

Impact of breast density on cancer detection: observations from digital mammography test sets

Abstract

Introduction: The aim of this study is to investigate the impact of breast density on the diagnostic efficacy of 273 breast screening radiologists reading 1 of 5 test sets of digital mammograms within the BREAST program.

Methods: Retrospective data was collected from two hundred and seventy-three breast screening radiologists who participated in BREAST test sets between 2012 and 2017. Radiologists reviewed one of five test sets (labeled T1-T5) each containing 60 digital mammographic cases with 20 cancers and 40 normal cases. The cases also had varying mammographic densities based on BI-RADS density ratings. Cases were grouped into low-density (LD) and high-density (HD), and sensitivity, lesion sensitivity, specificity, receiver operating characteristic (ROC) and jackknifing free response operating characteristic (JAFROC) figures of merit were compared using Mann-Whitney U or unpaired t-tests.

Results: Readers in three out of five test sets showed better case sensitivity and lesion sensitivity in LD compared with HD cases (T2, T4, and T5: $P \leq 0.001$). One out of five test sets showed the same trend with specificity (T4: $P < 0.0001$) and another set followed this pattern for ROC (T5: $P < 0.0001$) and JAFROC (T5: $P < 0.0001$) values. One test set (T3) demonstrated better performance in HD versus LD cases (ROC $P < 0.0001$ and JAFROC $P < 0.0001$). Reader experience, training, and set features may account for the variation in the results between the test sets.

Conclusion: Overall, radiologists taking the BREAST test sets perform better on cases with low mammographic density thus highlighting diagnostic advantage for women with lower breast density and reiterating the challenges presented by mammograms of dense breasts.

Keywords: breast imaging, mammography, breast cancer, breast density, clinical radiology

Volume 7 Issue 2 - 2020

Kriscia A Tapia,¹ Mary T Rickard,^{1,2} Mark F McEntee,^{1,3} Gail Garvey,^{1,4} Lorraine Lydiard,⁵ Patrick C Brennan¹

¹Department of Medicine and Health, University of Sydney, Australia

²Department of Medicine and Health, BreastScreen New South Wales, Australia

³Department of Medicine, University College Cork, Ireland

⁴Menzies School of Health Research, Charles Darwin University, Australia

⁵Department of Medicine and Health, BreastScreen Northern Territory, Australia

Correspondence: Kriscia A Tapia, department of Medical Imaging Sciences, Faculty of Health Sciences, University of Sydney, 75 East Street, Lidcombe, NSW, 2141, Australia, Tel +61 422610330, Email kiscia.tapia@sydney.edu.au

Received: February 25, 2020 | **Published:** March 05, 2020

Introduction

Mammographic density presents a challenge for breast screening radiologists with strong evidence of increased breast cancer risk and false negative outcomes in women with higher breast density levels. The risks are associated with breast tissue composition and histology linked to breast cancer susceptibility,^{1,2} and reduced mammographic accuracy with lower positive predictive values, higher risk of interval cancers, and higher recall rates among dense cases.³⁻⁵ The effect known as “masking”, where malignancies are obscured in dense areas of the breast rendering them difficult to detect, is reported to be a key factor in errors made by radiologists when examining mammograms.⁶⁻⁸

The majority of evidence pertaining to diagnostic errors and radiologic performance in dense breasts have in the past involved studies using film-screen systems.⁹⁻¹¹ More recently, researchers have focused on the now widely used digital mammography detectors,^{12,13} the findings of which support the results achieved in film mammography in that cases with low density versus high density yielded better mammographic accuracy. However, the result of one study,¹⁴ diverged from those findings in that the performance of six radiologists reading digital mammograms in a test set was better on high density breasts compared with low density breasts. Further work is therefore needed to investigate the impact of breast density on cancer detection in the digital era using rigorous test set approaches involving high quality images and many expert observers. BreastScreen Reader Assessment Strategy (BREAST) is a radiology reader assessment program used in Australia and New Zealand by BreastScreen radiologists. It assesses the performance of clinicians on enriched sets of 60 digital

mammography cases via a web-based application.^{15,16} Due to the large number of participants and data generated by BREAST, it provides an opportunity to validate the effect of breast density on cancer detection. The aim of this work is to utilize BREAST-generated data to investigate the effect of breast density on the diagnostic efficacy of screening radiologists reading digital mammograms.

Materials and methods

Data collection

Ethical approval was obtained for this retrospective study (HREC 2017/028) and electronic informed consent was collected from radiologists prior to participating in the test sets. A computer program collected test set performances from $n=273$ radiologists from 2012 to 2017. During this time radiologists may have done more than 1 test set however we only included the first test set completed. Study participants were employees of BreastScreen Australia or Aotearoa (New Zealand) at the time of completing a test. Performance metrics on five test sets were collected and for this study were named, test 1 (T1), test 2 (T2), test 3 (T3), test 4 (T4), and test 5 (T5). Each test set has 60 anonymized digital mammography cases compiled from the national screening programs’ image libraries. The BREAST program provided the software, which is routinely used in Australia and New Zealand by BreastScreen radiologists. Participation in BREAST is an activity that meets the National Accreditation Standards (NAS)¹⁷ and it is one of the largest continuing observer studies in digital mammography reading in the world.

Test sets

Five test sets or a total of 300 mammography cases from a screening population were used. Each test set includes 20 cases with biopsy proven malignancies and 40 normal cases. Cases were confirmed by two experienced radiologists and follow-up negative screening mammograms obtained in the succeeding screening round. The 20 positive cases in each test set contain a variety of lesion sizes and malignancy appearance. The cases also have a range of breast densities described using the categories 1 to 4 of the Breast Imaging Reporting and Data System (BI-RADS) version 4 (2003) and evaluated by at least two expert radiologists. This version of BI-RADS was in widespread use in Australia at the time of test set development by the BREAST program and generally describes volume of density. The categories are: 1=almost entirely fatty, approximately 0-25% density; 2= scattered areas of fibroglandular density, approximately 25-50% density; 3= heterogeneously dense, approximately 50-75% density; and 4= extremely dense, approximately 75-100% density.¹⁸ For this study we grouped BI-RADS 1&2 as “LD” (LD) and BI-RADS 3&4 as “HD” (HD). Where available, cases were presented with mammograms from the previous screening round.

Performance data

The clinicians (“readers”) who participated in the study viewed the test sets via the BREAST platform either at a conference workshop in a simulated reading room, or online in their usual clinical setting. Readers spent between 1.5 to 2 hours completing a test set. Comparable viewing conditions were available to readers in both settings. Images were displayed on pairs of 5 megapixel medical-grade monochrome liquid crystal display monitors with a resolution of 2.049 by 2.560 pixels. Ambient light levels were maintained at 25–40 Lux¹⁹ throughout the study. Readers were firstly asked to complete a survey with demographic and work experience questions prior to starting a test. Test instructions were then provided to readers in which they were asked to examine each case as per their usual clinical reporting and mark as many lesions as they could identify on both cranio-caudal and medio-lateral oblique views. Readers could mark a lesion with a double mouse click and allocate a confidence score ranging from 2 to 5 where: 2=benign; 3=equivocal; 4=suspicious; 5=malignant. Scores of 3, 4 and 5 were used to demonstrate increasing confidence in the malignancy of the marked lesion. Where no significant abnormality was detected on a case, readers were asked to click on the “next case”

icon which is saved by the computer program as 1= normal. When readers complete the final case in the set, they submit their answers and immediately receive feedback on their performance with scores on sensitivity and specificity (sensitivity and specificity were calculated by combining reader confidence score of 1 and 2 as a normal result and 3, 4 and 5 as abnormal), lesion sensitivity, receiver operating characteristic (ROC) and jackknife free-response receiver operating characteristic (JAFROC) figures of merit (FOM). JAFROC FOM is acquired by combining lesion sensitivity, specificity and confidence score³⁻⁵ with a minimum value of 0.00 and maximum value of 1.00).

Data analysis

Firstly, the characteristics of readers in each test set were explored based on the information survey-collected including age, gender, training and work-experience. We then separated the cases in each of the test sets, T1, T2, T3, T4, and T5 into two density groupings, LD and HD, and further divided the cancer cases from the cancer-free cases. We therefore had each test set stratified by densities (low and high) and types (abnormal and normal). Next, we measured performance metrics of individual readers in each test set within the LD and HD categories. We then calculated for each reader, sensitivity, lesion sensitivity and specificity scores out of 100%. ROC and JAFROC FOM were calculated using the BREAST software algorithm, which provided resultant figures out of 1.0. We compared the means (for parametric data) or medians (for non-parametric data) of the scores between LD and HD cases in each test set. Unpaired t-test and Mann-Whitney test was used to determine p-values, and $p \leq 0.05$ was considered significant.

Results

Reader profiles

There were 273 individual radiologists in this sample. The numbers of readers per test set are described in Table 1. Readers had an average age of 50 years and were 46% males and 54% females. Readers on average had 13 years’ experience with reading mammograms and a median of 2,640 cases were being read per year. Only 25% of readers said that they had completed a fellowship in mammography reading, and there were a larger proportion of radiologists (80%) currently reading mostly digital mammograms as opposed to film. The responses to questions about training and work experiences are shown in Table 1.

Table 1 Radiologists’ characteristics and work experience at the time of completing a test set

Reader characteristics	All tests (n=273)	T1 (n=40)	T2 (n=108)	T3 (n=31)	T4 (n=64)	T5 (n=30)
Mean age (y)	50 (±11)	47 (±13)	51 (±11)	48 (±12)	51 (±12)	50 (±8)
Gender (male, female)	46%, 54%	45%, 55%	35%, 65%	52%, 48%	67%, 33%	40%, 60%
Mean no. of years reading mammograms	13 (±10)	8 (±9)	15 (±11)	12 (±9)	15 (±10)	13 (±9)
Median no. of annual cases read (Q1, Q3)	2640 (480, 8640)	960 (480, 3240)	3840 (1140, 8640)	1140 (480, 6240)	1140 (480, 8640)	3840 (1200, 8880)
% of readers who completed a fellowship lasting 3 to 6 months	25%	33%	29%	7%	18%	37%
% readers who read mostly digital mammograms	80%	70%	86%	81%	70%	93%

Mean are presented with standard deviations (±) and Medians are presented with 25th percentile (Q1) and 75th percentile (Q3)

Test set characteristics

While the tests each contained 20 cancer cases, T1, T2, and T3 contained one case with 2 lesions. Therefore, the total number of lesions were $n=103$ with 47 lesions in the LD category and 56 lesions in the HD category. T3 contained the lowest number of HD

cases ($n=16$ or 27% dense), and T4 had the most ($n=38$ or 63%). The average lesion sizes for LD and HD cases respectively were 12.2mm and 11.7mm. The smallest average lesion within LD cases was in T5 (9.1mm) and the largest was within T4 (16.5mm). For HD cases, the smallest average lesion was within T1 (10mm) and the largest was within T4 (14.8mm). These are shown in Table 2.

Table 2 Numbers of lesions and normal cases in LD and HD mammograms, and average lesion sizes in each of the five test sets

Low density				High density		
Test set	No. of lesions	Lesion size (mm)	No. of normal cases	No. of lesions	Lesion size (mm)	No. of normal cases
T1*	9	9.9	19	12	10	21
T2*	7	10.6	17	14	11.9	23
T3*	14	14.8	31	7	10.3	9
T4	6	16.5	16	14	14.8	24
T5	11	9.1	20	9	11.6	20
Total	47	12.2	103	56	11.7	97

*Test sets with one cancer case containing two lesions (all other test sets have only one lesion per cancer case)

Readers' performances

The sensitivity, lesion sensitivity, and specificity scores of all test sets had non-normal distributions, while ROC and JAFROC FOM were normally distributed, except for T2 and T3 scores. Therefore, medians and interquartile ranges were assessed for sensitivity, lesion sensitivity, and specificity scores on each test set, and means were tested for ROC and JAFROC FOM, except for T2 and T3 in which means and standard deviations were used. All test sets except T3 showed that performance was better in LD cases than HD cases. This was apparent in sensitivity, lesion sensitivity, specificity, ROC and

JAFROC FOM. However, these were only significant for sensitivity and lesion sensitivity in T2, T4 and T5 ($p<0.05$), specificity in T4 ($p<0.0001$), ROC and JAFROC FOM in T5 ($p<0.0001$). Although there were marginal improvements in lesion sensitivity for T1 in LD versus HD cases (approx. 0.04% difference), and in specificity for T2 LD versus HD cases (approx. 1.8% difference), these were not statistically significant. For T3, performance was significantly better in HD cases compared with LD cases in sensitivity ($p<0.001$), lesion sensitivity ($p\leq 0.001$), ROC ($p<0.05$), JAFROC FOM, ($p<0.05$), but non-significant for specificity scores. These are shown in Table 3.

Table 3 Differences between scores of readers in low and HD cases in each test set

Test set	Low density cases	High density cases	p-value
Sensitivity (%)			
T1	75 (53.1, 87.5)	72.7 (54.5, 81.8)	0.478
T2	71.4 (57.1, 85.7)	69.2 (61.5, 84.6)	0.009*
T3	53.8 (46.2, 69.2)	71.4 (57.1, 85.7)	0.001*
T4	66.7 (50, 83.3)	57.1 (35.7, 71.4)	0.001*
T5	90.9 (81.8, 90.9)	77.8 (66.7, 88.9)	0.001*
Lesion Sensitivity (%)			
T1	72.3 (55.6, 88.9)	72.7 (54.5, 81.8)	0.688
T2	71.4 (57.1, 85.7)	71.4 (51.7, 78.6)	0.001*
T3	50 (42.9, 71.4)	71.4 (57.1, 85.7)	0.001*
T4	66.7 (50, 83.3)	57.1 (35.7, 71.4)	0.006*
T5	90.9 (81.8, 90.9)	77.8 (66.7, 88.9)	0.001*

Table Continued...

Test set	Low density cases	High density cases	p-value
Specificity (%)			
T1	84.2 (68.4, 89.5)	73.4 (66.8, 85.5)	0.063
T2	76.5 (66.2, 82.4)	78.3 (69.3, 91.3)	0.099
T3	80.6 (71, 90.3)	77.8 (66.7, 88.9)	0.611
T4	87.5 (81.3, 93.8)	79.2 (70.8, 87.5)	<0.0001*
T5	80 (70, 85)	77.5 (65, 80)	0.064
ROC (max. is 1.0)			
T1*	0.83 ±0.10	0.83 ±0.08	0.912+
T2	0.84 (0.77, 0.90)	0.86 (0.80, 0.90)	0.273
T3	0.78 (0.72, 0.84)	0.85 (0.79, 0.93)	0.004*
T4*	0.74 ±0.11	0.73 ±0.10	0.595+
T5*	0.90 ±0.05	0.80 ±0.08	<0.0001*+
JAFROC FOM (max. is 1.0)			
T1*	0.68 ±0.15	0.66 ±0.11	0.523+
T2	0.72 (0.62, 0.81)	0.72 (0.63, 0.78)	0.225
T3	0.60 (0.50, 0.71)	0.69 (0.57, 0.84)	0.021*
T4*	0.63 ±0.15	0.62 ±0.14	0.569+
T5*	0.82 ±0.07	0.80 ±0.08	<0.0001*+

Note: P values with an asterisk (*) denotes significance. Where there is a plus symbol (+) mean values and standard deviations were used, and P values were derived from unpaired t tests. All other data represent median values with first and third quartiles, and P values were derived from Mann-Whitney U tests

Discussion

It is well-reported that radiologic performance with mammograms decreases with increasing breast density.^{7,8,20–22} We show a similar trend in four out of five tests with readers having improved cancer detection as well as lesion localization in LD compared with HD cases. In both ROC and JAFROC FOM analyses, one out of five tests reflected the beneficial effect of LD on reader performance, a finding consistent with the literature on diagnostic efficacy when accounting for density in digital mammograms.^{12,13} The anomalous test was T3 in that sensitivity, lesion sensitivity and JAFROC FOM were significantly better in HD rather than LD cases. While this result is irregular, it is not unheard of, with a small number of studies reporting similar findings particularly in digital mammography.^{22–24} The possible reasons for the results seen in T3 will be explored below. For specificity, our results show a similar trend in the majority of tests including T3 with decreased specificity in HD versus LD cases. Our findings reiterate the challenge that breast density presents for screening radiologists within a population-based program, for whom cancer-free cases consist the bulk of their caseload and recall rates are closely monitored. Our overall specificity results corroborate previously reported evidence that women with dense breasts are more likely to be recalled and attract false positive findings.^{24,25}

Similar to the nonconformity of T3's results a recent study using digital mammograms of Australian women (14) demonstrated that six expert radiologists, that is, observers who had annual case-loads of over 2000 mammograms, produced significantly higher JAFROC FOM when reading high (BI-RADS 3&4) versus low (BI-RADS 1&2) density cases. However, observers in that previous work average more than 6000 annual reads while the readers of T3 in this current work had a median annual case-load of 1140, much like the readers in T4 where sensitivity and specificity were lower in HD cases. Indeed, T3 shared many similarities with other tests in the current work such as readers' reading experience and usage of digital mammography as opposed to film, giving little clues as to possible associations with performance. The only notable variation is that T3 readers had the lowest rate of breast fellowship completions among all tests. This suggests that more readers in T3 were likely to have experienced generalized training or alternative forms of breast instruction such as on-the-job training or a rotation through BreastScreen. The influence of fellowship training upon readers' performance on mammographically dense breasts should be the focus of future work.

A test set feature which may explain T3's results is that T3 contained the lowest proportion of dense abnormal and cancer-free cases compared to the other tests. This suggests that encountering

fewer difficult cases within one reading experiment session potentially enhances diagnostic ability when presented with challenging cases and may heighten radiologists' visual attention to abnormalities.¹⁴ While this effect, if true, should influence all other tests and not just one, it could be that the smaller number of challenging cases within T3 and the restricted time-frame for readers to complete the test set (i.e., 1.5 to 2 hours) might have provided some advantage to readers. The validity of this claim should be verified using test sets of similar density characteristic within the BREAST program. While the current work confirms the need for optimizing the diagnostic efficacy of radiologists reading mammograms of dense breasts, it also sheds light on the increased benefits of breast screening women with lower mammographic densities as their cancers are more easily detectable. It is estimated that 57% of women in the United States' undergoing breast screening have scattered density or almost entirely fatty breasts,²⁵ and a study in the Northern Territory of Australia found that 53% of screened women were allocated BI-RADS 1 or 2 density ratings by pairs of radiologists.²⁶ While all women benefit from early detection, especially women with high density measures and increased breast cancer risk, screening sections of the population with known lower density, for example, older women and women of diverse ethnicities poses an advantage.

The increased diagnostic efficacy of radiologists reviewing cases that represent the majority of women in the screening program and sections of the population where attendance rates are the lowest, should work to encourage greater participation. In Australia, the participation rate of Indigenous women to breast screening is 15% lower than the general populace and the mortality from breast cancer is two-fold higher.^{27,28} With breast density in Indigenous Australians described as predominantly adipose²⁹ and reported to be significantly lower compared with non-Indigenous women^{26,30} the current work suggests that Indigenous women would particularly benefit from early breast cancer detection and low false positives provided through engagement with the screening program in the recommended schedule.

We acknowledge the limitations of the current work including the potential selection bias in the test sets. While test cases were selected with the intent to maintain a consistent level of difficulty across all tests,¹⁶ the possibility that difficulty levels still varied between cases in LD and HD categories cannot be eliminated. Another critical limitation was the use of a single method for assessing mammographic density. While BI-RADS density measurement is a widely used method, it is visually and qualitatively assessed. Automated measurement tools currently available in the market would increase the accuracy and reproducibility of density measurements within the test sets.^{31,32} The current work reaffirms the beneficial effect of low mammographic density on reader performance, however the trend is not completely consistent suggesting that there may be other factors that affect performance such as radiologists' training and test set characteristics. Further work to identify these agents is required so a more precise understanding of the impact of mammographic density on performance is available.

Acknowledgments

The authors thank the BreastScreen Reader Assessment Strategy (BREAST) program for access to data. BREAST is a member of the Medical Image Optimisation and Perception Group (MIOPeG) at the University of Sydney and is funded by the Royal Australian and New

Zealand College of Radiologists, National Breast Cancer Foundation, Australian Department of Health, New Zealand Ministry of Health, and Cancer Institute NSW. The authors wish to thank Dr Robert Heard and for his assistance in the statistical analysis of the data, and Dr Phuong Dung Trieu for her assistance in data-collection.

Funding

My research project was not partially or fully sponsored by a sponsoring agency.

Conflicts of interest

Author declares that there is no conflict of interest.

References

- McCormack VA, Silva IDS. Breast density and parenchymal patterns as markers of breast cancer risk: A meta-analysis. *Cancer Epide Biomar*. 2006;15(6):1159–1169.
- Boyd NF, Martin LJ, Bronskill M, et al. Breast Tissue Composition and Susceptibility to Breast Cancer. *JNCI J Natl Cancer Inst*. 2010;102(16):1224–1237.
- Posso M, Louro J, Sanchez M, et al. Mammographic breast density: How it affects performance indicators in screening programmes? *Eur J Radiol*. 2019;110:81–87.
- Lee CI, Chen LE, Elmore JG. Risk-based Breast Cancer Screening: Implications of Breast Density. *Med Clin North Am*. 2017;101(4):725–741.
- Yeom YK, Chae EY, Kim HH, et al. Screening mammography for second breast cancers in women with history of early-stage breast cancer: factors and causes associated with non-detection. *BMC Med Imaging*. 2019;19(1):2.
- Boyd NFMDD, Guo HM, Martin LJP, et al. Mammographic Density and the Risk and Detection of Breast Cancer. *N Engl J Med*. 2007;356(3):227–236.
- Holland K, van Gils CH, Mann RM, et al. Quantification of masking risk in screening mammography with volumetric breast density maps. *Breast Cancer Res Treat*. 2017;162(3):541–548.
- Oiwa M, Endo T, Suda N, et al. Can quantitative evaluation of mammographic breast density, “volumetric measurement”, predict the masking risk with dense breast tissue? Investigation by comparison with subjective visual estimation by Japanese radiologists. *Breast Cancer*. 2019;26(3):349–358.
- Mandelson MT, Oestreicher N, Porter PL, et al. Breast Density as a Predictor of Mammographic Detection: Comparison of Interval- and Screen-Detected Cancers. *J Natl Cancer Inst*. 2000;92(13):1081–1087.
- Roubidoux MA, Bailey JE, Wray LA, et al. Invasive Cancers Detected after Breast Cancer Screening Yielded a Negative Result: Relationship of Mammographic Density to Tumor Prognostic Factors. *Radiology*. 2004;230(1):42–48.
- Kolb TM, Lichy J, Newhouse JH. Comparison of the Performance of Screening Mammography, Physical Examination, and Breast US and Evaluation of Factors that Influence Them: An Analysis of 27,825 Patient Evaluations. *Radiology*. 2002;225(1):165–175.
- Wanders JO, Holland K, Veldhuis WB, et al. Volumetric breast density affects performance of digital screening mammography. *Breast Cancer Res Treat*. 2017;162(1):95–103.
- Weigel S, Heindel W, Heidrich J, et al. Digital mammography screening: sensitivity of the programme dependent on breast density. *Eur Radiol*. 2017;27(7):2744–2751.

14. Al Mousa DS, Mello-Thoms C, Ryan EA, et al. Mammographic density and cancer detection: does digital imaging challenge our current understanding? *Acad Radiol*. 2014;21(11):1377–1385.
15. Brennan P, Lee W, Tapia K. Breast Screen Reader Assessment Strategy (BREAST): A Research Infrastructure with a Translational Objective. In: Samei E, Krupinski E, editors. *The Handbook of Medical Image Perception and Techniques*. 2nd ed. Cambridge: Cambridge University Press; 2018. p. 343–356.
16. Suleiman WI, Rawashdeh MA, Lewis SJ, et al. Impact of Breast Reader Assessment Strategy on mammographic radiologists' test reading performance. *J Med Imaging Radiat Oncol*. 2016;60(3):352–358.
17. BreastScreen Australia. *Breastscreen Australia National Accreditation Standards, October 2015*. Canberra ACT: Cancer Australia; 2016.
18. Samuel J Magny, Rachel Shikhman, Ana L Kepcke. Breast Imaging Reporting and Data System Atlas (BI-RADS (Atlas). Reston, VA: ACR; 2003.
19. McEntee M, Brennan P, Evanoff M, et al. Optimum ambient lighting conditions for the viewing of softcopy radiological images. *SPIE*. 2006.
20. Finder CA, White D, White E, et al. Breast Density as a Predictor of Mammographic Detection: Comparison of Interval- and Screen-Detected Cancers. *J Natl Cancer Inst*. 2000;92(13):1081–1087.
21. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and Combined Effects of Age, Breast Density, and Hormone Replacement Therapy Use on the Accuracy of Screening Mammography. *Ann Intern Med*. 2003;138(3):168–175.
22. Al Mousa DS, Ryan EA, Mello-Thoms C, et al. What effect does mammographic breast density have on lesion detection in digital mammography? *Clin Radiol*. 2014;69(4):333–341.
23. Pisano ED, Hendrick RE, Yaffe MJ, et al. Diagnostic Accuracy of Digital versus Film Mammography: Exploratory Analysis of Selected Population Subgroups in DMIST. *Radiology*. 2008;246(2):376–383.
24. Kerlikowske K, Hubbard RA, Miglioretti DL, et al. Comparative Effectiveness of Digital Versus Film-Screen Mammography in Community Practice in the United States: A Cohort Study. *Ann Intern Med*. 2011;155(8):493–502.
25. Sprague BL, Gangnon RE, Burt V, et al. Prevalence of mammographically dense breasts in the United States. *J Natl Cancer Inst*. 2014;106(10):255.
26. Tapia KA, Garvey G, McEntee MF, et al. Mammographic densities of Aboriginal and non-Aboriginal women living in Australia's Northern Territory. *Int J Public Health*. 2019;64(7):1085–1095.
27. Australian Institute of Health and Welfare. *BreastScreen Australia monitoring report 2018*. Canberra: AIHW; 2018. p. 109.
28. Australian Bureau of Statistics. *4727.0.55.001-Australian Aboriginal and Torres Strait Islander Health Survey: First Results, Australia, 2012-13*. Canberra, ACT: ABS; 2013.
29. Pape R, Spuur KM, Currie G, et al. Mammographic parenchymal patterns and breast cancer risk in New South Wales North Coast Aboriginal and Torres Strait Islander women. *J Med Radiat Sci*. 2016;63(2):81–88.
30. McLean K, Darcey E, Cadby G, et al. The distribution and determinants of mammographic density measures in Western Australian aboriginal women. *Breast cancer research*. 2019;21(1):33.
31. Kerlikowske K, Ma L, Scott CG, et al. Combining quantitative and qualitative breast density measures to assess breast cancer risk. *Breast cancer research*. 2017;19(1):97.
32. Astley SM, Harkness EF, Sergeant JC, et al. A comparison of five methods of measuring mammographic density: a case-control study. *Breast Cancer Res*. 2018;20(1):10.