Journal of Medical Imaging

SPIEDigitalLibrary.org/jmi

Number of mammography cases read per year is a strong predictor of sensitivity

Wasfi I. Suleiman
Sarah J. Lewis
Dianne Georgian-Smith
Michael G. Evanoff
Mark F. McEntee

SPIE

Number of mammography cases read per year is a strong predictor of sensitivity

Wasfi I. Suleiman,^{a,*} Sarah J. Lewis,^a Dianne Georgian-Smith,^b Michael G. Evanoff,^c and Mark F. McEntee^a

^aUniversity of Sydney, Medical Imaging Optimisation and Perception Group, and the Brain and Mind Institute, 94 Mallet Street, Camperdown, NSW 2050, Australia

^bHarvard Medical School, Brigham and Women's Hospital, Department of Radiology, 75 Francis Street, RA 020, Boston, Massachusetts 02115

^cThe American Board of Radiology, 5441 East William Boulevard, Suite 200, Tucson, Arizona 85711

Abstract. Early detection of breast cancers affects the 5-year recurrence rates and treatment options for diagnosed patients, and consequently, many countries have instituted nationwide screening programs. This study compared the performance of expert radiologists from Australia and the United States in detection of breast cancer. Forty-one radiologists, 21 from Australia and 20 from the United States, reviewed 30 mammographic cases containing two-view mammograms. Twenty cases had abnormal findings and 10 cases had normal findings. Radiologists were asked to locate malignancies and assign a level of confidence. A jackknife free-response receiver operating characteristic, figure of merit (JAFROC, FOM), inferred receiver operating characteristic, area under curve (ROC, AUC), specificity, sensitivity, and location sensitivity were calculated using Ziltron software and JAFROC v4.1. A Mann-Whitney U test was used to compare the performance of Australian and U.S. radiologists. The results showed that when experience and the number of mammograms read per year were taken into account, the Australian radiologists sampled showed significantly higher sensitivity and location sensitivity ($p \leq 0.001$). JAFROC (FOM) and inferred ROC (AUC) analysis showed no difference between the overall performance of the two countries. ROC (AUC) and location sensitivity were higher for the Australian radiologists who read the most cases per year. © 2014 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.1.1.015503]

Keywords: radiologists performance; mammography; experience; sensitivity; breast cancer.

Paper 13030PRR received Dec. 24, 2013; revised manuscript received Mar. 26, 2014; accepted for publication Apr. 7, 2014; published online May 7, 2014.

1 Introduction

Breast cancer is the second leading cause of cancer-related death in U.S. and Australian females.^{1,2} About 1 in 8 (12%) women in the United States and 1 in 11 women in Australia will develop this disease during their lifetime.^{2,3} Early detection affects the 5-year recurrence rates and treatment options for diagnosed breast cancer patients, and consequently, many countries have instituted nationwide screening programs.^{4–9} In 1985, the first results from the Swedish Two-County Trial showed a significant 30% reduction in breast cancer mortality between women given regular invitations to screening mammography versus those with no such invitation.¹⁰ After an 11-year follow-up, this study further demonstrated a benefit in screening for small-size cancers at the early stage.¹⁰ Success of early breast cancer detection depends on, among other things, the effectiveness of the breast screening program. However, breast screening programs operate differently in each country (Table 1).^{11–13}

In Australia, the national program for the early detection of breast cancer, known as BreastScreen Australia, has been screening women through free mammography since 1991.²¹ In the United States, the Food and Drug Administration, through the Mammography Quality Standards Act (1992), is the responsible body for the technical aspects and employee qualifications in mammography and oversees the compliance of mammography practices through audits.²² Although there is no national health service, there are coordinated health promotion programs

in place for breast health,²³ including advertising and support by the American Cancer Society and other professional societies. Promotion is done annually throughout the month of October, where health awareness programming is on TV, radio, work programs, and news organizations. The American Cancer Society has documented the continued decrease in breast cancer mortality since 2001, indicating the benefits of the screening program.

Previous researches have looked at the effect of many factors on diagnostic accuracy, such as reader characteristics (age, years since receipt of medical degree, average annual volume of mammogram reading, and ratio of screening to diagnostic mammographic interpretation), environmental characteristics (room light, temperature, monitor cleanliness, reflections, luminance measures, and resolution), and lesion characteristics [calcification, mass, ductal carcinoma *in situ* (DCIS), and architectural distortion (AD)].^{17–20,24,25} However, most previous studies considered the impact of factors affecting radiologist performance as a closed entity within the same country (such as United States,^{18,25} United Kingdom,¹⁹ and Australia^{17,20,24}), and there have been no studies undertaken on comparisons of radiologists' performance between two countries that may share similar cancer burden statistics and screening objectives.

To our knowledge, no studies have undertaken international comparisons in the performance of Australian and U.S. radiologists. Such a comparison may help shed light on whether differing practice requirements produce different outcomes in radiologists' accuracy. Women with the same disease patterns, imaged under similar conditions, should have equal opportunity

*Address all correspondence to: Wasfi I. Suleiman, E-mail: wsul5363@uni.sydney.edu.au

for an accurate outcome, and if they do not, then investigating the reasons for any difference may help to improve breast cancer detection through mammography. The aim of the current work is to compare the diagnostic performance of radiologists from Australia and the United States in mammography and to evaluate work practices, such as number of cases read per year. A specific aim was to place this study in the context of finding malignancies via the appearance of AD. AD is the third most common mammographic sign of nonpalpable breast cancer and it is an important false-negative (FN) finding during screening and is often missed because of its variability and subtlety.^{26–28}

2 Methods

This institutional review board-approved study involved two samples of mammography radiologists considered experienced by their own national standards.^{14,16}

2.1 Radiologists

A total of 21 Australian and 20 U.S. radiologists gave informed consent, and demographic data were collected. This included age, number of years since certification or qualification, and number of mammograms read per year. Australian radiologists were sampled from the Breast Imaging Group of Royal Australia and New Zealand College of Radiologists. The U.S. radiologists were sampled from the breast radiology examiners of the American Board of Radiology.

2.2 Cases

The radiologists independently reviewed the same 30 digital mammographic cases containing two views, a cranio-caudal, and a medio-lateral oblique projection of each breast. Twenty cases had abnormal findings and 10 cases were normal. The normal cases were reviewed and confirmed by two independent radiologists, and a follow-up normal screening mammogram was obtained 2 years later. All 20 abnormal cases contained

a biopsy-proved malignancy, 10 with the appearance of AD and 10 with appearances other than AD, such as masses and DCIS. The abnormal cases had 19 single and 1 multicentric cancers. The images came from the archives of BreastScreen New South Wales (NSW) and the Brigham and Women's Hospital, Boston, Massachusetts. The images were selected by the NSW State Radiologist (W.L.) and the associate professor of radiology Brigham and Women's (D.G.S.).

2.3 Viewing Conditions

Two Eizo-Radiforce GS510 monochrome LCD diagnostic monitors (resolution: 2048×2560) were used to display images in both experiments. The monitors were calibrated according to the Digital Imaging and Communications in Medicine Part 14 Standard using an EIZO photometer (UXI Sensor) and EIZO Quality Control Software (RadiCS). Average viewing distance was ~70 cm from the display, and the ambient light in the reading room ranged from 12 to 20 lux. The same general procedure was used for both experiments.

2.4 Image Scoring

Radiologists were asked to visualize and mark any malignancy, assigning a confidence level from 2 to 5. When normal or benign images were marked with the number 1, this progressed the radiologist to the next case. The instructions were displayed on screen before the experiment began. Radiologists were not told the exact prevalence of disease but were told that the case set was highly enriched.

Radiologists were allowed to digitally manipulate the images, such as by panning, zooming, and windowing. The radiologists advanced to the next case with a mouse-click when they were certain that they had marked all lesions. There was no restriction on search time and number of mouse-clicks. Ziltron software (Dublin, Ireland) was used in this experiment to display images and collect the mark-rating pairs, as the mark is

Table 1 Summary of the breast cancer screening programs in Australia and the United States.^{14–21}

	Australia	USA
	BreastScreen Australia	American College of Radiology
Accreditation		
Minimum number of mammograms read	2000 per year	240 with direct supervision then 960 every 2 years
Recall rate	11%	10 to 13.3%
Subsequent exams recall rate	4%	5.0 to 6.6%
Target	Women 50 to 69 years old	Women >40 years old
Imaging	Bilateral two projections mammography	Bilateral two projections mammography
Cost	Free	Paid, with reimbursement from private or public insurance coverage, privately or government funded
Breast screen policy	Double reading of the screening mammograms by two screen readers independently. If the two reports are not concordant, then a third radiologist is required.	One reading, perhaps in conjunction with computer aided detection, where available

Note: 14,940 Australian women from a population of 23.28 million and 232,340 American women from a population of 313.9 million have been diagnosed with breast cancer in 2013.^{2,3}

the indicated location of the suspicious region and the rating is the radiologist's confidence in the mark. A jackknife free-response receiver operating characteristic, figure of merit (JAFROC, FOM) analysis and inferred receiver operating characteristic, area under curve (ROC, AUC) were used. The JAFROC (FOM) is the nonparametric estimate of the area under the alternative AFROC curve, except that in JAFROC (FOM) only normal images are used to estimate the x -coordinate false-positive fraction of the AFROC curve; it is a trapezoidal estimate of the area under the AFROC curve.²⁹ Results for JAFROC (FOM) and ROC (AUC) were calculated using JAFROC v.4.1 software for all radiologists. The ROC (AUC) was used to facilitate comparison with previously published work; it is a binary paradigm based on whether the case does or does not have cancer and is the area under the trapezoidal ROC curve.

Specificity measures the proportion of truly negative cases that are correctly identified as true negative (TN) by the radiologist (specificity = $TN / (TN + FP)$), where FP indicates false-positive rate. Sensitivity, or true-positive (TP) rate, measures the proportion of actual positive cases that are correctly defined as such (sensitivity = $TP / (TP + FN)$), where FN indicates false-negative rate. Location sensitivity is the proportion of TP marks identified in the correct location as defined by a 50-pixel radius from the center of the lesion.³⁰ These were calculated using the Ziltron software, and radiologists received an accuracy score at the end of the study.

Using SPSS Software (version 21.0; SPSS, Chicago, Illinois), a Mann-Whitney U test was used to compare the overall performance of both Australian and U.S. radiologists. Furthermore, an independent sample t -test was used to compare the radiologists' number of years of experience and cases read per year, as previous research has indicated this may be an important consideration in performance.^{17,18} Results with a $p \leq 0.05$ were deemed to represent significant differences.

Table 2 Compares JAFROC (FOM), ROC (AUC), specificity, sensitivity, location sensitivity, years of experience reading mammograms, and mammograms read per year between Australia and U.S. radiologists.

Score type	Australia mean (sd)	USA mean (sd)	p value
JAFROC (FOM) overall treatment	0.72 (0.1)	0.72 (0.1)	0.90
ROC (AUC) overall treatment	0.83 (0.08)	0.82 (0.07)	0.49
Specificity (true negative rate)	0.54 (0.18)	0.55 (0.22)	0.89
Sensitivity (true positive rate)	0.91 (0.07)	0.88 (0.08)	0.17
Location sensitivity	0.82 (0.08)	0.76 (0.07)	0.05
Years of reading mammograms	15.9 (8.93)	23.3 (8.58)	0.01
Mammograms read per year	11,027 (9356)	6288 (3492)	0.03

3 Results

A total of 1230 readings were made, with 41 radiologists reading 30 cases. The 21 Australian radiologists had a mean number of years experience of 15.9 (sd 8.93), which was significantly lower than that of the 20 U.S. radiologists sampled at 23.3 (sd 8.58) ($p \leq 0.01$). The Australian radiologists read an average of 11,027 (sd 9356) cases per year, which was significantly higher than the U.S. radiologists, who read 6288 (sd 3492) cases per year (Table 2).

An overall comparison of the two groups indicated no significant differences in performance between JAFROC (FOM) ($p \leq 0.9$) or inferred ROC (AUC) ($p \leq 0.49$); however, the Australian radiologists overall demonstrated higher location sensitivity ($p \leq 0.05$) (Table 2). The performance was analyzed on the basis of low and high numbers of years of experience and number of mammograms read per year. When the samples were split into those higher and lower than the mean years of experience and the mean numbers of mammograms read per year, differences are evident, with the most experienced Australian radiologists demonstrating higher performance in sensitivity ($p \leq 0.01$) and location sensitivity ($p \leq 0.001$) (Table 3).

Table 3 Demonstrates the performance characteristics for Australian and U.S. radiologists and indicates the statistical comparison of the groups split into more ($\geq$ mean) and less ($<$ mean) number of years of experience. It also compares performance characteristics between Australia and U.S. low and high years of experience.

Score type	Australia (mean = 15.9)	USA (mean = 23.3)
	Mean (sd)	Mean (sd)
JAFROC (FOM) more experience	0.76 (0.01)	0.71 (0.10)
JAFROC (FOM) less experience	0.70 (0.09)	0.72 (0.10)
ROC (AUC) more experience	0.87 (0.07)	0.80 (0.07)
ROC (AUC) less experience	0.81 (0.07)	0.83 (0.07)
Specificity more experience	0.51 (0.20)	0.61 (0.20)
Specificity less experience	0.57 (0.17)	0.51 (0.23)
Sensitivity more experience	0.95 (0.02)*	0.84 (0.09)*
Sensitivity less experience	0.89 (0.09)	0.91 (0.07)
Location sensitivity more	0.86 (0.05)#	0.73 (0.07)#
Location sensitivity less	0.79 (0.09)	0.78 (0.07)
FP more experience	0.49 (0.2)	0.39 (0.2)
FP less experience	0.4 (0.15)	0.49 (0.23)

Note: * and # indicate that the identified pairs were significantly different ($p \leq 0.05$).

Table 4 Demonstrates the performance characteristics for Australian and U.S. radiologists and indicates the statistical comparison of the groups split in high number of mammograms ($\geq$ mean) and low number of cases ($<$ mean) read per year. It also compares performance between those with a low and high number of mammograms read per year.

	Australia (mean = 11,027)	USA (mean = 6288)
Score type	Mean (sd)	Mean (sd)
JAFROC (FOM) high cases	0.77 (0.09)	0.72 (0.10)
JAFROC (FOM) low cases	0.69 (0.09)	0.72 (0.11)
ROC (AUC) high cases	0.87 (0.07)*	0.82 (0.08)
ROC (AUC) low cases	0.80 (0.07)*	0.81 (0.07)
Specificity high cases	0.52 (0.19)	0.54 (0.19)
Specificity low cases	0.57 (0.17)	0.56 (0.25)
Sensitivity high cases	0.94 (0.03)	0.9 (0.09)
Sensitivity low cases	0.89 (0.09)	0.87 (0.07)
Location sensitivity high cases	0.86 (0.05)^	0.76 (0.09)#
Location sensitivity low cases	0.79 (0.10)^	0.77 (0.06)#
FP high cases	0.48 (0.19)	0.46 (0.19)
FP low cases	0.43 (0.17)	0.44 (0.26)

Note: *, ^, and # indicate that the identified pair were significantly different ($p \leq 0.05$).

The most experienced Australian radiologists had higher performance than the most experienced U.S. radiologists (Table 3), although this pattern was nonsignificant; in the less experienced groups, the U.S. radiologists performed better than the Australian radiologists. The sensitivity and location sensitivity of more experienced Australian radiologists were significantly higher than the more experienced U.S. radiologists ($p \leq 0.05$). A similar pattern was seen when the number of mammograms read per year was used to separate the groups. In the second analysis, those with more mammograms read per year for that country were considered more experienced and those with less than the mean were considered less experienced (Table 4).

4 Discussion

The American radiologists sampled had more years of experience ($p \leq 0.01$); however, they read fewer mammograms per year ($p \leq 0.03$) compared to the Australian radiologists sampled. There was no difference in the JAFROC (FOM) and ROC (AUC) performance measures for both countries. Previous research has indicated that radiologists' sensitivity performance in reading mammographic images is closely linked to the number of mammographic cases they read per year and less linked to their number of years of experience.¹⁷⁻²⁰ Our work has confirmed this, with radiologists who have the highest number of cases per year performing better than those with lower numbers, while those with a higher number of years' experience did not

perform better than those with less. In fact, even though the American radiologists had the highest number of years of experience, their performance was lower in sensitivity and location sensitivity when compared to the Australian sample.

Comparison of multiple measures of performance was carried out; however no *post hoc* correction was performed. Adjusting for multiple comparisons increases the risk of type II errors, which can make results difficult to interpret. It is accepted that p values do not need to be adjusted in exploratory studies.³¹ The results showed no difference in the performance of Australian and U.S. radiologists when viewed overall; however, when the number of cases per year is used as a determinant of experience, the results demonstrated that the most experienced Australian radiologists outperform all others in location sensitivity. The Australian radiologists who read the most cases per year had a better ROC and location sensitivity than the least experienced radiologists, and this suggests that reading more cases per year increases location sensitivity. This confirms the findings of other authors¹⁷⁻²⁰ who have suggested that reading >2000 cases per year improves performance.

There was a significant increase in location sensitivity between the experienced Australian radiologists (mean experience ≥ 15.9 years) and the experienced U.S. radiologists (mean experience ≥ 23.3), and this could be regarded as an unexpected finding. However, this finding may, in part, be due to Australian radiologists' operating under a system of double reading that promoted location of lesions as well as adhering to a lower national standard of recall rate. On the other hand, U.S. radiologists make singular decisions on diagnosis, and official reports indicate a higher recall rate between 10 and 13.3% in the United States for initial screening,^{15,16} while in Australia it is 11%.²¹

Increasing the number of cases read per year does not seem to have the same effect on specificity, with no significant differences seen between groups. Using the mean number of years as a determinant of experience shows a similar result for the international comparison, with the Australian radiologists having higher location sensitivity than the U.S. radiologists ($p \leq 0.001$). Again, specificity is not affected. However, unlike increasing the number of cases read per year, increasing the number of years of experience does not result in a significant difference between the least and most experienced radiologists. The difference in performance within the U.S. and Australian groups was most clearly demonstrated using the number of cases per year rather than by the number of years of experience.

Previous studies of the effect of radiologists' characteristics, training, and clinical experience on performance have generated conflicting results, even when the same populations of radiologists were studied.^{12,32} Uncertainty about this effect is caused, in part, by the often limited number of radiologists and mammographic examinations in published studies and the different measures used to define such characteristics as clinical experience. Radiologists' characteristics that have been studied in the United States and Australia include number of years postgraduation as radiologist and number of years of reading mammograms, fellowship training in breast imaging, and various measures of mammographic interpretive volume.^{12,19,20,33}

We acknowledge that there were limitations in this work. First, the data set used in the current work was relatively small, and the cases were not typical of a screening environment due to enrichment. Second, the work practices of the two samples are very different. No attempt was made to correct for this as, to a large extent, the current work seeks to assess the impact

of this difference. Third, the samples of radiologists were chosen for their convenience rather than as a random sample of Australian or American population of radiologists. Fourth, dividing the samples into the less and more experienced radiologists on the basis of the sample mean is an artificial measure of the actual population mean, particularly as the American sample contained only highly experienced radiologists in terms of years. Finally, it is important to note that in the research setting, radiologists overcall, which decreases the specificity. This may also increase sensitivity because more actual positive cases are also being called back when radiologists increase recalls. Both the FP fraction and the TP fraction increase along the ROC curve as radiologists change their decision thresholds. This was confirmed by a study in New South Wales (Australia),³² which assessed the ability of test conditions to represent performance in actual clinical reporting in screening mammography. These authors concluded that an acceptable level of agreement between actual clinical reporting and test set conditions can be achieved (Kendall's $W = 0.69$ to 0.72 , $p < 0.01$) when prior images are used and a greatest decrease in agreement in tests versus clinical when prior images were not provided to the radiologists.

Conversely, the considered nature of the test bank, the involvement of a relatively large number of experienced breast radiologists, and the controlled and realistic viewing and ambient conditions mean the results are worthy of consideration. The breast cancer cases chosen for this work included 10 with AD and 10 with other presentations of breast cancer, including masses. Thus, this test set may have presented as more challenging than test sets used in other studies, and this needs to be taken into consideration when interpreting the results. Further work will investigate the performance of radiologists in the detection of AD versus other breast cancer appearances and examine the causes of missed cancers, particularly whether they are perceptual or decision-making errors. The controlled experimental design means that any differences demonstrated were observer-related rather than a result of differences in technology.

5 Conclusion

In this study, the number of mammographic cases read per year was the most important factor in predicting radiologists' increased sensitivity performance. The highest performances were seen in those who read the highest number of cases per year, with sensitivity demonstrating this most clearly.

Acknowledgments

We would like to acknowledge the assistance of Dr. Warrick Lee and Mohammad Rawashdeh for selecting the normal and abnormal (nonarchitectural) distortion images. Thanks also go to Dr. John Ryan (UCD Dublin and Ziltron) for his assistance with software and data analysis. Statistical support was provided for this study by Professor Jenny Peat, Honorary Professor, Australian Catholic University.

References

- AIHW & AACR, "Cancer in Australia: an overview 2012," 2012, <http://www.aihw.gov.au/cancer/cancer-in-australia-overview-2012/> (8 April 2014).
- ACS, "What are the key statistics about breast cancer," 2013, <http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-key-statistics> (20 September 2013).
- Breast Cancer Care Western Australia, "What is breast cancer?," 2013, <http://www.breastcancer.org.au/about-breast-cancer/what-is-breast-cancer/statistics.aspx> (20 September 2013).
- A. Mencattini, M. Salmeri, and P. Casti, "Bilateral asymmetry identification for the early detection of breast cancer," in *IEEE Int. Workshop on Medical Measurements and Applications Proceedings (MeMeA)*, pp. 613–618 (2011).
- M. A. Schneider, "Better detection: improving our chances," in *Digital Mammography: 5th International Workshop on Digital Mammography IWDM*, pp. 3–6, Wis. Medical Physics Publishing, Madison (2000).
- S. W. Duffy et al., "The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties," *Cancer* **95**(3), 458–469 (2002).
- K. Kerlikowske, "Efficacy of screening mammography among women aged 40 to 49 years and 50 to 69 years: comparison of relative and absolute benefit," *JNCI Monogr.* **1997**(22), 79–86 (1997).
- L. Nyström et al., "Long-term effects of mammography screening: updated overview of the Swedish randomised trials," *Lancet* **359**(9310), 909–919 (2002).
- R. Taylor et al., "Mammography screening and breast cancer mortality in New South Wales, Australia," *Cancer Causes Control* **15**(6), 543–550 (2004).
- L. Tabár et al., "The Swedish Two-County Trial twenty years later: updated mortality results and new insights from long-term follow-up," *Radiol. Clin. North Am.* **38**(4), 625–651 (2000).
- B. S. Monsees et al., "The Mammography Quality Standards Act: an overview of the regulations and guidance," *Radiol. Clin. North Am.* **38**(4), 759–772 (2000).
- R. Smith-Bindman et al., "Comparison of screening mammography in the United States and the United Kingdom," *JAMA* **290**(16), 2129–2137 (2003).
- L. Kan et al., "Standardized abnormal interpretation and cancer detection ratios to assess reading volume and radiologist performance in a breast screening program," *Radiology* **215**(2), 563–567 (2000).
- BreastScreen Australia, "National accreditation standards," 2008, [http://cancerscreening.gov.au/internet/screening/publishing.nsf/Content/br-accreditation/\\$File/standards.pdf](http://cancerscreening.gov.au/internet/screening/publishing.nsf/Content/br-accreditation/$File/standards.pdf) (19 September 2013).
- Breast Cancer Surveillance Consortium, "Evaluating screening performance in practice," 2004, <http://breastscreening.cancer.gov/espp.pdf> (27 January 2014).
- R. Smith-Bindman et al., "Comparing the performance of mammography screening in the USA and the UK," *J. Med. Screen.* **12**(1), 50–54 (2005).
- M. A. Rawashdeh et al., "Markers of good performance in mammography depend on number of annual readings," *Radiology* **269**(1), 61–67 (2013).
- J. G. Elmore et al., "Does diagnostic accuracy in mammography depend on radiologists' experience," *J. Women's Health* **7**(4), 443–449 (1998).
- L. Esserman et al., "Improving the accuracy of mammography: volume and outcome relationships," *J. Natl. Cancer Inst.* **94**(5), 369–375 (2002).
- W. Reed et al., "Malignancy detection in digital mammograms: important reader characteristics and required case numbers," *Acad. Radiol.* **17**(11), 1409–1413 (2010).
- AIHW, "BreastScreen Australia monitoring report 2010–2011," 2013, <http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129544880> (8 April 2014).
- Food and Drug Administration, "Quality mammography standards; correction—FDA: final rule; correction. 62," Federal Register, pp. 60613–60632 (1997).
- A. I. Scherl-Slusher, "Screening mammography: a comparison between US and Australian healthcare delivery," 2000, https://etd.ohiolink.edu/ap:0:0:APPLICATION_PROCESS%3DDOWNLOAD_ETD_SUB_DOC_0:0:ACCNUM::F1501_ID:ysu1007583885%2Cinline (21 November 2013).
- B. Soh et al., "Breast screen new South Wales generally demonstrates good radiologic viewing conditions," *J. Digit. Imaging* **26**(4), 759–767 (2013).
- E. A. Krupinski and H. Roehrig, "Recent developments in evaluating soft-copy displays for medical images," *Recent Res. Dev. Opt. Eng.* **3**, 35–49 (2000).
- A. M. Knutzen and J. J. Gisvold, "Likelihood of malignant disease for various categories of mammographically detected nonpalpable breast lesions," *Mayo Clin. Proc.* **68**(5), 454–460 (1993).
- H. C. Burrell et al., "Screening interval breast cancers: mammographic features and prognosis factors," *Radiology* **199**(3), 811–817 (1996).

28. S. R. Banik, R. M. Rangayyan, and J. L. Desautels, "Detection of architectural distortion in prior mammograms using fractal analysis and angular spread of power," *Proc. SPIE* **7624**, 762408 (2010).
29. DevChakraborty, "JAFROC," <http://www.devchakraborty.com/JAFROC.html> (2011).
30. T. M. Haygood et al., "On the choice of acceptance radius in free-response observer performance studies," *Br. J. Radiol.* **86**(1021), 42313554 (2013).
31. R. Bender and L. Stefan, "Adjusting for multiple testing—when and how?," *J. Clin. Epidemiol.* **54**(4), 343–349 (2001).
32. B. P. Soh et al., "Screening mammography: test set data can reasonably describe actual clinical reporting," *Radiology* **268**(1), 46–53 (2013).
33. C. A. Beam, P. M. Layde, and D. C. Sullivan, "Variability in the interpretation of screening mammograms by US radiologists: findings from a national sample," *Arch. Intern. Med.* **156**(2), 209 (1996).

Wasfi I. Suleiman graduated in electrical engineering. He completed his master's degree at the University of Technology, Sydney. He has joined the Faculty of Health Sciences at the University of Sydney in the discipline of medical radiation sciences, as a PhD student. His topic investigates the factors that affect diagnostic performance in mammography.

Sarah J. Lewis has research interests in the areas of ethical and optimal imaging. As a clinical radiographer, clinical academic, and researcher, her areas are in the articulation of clinical ethics in the

radiography/radiology service setting and the impact of ethics and professional boundaries upon the role of the radiographer. She has a strong interest in optimal imaging in CT, mammography, and pediatrics, both from a radiation dose and best practice perspective.

Dianne Georgian-Smith is a breast imager at Brigham and Women's Hospital and Boston Medical Center and associate professor of radiology at Harvard Medical School as well as an adjunct associate professor of radiology at Boston University School of Medicine. Her clinical interests include breast imaging, magnetic resonance imaging, mammography, sonography, ultrasound, and women's imaging—ultrasound, mammography.

Michael G. Evanoff is the division director of the Digital Imaging Division and Facilities Management of the American Board of Radiology. He has many years of experience in perception studies and has facilitated over 20 collaborations among the University of Sydney, the American Board of Radiology, and UCD Dublin.

Mark F. McEntee is a researcher and teacher in radiography with a special focus on breast cancer. He became a senior lecturer in the University of Sydney in 2011. 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.