

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Mammographic density measurements are not affected by mammography system

Christine N. Damases
Patrick C. Brennan
Mark F. McEntee

Mammographic density measurements are not affected by mammography system

Christine N. Damases,^{a,b,*} Patrick C. Brennan,^a and Mark F. McEntee^a

^aUniversity of Sydney, Faculty of Health Sciences, Discipline of Medical Radiation Sciences and Brain and Mind Research Institute, M205, Cumberland Campus, East Street, Lidcombe, New South Wales 2141, Sydney

^bUniversity of Namibia, Faculty of Health Sciences, Department of Allied Sciences, M-Block, Room M-105, Mandume Ndemufayo Avenue, Private bag 13310, Windhoek 9000, Namibia

Abstract. Mammographic density (MD) is a significant risk factor for breast cancer and has been shown to reduce the sensitivity of mammography screening. Knowledge of a woman's density can be used to predict her risk of developing breast cancer and personalize her imaging pathway. However, measurement of breast density has proven to be troublesome with wide variations in density recorded using radiologists' visual Breast Imaging Reporting and Data System (BIRADS). Several automated methods for assessing breast density have been proposed, each with their own source of measurement error. The use of differing mammographic imaging systems further complicates MD measurement, especially for the same women imaged over time. The purpose of this study was to investigate whether having a mammogram on differing manufacturer's equipment affects a woman's MD measurement. Raw mammographic images were acquired on two mammography imaging systems (General Electric and Hologic) one year apart and processed using VolparaDensity™ to obtain the Volpara Density Grade (VDG) and average volumetric breast density percentage (AvBD%). Visual BIRADS scores were also obtained from 20 expert readers. BIRADS scores for both systems showed strong positive correlation ($\rho = 0.904$; $p < 0.001$), while the VDG ($\rho = 0.978$; $p < 0.001$) and AvBD% ($\rho = 0.973$; $p < 0.001$) showed stronger positive correlations. Substantial agreement was shown between the systems for BIRADS ($\kappa = 0.692$; $p < 0.001$), however, the systems demonstrated an almost perfect agreement for VDG ($\kappa = 0.933$; $p < 0.001$).

© 2015 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.2.1.015501](https://doi.org/10.1117/1.JMI.2.1.015501)]

Keywords: mammographic density; General Electric; Hologic; Volpara; Breast Imaging Reporting and Data System.

Paper 14140RR received Oct. 22, 2014; accepted for publication Feb. 5, 2015; published online Mar. 4, 2015.

1 Introduction

Women with dense breasts have a risk of breast cancer four to six times higher than women with fatty breasts.¹ Fat and fibroglandular tissues have different radiographic appearances on a mammogram.¹ Fat appears as regions of darkness, whereas fibroglandular tissue appears as regions of brightness; this bright tissue is referred to as mammographic density (MD).² Increased MD is linked to higher breast cancer risk and it can affect mammography screening sensitivity through masking effects making diagnosis of cancer more difficult.³ Measuring MD is important for breast cancer risk prediction and for recommending an appropriate imaging pathway. Women with increased MD need further imaging, such as ultrasound or magnetic resonance imaging (MRI), to better visualize structures within the dense tissue.^{4,5} However, a number of MD features are not standardized including method of measurement, use of American College of Radiology Breast Imaging Reporting and Data System (ACR BIRADS) density definitions, and legislative requirements around informing the woman.

MD as apparent to an observer is estimated by evaluating the bright and dark areas of the image, which represent the fibroglandular and fatty tissues, respectively.⁶ The appearance of MD is due to variations in the breast tissue composition and the x-ray attenuation characteristics of these tissues.² The attenuation characteristics may also be affected by the imaging

technology and technique used. For example, increased noise can increase brightness and may affect the perception of MD.^{7,8} Reduction in x-ray penetration will increase contrast and thus areas of brightness resulting in increased apparent MD.

The brightness of the image is also affected by acquisition parameters such as the exposure factors and the detector characteristics.^{9,10} Each manufacturer uses different anode-filter combinations; these impact upon the penetration of dense tissue, dose and image quality.¹¹ Dance et al. found a 20% improvement in image contrast at low dose using molybdenum/rhodium (Mo/Rh) or rhodium/rhodium (Rh/Rh) for dense breasts, whereas tungsten/rhodium or rhodium/aluminum had more than a 50% dose reduction while maintaining contrast.¹² Improved image contrast results in more differentiation of dark and bright regions on the mammogram.

Different manufacturers' digital detectors differ in their signal-to-noise ratio (SNR) and detective quantum efficiency (DQE) properties.^{8,10} Direct conversion amorphous selenium (a-Se) mammography detectors have higher DQE than indirect conversion amorphous silicon cesium iodide (a-S:CsI) detectors.^{8,10} DQE measures SNR transfer through the system as a function of spatial frequency. At an equal dose, decreasing DQE translates to less signal and more noise. As noise is a high frequency high brightness image attribute, it affects the mean brightness of the image, particularly in difficult to penetrate areas such as fibroglandular tissue.^{8,10} This difference

*Address all correspondence to: Christine N. Damases, E-mail: cdam2504@uni.sydney.edu.au

in brightness characteristics of the images may affect the perceived or measured MD.⁷⁻¹⁰

Two of the largest mammography equipment manufacturers have chosen different anode filter combinations and detector materials. The General Electric (GE) manufactured “Senographe DS and Essential” full-field digital mammography (FFDM) systems (GE Healthcare, Fairfield, Connecticut) are equipped with anode filter combinations of molybdenum (Mo) and rhodium (Rh).^{7,12,13} These systems use a-Si:CsI flat panel arrays.⁸⁻¹⁰ The DQE ranges from 35% to 45% and is measured at 18% DQE for (5 mm⁻¹).^{7,8,14-16} Both the GE Senographe Essential and Senographe DS have the same detector element size (100 μm) but different detector sizes: Senographe Essential, 24 × 30.7 cm² and Senographe DS (23 × 19.2 cm²).^{8,9,13} The Hologic manufactured “Lorad Selenia” is an FFDM system using a-Se flat panel detectors (Hologic, Inc., Belford, Massachusetts).^{7-9,16} It is equipped with a tungsten (W) anode and a choice of rhodium (Rh) or silver (Ag) filtration. Previous literature reports that the combination of W/Rh is the best option for high performance in terms of contrast and brightness at a lower dose.¹¹ The range of DQE for this system is from 27% to 47% and is measured at 20% DQE for (5 mm⁻¹).¹⁶ The Lorad Selenia detector size is 24 × 29 cm².

Last, but maybe most importantly, different manufacturers apply different processing algorithms to change the raw data into an image suitable for cancer detection. These algorithms can be highly nonlinear and are highly proprietary. The effect of this on MD was assessed recently by Brand et al.¹⁷ They conducted a comparative study using Volpara on five different mammographic imaging systems assessing breast cancer risk with automated measurement. The authors reported that with the exclusion of the small GE detector, minor differences were noted between GE, Phillips, and Sectra systems, with the largest relative difference in percent dense volume being 3%.¹⁷ They conclude that “automated measurement of volumetric MD can be used as part of screening programs to provide risk and masking information that could be used to alter women’s clinical management.” The same certainty is needed for the most commonly used MD assessment system, BIRADS. Therefore, the purpose of this study is to investigate whether MD measurement using BIRADS and Volpara differs for images produced on GE and Hologic systems.

2 Methods

Institutional Review Board approval was obtained for the study (IRB 2013/448) in which two mammographic imaging systems were compared for their impact on MD measurement.

2.1 Selection of Images

The data set is composed of 40 cases, each containing a combined image of the left craniocaudal (LCC) and left mediolateral oblique (LMLO). These images were obtained from 20 women aged between 42 and 89 years. The images were acquired on three imaging systems (GE Senographe Essential/DS and Hologic Lorad) 1 year apart. From the 20 women, 11 were imaged on GE first and 9 on Hologic first. As the images were obtained at the same clinic, under the same protocol, we can have some confidence that there will be some similarities in the production of the images. All images were produced with the same imaging quality criteria such as those presented in the

Mammographic Quality Standards (MQSA) in breast cancer screening and diagnosis.¹⁸

The images for the study were obtained from Volpara. These women had given consent for their images to be used for research and all the images included in the study are of healthy women. Volpara selected the images to represent the range of VDGs and AvBD% found across the population. The images were selected to enable a comparison of the VDGs for women whose images were produced 1 year apart and also to have a comparable number of cases for each of the four VDG categories.

2.2 Image Display and MD Quantification Using BIRADS

Images were displayed on a single five-megapixel diagnostic quality monitor (Eizo, Japan) using ViewDEX software version 2.0.¹⁹ Twenty American Board of Radiology (ABR) examiners were presented with 40 cases. Radiologists had the ability to adjust the window width and level as well as to pan and zoom the image. A random order of case presentation was generated and the same random order was used for all observers. The images from each discrete mammographic examination for each woman were kept together and presented in the same sequence. For each case, there were three images, first the LCC followed by the LMLO and finally the combined LCC and LMLO presented together. A score on the ACR BIRADS density scale of 1 to 4 was recorded with 1 describing an entirely fatty breast and 4 representing an extremely dense breast.²⁰⁻²² To replicate clinical practice, the radiologist made one overall judgment from the combined image of LCC and LMLO for each of the 40 cases. A total of 800 MD judgments were made (400 GE and 400 Hologic). During evaluation of the images, the ambient lighting was kept constantly between 25 and 35 lux as confirmed by a calibrated photometer (model 07-621, Nuclear Associates).²³

2.3 MD Quantification Using Volpara Automated Software

The combined image of the left breast was categorized using Volpara imaging software version 1.4.3 (Matakina, Wellington, New Zealand) into Volpara Density Grade (VDG) categories 1 to 4, with these density grades designed to simulate the ACR BIRADS density scales. The software generates automatic measurement of volumetric MD values reported as the average breast density percentage (AvBD%). Using preset thresholds, the percentages are classified as VDGs (VDG 1 = 0% to 4.5%, VDG 2 = 4.5% to 7.5%, VDG 3 = 7.5% to 15.5% and over 15.5% is VDG 4).²⁴ For 20 women with two density measurements 1 year apart, 40 VDG and AvBD% data points were created.

2.4 Statistics

Statistical analyses were performed by using SPSS 21.0 (SPSS, Chicago, Illinois). A Mann-Whitney *U*-test was used to compare the medians of GE versus Hologic for BIRADS, VDG, and mean values for AvBD%. Spearman’s rank coefficient of correlation (ρ) was used to examine relationships between the systems for BIRADS, VDG, and AvBD%. The intrareader agreement on BIRADS categorization between the 20 readers was compared and expressed as Cohen’s Kappa (κ) with the use of a 20 × 20 matrix. Results were considered to be statistically significant at $p < 0.05$.

3 Results

3.1 Comparison of Median Values Between GE and Hologic System for BIRADS, VDG, and AvBD%

Although both GE and Hologic had a median BIRADS value of 2.0, the absolute difference in medians was 0.225 with GE resulting in a significantly higher BIRADS ($U = -2.03$; $p < 0.043$). The absolute difference is calculated by averaging the difference in median BIRADS score for GE and Hologic for each individual reader. The absolute difference in medians for VDG between the systems was 0.05 and was not significant ($U = -0.15$; $p < 0.877$). Mann-Whitney tests showed no significant difference between the median AvBD%, with GE having a median value of 6.51 and Hologic 6.79 ($U = -0.81$; $p < 0.935$) see Table 1.

3.2 Correlation Between GE and Hologic for BIRADS, VDG, and AvBD%

BIRADS for Hologic versus GE showed strong positive correlation ($\rho = 0.904$; $p < 0.001$) (Fig. 1), however, the VDG and

Table 1 Comparison of median values between GE and Hologic for BIRADS, VDG and AvBD%.

	N	Minimum	Maximum	Median
BIRADS GE	400	1.00	4.00	2.00
BIRADS HOL	400	1.00	4.00	2.00
VDG GE	20	1.00	4.00	2.00
VDG HOL	20	1.00	4.00	2.00
AvBD% GE	20	2.54	24.64	6.51
AvBD% HOL	20	3.47	22.92	6.79

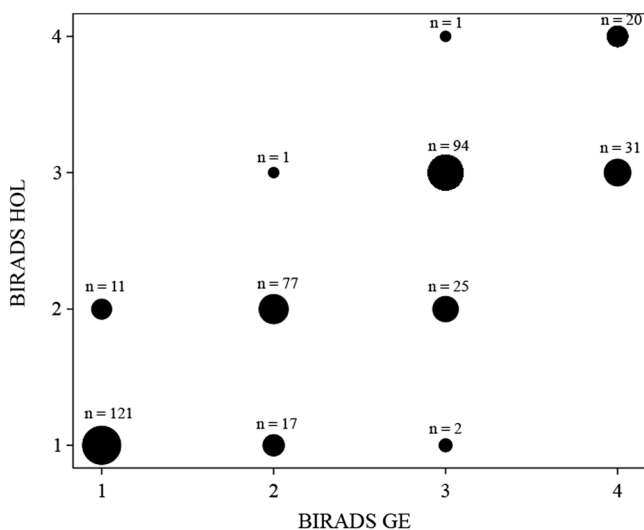


Fig. 1 A scatter plot demonstrating BIRADS correlation for GE versus Hologic imaging systems. The systems showed strong positive correlation ($\rho = 0.904$; $p < 0.001$). The size of the dot indicates the number of data points it contains.

AvBD% between the two systems had even stronger positive correlations ($\rho = 0.978$; $p < 0.001$) and ($\rho = 0.973$; $p < 0.001$), respectively (Figs. 2 and 3).

3.3 Agreement Between GE and Hologic Systems on BIRADS Categorization and VDG

For BIRADS, there was substantial agreement between the systems with a kappa of 0.692 $p < 0.001$. The inter-reader agreement for radiologists using the BIRADS four point scale had an average Kappa of 0.564 (95% CI = 0.519 to 0.610) and ranged between 0.328 and 0.669, while the inter-reader agreement for radiologists using BIRADS on the binary scale had an average kappa of 0.855; (95% CI = 0.824 to 0.866) and ranged between 0.656 and 0.901. There was an almost perfect agreement

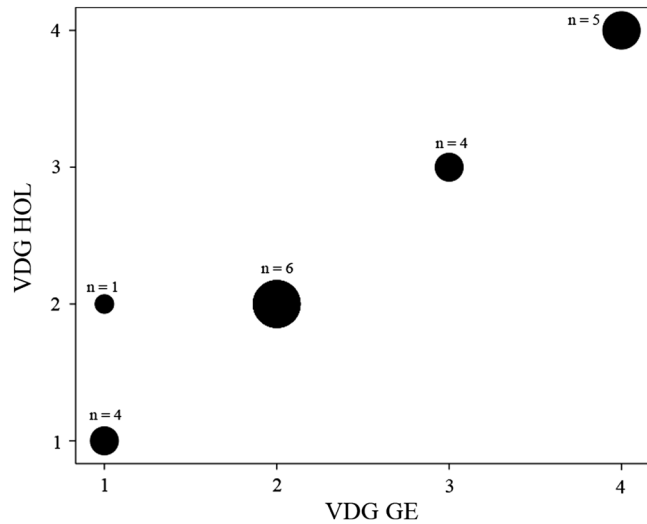


Fig. 2 A scatter plot demonstrating VDG correlation for GE versus hologic imaging systems. The systems showed stronger positive correlation ($\rho = 0.978$; $p < 0.001$) for VDG than BIRADS. The size of the dot indicates the number of data points it contains.

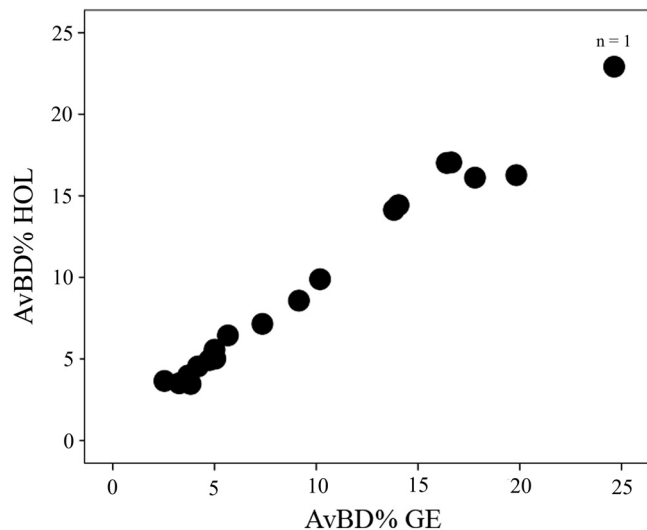


Fig. 3 A scatter plot demonstrating AvBD% correlation for GE versus Hologic imaging systems. The systems showed stronger positive correlation ($\rho = 0.973$; $p < 0.001$) for AvBD% than BIRADS. The size of the dot indicates the number of data points it contains.

between GE and Hologic imaging systems for VDG ($\kappa = 0.933$; $p < 0.001$).

4 Discussion

MD measurement is increasingly important in clinical management of patients. However, differing mammographic imaging systems are used for acquiring mammograms. Although having the same objective, the GE Senographe Essential/DS and Hologic Lorad Selenia mammographic imaging systems use different detector and filter materials and different postprocessing algorithms. Postprocessing does not affect VDG and AvBD% as this is applied to the raw image, however, BIRADS is affected by postprocessing. These differences could result in different MD measurements. Crucially, differences in MD measurement could affect a woman's imaging pathway. The current study investigated the impact of two mammographic imaging systems on MD measurement using BIRADS and Volpara.

Current result indicates that the BIRADS density assessment on the GE system is higher than on the Hologic system. Our finding is similar to that of Vinnicombe et al. who suggested that human visual assessments of density were more likely to be higher on GE than the Hologic system.²⁵ When the systems were compared for the absolute difference in medians for VDG, there was no significant difference between the systems, a result which demonstrated that with automated measurement, the impact of the mammographic imaging system on BD is negligible. Our finding is contrary to previous studies which reported that the Hologic mammographic imaging system is less accurate for volumetric measurement than the GE system.^{26,27} Tyson et al.²⁷ reported that the Hologic system uses a tilting compression plate which results in variation of breast thickness measurement.

The lower correlation between BIRADS MD measurements on different mammographic imaging systems is likely due to the larger intraobserver variations common with qualitative reporting measures. The systems demonstrated a substantial agreement for BIRADS, whereas an almost perfect agreement was shown between the systems for VDG. The comparison of BIRADS measures over 20 radiologists demonstrates that there is a large variation in BIRADS reading for the same women. The intrareader variation with BIRADS is possibly due to a reduced BIRADS correlation between the imaging systems, different technology, and varying postprocessing algorithms.

For the current study, intrareader variation with BIRADS was assessed using Cohen's Kappa. This assumes that errors associated with the readers' ratings are independent.²⁸⁻³⁰ In the current study, as with all measures of intrareader reliability, initial ratings may influence subsequent ones, and this threatens the assumption of independence. One solution to this is to separate readings by a number of days. However, this would not be practical where 40 ratings are involved. Although Kappa can be adjusted by methods such as Prevalence and Bias Adjusted Kappa (PABAK),³¹ Hoehler is of the opinion that the effects of bias and prevalence on the scale of Kappa are useful and, therefore, should not be adjusted.³²

The separation of the mammographic cases for the same woman by 1 year may not have impacted upon the density of the latter case. The average difference between our two screening studies was 12.83 months, with a range of 12.03 to 16.01. A study by Kerlikowske et al.³³ demonstrate that "among women without breast cancer, those aged 70 years

and older were four times more likely to be assigned to the lowest breast density category (category 1) and six times less likely to be assigned to the highest breast density category (category 4) than were women aged 30-39 years." Thus, breast density decreased with age. Their work shows that this change in density takes many years to be measurable and significant. Likewise, Vachon et al.³⁴ reported that breast density decreased by at least one BIRADS category over ~6 years.³⁴ Thus, while breast density decreases with age, the effects are likely to be imperceptible over 1 year. However, volumetric studies are likely to be more sensitive than area based studies of MD. The only study investigating the impact of time on volumetric MD was by Holland et al.³⁵ The mean screening interval between their two studies was 22.65 months. The authors reported that in 89.7% of the cases, MD remained in the same category. In 3.2% of the pairs, an increase in percentage density was actually reported, resulting in a change from the nondense to dense category. They report that this effect may have been due to differences in the breast thickness measurement as reported in the DICOM headers. For the current study, the impact of the passing of 1 year on the AvBD% was assessed and a perfect correlation for AvBD% was shown between year 1 and 2 ($r = 0.997$; $p \leq 0.001$).³⁶

The study showed that the impact of differing imaging systems on volumetric MD measurements was negligible, thus demonstrating that volumetric measures have higher reproducibility than BIRADS. VDG and AvBD% results are very promising as they indicate not only that mammographic systems do not seem to have an influence on MD measurement but also that other circumstances, such as slightly different positioning and most likely different technicians do not seem to matter. For the current study, cases were selected to show the temporal stability with Volpara, and other studies have shown Volpara to be robust to differing manufacturer mammography systems for women screened 1 year apart.²⁵

Even though our results are clear, there is no known truth for MD measurement, although people have been using MRI for this and other groups are now looking at CT measurements.

The current study had some limitations. Our study only investigated the BIRADS reading on a left breast; it did so only with American radiologists. Different intrareader variability may exist in other countries, and this needs to be assessed. The radiologists may not have been familiar with the presentation state of the images and may be used to a different look, which may have affected their perception of density. It is also not known whether inclusion of the right breast would have increased or reduced the intrareader variability for BIRADS MD. The results of this paper pertain to FFDM, however, further work may be needed to investigate whether differences exist for other systems such as CR and DBT.

This work demonstrates that automated measures of MD have higher reproducibility than BIRADS. However, implementation of automated systems is not universal and the final decision on MD will continue to be the responsibility of the radiologists. Therefore, further work will investigate the effect of factors that influence radiologists' MD decision such as prevalence of higher BMI in the patient population, number of years of experience and the number of cases read per year.

5 Conclusion

Stronger positive correlations in VDG and AvBD% scores compared with BIRADS values were demonstrated between

imaging systems. The effect of using two different mammographic imaging systems on MD measurement was negligible.

Acknowledgments

We would like to acknowledge Dr. Ralph Highnam and Dr. Ariane Chan of Volpara Solutions Ltd. for their assistance with data analysis and statistical support for this study. We would also like to acknowledge Carestream for funding travel for data collection and the ABR for facilitating the reading of mammographic images.

References

- N. F. Boyd et al., "Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study," *J. Natl Cancer Inst.* **87**(9), 670–675 (1995).
- N. F. Boyd et al., "Mammographic density and breast cancer risk: current understanding and future prospects," *Breast Cancer Res.* **13**(6), 223 (2011).
- N. F. Boyd et al., "Mammographic density and the risk and detection of breast cancer," *N. Engl. J. Med.* **356**(3), 227–236 (2007).
- W. A. Berg et al., "Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer," *J. Am. Med. Assoc.* **299**(18), 2151–2163 (2008).
- W. A. Berg et al., "Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk," *J. Am. Med. Assoc.* **307**(13), 1394–1404 (2012).
- S. W. Duffy et al., "Interaction of dense breast patterns with other breast cancer risk factors in a case-control study," *Br. J. Cancer* **91**(2), 233–236 (2004).
- J. B. McCullagh, P. Baldelli, and N. Phelan, "Clinical dose performance of full field digital mammography in a breast screening programme," *Br. J. Radiol.* **84**(1007), 1027–1033 (2011).
- S. Rivetti et al., "Comparison of different commercial FFDM units by means of physical characterization and contrast-detail analysis," *Med. Phys.* **33**(11), 4198–4209 (2006).
- M. Mahesh, "AAPM/RSNA physics tutorial for residents—digital mammography: an overview," in *89th Scientific Assembly and Annual Meeting of the Radiological-Society-of-North-America (RSNA)*, **24**, pp. 1747–1760, Chicago, Illinois (2004).
- J. Shaw et al., "Enhanced a-Si/CsI-based flat panel X-ray detector for mammography," *Proc. SPIE* **5368**, 370–378 (2004).
- P. Baldelli, N. Phelan, and G. Egan, "Investigation of the effect of anode/filter materials on the dose and image quality of a digital mammography system based on an amorphous selenium flat panel detector," *Br. J. Radiol.* **83**(988), 290–295 (2010).
- D. R. Dance et al., "Influence of anode/filter material and tube potential on contrast, signal-to-noise ratio and average absorbed dose in mammography: a Monte Carlo study," *Br. J. Radiol.* **73**(874), 1056–1067 (2000).
- C. Ghetti et al., "Physical characteristics of GE senographe essential and DS digital mammography detectors," *Med. Phys.* **35**(2), 456–463 (2008).
- D. Albagli et al., "Performance of advanced a-Si/CsI-based flat panel x-ray detectors for mammography," *Proc. SPIE* **5030**, 553–563 (2003).
- S. Vedantham et al., "Full breast digital mammography with an amorphous silicon-based flat panel detector: physical characteristics of a clinical prototype," *Med. Phys.* **27**(3), 558–567 (2000).
- P. Monnin et al., "A comparison of the performance of digital mammography systems," *Med. Phys.* **34**(3), 906–914 (2007).
- J. S. Brand et al., "Automated measurement of volumetric mammographic density: a tool for widespread breast cancer risk assessment," *Cancer Epidemiol. Biomarkers Prev.* **23**(9), 1764–1772 (2014).
- USDoHaH Services, "Mammography Quality Standards Act and Program," www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfm/QSA/mail/listman.cfm (3 December 2014).
- S. Börjesson et al., "A software tool for increased efficiency in observer performance studies in radiology," *Radiat. Prot. Dosimetry* **114**(1–3), 45–52 (2005).
- C. J. D'Orsi, L. W. Bassett, and W. A. Berg, *BI-RADS Mammography in, 4th edition: D.Orsi CJ, Mendelson FB, Ikeda DM et al: Breast Imaging and Reporting and Data System: ACR BI-RADS-Breast Imaging Atlas*, 4th ed., American College of Radiology, Reston, VA (2003).
- ACo Radiology, "BI-RADS Mammography 2013-ACR BI-RADS Atlas, 5th Edition," <http://www.acr.org/Quality-Safety/Resources/BIRADS> (17 March 2014).
- ACo Radiology, "The American College of Radiology BIRADS ATLAS and MQSA: Frequently asked questions," <http://www.acr.org> (25 September 2012).
- P. C. Brennan et al., "Ambient lighting: Effect of illumination on soft-copy viewing of radiographs of the wrist," *Am. J. Roentgenol.* **188**(2), W177–W180 (2007).
- Volparasolutions, "Volpara clinical breast density and its implications for your patients," <http://www.volparadensity.com/clinicians/> (27 March 2013).
- S. J. Vinnicombe et al., "Visual and automated volumetric assessment of density (MD): do measurements depend on the digital mammography unit?," in *European Congress of Radiology*, Austria, Vienna (2014).
- G. E. Mawdsley et al., "Accurate estimation of compressed breast thickness in mammography," *Med. Phys.* **36**(2), 577–586 (2009).
- A. H. Tyson, G. E. Mawdsley, and M. J. Yaffe, "Measurement of compressed breast thickness by optical stereoscopic photogrammetry," *Med. Phys.* **36**(2), 569–576 (2009).
- W. D. Thompson and S. D. Walter, "A reappraisal of the kappa-coefficient," *J. Clin. Epidemiol.* **41**(10), 949–958 (1988).
- M. M. Shoukri, *Book Measures of Interobserver Agreement.*, Chapman & Hall/CRC, Boca Raton, Florida (2004).
- R. L. Brennan and D. J. Prediger, "Coefficient kappa—some uses, misuses, and alternatives," *Educ. Psychol. Meas.* **41**(3), 687–699 (1981).
- J. Sim and C. C. Wright, "The kappa statistic in reliability studies: Use, interpretation, and sample size requirements," *Phys. Ther.* **85**(3), 257–268 (2005).
- F. K. Hoehler, "Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity," *J. Clin. Epidemiol.* **53**(5), 499–503 (2000).
- 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).
- C. M. Vachon et al., "Longitudinal breast density and risk of breast cancer," in *Proceedings of the American Association for Cancer Research Annual Meeting*, Vol. 51, pp. 1174–1174 (2010).
- K. Holland et al., "Stability of volumetric tissue composition measured in serial screening mammograms," in *Breast Imaging*, H. Fujita, T. Hara, and C. Muramatsu, Eds., Vol. 8539, pp. 239–244, Springer International Publishing, Switzerland (2014).
- M. F. McEntee and C. N. Damases, "Mammographic density measurement: A comparison of automated volumetric density measurement to BIRADS," *Proc. SPIE* **9037**, 90370T (2014).

Christine N. Damases is a PhD candidate at the University of Sydney, Faculty of Health Sciences, Medical Image Optimization and Perception research group (MIOPeG). She received her MTech (2006) and bachelor degrees (2001) from Durban University of Technology. Her current research interest includes medical image optimization and perception.

Patrick C. Brennan's research involves exploring novel technologies and techniques that enhance the detection of clinical indicators of disease, whilst minimizing risk to the patient. His research has involved most major imaging modalities including x-ray, computerized tomography, ultrasound, and magnetic resonance imaging, with a particular focus on breast and chest imaging. His research findings have translated into improved diagnosis and management of important disease states such as cancer, musculoskeletal injury, arthritis, and multiple sclerosis.

Mark F. McEntee is a researcher and lecturer in radiography with special focus on breast cancer. He became a senior lecturer in the University of Sydney in 2011. His current research interests mainly revolve around perception in medical imaging, receiver operating characteristic analysis of performance, human performance, and performance errors—particularly in medical imaging interpretation, as well as dose and image quality analysis.