

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Consistency of visual assessments of mammographic breast density from vendor-specific “for presentation” images

Mohamed Abdoell
Kaitlyn Tsuruda
Christopher B. Lightfoot
Eva Barkova
Melanie McQuaid
Judy Caines
Sian E. Iles

SPIE.

Consistency of visual assessments of mammographic breast density from vendor-specific “for presentation” images

Mohamed Abdolell,^{a,b,*} Kaitlyn Tsuruda,^b Christopher B. Lightfoot,^{a,b} Eva Barkova,^c Melanie McQuaid,^d Judy Caines,^{a,b,e} and Sian E. Iles^{a,b}

^aDalhousie University, Faculty of Medicine, Department of Diagnostic Radiology, 1276 South Park Street, Room 3212, Dickson Building, Halifax, Nova Scotia B3H 2Y9, Canada

^bNova Scotia Health Authority, Department of Diagnostic Imaging, 1276 South Park Street, Room 3212, Dickson Building, Halifax, Nova Scotia B3H 2Y9, Canada

^cSouth Shore Regional Hospital, Department of Diagnostic Imaging, 90 Glen Allan Drive, Bridgewater, Nova Scotia B4V 3S6, Canada

^dQueen Elizabeth Hospital, Department of Diagnostic Imaging, 60 Riverside Drive, PO Box 6600, Charlottetown,

Prince Edward Island C1A 8T5, Canada

^eNova Scotia Breast Screening Program, 603L-7001 Mumford Road, Halifax, Nova Scotia B3L 2H8, Canada

Abstract. Discussions of percent breast density (PD) and breast cancer risk implicitly assume that visual assessments of PD are comparable between vendors despite differences in technology and display algorithms. This study examines the extent to which visual assessments of PD differ between mammograms acquired from two vendors. Pairs of “for presentation” digital mammography images were obtained from two mammography units for 146 women who had a screening mammogram on one vendor unit followed by a diagnostic mammogram on a different vendor unit. Four radiologists independently visually assessed PD from single left medio-lateral oblique view images from the two vendors. Analysis of variance, intra-class correlation coefficients (ICC), scatter plots, and Bland–Altman plots were used to evaluate PD assessments between vendors. The mean radiologist PD for each image was used as a consensus PD measure. Overall agreement of the PD assessments was excellent between the two vendors with an ICC of 0.95 (95% confidence interval: 0.93 to 0.97). Bland–Altman plots demonstrated narrow upper and lower limits of agreement between the vendors with only a small bias (2.3 percentage points). The results of this study support the assumption that visual assessment of PD is consistent across mammography vendors despite vendor-specific appearances of “for presentation” images. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMI.3.1.011004](https://doi.org/10.1117/1.JMI.3.1.011004)]

Keywords: digital mammography; mammographic density; agreement; reader variability.

Paper 15147SSRR received Jul. 15, 2015; accepted for publication Sep. 23, 2015; published online Oct. 30, 2015.

1 Introduction

Percent mammographic breast density (PD) determined from “for presentation” mammography images is a well-known risk factor for breast cancer.^{1–4} Discussion of PD in the medical literature, particularly that pertaining to multisite studies, implicitly assumes that visual assessments of PD are comparable across mammograms generated by different vendors. However, all vendors apply unique proprietary image processing algorithms to “for processing” mammography images in order to optimize the contrast of the displayed image for cancer detection.⁵ Although little is known about the details of the specific algorithms used, such processing may be pixel-based, cluster-based, or global and results in vendor-specific differences in the appearance of “for presentation” mammography images that may be important in distinguishing lesions from normal tissue.⁶

Differences in vendor-specific image acquisition technology can result in images from some vendors having a wider range of pixel intensities and darker appearance in fatty regions and brighter appearance in dense tissue regions compared to other vendors even when variations in positioning are minimized.⁷

The resulting differences in display may affect perception of the amount and distribution of breast tissue and, therefore, the visual assessment of PD. As such, vendor-specific differences in the appearance of the “for presentation” mammography images routinely reviewed in clinical care may contribute to the unreliability of visual assessments of PD, particularly in cases when data are pooled across multiple sites.

To date, research has focused on comparing potential differences in the presentation of PD from analog and digital mammograms, as well as from “for processing” and “for presentation” digital mammograms, to determine how the different formats impact the reporting of PD.^{4,8,9} To the best of our knowledge, only one study has investigated the differences that vendor-specific postprocessing may have on the visual assessment of PD from “for presentation” mammography images at the time of writing.¹⁰ This study found a minimal difference in visually assessed density between vendors using visually reported PD and BI-RADS density categories from two major mammography equipment vendors (GE and Hologic).

Some researchers have investigated the reliability of automated solutions to measure breast density from consecutive mammograms across different vendors.^{7,11,12} The commercially available algorithms used in these research studies analyze “for

*Address all correspondence to: Mohamed Abdolell, E-mail: mo@dal.ca

processing” images that are not for clinical use and are not routinely stored in picture archiving communication systems. While such algorithms aim to generate density assessments that agree with radiologists’ visual assessments of density, the manner in which radiologists visually process “for presentation” images is a complex visual perception task that is fundamentally different from the varied algorithmic approaches implemented in automated software solutions. As such, the reliability results from software algorithms across vendors do not extend to the task of visual assessment of breast density by radiologists.

As of September 2015, legislation in 24 U.S. states covering over 65% of the population requires women to be notified if they have dense breast tissue, and often suggests that supplemental imaging be discussed.¹³ In this context, vendor-specific differences in assessments of PD have the potential to affect a woman’s follow-up care, particularly where visual assessments of PD are used.

This study examines the extent to which visual assessments of PD differ between mammograms acquired from two different vendors (Siemens and Hologic) within a 12-month timeframe.

2 Methods

Institutional review board (IRB) approval was obtained for the study (RS/2015-158) in which breast density was assessed on and compared between mammography studies that were acquired using full field digital mammography (FFDM) units from two vendors. All personal identifiers were removed from the images, and the requirement for informed consent was waived by the IRB.

2.1 Image Selection

The data set was composed of 146 pairs of vendor-matched left mediolateral oblique (MLO) mammogram images. Mammograms were obtained using imaging systems from two major mammography equipment vendors (Siemens Healthcare GmbH, Erlangen, Germany, and Hologic Inc., Belford, Massachusetts) from 146 women who had a screening mammogram on a Siemens unit between November 5, 2013, and December 15, 2014, followed by a diagnostic mammogram on a Hologic unit within a 12-month period. All women were imaged first using a Siemens mammography unit.

The Siemens mammography unit models used in this study were the Mammomat Novation and Mammomat Inspiration. All images were acquired with automatic exposure control (AEC), where the peak kilovoltage (kVp) was selected based on patient thickness. Both models use a tungsten target and rhodium filter. The detector for the Novation uses 0.07 μm pixel spacing, and the detector for the Inspiration uses 0.085 μm pixel spacing. The Hologic model used in this study was the Selenia Dimensions. All images were acquired using an AEC mode with autofilter option that uses a prepulse from the machine to determine the filter type, kVp, and milliampere second for each image. While this model uses a tungsten target and a rhodium, silver, or aluminum filter material, the mammograms in this study were acquired using either the rhodium or silver filters. The detector for the Selenia Dimensions uses 0.07 μm pixel spacing.

All images included in this study were obtained within a single organized breast screening program following the practice guidelines and technical standards for breast imaging required by the Canadian Association of Radiologists (CAR) for breast cancer screening and diagnosis.¹⁴

2.2 Mammographic Density Assessment

Mammography images were reviewed on a clinical workstation with either 3- or 5-megapixel Barco monitors that are maintained according to CAR quality guidelines and manufacturer specifications. MediCal QAWeb (Barco, Kortrijk, Belgium) is used within the clinical facility to run automated reports that confirm calibration, and any failures are sent to quality control technologists automatically. Additional quality control testing is done using the American Association of Physicists in Medicine TG18 QC phantom to verify luminance response, linearity, and visual performance. All display monitors are DICOM Grayscale Standard Display Function calibrated and maintained at a luminance between 400 cd/m^2 (minimum) and 420 cd/m^2 (maximum). Lighting conditions for the density assessments were consistent with those used in accredited clinical conditions, and ambient light was held between 25 and 40 Lux. The radiologists were able to pan, zoom, and adjust the window level as desired.

Four radiologists each independently reviewed a set of single standard left MLO images from one vendor and visually assessed PD. Two vendor-specific worklists were created. The order of images within each worklist was fixed; however, the order of subjects differed between the two vendor worklists. The worklists could be read in any order, but radiologists were blinded to their previous assessments of PD when reading the second worklist. All four radiologists visually assessed PD for each image in the data set.

2.3 Statistical Analysis

Vendor- and rater-stratified descriptive statistics were calculated, and box-and-whisker plots were used to visualize the distribution PD assessments. A two-way, type III, mixed model analysis of variance (ANOVA) was performed to determine the effect of vendor and rater on PD assessments where vendor was a fixed effect, and rater and vendor by rater interaction was a random effect. Between-rater agreement of PD assessments within vendors were measured using the intra-class correlation coefficient (ICC) and considered alongside the ANOVA results to determine whether a consensus measure of PD could be used to evaluate the reliability of PD measurements between vendors. Additionally, the bias between raters’ PD assessments was calculated as the absolute value of the mean of the differences in PD assessments for each pair of radiologists and stratified by vendor.

Box-and-whisker plots and histograms were used to graphically display vendor-specific distributions of PD.

The reliability of visual PD assessments between vendors was evaluated using the Pearson’s correlation coefficient (PCC) to assess the strength of the linear relationship between PD assessments, the ICC to measure agreement between the PD assessments, and a scatter plot to graphically display the results. Although the interpretation of the ICC can vary depending on the context, the ICC is mathematically equivalent to the quadratically weighted kappa statistic, and as such the guideline proposed by Landis and Koch for qualitative interpretation of the kappa statistic was used to interpret the ICC results presented in this study.^{15,16} Using this interpretation scale, an ICC < 0 indicates poor agreement, a value between 0 and 0.2 indicates slight agreement, a value between 0.21 and 0.40 indicates fair agreement, a value between 0.41 and 0.60 indicates moderate agreement, a value between 0.61 and 0.80 indicates substantial

agreement, and a value between 0.81 and 1.00 indicates almost perfect agreement.

A Bland–Altman disagreement plot was used to evaluate the agreement between visual PD assessments made on consecutive mammograms from the two vendors as well as to quantify any bias observed between the visual PD assessments made on consecutive mammograms from the two vendors’ mammography units.¹⁷ The disagreement plot shows the difference between the PD values for each woman assessed on both of the vendor’s mammography units against the average of the PD values from both vendors. On the vertical axis, the mean difference provides an estimate of bias, and the mean difference ± 1.96 standard deviations of the difference provides upper and lower limits of agreement that indicate how far apart PD measurements from the two different vendors are most likely to be for paired mammograms. A small bias and narrow limits of agreement are preferred; the interpretation of these Bland–Altman plot statistics is essentially clinically driven and context dependent.

Statistical analyses were performed using R version 3.0.2 for Linux using the car, irr, and ggplot2 packages.^{18–21} ANOVA

analysis was performed using SAS version 9.3 for Windows using proc mixed.

3 Results

For the analysis, 146 vendor- and subject-matched left MLO image pairs were available. The women were aged 40 to 82 years (mean 54 years) at the time of the screening mammogram on the Siemens unit.

3.1 Vendor- and Rater-Stratified Percent Breast Density Assessments

Figure 1(a) shows a boxplot of vendor-stratified PD assessments (all raters). The overall range of PD assessments for both vendors was similar, as was the mean PD assessment (38% for Siemens, 35% for Hologic; vendor fixed main effect $p = 0.090$ from mixed effects ANOVA). Figure 1(b) shows a boxplot of rater-stratified PD assessments (both vendors). Some variability was observed in the distribution of PD assessments between raters; however, the mean PD assessments were similar across

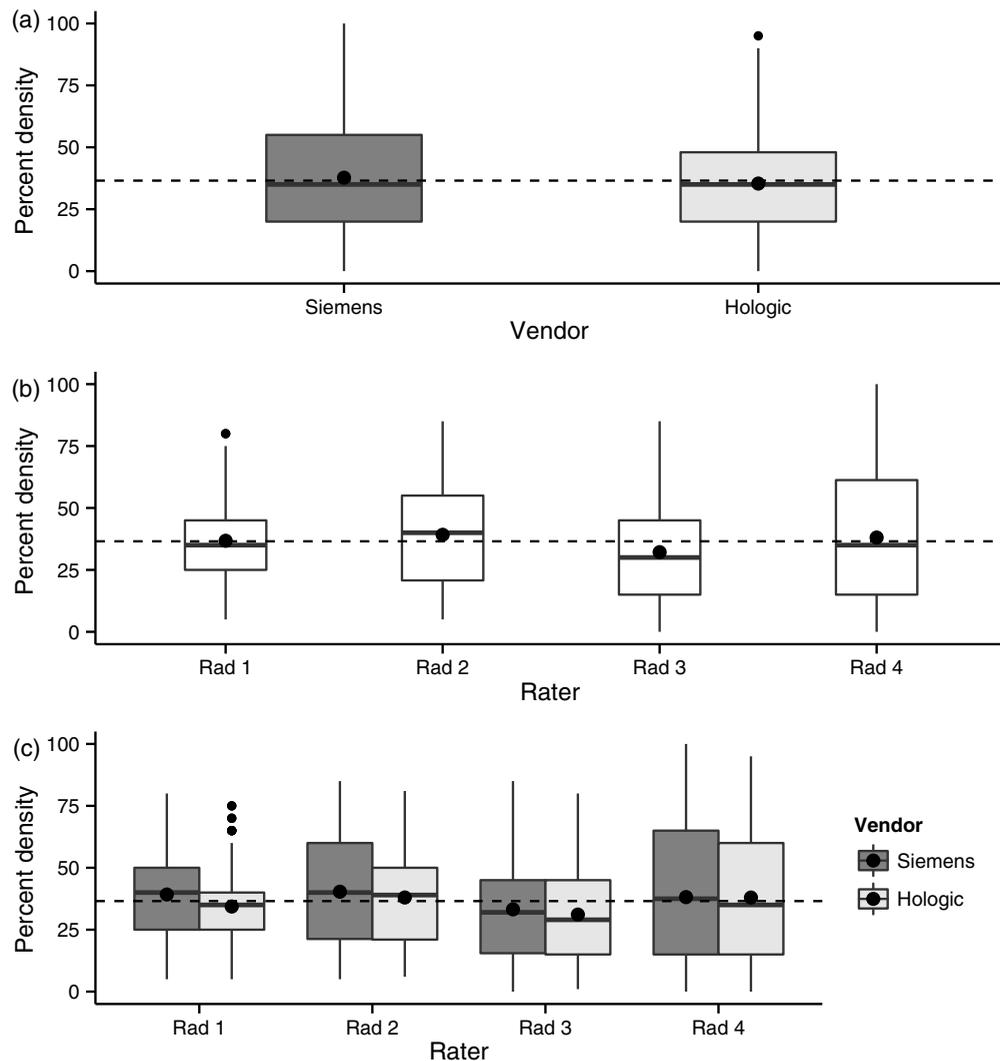


Fig. 1 Box-and-whisker plots show the distribution of percent breast density (PD) (a) by vendor (all raters), (b) by rater (both vendors), and (c) stratified by both rater and vendor. The black dots within the boxes indicate the stratified mean, while the dashed line indicates the grand mean.

raters (37, 39, 32, and 38% for raters 1 through 4, respectively; rater random main effect $p = 0.043$ from mixed effects ANOVA). Figure 1(c) shows a boxplot of vendor- and rater-stratified PD. While the PD assessments for Siemens images were marginally higher than those of Hologic images across all raters, the mixed effects ANOVA demonstrated that this effect did not differ across raters (rater by vendor interaction term random effect $p = 0.594$). Additionally, while some variability in the distribution of vendor-specific PD assessments was observed across raters, agreement among the four raters was excellent to almost perfect for Siemens images with an overall ICC of 0.91 [95% confidence interval (CI): 0.88 to 0.93] as well as for Hologic images with an overall ICC of 0.85 (95% CI: 0.82 to 0.89).

There was a small amount of bias observed between raters, ranging from 0.02 percentage points to 7.1 percentage points across both vendors (1.0 to 7.1 percentage points for Siemens images and 0.02 to 6.9 percentage points for Hologic images).

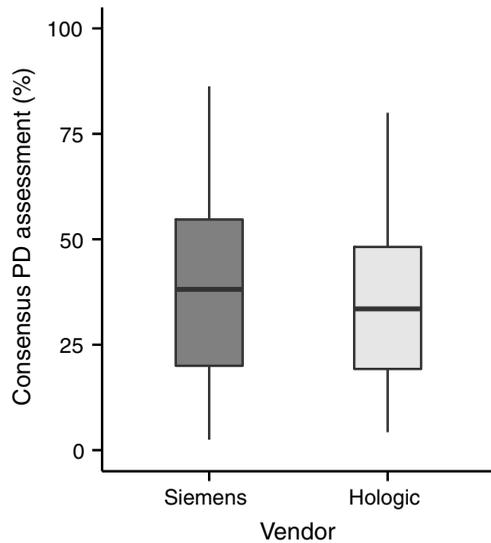


Fig. 2 Box-and-whisker plots showed the distribution of the consensus PD measure was similar for the two vendors.

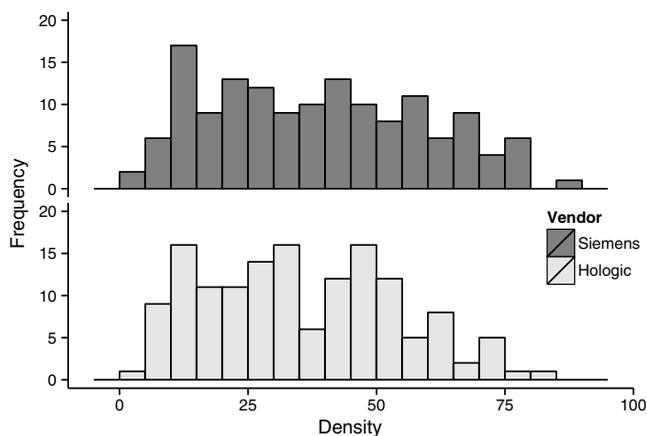


Fig. 3 Histograms show a similar distribution of consensus PD measures between vendors.

Because within-vendor agreement was excellent for both vendors, the mean PD assessment for each image was used as a consensus PD measure per image.

3.2 Reliability of Percent Breast Density Assessments Between Vendors

Using the mean density as a consensus PD measure between radiologists’ density assessments of each image, box-and-whisker plots showed a similar distribution of PD between vendors, although the median consensus PD was slightly higher for Siemens images (38.1 versus 33.5%, Fig. 2). A histogram also showed that the distribution of consensus PD between the two vendors was similar (Fig. 3).

There was a strong linear correlation between vendor PD assessments ($PCC = 0.96$), and overall agreement of the consensus PD assessments was almost perfect between the two vendors with an ICC of 0.95 (95% CI: 0.93 to 0.97). A scatter plot reinforced this finding (Fig. 4). A Bland–Altman plot demonstrated narrow upper and lower limits of agreement between

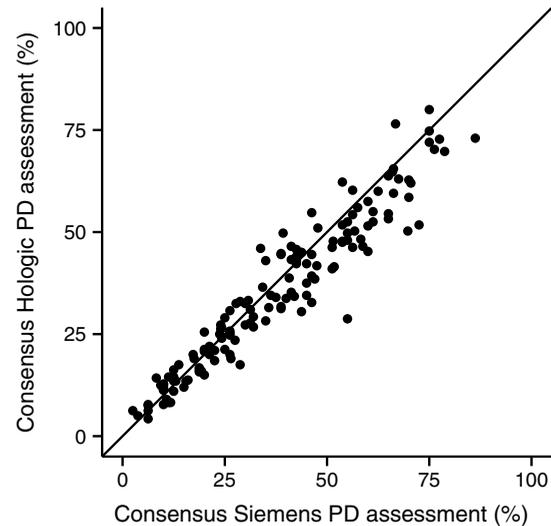


Fig. 4 A scatter plot shows that the consensus PD measure was similar between vendors.

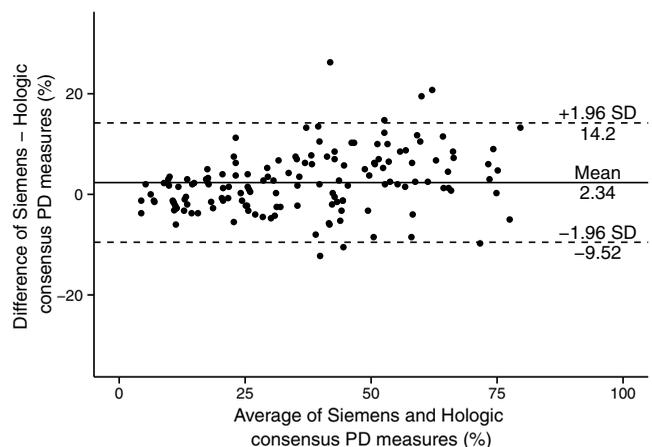


Fig. 5 A Bland–Altman plot shows little bias between consensus PD measures from paired images acquired by the two vendors’ mammography units.

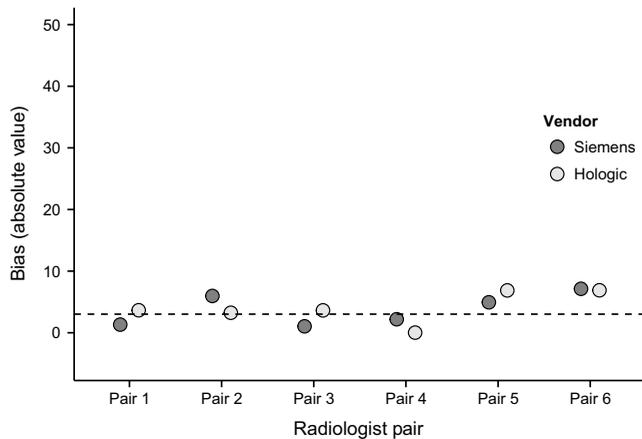


Fig. 6 The bias observed between vendors (dashed line) was small and within the range observed between pairs of radiologists reading the same mammograms. Bias was calculated as the absolute value of the mean of the differences in PD assessments either between vendors or between radiologists.

the vendors with a small bias (2.3 percentage points), indicating that consensus PD assessments from Siemens images were marginally higher than those from Hologic images (Fig. 5). The level of bias observed was not clinically meaningfully different from the observed bias between pairs of radiologists reading the same mammograms from the same mammography units (Fig. 6).

4 Discussion

This study investigated the magnitude of vendor-specific differences in visually assessed PD. On average, it was found that visual consensus PD assessments from Siemens images were 2.3 percentage points higher than Hologic consensus PD assessments taken from a mammogram of the same woman within a 12-month timeframe. Such a small between-vendor bias is unlikely to be clinically significant or alter the course of a woman’s follow-up care. Additionally, such a small difference is unlikely to be a source of bias in multivendor studies assessing PD, particularly as PD is often assessed in the more broad BI-RADS density categories, which typically span a 25%-wide category of PD.²² Furthermore, two percentage points fall within the 5% levels that are consistent with radiologists’ internal rating scales of PD.^{23,24} Additionally, the upper and lower limits of agreement, capturing how far apart 95% of the measurements on vendor-paired mammograms are, are very narrow (± 12 percentage points around the bias) and suggest that the two vendors’ mammography units may be used interchangeably in assessing breast density. In combination with the very strong correlation ($PCC = 0.96$) and agreement statistic ($ICC = 0.95$), the small bias and narrowness of the upper and lower limits of agreement support the argument that visually assessed PD by radiologists agrees across the two mammography device vendor units and, therefore, that radiologists’ visual assessments of PD may be reliable across different mammography device vendors.

The magnitude of the difference in consensus PD assessments between different vendors found in this study is similar to that reported by Vinnicombe et al.¹⁰ In their study, GE images were reported, on average, to have higher PD than Hologic “for presentation” images acquired within a one-year period reported both using PD measurements and BI-RADS density categories.

Based on our results and those of Vinnicombe et al., visual assessments of PD may differ minimally between vendors despite the fact that the visual appearance of PD is affected by vendor-specific postprocessing of “for processing” mammography images.

A major strength of this study is the availability of subject-matched, vendor-paired images acquired during a short time-frame within a single population-based, accredited screening program. A study design that requires women to be subjected to consecutive mammograms in a short period of time using two different mammography vendor units without clinical necessity would normally be infeasible and considered unethical. The opportunity to perform this research emerged from the natural experiment resulting from the introduction of a Hologic tomosynthesis unit into the hospital breast imaging department, such that women seen in screening mammography on Siemens mammography units were referred to diagnostic workup using tomosynthesis and standard mammography using the Hologic tomosynthesis unit, resulting in consecutive mammograms in a short time period.

In this study, Siemens images were acquired in a screening setting, and the paired Hologic images were acquired in a diagnostic setting. In the Nova Scotia Breast Screening Program the standard CC and MLO screening views are repeated in diagnostic workup in addition to spot views of the area of concern or tomosynthesis, which started on October 31, 2014. Both screening and diagnostic imaging occur within a single department under the direction of a single medical director, technical manager, and quality assurance officer. Furthermore, the same group of mammography-certified technologists is responsible for acquiring both screening and diagnostic images. These factors result in a consistent image quality across the screening and diagnostic settings and make it unlikely that the use of screening or diagnostic images would differentially affect the appearance of PD between vendors.

A source of potential variability in the visual assessment of PD is the monitor used to read density: Some radiologists used 5 MP mammography monitors and others used 3 MP general radiology monitors to assess PD. It was the opinion of the radiologists involved in the study that the level of detail displayed on a 3 MP monitor would be sufficient to reliably assess PD as it is a global feature of breast composition. This assumption by the radiologists was borne out by excellent to almost-perfect vendor-specific reliability in PD assessment across raters despite the use of two different monitor resolutions ($ICC = 0.91$ for Siemens images and 0.85 for Hologic images) and is unlikely to have biased the observed results.

It is possible that the density observed in our study from Hologic images was less than that observed from Siemens images due to naturally occurring changes in density as a woman ages. However, the mean and median time between images was 7.1 and 4 weeks, respectively, and it is unlikely that the PD of the women in this study would have perceptibly and significantly changed during this short amount of time.^{7,25,26} Furthermore, it is possible that changes in positioning technique could affect the observed density between the Siemens and Hologic images for a given woman; however, such differences are unlikely to be systematic based on screening or diagnostic imaging status.

A limitation of this study is that it considered only two major digital mammography vendors (Siemens and Hologic). While it would be of interest to evaluate subject-matched images

acquired from all major vendors within a short period of time, the feasibility and ethical considerations of developing such a study make this impracticable. Nevertheless, the results of this study suggest that visually assessed breast density is similar between these two vendors, and the results of the study by Vinnicombe et al. additionally suggest that visually assessed breast density is similar between GE and Hologic FFDM images. The results of both studies, when considered together, appear to suggest that radiologists’ visual assessments of PD may be generalizable across three of the major digital mammography unit vendors and potentially generalizable across all digital mammography unit vendors. Furthermore, these combined results suggest that radiologists may self-adjust or self-calibrate when they visually assess PD on digital mammograms from different vendors: despite the distinctly different appearance of the paired “for presentation” images, radiologists are able to reliably discern the dense tissue from the fatty tissue in the images. Additional research is needed to investigate the underlying visual perception processes that enable this to happen.

5 Conclusion

The results of this study suggest that while vendor-specific post-processing of “for processing” digital mammograms affects the appearance of dense breast tissue in “for presentation” images, the magnitude of the difference between visually assessed PD between vendors is not clinically significant.

Acknowledgments

The authors would like to thank Nina Reddick and Melissa Butler for helping with image and density data acquisition. The authors would also like to thank Stephanie Schofield for providing information pertaining to the FFDM units and viewing workstations.

References

1. N. F. Boyd et al., “Mammographic density and the risk and detection of breast cancer,” *N. Engl. J. Med.* **356**(3), 227–236 (2007).
2. J. Brisson, C. Diorio, and B. Mâsse, “Wolfe’s parenchymal pattern and percentage of the breast with mammographic densities redundant or complementary classifications?,” *Cancer Epidemiol. Biomarkers Prev.* **12**(8), 728–732 (2003).
3. V. A. McCormack and I. dos Santos Silva, “Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis,” *Cancer Epidemiol. Biomarkers Prev.* **15**(6), 1159–1169 (2006).
4. C. M. Vachon et al., “Comparison of percent density from raw and processed full-field digital mammography data,” *Breast Cancer Res.* **15**(1), R1 (2013).
5. P. Sprawls, *Physical Principles of Medical Imaging*, 2nd ed., Medical Physics Pub Corp, Madison, Wisconsin (1995).
6. E. B. Cole et al., “The effects of gray scale image processing on digital mammography interpretation performance,” *Acad. Radiol.* **12**(5), 585–595 (2005).
7. C. N. Damases, P. C. Brennan, and M. F. McEntee, “Mammographic density measurements are not affected by mammography system,” *J. Med. Imaging* **2**(1), 015501 (2015).
8. B. M. Keller et al., “Reader variability in breast density estimation from full-field digital mammograms: the effect of image postprocessing on relative and absolute measures,” *Acad. Radiol.* **20**, 560–568 (2013).
9. J. A. Harvey et al., “Reported mammographic density: film-screen versus digital acquisition,” *Radiology* **266**(3), 752–758 (2013).
10. S. J. Vinnicombe et al., “Visual & automated volumetric assessment of mammographic density: do measurements depend on the digital mammography unit,” in *European Congress of Radiology*, Austria, Vienna (2014).
11. X. Lin, N. Sauber, and R. Highnam, “Assessing breast density changes over time,” 2013, http://posterng.netkey.at/esr/viewing/index.php?module=viewing_poster&doi=10.1594/ecr2013/C-1770 (30 September 2015).
12. F. Engelken et al., “Volumetric breast composition analysis: reproducibility of breast percent density and fibroglandular tissue volume measurements in serial mammograms,” *Acta Radiol.* **55**(1), 32–38 (2014).
13. Dense Breast Info, “Legislation and regulations—what is required?,” <http://densebreast-info.org/legislation.aspx> (1 September 2015).
14. Canadian Association of Radiologists, “CAR practice guidelines and technical standards for breast imaging and intervention,” http://www.car.ca/uploads/standards%20guidelines/20131024_en_breast_imaging_practice_guidelines.pdf (13 June 2015).
15. J. L. Fleiss and J. Cohen, “The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability,” *Educ. Psychol. Meas.* **33**(3), 613–619 (1973).
16. J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *Biometrics* **33**(1), 159–174 (1977).
17. J. M. Bland and D. G. Altman, “Statistical methods for assessing agreement between two methods of clinical measurement,” *Int. J. Nurs. Stud.* **47**(8), 931–936 (2010).
18. R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2013).
19. J. Fox and S. Weisberg, *An R Companion to Applied Regression*, 2nd ed., Sage, Thousand Oaks, California (2011).
20. M. Gamer et al., “irr: various coefficients of interrater reliability and agreement,” R package version 0.84, 2012, <https://cran.r-project.org/web/packages/irr/index.html> (30 September 2015).
21. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, 1st ed., Springer-Verlag, New York (2009).
22. E. A. Sickles et al., “ACR BI-RADS® mammography,” in *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*, ACR BI-RADS Committee, Ed., pp. 179–180, American College of Radiology, Reston, Virginia (2013).
23. L. Hadjiiski et al., “Quasi-continuous and discrete confidence rating scales for observer performance studies,” *Acad. Radiol.* **14**(1), 38–48 (2007).
24. A. Sukha et al., “Visual assessment of density in digital mammograms,” *Lec. Notes Comput. Sci.* **6136**, 414–420 (2010).
25. N. Boyd et al., “A longitudinal study of the effects of menopause on mammographic features,” *Cancer Epidemiol. Biomarkers Prev.* **11**(10 Pt 1), 1048–1053 (2002).
26. K. Kerlikowske et al., “Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk,” *J. Natl. Cancer Inst.* **99**(5), 386–395 (2007).

Mohamed Abdolell is an associate professor at Dalhousie University, Diagnostic Radiology Department. He received his BSc degree in applied mathematics and statistics and his MSc degree in biostatistics from the University of Toronto in 1991 and 1995, respectively. He is an accredited professional statistician (P.Stat.) with the Statistical Society of Canada. His current research interests include breast screening, mammographic density, breast cancer risk, and medical informatics.

Biographies for the other authors are not available.