widely validated, interactive, image-thresholding software tool for breast density estimation (Cumulus, version 4.0, University of Toronto) (22,25) with which both readers have had prior experience (15). First, each digital mammogram was adjusted by the radiologist to a window level to optimize the display for density assessment. This was followed by application of a manually determined intensity threshold to identify and subsequently exclude background air and to identify the breast edge. The pectoral muscle edge was then manually delineated and excluded from subsequent analyses. The remainder of the image was designated as the breast tissue region of interest, and the total breast area, B_A , was computed automatically by the software. After the breast tissue area identification, a second gray-level intensity threshold was interactively chosen by the radiologist to identify the relatively bright fibroglandular tissue from the remaining, relatively darker adipose tissue. Pixels within the segmented breast area with a higher gray-level value than the selected threshold were designated as dense tissue. The absolute area of the dense tissue, D_A , was then recorded and used to compute PD% as:

$$PD\% = \frac{D_A}{B_A} \quad (1)$$

Both raw and processed images were analyzed with this process, except in the case of the raw digital images; because they have not been optimized for clinical visualization and interpretation, the radiologist had to manually window and level the image to improve local image contrast before each of the segmentation and thresholding steps described previously was possible.

Statistical Analysis

All statistical analyses in this work were performed using MATLAB (Matlab 2012b, Mathworks, Natick, MA); all statistical tests were two-sided and performed at an $\alpha = 0.05$ significance level.

Per-woman, Inter-breast Repeatability of Density Estimates

Agreement between each reader's left and right breast density estimates for an individual woman was assessed via the Pearson correlation (26). Student's paired *t*-test was applied to determine if there were systematic differences between left and right breast density estimates. The Bland-Altman 95% limits of agreement (27) were calculated to demonstrate the level of agreement between left and right breast density estimates, and F-test for equal variances was applied to determine statistical significance between this spread of measure disagreement. These analyses were independently performed on each reader's raw and processed DM breast density estimates.

Reader Agreement in Density Estimates from Raw and Processed DM Images

Agreement between each reader's raw and processed breast density estimates made on individual breasts was assessed via the Pearson correlation, as above. Student's paired *t*-test was applied to determine if there were systematic differences between raw and processed breast density estimates. Bland-Altman limits of agreement were calculated to demonstrate the level of agreement between a single reader's breast density estimates made on raw and processed DM images, and F-test for equal variances was applied to determine if the two readers varied statistically in terms of their raw-processed density estimate agreement.

Interreader Agreement of Continuous Breast Density Estimates

To assess interreader agreement of continuous estimates of absolute and percentage breast density, the average of each individual reader's left and right breast density estimates for the same woman was computed to create a per-woman density score, as is commonly done in clinical practice. Agreement between the two reader's raw and processed, per-woman breast density estimates was assessed via the Pearson correlation as well as linear regression, the latter of which provides an estimate of the systematic difference between the scales of reported density estimates for the two readers. Student's paired *t*-test was applied to determine if there were systematic differences between the breast density estimates made by the two readers and Bland-Altman limits of agreement were calculated to demonstrate the level of agreement between the two readers. The F-test for equal variances was applied to determine if reader variability was affected by whether raw or processed DM image data were used to generate each woman's density estimates.

Intra- and interreader Agreement of Categorical BIRADS Density Estimates

Finally, we investigated reader agreement of categorical breast density estimates using quadratic-weighted Cohen's kappa, κ_w , statistics (28). Categorical estimates of breast density were generated by converting the each reader's continuous PD% estimates into the four American College of Radiology BIRADS density categories (9) using the standard category ranges of 0% to 25%, 26% to 50%, 51% to 75%, and 76% to 100%. Intrareader agreement was assessed by comparing each reader's categorical raw and processed density estimates. Interreader agreement was assessed by comparing the two reader's raw versus processed categorical density estimates.

RESULTS

Per-woman, Inter-breast Repeatability of Density Estimates

Each reader's left and right breast density estimates for the same woman were highly correlated by Pearson correlation (all correlations $r > 0.97$, $P < .001$). Student's paired t -test showed no statistically significant differences ($P > .1$) between a woman's left and right PD % or absolute dense area for either reader or image presentation method (ie, raw or processed images). It was observed that each reader's density estimates made on the raw DM images had relatively narrower Bland-Altman 95% limits of agreement when compared to estimates made on processed DM images, although this decreased variability was only statistically significant in the case of a single reader's PD% estimates (F-test: $P < .001$). A summary of the per-woman, inter-breast reproducibility metrics is provided in Table 1.

Reader Agreement in Density Estimates from Raw and Processed DM Images

Each readers' raw and processed DM breast density estimates were found to be highly correlated ($r > 0.96$; $P < .001$). Some systematic differences between the raw and processed estimates, although small in magnitude, were observed. In particular, one reader's PD% estimates from processed images were significantly smaller when compared to those from raw images were used (mean difference: -1.23% ; paired t -test: $P = .002$). In contrast, the second reader's absolute dense tissue area estimates from processed images were significantly larger when compared to those from raw images (mean difference: 0.96 cm^2 ; paired t -test: $P < .001$). The readers were also found to vary significantly in terms of the spread of their raw and processed density estimate disagreement, both for PD% (F-test: $P < .001$) and absolute dense tissue area (F-test: $P < .001$). Table 2 provides a summary of the intrareader, raw versus processed density agreement analysis statistics.

Interreader Agreement of Continuous Breast Density Estimates

The two readers' density estimates were found to be strongly correlated to one another ($r > 0.84$, $P < .001$) for both raw and processed DM images. Figures 2 and 3 provide scatter and Bland-Altman plots showing the association between the two readers' PD% and dense tissue area estimates. It was observed that, although there is a strong linear relationship between the readers' density estimates, Student's pair t -test found that both the PD% estimates and dense tissue area estimates were statistically different between the two readers ($P < .001$) for both raw and processed images. Finally, the variability of interreader differences in density assessment was not found to be affected by the image presentation style (ie, raw or processed, F-test: $P > .5$). A summary of the between reader analysis statistics is provided in Table 3.

Interreader Agreement of Categorical BIRADS Density Estimates

Both readers were found to have very high agreement between their respective raw and processed BIRADS density assessments ($\kappa_w = 0.84$, $P < .001$). Interreader agreement was

observed to be similar independent of whether density assessments were made on raw or processed ($\kappa_w = 0.760$, $P < .001$) DM images. Table 4 provides the relationship between two readers' categorical density estimates. In general, it was observed that there was a higher level of reader disagreement when categorically assessing higher density women (ie, BIRADS III and IV) versus lower density women, consistent with what was observed for continuous density estimates, as can be seen in Figures 2 and 3.

DISCUSSION

In this study, we assessed reader variability and the magnitude of reader disagreement in estimating mammographic breast density from raw versus processed DM images. We found that readers had a similar level of interreader variability regardless of whether raw or processed digital mammograms were used for density estimation ($P > .5$). We also observed that interreader differences can have approximately two to 10 times the variability as intrareader differences, as can be seen when comparing the Bland-Altman limits of agreement presented in Tables 1, 2, and 3. We also found that the readers were as consistent in categorical assessment as has been seen in previous studies (weighted- κ : 0.61–0.84) (12,19,29). Finally, we also observed that, for the same reader, both relative and absolute density estimates are consistent across a woman's two breasts. This implies that density estimates made on a single breast image, such as the contralateral breast in women with breast cancer, are reflective of a woman's inherent breast density.

Overall, our results show that in general, individual readers rank women in the same general order from low density to high density, as evidenced by the strong correlation seen between their density estimates. However, the readers were found to report different absolute magnitudes for each woman's breast density, leading to the interreader variation in density estimates. Furthermore, we found that one reader underestimated breast density overall as compared to the second reader, especially in the case of denser breasts, as evidenced by the regression and Bland-Altman plots presented in Figure 3 and the Bland-Altman limits of agreement in Table 3. These two findings may have implications in the use of breast density in breast cancer risk assessment, because different interpretations of breast density may lead to different risk profiles for the same woman. In addition, we found that this variation may be due to the visual perception of what constitutes "dense tissue" by the individual readers, as illustrated in Figure 4. The use of either training and standardization protocols or fully automated, computer-aided density assessment tools (30–33) could therefore be warranted as a means to alleviate the biases introduced by interreader variability.

One interesting finding from this study is that the use of raw or processed digital mammograms was not found to significantly affect the reader estimates of relative percentage or absolute breast density nor the interreader variability of these estimates (F-test: $P > .5$). Previous studies have suggested that raw mammograms should be used for breast density assessment because they maintain the original relationship between the physical properties of the breast tissue and x-ray attenuation levels (22). Our findings suggest that breast density estimates made on processed digital mammograms could provide reliable density estimates for those cases in which the original, raw mammograms are not available for analysis.

The ability to consistently measure breast density, whether PD% or dense tissue area, has become increasingly important for both personalized risk assessment and for clinical practice. Previous studies investigating reader agreement in breast density assessment have primarily focused on digitized screen-film mammograms (12,19–21,29). In general, categorical assessment of breast density made using digitized film mammograms has shown strong levels of intrareader agreement (12,20). Moderate to strong interreader agreement has

also been reported, with estimated weighted Cohen's kappa's ranging from 0.61 to 0.84 (12,19,29). Continuous measures of PD% made from digitized film screen mammograms are also generally associated with high interreader agreement (20,34), although substantial disagreements of up to 25% between individual readers have also been reported (20).

Of the few studies that have investigated the reproducibility of breast density estimates in DM, most have focused on characterizing the categorical variability of density estimates (18,35). For example, Pérez-Gómez et al found high intra- and interrater agreement when using the six-class Boyd percent density scale (18). Studies that investigate the reproducibility of fully continuous measures of breast density have primarily focused on measuring the correlations between reader estimates of PD% only in vendor postprocessed DM images (15,25). Our study contributes to this current understanding of breast density estimate variability in DM by comparing reader agreement using both raw and vendor postprocessed DM images. Furthermore, this work also investigates differences in the assessment of absolute dense tissue area in addition to the relative (eg, percentage) measures, providing further insight into the sources of reader disagreement.

One limitation of this study was that only two readers were available to provide breast density estimates, preventing a more systematic analysis of the generalizability of our findings. Considering this report as a proof-of-concept evaluation, future larger studies will seek to investigate inter- and intrareader variability using additional readers, which would also allow for the assessment of the reproducibility of reader density estimates across other additional factors such as clinical experience. Furthermore, as only a single read was performed by each reader, the reproducibility of the density estimates over time could not be assessed. Additionally, in this study, the density assessments were made on digital mammograms acquired using one model from a single manufacturer. Although it may be expected that raw images may offer a more consistent modality for assessing breast density across different device manufacturers, the use of different vendor postprocessing algorithms may alter the appearance of "for presentation" mammograms, the more commonly type of read mammogram, leading to the possibility of additional variability in density. This could have an impact on multisite, multivendor analysis of breast density estimates obtained from processed images, especially in those cases where specific sites also use a single model of DM system. Furthermore, this may also affect the longitudinal assessment of an individual woman's breast density, particularly in those cases where women are imaged over time using DM systems from different vendors.

In conclusion, the results of our study suggests that 1) the primary source of density variability comes from the subjectivity of the individual reader, rather than differences in the DM imaging format, and that 2) both absolute and relative percent breast density may be measured by clinical readers with equal intrareader reliability from either raw or vendor postprocessed DM data. Given that most imaging clinics routinely use and store only the postprocessed DM images because of cost and storage limitations, breast PD% estimation from the postprocessed images may provide a means for easier integration of breast density measures in breast cancer risk stratification in clinical practice. Further studies are under way to also assess intrareader reproducibility of density estimates from raw and processed images as compared to fully automated measures (32), which could be useful in alleviating the effect of reader subjectivity. The results of our investigation could have implications on the appropriate use of digital mammographic images for breast density assessment in breast cancer risk estimation.

Acknowledgments

This work was supported in part by the American Cancer Society Grant No. RSGHP-CPHPS-119586 (Investigating the role of quantitative breast imaging in predicting breast cancer risk), the National Institutes of Health PROSPR Grant No. 1U54CA163313-01 (Investigating the role of quantitative breast imaging in personalizing and communicating breast cancer risk assessment information to individual women) and Grant No. R01 CA161749 (Investigating the effect of breast density on screening recall).

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin.* 2012; 62:10–29. [PubMed: 22237781]
2. Hayat MJ, Howlader N, Reichman ME, et al. Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) Program. *Oncologist.* 2007; 12:20–37. [PubMed: 17227898]
3. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst.* 1989; 81:1879–1886. [PubMed: 2593165]
4. McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ.* 2000; 321:624–628. [PubMed: 10977847]
5. Lee CH, Dershaw DD, Kopans D, et al. Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *J Am Coll Radiol.* 2010; 7:18–27. [PubMed: 20129267]
6. Tice JA, Cummings SR, Ziv E, et al. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res Treat.* 2005; 94:115–122. [PubMed: 16261410]
7. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev.* 2006; 15:1159–1169. [PubMed: 16775176]
8. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med.* 2007; 356:227–236. [PubMed: 17229950]
9. D'Orsi, CJ.; Bassett, LW.; Berg, WA., et al. Breast imaging reporting and data system: ACR BI-RADS, Mammography. 4th ed.. Reston, VA: American College of Radiology; 2003.
10. Byng JW, Boyd NF, Fishell E, et al. The quantitative analysis of mammographic densities. *Phys Med Biol.* 1994; 39:1629–1638. [PubMed: 15551535]
11. Mandelblatt JS, Stout N, Trentham-Dietz A. To screen or not to screen women in their 40s for breast cancer: is personalized risk-based screening the answer? *Ann Intern Med.* 2011; 155:58–60. [PubMed: 21727294]
12. Spayne MC, Gard CC, Skelly J, et al. Reproducibility of BI-RADS Breast Density Measures Among Community Radiologists: a prospective cohort study. *Breast J.* 2012; 18:326–333. [PubMed: 22607064]
13. Harvey JA. Quantitative assessment of percent breast density: analog versus digital acquisition. *Technol Cancer Res Treat.* 2004; 3:611–616. [PubMed: 15560719]
14. Yaffe MJ. Mammographic density. Measurement of mammographic density. *Breast Cancer Res.* 2008; 10:209–218. [PubMed: 18598375]
15. Conant E, Li D, Gavenonis S, et al. A comparative study of the inter-reader variability of breast percent density estimation in digital mammography: potential effect of reader's training and clinical experience. *Dig Mammogr.* 2010:114–120.
16. Chen J, Pee D, Ayyagari R, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *J Natl Cancer Inst.* 2006; 98:1215–1226. [PubMed: 16954474]
17. Stone J, Ding J, Warren RM, et al. Predicting breast cancer risk using mammographic density measurements from both mammogram sides and views. *Breast Cancer Res Treat.* 2010; 124:551–554. [PubMed: 20544272]

18. Perez-Gomez B, Ruiz F, Martinez I, et al. Women's features and inter-/intra-rater agreement on mammographic density assessment in full-field digital mammograms (DDM-SPAIN). *Breast Cancer Res Treat.* 2012; 132:287–295. [PubMed: 22042363]
19. Ooms EA, Zonderland HM, Eijkemans MJ, et al. Mammography: interobserver variability in breast density assessment. *Breast.* 2007; 16:568–576. [PubMed: 18035541]
20. Gao J, Warren R, Warren-Forward H, et al. Reproducibility of visual assessment on mammographic density. *Breast Cancer Res Treat.* 2008; 108:121–127. [PubMed: 17616811]
21. Ciatto S, Houssami N, Apruzzese A, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. *Breast.* 2005; 14:269–275. [PubMed: 16085233]
22. Li D, Gavenonis S, Conant E, Kontos D. Comparison of breast percent density estimation from raw versus processed digital mammograms. *Proc SPIE.* 2011; 7963:79631X.
23. Stone J, Warren RM, Pinney E, et al. Determinants of percentage and area measures of mammographic density. *Am J Epidemiol.* 2009; 170:1571–1578. [PubMed: 19910376]
24. Kontos D, Ikejima LC, Bakic PR, Troxel AB, Conant EF, Maidment AD. Analysis of parenchymal texture with digital breast tomosynthesis: comparison with digital mammography and implications for cancer risk assessment. *Radiology.* 2011; 261:80–91. [PubMed: 21771961]
25. Bakic PR, Carton AK, Kontos D, et al. Breast percent density: estimation on digital mammograms and central tomosynthesis projections. *Radiology.* 2009; 252:40–49. [PubMed: 19420321]
26. Pearson K. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character.* 1896; 187:253–318.
27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986; 1:307–310. [PubMed: 2868172]
28. Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull.* 1968; 70:213. [PubMed: 19673146]
29. Martin KE, Helvie MA, Zhou C, et al. Mammographic density measured with quantitative computer-aided method: comparison with radiologists' estimates and BI-RADS categories. *Radiology.* 2006; 240:656–665. [PubMed: 16857974]
30. Heine JJ, Cao K, Rollison DE, et al. A quantitative description of the percentage of breast density measurement using full-field digital mammography. *Acad Radiol.* 2011; 18:556–564. [PubMed: 21474058]
31. Shepherd JA, Kerlikowske K, Ma L, et al. Volume of mammographic density and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2011; 20:1473–1482. [PubMed: 21610220]
32. Keller BM, Nathan DL, Wang Y, et al. Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. *Med Phys.* 2012; 39:4903–4917. [PubMed: 22894417]
33. Keller BM, Nathan DL, Conant EF, et al. Model-based estimation of breast percent density in raw and processed full-field digital mammography images from image-acquisition physics and patient-image characteristics. *SPIE Proc.* 2012; 8313
34. Byng JW, Yaffe MJ, Jong RA, et al. Analysis of mammographic density and breast cancer risk from digitized mammograms. *Radiographics.* 1998; 18:1587–1598. [PubMed: 9821201]
35. Lobbes MB, Cleutjens JP, Lima Passos V, et al. Density is in the eye of the beholder: visual versus semi-automated assessment of breast density on standard mammograms. *Insights Imaging.* 2012; 3:91–99. [PubMed: 22696002]

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.

Figure 2.

Inter-reader agreement of breast percent density (PD%) estimates for raw (*left*) and processed (*right*) digital mammography images via linear regression (*top*) and Bland-Altman difference (*bottom*) plots. For the scatter plots, the regression equation, regression-line (*solid*) and unity-line (*dashed*) are provided as reference. Bland-Altman plots are annotated with horizontal lines providing the mean difference (*solid*) and 95% limits of agreement (*dashed*).

Figure 3. Inter-reader agreement of absolute dense tissue area (cm^2) estimates for raw (*left*) and processed (*right*) digital mammography (DM) images via linear regression (*top*) and Bland-Altman difference (*bottom*) plots. For the scatter plots, the regression equation, regression-line (*solid*) and unity-line (*dashed*) are provided as reference. Bland-Altman plots are annotated with horizontal lines providing the mean difference (*solid*) and 95% limits of agreement (*dashed*).

Figure 4.

Vendor postprocessed digital mammography (DM) images segmented using the Cumulus software tool illustrating variable levels of between reader averages (μ) and differences (Δ) in breast percent density (PD%) and absolute dense tissue area assessment. Red regions denote areas of the digital mammogram which contain non-breast tissue regions (ie, background and the pectoralis muscle); green regions indicate dense breast tissue as identified and segmented by the two readers. **(a,b)** Sample mediolateral oblique view mammogram from a 48-year-old woman for which there was a relatively high level of interreader agreement (interreader μ PD%: 38.7%, Δ PD%: 2.1%; μ Dense-area: 27.3 cm²,

Δ Dense-area: 1.7 cm²). **(c,d)** Sample mediolateral oblique view mammogram from a 55-year-old woman for which there was a relatively low level of interreader agreement (interreader μ PD%: 43.9%, Δ PD%: 23.8%; μ Dense-area: 90.6 cm², Δ Dense-area: 51.0 cm²).

TABLE 1

Intrareader Agreement between Individual Women's Left and Right Breast Density Estimates as Measured by Pearson Correlation and Bland-Altman Statistics

	Breast Percent Density (%)			
	Reader 1		Reader 2	
	Raw	Processed	Raw	Processed
Pearson r	0.99	0.99	0.99	0.99
Mean difference	0.03	-0.22	0.02	-0.13
95% limits of agreement	(-4.6, +4.7)	(-5.6, +5.1)	(-1.0, +1.0)	(-3.7, +3.4)
	Dense Tissue Area (cm ²)			
	Reader 1		Reader 2	
	Raw	Processed	Raw	Processed
Pearson r	0.98	0.97	0.99	0.98
Mean difference	0.61	0.25	0.54	0.50
95% limits of agreement	(-8.7, +9.9)	(-10.1, +10.6)	(-5.1, +6.1)	(-6.2, +7.2)

Top, breast percent density; bottom, dense tissue area. All correlations are statistically significant ($P < .001$).

TABLE 2

Intrareader Agreement between Individual Readers' Raw and Processed Breast Percent Density and Absolute Dense Tissue Area Estimates as Measured by Pearson Correlation and Bland-Altman Statistics

	Reader 1		Reader 2	
	Breast Percent Density (%)	Dense Tissue Area (cm ²)	Breast Percent Density (%)	Dense Tissue Area (cm ²)
Pearson <i>r</i>	0.97	0.97	0.99	0.99
Mean difference	-1.24	-0.81	-0.05	0.96
95% limits of agreement	(-9.5, +7.0)	(-12.2, +10.6)	(-1.8, +1.7)	(-2.5, +4.4)

All correlations are statistically significant ($P < .001$).

TABLE 3

Interreader Agreement between Readers' Per-woman Density Estimates as Measured by Pearson Correlation and Bland-Altman Statistics

	Breast Percent Density (%)		Dense Tissue Area (cm ²)	
	Raw	Processed	Raw	Processed
Pearson <i>r</i>	0.90	0.89	0.87	0.85
Mean difference	-5.05	-3.94	-7.05	-5.28
95% limits of agreement	(-20.6, +10.5)	(-18.8, +11.0)	(-29.2, +15.1)	(-28.3, +17.7)

All correlations are statistically significant ($P < .001$).

Categorical BIRADS Density Assignment by Reader Made on Raw (Left) and Processed (Right) DM Images

TABLE 4

		Reader 1 (D.N.)				Reader 1 (D.N.)					
		I	II	III	IV	I	II	III	IV		
Reader 2 (S.G.)	I	38	9	0	0	Reader 2 (S.G.)	I	39	9	0	0
	II	1	19	9	0	II	2	19	7	0	
	III	0	0	3	2	III	0	0	3	2	
	IV	0	0	0	0	IV	0	0	0	0	