Impact of prior mammograms on observer performance in screening mammography

Judith David Akwo

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

THE UNIVERSITY OF SYDNEY

Faculty of Medicine and Health

2025

Statement of originality

This is to certify that the content of this thesis is my own work completed under the University of Sydney Human Research Ethics Committee Protocol number 2023/101. I certify that the co-authors of the published pieces of work included in this thesis were my supervisors or assisted at various stages of the work including study design, data collection, and provided feedback on manuscript drafts. This thesis contains no material that has been previously published or written by another person except where acknowledged. I also certify that no part of this thesis has been submitted or is being considered for any other degree or purpose.

I certify that the intellectual content of this thesis is the product of my own work, and that all assistance received in preparing this thesis and all sources have been acknowledged.

I acknowledge that this thesis will be deposited in the University of Sydney Library if it is awarded a PhD degree. I understand that the Librarian or Head of the Faculty may provide a copy of this thesis to a third party for academic or research purposes.

Judith David Akwo

June 2025

Keywords

Prior mammograms, Previous mammograms, Screening mammography, Observer performance, Diagnostic performance, Radiologist, Radiographer

Supervisor's Statement

As the primary supervisor of Judith David Akwo, I certify that I consider this thesis "**Impact of prior mammograms on Observer performance**" suitable for examination.

Name: Tess Reynolds

Date: 25th June2025

Abstract

Aims: Early detection of breast cancer through screening mammography has been credited with 20 – 40% reduction in mortality from the disease. However, screening mammography misses some cancers and some women who have no cancer are thought to have cancer because of suspicious features identified on their mammograms. Strategies that improve the early detection of breast cancer without increasing the number of women incorrectly recalled for additional testing may improve and expand screening services. Previous screening mammograms provide an important baseline to assess whether abnormal features identified in screening mammograms from the current screening round are benign or represent malignancy. The few previous studies that have investigated the influence on prior mammograms on the efficacy of screening mammography interpretation were mostly based on film-screen technology, used older observer performance methodologies and did not examine factors that may confound the effect of prior mammograms on performance, and involved radiologists. Radiographers are also increasingly participating in mammography image interpretation, but the influence of prior mammograms on their diagnostic efficacy has not been explored.

This thesis addresses these limitations to provide informed knowledge of the influence of prior mammograms on the efficacy of screening mammography interpretation. First, it scopes the literature to examine the impact that previous mammograms had on diagnostic performance. Thereafter, three original studies that examined the impact of prior mammograms to radiologists, radiographers, and the interpretation of digital breast tomosynthesis (DBT) were conducted. The first study established the impact that access to prior mammograms had on accurate interpretation of screening mammograms and the characteristics of radiologists and breast compositions that benefit from prior mammograms. The second extended the methodology to radiographers and established the impact of providing access to the previous

screening mammograms on the interpretation of the current screening round by breast imaging radiographers. The third study examined the impact of prior digital mammograms on the detection and classification of different breast cancer lesion types.

Methods: The three original studies were conducted in three phases using cases from the BreastScreen Reader Assessment Strategy (BREAST) platform. In the first phase, eight radiologists independently interpreted 72 screening mammography cases containing 32 cancer cases in two reading sessions using the Royal Australian and New Zealand College of Radiologists' (RANZCR) classification: 1 (normal); 2 (benign); 3 (indeterminate); 4 (suspicious); 5 (highly suspicious). In the first reading session the current screening mammograms and the mammograms of these women acquired 2 to 4 years prior were available. In the second reading session the radiologists interpreted the same 72 cases using only the current mammograms of these women. A Paired T-Test was used to compare performance in terms of false positives, specificity, sensitivity, lesion sensitivity, Receiver Operating Characteristics (ROC) Curve, and Jacknife Alternative Free-Response Receiver Operating Characteristics (JAFROC) scores with and without prior mammograms. Independent sample T-test was used to compare the performance of groups of readers with different characteristics and the impact that breast density had on performance when prior mammograms were available. A relative risk analysis (RR) analysis was also performed to assess the probability of false positive and false negative outcomes when prior mammograms were available.

In the second phase, 13 radiographers interpreted two sets ("Prior" and "No prior") of mammography cases of the same 28 women containing nine cancer cases. The first reading session involved the interpretation of the "No prior test-set" (no previous mammograms available) while second readings session involved the interpretation of the "Prior test-set"

(mammograms from the most recent screening round and the previous mammograms of these women). The performance of these radiographers with and without prior mammograms were compared using metrics such as false positives, specificity, sensitivity, lesion sensitivity, Receiver Operating Characteristics (ROC) Curve, and Jacknife Alternative Free-Response Receiver Operating Characteristics (JAFROC).

The third phase involved the secondary analysis of the radiologists/breast physicians' data collected in the first and second phases of the project. For this phase, the marks assigned to all the lesions detected in first and second phases (spiculated lesion, discrete mass, calcification, non-specific density, architectural distortion, or stellate lesion) were analysed. Performance analysis in each of the test sets used in both phases was conducted for each lesion type. Performance analyses for each lesion type were conducted separately for radiologists/breast physicians and radiographers. For each of these phases, the performance of the readers for each lesion type when prior mammograms were available were compared to the performance of the same cohort of readers when prior mammograms were not available. Statistical significance was defined as $p \leq 0.05$.

Results: In the first phase involving radiologists, access to prior mammograms improved specificity in dense and non-dense breasts ($p \leq 0.01$) and reduced false positives ($p=0.01$) but had no effect on sensitivity ($p=0.37$), lesion sensitivity ($p=0.67$), ROC ($p=0.16$), and JAFROC ($p=0.24$). Access to prior mammogram also reduced the probability of false positives (RR=0.38; 95%CI:0.26-0.57; $p<0.0001$) without affecting the false negative rate (RR=1.14; 95%CI:0.88-1.49; $p=0.30$). The impact of prior mammograms on performance was not influenced by breast density or radiologists' characteristics.

In the second phase involving radiographers, specificity [81(range:58-95) vs. 60(range:37-79); p=0.002], ROC [91(range:80-99) vs. 82 (range:57-91); p=0.003], and JAFROC 87(range:73-99) vs. 79 (range:52-91); p=0.01] were higher when radiographers read with prior mammograms. False positives (p=0.002) were reduced when prior mammograms were available. No differences were observed between readings with and without prior mammograms in terms of sensitivity (p=0.70) and lesion sensitivity (p=0.82). Years qualified as a radiographer did not modify the influence of prior mammograms on specificity, ROC, and false positives. Years specialised as breast imaging radiographers slightly modified the influence of prior mammograms in radiographers with ≥25 years of experience but not those with <25 years of experience as breast imaging radiographers.

In the third phase, no differences were observed in radiologists' sensitivity when reading occurred with compared to without prior mammograms for architectural distortions (p=0.48), calcifications (p=0.85), discrete masses (p=0.45), and non-specific density (p=0.22). However, prior mammograms improved the classification of spiculated and stellate lesions (p=0.05). Specificity was significantly higher when prior mammograms were available (p<0.01 for all). No differences were observed in lesion sensitivity for all lesion types (p≥0.16). Diagnostic accuracy did not differ between reading with and without prior mammograms when calcifications, discrete masses, spiculated and stellate lesions, and non-specific density lesions were examined independently (p≥0.16). However, the availability of prior mammograms significantly improved diagnostic accuracy for architectural distortions (p=0.006). When the assessments of radiographers were considered, no differences were observed in sensitivity and lesion sensitivity with compared to without prior mammograms for all lesion types (p>0.05 for all). However, overall accuracy, which represents performance in correctly identifying the

presence or absence of each lesion type was significantly higher ($p \leq 0.002$) when prior mammograms were available to the radiographers

Conclusion: Access to prior mammograms improves the ability to discriminate between normal and abnormal mammograms and reduce false positive decisions without affecting the detection of breast cancer, regardless of the characteristics of the mammogram reader and breast density. Prior mammograms improve the detection of spiculated lesions but not the detection of other breast lesion types. Therefore, strategies to store, access and understand the cognitive effects of reference to previous screening mammograms is a useful strategy to improve performance in mammography interpretation and reduce false positive rates.

Preface

Breast cancer is the most common cancer and the leading cause of deaths in women worldwide. The establishment of screening programs to facilitate early detection of breast cancer has been credited with reduction in deaths from the disease. However, the interpretation of screening mammograms can be challenging and lead to about 30% of breast cancer not detected early or a significant proportion of women participating in screening mammography wrongly adjudged to have breast cancer. These challenges have necessitated the need for strategies to improve early detection while simultaneously reducing the number of women incorrectly recalled for additional assessment because of suspicious findings in their mammograms.

This thesis presents the outcomes of three original pieces of research that investigated the influence of prior mammograms of women participating in breast cancer screening on the accuracy of interpretation of mammograms and detection of different lesion types by radiologists and breast radiographers. It also highlights the impact that breast density and the practice-related characteristics of radiologists and breast radiographers have on the influence of prior mammograms on diagnostic efficacy. The findings highlight the relevance of prior screening examinations in improving the interpretation of women whose mammograms do not contain cancer thereby reducing the number of women incorrectly called back for additional testing. The outcomes have implications for policy and practice. First, the findings highlight the need for screening programs to store prior mammograms for future reference and for the establishment of national databases of screening mammograms to support one-stop breast clinics and to account for population mobility across states. Second, the findings can be used to support guidelines around the storage, retrieval, and reference to previous screening mammography examination when interpreting mammograms from the current screening round, and by extension, the relevance of previous imaging examinations in radiology.

Acknowledgement

I would like to express my gratitude to God for life, and for his grace upon me throughout my PhD journey. His blessings, and gift of strength, wisdom, and good health ensured the successful completion of my PhD studies.

I am immensely thankful to my supervisors whose wisdom, perseverance, guidance, and invaluable support put me on target to complete the pieces of work reported in this thesis. To Prof. Sarah Lewis, I am incredibly grateful to you for accepting to supervise my PhD project, your mentorship, and the wealth of knowledge you have impacted on me during the short period of working with you. Your passion, dedication, feedback, and support for me and my project have been outstanding and were catalysts to the outcomes recorded in my thesis. I would also like to convey my special gratitude to Dr Phoung Dun (Yun) Trieu for holding my hand each step of the way, from assembling the image dataset used for the project to creating the test-sets, supporting with participant recruitment and data collection through to extracting and curating the observer performance data and providing feedback on my statistical analyses. Thank you for always being available to provide clarification and to cheer on me during each presentation. I am also grateful to Dr Tess Reynolds for stepping in as lead supervisor following Prof. Lewis' departure from the University of Sydney. Your feedback and valuable insights into my PhD project also helped me get to the finish line. It has been an amazing experience working with you all.

The pieces of research reported in this thesis would not have been possible without the generous support of the BreastScreen Reader Assessment Strategy (BREAST) at the University of Sydney. BREAST is funded/supported by the Commonwealth Government Department of Health and Aged Care, with their images provided by BreastScreen Australia and the Cancer Institute New South Wales (CINSW). I would like to express my appreciation to BREAST and

to Dr Melissa Robinson, BREAST manager for her support with participant recruitment and data collection.

This research reported in this thesis was supported by the award of a Research Training Program scholarship to the PhD Candidate. I would like to thank the Australian Government for the RTP fee offset which enabled me to undertake and complete my PhD studies. Without the RTP fee offset, I would not have been able to enrol for a PhD at a university as prestigious as the University of Sydney. The scholarship has not only enabled me to complete a PhD degree but to generate an important body of research that has potential to support policy and practice around breast screening. It has also given me the opportunity to acquire research education and tools, which are crucial to a successful career in academia and research. For this, I will forever remain grateful.

I am eternally thankful to my beloved parents, David Akwo (Assistant Comptroller General of Immigration, mni Rtd) and Mrs Imelda Akwo for birthing me and your unconditional love, prayers, and faith in me. You make my existence easy and memorable and have continued to support my ambitions. To my parent in-laws, Peter Ekpo Usang (Assistant Superintendent of Police Rtd) and Mrs Grace Peter Usang, I am grateful for your love, support and prayers. Thank you for bringing my beloved husband to the world and raising him to be an amazing human being. I would like to acknowledge my dear siblings – Dr Violet Akwo Nelson, Barrister Sylvia Akwo, Miss Ruth Akwo, and Mr David Akwo Jnr for their prayers, support, and love. I would like to thank my adorable nephews (Jason and Jayden Nelson), niece (Alexis Nelson), and brother-in-law, Mr Ime Nelson for regularly checking in to ensure that things are going well. I would also like to appreciate my dear friends, Engineer Aniekeme and Dr Omolara Ukpong for their constant prayers and encouragement throughout my studies.

I have also been blessed with pleasurable distractions throughout my PhD studies– my lovely children: Atara, Elora, Kaira, and Zayne. Your beautiful smiles, peck kisses, kicks, and cuddles illuminated every moment and path of my PhD journey. I feel blessed being your mother. Finally, I would like to express my deep appreciation to my one in a billion husband, heartthrob, better half, king, crown, favourite person, best friend, forever partner and absolute love of my life – Prof. Ernest Ekpo for encouraging me to embark on this journey and equipping me with all the tools needed to successfully navigate every aspect of my studies. I am extremely grateful for his criticism, feedback, encouragement, and general support. Thank you for your kindness, love, perseverance, patience, and for taking care of me and the kids during this busy period of my life –you made everything easy. I could not have done marriage with anyone else but you. I remain thankful to God for making our parts to cross.

Artificial Intelligence Statement

No content produced by generative AI tools has been used as part of the research project or in the preparation of this thesis.

Author Attribution Statement

The following published papers are included in the thesis

- Chapter 2 has been published as "Akwo JD, Trieu PD, Lewis SJ (2023). Does the availability of prior mammograms improve radiologists' observer performance?- a scoping review. BJR Open. 2023; 5(1): 20230038. I searched the literature, extracted all the articles included in the review, and wrote the drafts of the manuscript.
- Chapter 3 has been published as "Akwo JD, Trieu PD, Barron M, Reynolds T, Lewis SJ (2024). Access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance. Clinical Radiology; 79(12): e1549-e1556". I co-designed the study with my supervisors (Phuong Dung (Yun) Trieu, Melissa L. Barron, Tess Reynold, and Sarah J. Lewis) who supported data collection and extraction from the BREAST platform. I conducted the analysis and interpretation of the data and wrote the drafts of the manuscript.
- Chapter 4 is published as "Akwo JD, Trieu PD, Barron M, Reynolds T, Lewis SJ (2024). Does access to prior mammograms improve the performance of radiographers in interpreting screening mammograms? Radiography; 31(1): 247-253. I co-designed the study with my supervisors (Phuong Dung (Yun) Trieu, Melissa L. Barron, Tess Reynold, and Sarah J. Lewis) who supported data collection and extraction from the BREAST platform. I conducted the analysis and interpretation of the data and wrote the drafts of the manuscript.

- Chapter 5 has been published as "Akwo JD, Trieu PD, Barron M, Reynolds T, Lewis SJ (2025). Impact of prior mammograms on Radiographers and Radiologists' detection of different breast cancer lesion types" in the Journal of Medical Radiation Sciences. 0:1–7. https://doi.org/10.1002/jmrs.70015
- I co-designed the study with my supervisors (Phuong Dung (Yun) Trieu, Melissa L. Barron, Tess Reynold, and Sarah J. Lewis) who supported data collection and extraction from the BREAST platform. I conducted the analysis and interpretation of the data, wrote the drafts of the manuscript, and served as corresponding author.

In addition to the authorship attribution statements above, I am the corresponding author of all publications included in the thesis, and permission to include these publications has been granted by all co-authors.

Judith David Akwo

Date: 25th June 2025

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Name: Tess Reynolds

Date: 25th June 2025

Publications and presentations

The current work has been published and presented in the following forums:

Published Peer-Reviewed Journal Articles

- Akwo JD, Trieu PD, Lewis SJ (2023). Does the availability of prior mammograms improve radiologists' observer performance?-a scoping review. BJR Open. 2023; 5(1): 20230038. 10.1259/bjro.20230038
- Akwo JD, Trieu PD, Robinson M, Reynolds T, Lewis SJ (2024). Access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance. Clinical Radiology; 79(12): e1549-e1556. DOI: https://doi.org/10.1016/j.crad.2024.09.007
- Akwo JD, Trieu PD, Robinson M, Reynolds T, Lewis SJ (2024). Does access to prior mammograms improve the performance of radiographers in interpreting screening mammograms? Radiography; 31(1): 247-253. Doi: https://doi.org/10.1016/j.radi.2024.11.025.
- Akwo JD, Trieu PD, Barron M, Reynolds T, Lewis SJ (2025). Impact of prior mammograms on Radiographers and Radiologists' detection of different breast cancer lesion types. Journal of Medical Radiation Sciences. 0:1-7. https://doi.org/10.1002/jmrs.70015

Oral and Poster Presentations

- Does the availability of prior mammograms impact upon radiologists' performance in mammography interpretation? Symposium Mammographicum, Bournemouth 9 -11 July 2025. [Poster – accepted]
- Akwo J.D, Trieu P.D, Barron M.L, Reynolds T, Lewis S.J. Influence of prior mammograms on radiologists' performance and the detection of different lesion types. 14th General Breast Imaging Meeting, 31 March-4 April 2025. [Poster].
- Akwo JD, Trieu PD, Lewis SJ (2023). Does the availability of prior mammograms improve radiologists' observer performance? -a scoping review: An Overview. The Medical Image Optimisation and Perception Group (MIOPeG) Lab Meeting, The University of Sydney. May 2023. [Oral].
- Akwo JD, Trieu PD, Lewis SJ (2024). Does the availability of prior mammograms improve radiologists' observer performance? -a scoping review. ImageX Institute, The University of Sydney. May 2024. [Oral].
- Akwo J.D, Trieu P.D, Barron M.L, Reynolds T, Lewis S.J. Impact of prior mammograms on observer performance. The Medical Image Optimisation and Perception Group (MIOPeG) Lab Meeting, The University of Sydney. March 2025. [Oral].

Table of Contents

Glossary of terms

Abnormal Interpretation Rate	The number of abnormal findings that require additional follow-up or the number of mammograms with abnormal final interpretation.
Algebraic reconstruction technique	An iterative reconstruction technique used to reconstruct radiological images from a series of angular projections.
Artificial Intelligence	A computer program or software that performs tasks requiring human intelligence.
Benign	Harmless tumours.
Biopsy rate	The proportion of females had been recalled at screening and had a biopsy test.
Breast Imaging Reporting and Data Systems	Quality assurance and risk assessment tool for reporting breast imaging examinations.
Breast density	The proportion of the breast comprising of fibroglandular relative to fatty tissues.
Breast physician	Medical doctor who specialises in diagnosing and treating breast diseases.
BreastScreen Australia	A joint initiative of the Australian and state and territory governments that provide breast screening to Australian women.
BREAST	Breastscreen REader Assessment STrategy Educational self-assessment and training scheme for breast screening readers in Australia
BreastScreen Australia National Accreditation Standards	Guidelines to evaluate and maintain the quality of care offered to women participating in breast cancer screening in BreastScreen Australia services.
Cancer detection rate	The number of cancers detected per 1000 women screened.
Cancer Institute New South Wales	A New South Wales agency established to lessen the impact of breast cancer in New South Wales.
Dense breast	Breast composed of high fibroglandular tissue relative to fat.

Continuous Professional Development	Learning activities that build practical job applicable skills used to maintain professional certification.
Craniocaudal	A standard projection used in mammography where the X-ray tube is above the breast with the detector underneath the breast.
Current mammograms	Mammograms obtained in the most recent screening round.
Detective quantum efficiency	The efficiency of an imaging detector to convert X-ray energy into a usable signal.
Digital breast tomosynthesis	A type of mammographic equipment that uses low-dose X-rays to produce pseudo three-dimensional images of the breast.
Digital mammography	A type of mammographic equipment that uses electronic detectors to covert X-rays into digital images of the breast.
False negatives	The number of cases with cancer or positive cases that were incorrectly reported as negative.
False positives	The number of negative cases that were incorrectly reported as positive.
False positive rate	The proportion of negative cases that were incorrectly reported as positive.
Filtered back projection	A reconstruction technique that uses a convolution filter to remove blurring in radiological images.
Iterative reconstruction	Reconstruction techniques that use iterative algorithms that compare real-time values to assumed images and continuously adjust until both images are in concordance.
Jacknife Alternative Free-Response Receiver Operating Characteristics	A measure of performance in localising a lesion and simultaneously rating the level of malignancy vs identifying correct normal cases.
Laboratory effect	Changes in observer behaviour due to observer expectation and knowledge that they are being tested.
Lesion localised fraction	Fraction of lesions whose exact location within the breast are correctly established.

Lesion sensitivity	The number of cancer lesions correctly localised divided by the total number of lesions in the test-set or population whose mammograms are interpreted.
Local health districts (LHD)	Entities responsible for managing public hospitals and healthcare services.
Location receiver operating characteristic curve	A performance measure that quantifies readers ability to detect and locate an abnormality on the mammogram.
Malignant	Cancer lesions that have potential to spread and invade surrounding tissues.
Mammography Quality Assurance Program	A framework for assessing the quality mammography screening services to ensure that they meet the required standard.
Mammography quality standard act	An Act enacted by the United States congress to regulate the quality of care in mammography.
Mediolateral oblique	A standard projection used in mammography where the X-ray tube captures breast images from the medial aspect of the breast with the detector on the lateral part of the breast.
National Health Service Breast Screening Programme	A national breast cancer screening programme for early detection of breast cancer.
National Safety and Quality Medical Imaging (NSQMI) Standards	The Australian National Safety and quality standards for imaging providers to protect the public from radiation harm and improve medical image service quality.
Non-dense breast	Breasts comprising mostly of fatty tissue.
"No Prior test-set"	Test-set comprising only mammograms from the most recent screening round.
PERsonal PerFORmance in Mammographic Screening	Educational self-assessment and training scheme for breast screening readers in the United Kingdom.
Positive predictive value	The proportions of positive results that are true positive.

Picture archival and communication system	Software imaging technology for storing, retrieving, and sharing medical images securely.
Prior test-set	Test-set comprising both mammograms from the most recent and previous screening rounds.
Prior mammograms	Mammograms from the previous screening rounds
Radiographer	Healthcare personnel who use radiological equipment to acquires radiological images.
Radiologist	A medical practitioner who interprets radiological images
Royal Australian and New Zealand College of Radiologists	A professional organisation for the promotion of the science and practice of the medical specialties of clinical radiology (diagnostic and interventional radiology) and radiation oncology in Australia and New Zealand.
Recall rate	The rate of females recalled for follow-up imaging due to suspicious of cancer.
Receiver Operating Characteristics Curve	A measure of performance in correctly classifying normal and abnormal cases
Region of interest figure of merit (ROI FOM)	The empirical probability that a cancer containing region of interest is rated higher than a normal region of interest
Relative risk	A statistical test that compares the probability of an event occurring in an exposed relative to an unexposed group to test the strength of the association between exposure and outcome.
Sensitivity	The rate of true-positive (cancer) cases that were correctly identified
Simultaneous algebraic reconstruction technique	An iterative reconstruction technique used to reconstruct radiological images when projection data is limited.
Simultaneous iterative reconstruction technique	A technique that reconstructs images by comparing real-time values to assumed images and continuously adjust the image estimate by iteratively minimising the difference between the measured and estimated projections.
Satisfaction of search	premature termination of search due to identification and incorrect classification of features that mimic breast cancer

Specificity	The rate of true-negative (normal) cases that were correctly identified.
Surveillance, Epidemiology, and End Results	An American National Cancer Institute that collects published data on cancer incidence, survival, and mortality from various population-based cancer registries across the United States of America.
True positive	The number of cases with cancer that were correctly classified as positive.
True negative	The number of cases with no cancer that were correctly classified as negative.
World Health Organisation	An agency for the United Nations that coordinates response to internal public health issues and emergencies.

List of abbreviations

ART	Algebraic Reconstruction Technique
ACS	American Cancer Society
a-Se	amorphous Selenium
a-Si	amorphous Silicon
AI	Artificial Intelligence
BI-RADS	Breast Imaging Reporting and Data Systems
BRCA	Breast Cancer
BREAST	BreastScreen Reader Assessment STrategy
BD	Breast density
BSA	BreastScreen Australia
NAS	National Accreditation Standards
CDR	Cancer detection rate
CINSW	Cancer Institute New South Wales
CPD	Continuous Professional Development
CC	Craniocaudal
DQE	Detective Quantum Efficiency
DBT	Digital Breast Tomosynthesis
DM	Digital Mammography
EUSOBI	European Society of Breast Imaging
FN	False Negatives
FP	False Positives

FPR	False Positive Rate
FBP	Filtered Back Projection
IARC	Internation Agency for Research in Cancer
IR	Iterative Reconstruction
JAFROC	Jacknife Alternative Free-Response Receiver Operating Characteristics
FOM	Figure-of-Merit
LLF	Lesion Localised Fraction
LHDs	Local health districts
LROC	location receiver operating characteristic curve
L.Sen	Lesion sensitivity
MQAP	Mammographic Quality Assurance Program
MQSA	mammography Quality Standard Act
MLO	Mediolateral oblique
Mo-Rh	Molybdenum-Rhodium
MIOPeG	Medical Image Optimisation and Perception Group
MRI	Magnetic Resonance Imaging
NHBSP	National Health Service Breast Screening Programme
NSQMI	National Safety and Quality Medical Imaging Standards
PERFORMS	PERsonal PerFORmance in Mammographic Screening
PACS	picture archival and communication system
RANZCR	Royal Australian And New Zealand College of Radiologists
RR	Recall Rate
ROC Curve	Receiver Operating Characteristic Curve

AUC	Area Under the Curve
ROI	Region of interest
RR	Relative risk
Rh-Rh	Rhodium-Rhodium
RTP	Research Training Program
SART	Simultaneous Algebraic Reconstruction Technique
SIRT	Simultaneous Iterative Reconstruction Technique
SOS	Satisfaction Of Search
SEER	Surveillance, Epidemiology, and End Results
TP	True positive
TN	True negative
UK	United Kingdom
USA	United States of America
WHO	World Health Organisation
W-Rh	Tungsten-Rhodium

CHAPTER 1

INTRODUCTION

Breast cancer and screening mammography

Cancer is a global health challenge. The World Health Organisation (WHO) 2023 statistics ranked cancer as the top causal factor for deaths worldwide. Estimates indicate that 1 in 6 deaths worldwide is caused by cancer, and deaths are more common in patients with advanced stage cancer (1, 2). Cancer represents growth of abnormal cells in tissues and organs, with potential to spread beyond their boundaries and alter the body's physiological function. The International Agency for Research on Cancer (IARC) believe that cancer occurs because of the interaction between genetic signatures and physical, biological, and chemical carcinogens (3). Examples of physical carcinogens include ionising radiation from radon and medical imaging, and ultraviolet rays (4-6). Common chemical carcinogens include alcohol, asbestos, tobacco smoke, and substances that can be found in food (aflatoxin) and water (arsenic) (5, 7, 8). Genetic and lifestyle factors also increase the risk of cancer and account for the population and geographical variation in cancer incidence (5, 9, 10). For example, alcohol consumption, tobacco smoking, physical inactivity, and dieting are leading risk factors for cancer in the United States, Australia, and Europe (9, 10). In developing low-resource countries, chronic viral infections have been identified as the major determinants of cancer (11). The age-standardised incidence of cancer varies from 452.4 cases per 100,000 people in Australia to 40.4 cases per 100,000 people in Niger Republic (12). Cancer estimates from low-resource countries such as Niger Republic and the Gambia may not necessarily represent the number of cases per 100,000 population due to the absence of tailored screening programs (12). Conversely, cancer mortality rates are lower in developed countries with established screening

programs and higher in resource-poor countries. Prior to 2020, lung cancer was considered the most common cancer and leading cause of cancer deaths (1). However, recent data show that there are2.26 million new cases compared to 2.21 and 1.93 million new cases of lung and colorectal cancers respectively (1).

Breast cancer is a significant global health issue, representing 12.5% of new cases of cancer annually and 25.8% of new cases of female cancers (13, 14). However, like all cancers, the incidence rates vary across the world. The age-standardised rates of breast cancer vary from 113.2 per 100,000 population to 17.4 per 100,000 population (15). The top five countries with the highest cancer rates are Belgium, The Netherlands, Luxemburg, France, and France New Caledonia. Lower cancer rates have been reported for low-income countries such as Niger Republic and The Gambia. In 2020, it was estimated that there were 685,000 deaths globally from breast cancer (15, 16). Australia ranks 7th globally in the age-standardised breast cancer incidence chart, with 96.0 new cases per 100,000 population. Estimated cancer rate for Australian women was 20,428 in 2022, representing 12.7% of all cancers, and 3,214 Australian women were estimated to die from breast cancer (17). Fortunately, the 5-year survival rate of breast cancer has improved, and the death rates have reduced considerably (15, 18, 19). For example, 92 out of 100 women diagnosed with breast cancer survive for at least five years. Factors such as better awareness of risk factors and prevention strategies, early detection, and better therapeutic interventions have been credited with improved survival from breast cancer (20-22).

The effectiveness of treatment interventions depends on the stage at which the cancer is detected. The Surveillance, Epidemiology, and End Results (SEER) evidence demonstrate that localised breast cancer had the highest 5-year survival rate (99%). Regional disease has a 5-year survival rate of 86%, with the survival rate dropping drastically to 30% for cancers with

distant metastasis (23). The SEER evidence may explain the geographical and population variation in breast cancer survival and mortality rates, particularly between high- and low-income countries (23, 24). For example, countries with established screening programs such as Australia, USA, and many European countries can detect many breast cancers at early stages where treatment is more likely to be successful. Underserved countries such as those in Sub-Sahara Africa and South Asia demonstrate lower cancer incidence but high mortality rates due to the lack of screening programs (23, 25). Therefore, early detection is key to reducing deaths from the disease.

To ensure early detection and treatment of breast cancer, many high-income countries including Australia, USA, and European countries such as Netherlands and Denmark have established organised breast cancer screening programs (26-28). Mammography is the standard imaging tool for breast cancer screening, with digital breast tomosynthesis (DBT) also becoming more common in the last 5 years. The objective of screening mammography is to detect small-sized localised breast cancers in asymptomatic women. To achieve this objective, breast cancer screening programmes undertake five major tasks: 1) provide education and awareness about breast cancer and the relevance of early detection, and inviting eligible women for free breast screening; 2) provide easily accessible screening facilities including mobile screening vans to increase uptake of the screening service; 3) training and continuous professional development for radiographers and radiologists to improve their competence in acquiring and interpreting the screening images. Screening programs also undertake quality assurance and quality control audits to improve the effectiveness of screening the program; 5) Provide a medical image reporting system to ensure that women whose mammograms show features suggestive of breast cancer are appropriately recalled to assessment clinics and assessed. Thus, breast screening programmes play a crucial role in reducing mortality from

breast cancer from increasing women's uptake of screening practices to early detection through to assessment of women whose mammograms show features of breast cancer.

Age is the major eligibility criterion for invitation to participate in breast screening. However, the age recommendation for screening varies across international screening services. For example, the European Commission Initiative for Breast Cancer Screening and Diagnosis recommends the screening of women aged 45 to 74 years every two or three years (29). The Australian and US breast cancer screening programmes target women aged 50 to 74 years every two years (30, 31). It should be noted that the age criterion is for women at average risk of developing breast cancer. These include women with no personal or strong family history of breast cancer, not exposed to ionising radiation before the age of 30 years, and those with no breast cancer (BRCA) gene mutations. Women with these risk factors are at higher-than-average risk of breast cancer. The American Cancer Society (ACS) recommends optional annual screening digital mammogram for women aged 40 – 44 years and for women aged 45 – 54 years to get annual screening mammograms. The ACS also recommend biennial screening mammography for women aged 55 years and older (32, 33). Furthermore, the ACS recommends breast magnetic resonance imaging (MRI) for women with 20% or higher risk of breast cancer (33). It is important to note that while digital mammography (DM) is the recommended tool for breast cancer screening, particularly in women at average risk of breast cancer, missed cancer rates can be as high as 30% for some cases (34, 35). Consequently, DBT with advances in capturing multiple slices of the breast from different angles has been introduced for better assessment of lesion morphology. DBT is mostly used as an adjunct to mammography and assessment tool, but some practices in the USA have implemented this modality as a primary screening tool. Recently, the European Society of Breast Imaging (EUSOBI) recommended breast MRI for women with extremely dense breasts to detect cancers

that may have been missed on mammograms (36). In Australia, DM is still the preferred screening tool regardless of a woman's risk factor or breast density, but some practices use both DM and DBT. These two X-ray-based breast imaging tools have evolved since their introduction into the clinical setting and are discussed below.

Mammography: historical overview

The discovery of X-rays in 1895 provided opportunities to create images of the human tissue for screening and diagnostic purposes (37). The application of X-rays for breast imaging was first explored using mastectomies in 1913, where it was found to be useful for detecting breast calcifications and distinguishing between benign and malignant lesions (38). About 17 years later, general X-ray equipment was used to acquire images of actual women's breast and correctly detected cancer in 93% of these women (39). This led to the development of the first dedicated equipment for breast imaging by Gershon-Cohen in 1937 (40, 41).

Mammography uses low energy X-ray photons to acquire images of the breast. Photon energies used in mammography vary depending on the material used as the target-filter of the mammography X-ray tube. These energies are typically between 17.5 and 19.6 keV for Molybdenum (Mo) targets and 20.2 and 22.7 keV for Rhodium (Rh) targets (41). However, images produced by the mammographic system developed in 1937 were of poor quality and research into ways to optimise image quality began shortly thereafter (40). In 1950, it was discovered that compressing the breast reduced superimposition of breast tissue and improved image quality. Consequently, a breast compression technique was developed by Raul Leborgne (42, 43), which enabled the early clinical trials of mammography equipment as a breast cancer screening tool in 1956. However, it was until 1966 that mammography became widely used as

a breast cancer screening tool following a large-scale randomised controlled trial (44), which showed that screening mammography reduce mortality and improved treatment outcomes.

Early mammographic systems were based on non-screen technology (X-ray film and non-screen cassette), but this was replaced in 1967 with screen-film systems (45). The acquisition technology of screen-film systems involved an X-ray film enclosed within a cassette containing an intensifying screen and the X-ray equipment. When the breast is exposed to radiation, the X-ray photons are converted to light by the intensifying screen, creating a latent image on the X-ray film. The film is then passed through a developer solution to convert the silver bromide in the emulsion layer of the film to metallic silver. The developer creates different degrees of opacities on the X-ray film depending on the attenuating properties of the tissues traversed by the X-ray photons. Once the film is developed, it is then rinsed and passed through - a fixing solution to remove the excess silver halide and permanently fix the image to the film. The final processing step involves the washing and drying of the film before it can be interpreted by radiologists. With film-screen technology, optical density and contrast on the film are influenced by the amount of radiation exposure and processing (45). Studies have shown that screen-film mammography has low contrast resolution (which limit the visibility of image details), narrow dynamic range (exposure latitude), higher image noise, and is susceptible to film processing artifacts (46). The quality of the film-screen mammograms is further limited by exposure and processing parameters. To overcome these limitations, film-screen technology has been replaced with digital systems in many developed countries since the early 2000s and evidence shows that digital mammography (DM) detected more cancers than screen-film systems (47, 48).

Digital mammography (DM)

DM has become the frontline imaging tool for breast cancer screening in countries with organised screening programmes. In DM systems, the acquisition and display components are separate (Figure 1), and the images produced are stored in a picture archival and communication system (PACS). The acquisition system comprises of the X-ray tube and a detector usually linked to each other (49, 50). DM uses a motionless X-ray tube to obtain images of the breast. Technological innovations in the DM tube development have led to the combination of different target-filter materials by manufacturers: target/filter combinations (including Rh-Rh and W-Rh) (49, 50). These dual target materials allow for the energy of the X-ray beam used in DM to be optimised for different breast compositions. Current dual target systems with higher atomic numbers provide a harder X-ray beam, resulting in reduced image contrast. Similarly, digital detectors with higher detective quantum efficiency (DQE) are used in DM to balance image noise and spatial resolution. Available detector systems include flat-panel systems such as Charge-Couple Device, Scintillator/amorphous Silicon (a-Si), amorphous Selenium (a-Se), and Computed Radiography digital mammography systems (49). These detectors have a wide dynamic range in the region of ~1000:1, significantly higher than the latitude of screen-film mammography (40:1), allowing for post-processing to improve image quality (49, 50).

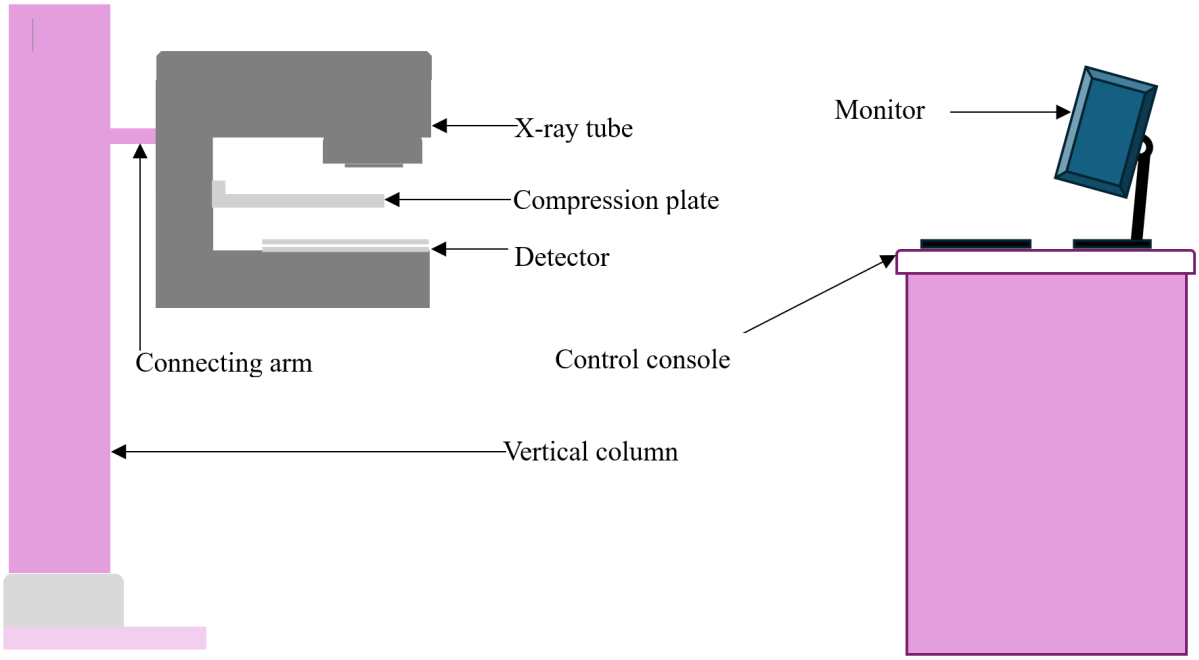


Figure 1: Digital mammography system configuration

Image formation in DM involves two processes: sampling and quantisation. When the X-ray photons emitted from the X-ray tube interact with the breast, their energy is attenuated (intensity is reduced) due to absorption and scattering. The transmitted photons are detected by the detector and converted to signal or light photons depending on whether the detector is a direct or indirect conversion system. The signal is sampled so that the information contained in the spatial signal is representative of the breast tissue traversed. The signal is then amplified and quantised (digitisation of the signal), allowing the information to be stored as picture elements (Pixel) and displayed as shades of Gray (bits of intensities) (50). Thus, the spatial resolution of DM images depends on the number of pixels per unit length of the image, while the intensity resolution and contrast are influenced by the bit dept (the number of bits per pixel).

Opportunities available in DM to change the signal sampling rate and the quantisation level allow post-processing of DM images to improve quality. Different image post-processing

techniques can be applied to digital mammograms to improve visualisation of anatomical details and the detection of breast cancer. These technological improvements have helped to overcome some of the limitations encountered in screen-film mammography. However, because the X-ray tube and imaging plate are stationary during screening, breast tissues along the path of the X-ray beam superimpose each other in the mammographic image produced. Also, anatomical information in the digital image represents the average of pixel information across the full thickness of the breast (50, 51). Solid cancers and calcifications have higher attenuation coefficient than normal breast tissue and appear brighter than both fibroglandular and fatty tissue. This attenuation differences allow radiologists to detect breast cancers in mammograms; however, some cancers have similar or low attenuation properties to fibroglandular tissue and may become invisible if contained within these tissues. The overlay of breast tissue and averaging of pixel information in two-dimensional images digital mammography cause anatomical noise, which may conceal cancer lesions (35, 52); these two phenomena limit the ability of digital mammography to detect cancer, particularly in women with dense breasts. To overcomes these challenges, imaging tools capable of producing three-dimensional images such as digital breast tomosynthesis were proposed to supplement mammography.

Digital breast tomosynthesis

Tomographic imaging was first proposed in the early 1930s (53, 54); however, it was not until the 1990s that tomosynthesis equipment was developed for breast imaging (54). Digital breast tomosynthesis (DBT), is an incomplete-angle tomographic X-ray imaging technology developed to reduce the superposition of breast tissue on X-ray images. DBT uses a rotating X-ray tube to obtain pseudo three-dimensional images of the compressed breast. During image acquisition, the X-ray tube undertakes angular rotations forming an arc-shaped motion across

the breast. During these angular rotations, multiple projection images of the breast are acquired at different depths across the breast (55, 56). While the principle of image formation is similar, the degree of X-ray tube motion, acquisition method, acquisition geometry, scan duration, number of projections, detector technology, and reconstruction algorithms vary across manufacturers. The acquisition of pseudo three-dimensional images is illustrated in Figure 2. In Figure 2a, tissues across the full thickness of the breast overlap at the center (see view 2). As the X-ray tube moves across the breast, these tissues are separated geometrically proportional to the shift of the tube angle (views 1 and 3). The tube movement ensures that projection images were acquired at different angles of the breast allowing tissues at different depths to be viewed separately in the reconstructed images (Figure 2B). The wider the range of tube motion, the better the tomographic information and separation of breast tissue (56)

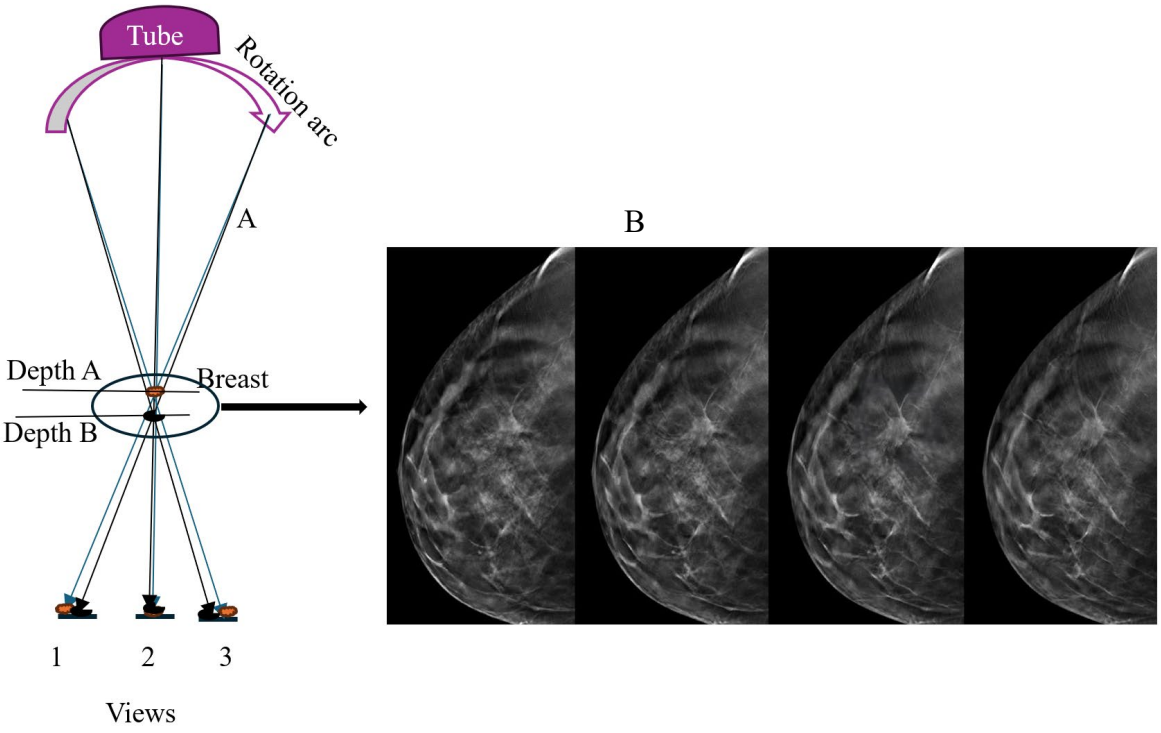


Figure 2: Principles of image acquisition in digital breast tomosynthesis showing image acquisition and separation of breast tissues as the X-ray tube moves through the compressed breast (Figure 2A) and how the acquired images are displayed (Figure 2 B).

All the projection images acquired from different breast depths are combined during image reconstruction. Reconstruction algorithms vary across vendors and include filtered back projection (FBP) and iterative reconstruction (IR) techniques such as simultaneous iterative reconstruction technique (SIRT), algebraic reconstruction technique (ART), and simultaneous algebraic reconstruction technique (SART) (57, 58). The reconstruction algorithms produce thin slices across different breast depths (Figure 2c). Breast tissues within the projected plane (in-focus) are enhanced to produce image slices while tissues which are not in the projected plane (out-of-focus) are blurred. Most of the algorithms produce in-focus slices of 1mm thickness at 1mm intervals; however, the slice thickness may vary between 0.5 – 1mm depending on the manufacturer (55, 56). The acquisition of multiple image slices allows breast images acquired by DBT to be viewed like the leaflet of a notebook. Thus, radiologists scroll through a stack of images to visualise different layers of the breast during image interpretation. This slice-by-slice assessment of breast tissue helps to overcome the challenges of tissue overlap and improves the visualisation of breast anatomy, and the detection and discrimination of breast lesions in dense breasts. Once the DBT acquisition creates the three-dimensional dataset, these can be projected onto a two-dimensional plane to generate two-dimensional images like digital mammography images (56, 59) These synthetic two-dimensional images such as C-view by Hologic, are generated without additional exposure to radiation. These synthetic images demonstrate better contrast and lesion conspicuity but have higher noise and lower resolution relative to DM images (60). DBT was initially used as an adjunct to digital mammography; however, due to newer systems synthesising two-dimensional images from the DBT acquisition and additional benefits of this technology, some screening services in Europe and the USA have recently implemented DBT as a screening tool (61, 62). In Australia, DBT is mostly used as an adjunct to digital mammography and for the assessment of women recalled

at screening. Therefore, it is important that Australian studies assessing the impact of prior mammograms on performance in Australia focus on the technology currently used in the Australian breast cancer screening service.

Breast cancer screening and image presentation

As described previously, in countries with organised screening programmes, eligible women are invited to undergo screening mammography at set intervals (30, 31). Screening mammography involves the mammographic imaging of breasts of asymptomatic women. The woman to be screened is positioned in front of the mammography equipment with the breast being screened placed between the compression plates. Thereafter, compression is applied to spread the breast tissue over the detector and reduce tissue superimposition. Compression improves the quality of the image and enhances the edges of the breast on the image. Usually, two standard projections of each breast are taken: craniocaudal (CC) view, where the X-ray tube is position above the breast with the detector under it. To optimise the CC view, the external portion of the breast is often rotated laterally to demonstrate the posterior edge of the pectoral muscle, retromammary space, and the nipple in profile. For the Mediolateral oblique view (MLO), the patient is positioned to face the mammography equipment with arm on the side being imaged moved away from the body. The breast is uniformly compressed to reduce skinfolds around the axilla. The X-ray tube and detector are rotated so that they are both at 45° to the breast, albeit the angle can vary between 40° and 60° depending on the woman's condition (Figure 3).

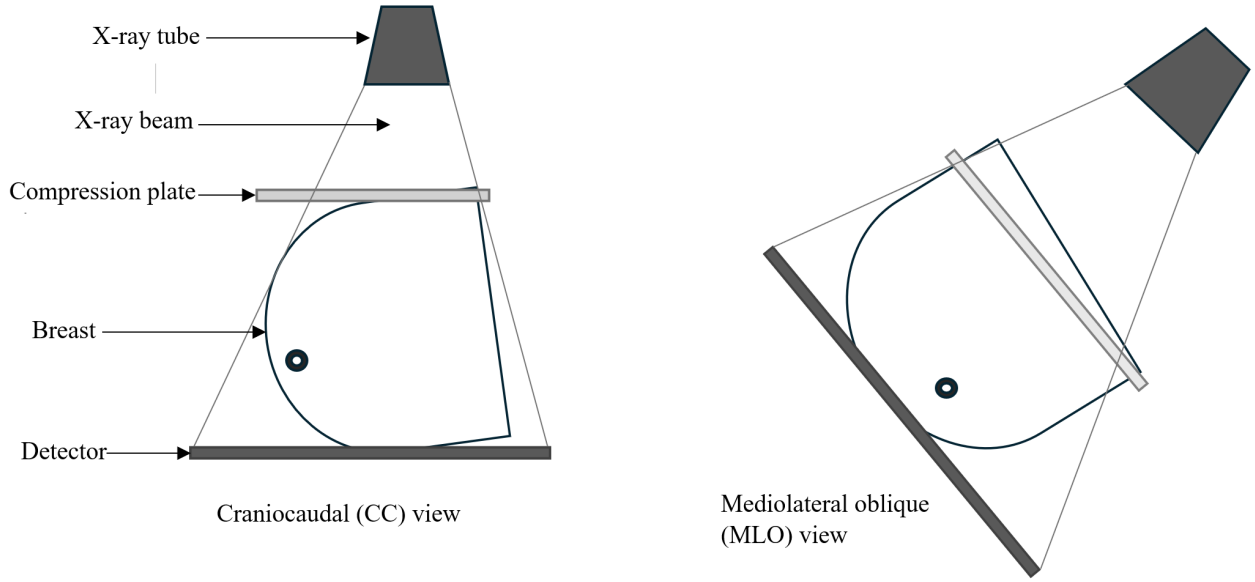


Figure 3: Illustration of patient positioning for the craniocaudal and mediolateral oblique views.

The mammograms acquired are displayed on digital monitors for interpretation under optimised ambient lighting conditions usually between 12-40 lux. The mammography quality standard act (MQSA) and other quality guidelines recommended that the mammograms should be displayed on a primary monitor with at least 5-megapixel resolution (63). Screening programmes require women participating in screening to re-attend screening every 1-3 years depending on the country's screening practices. This practice ensures that women have mammograms from many screening rounds, allowing radiologists access to prior mammograms of many women participating in breast cancer screening. With the digitisation of the screening process, it has become possible to display the prior mammograms of these women together with mammograms or DBT images from the current screening round. The simultaneous display of current and prior mammograms is to allow assessment of changes in the breast over time that may indicate malignancy (Figure 4). However, the large data size and

volume of digital images produced require large space for storage. Limited storage space sometimes challenges the storage, retrieval, and display of prior images alongside the current mammograms. It should also be noted that not all women have prior mammograms, and some women will have prior screening rounds from other geographical areas (such as across states, nations or have been imaged in private clinics). Therefore, not all women have screening mammograms from the current and previous rounds displayed simultaneously. There is contention around the relevance of prior mammograms to the outcomes of screening programmes (64). This thesis will address some of these issues.

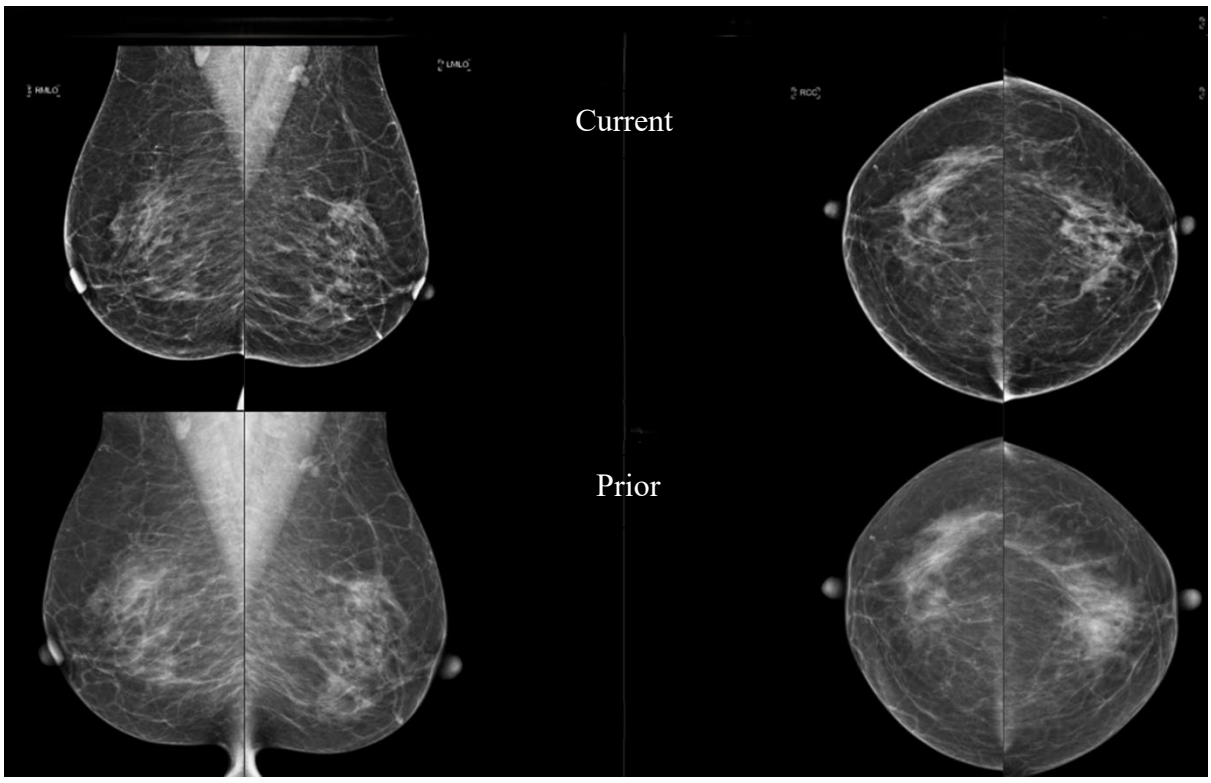


Figure 4: Current and prior mammograms displayed simultaneously.

Interpretation of screening mammograms

The images produced by both DM and DBT are often reviewed by radiologists and other screen readers (such as breast physicians in Australia, and advanced practice diagnostic radiographers in the UK) to visualise and interpret the features displayed in the images. The visual screening interpretation process involves search, perception, and evaluation of information perceived to make a diagnostic decision (35, 65, 66). Radiology search refers to the process of systematically scanning the different areas of the image to visualise and gather information necessary for the detection of abnormalities in radiological images. Search patterns differ between radiologists of different experience and workload characteristics (35, 65, 67). Non-visualisation of malignant features in mammograms may be due to inadequate sampling of mammograms (68) or premature termination of search due to identification and incorrect classification of features that mimic breast cancer, an error often referred to as "satisfaction of search (SOS)" (69, 70). Inadequate search and SOS have been estimated to be responsible for 42% of errors in the interpretation of screening mammograms (71, 72). Perception is the recognition and understanding of the significance of features identified. It involves the classification of the features as normal or abnormal, and understanding the severity of abnormal image presentations. Failure to perceive abnormal features can lead to cancer being missed in mammograms. Conversely, incorrect perception of normal image features may lead to false positive diagnosis. Errors due to faulty perception contribute to 60-80% of errors in the interpretation of radiological images (73-75). Diagnostic decision-making involves intuitive and analytical thought processes of using the image features identified to make a diagnostic decision.

Different disease processes affect the breast including inflammation, cysts, hormone-related breast discharge, and benign and malignant lesions (76, 77). Of these, breast malignancy is the breast disease commonly responsible for most breast-related mortality and morbidity. Therefore, the task of radiologists interpreting screening mammograms is to detect abnormal findings related to breast cancer in women participating in breast cancer screening. Different lexicons have been developed for the reporting of mammography findings. The most frequently used lexicon is the American College of Radiologists' Breast Imaging Reporting and Data Systems (BI-RADS) atlas (78). BI-RADS classifies mammography findings into seven assessment categories from 0 to 6 using criteria such as presence or absence of lesions in the breast, size, borders, and density of lesions, calcifications, asymmetry, and distortions. These categories are: BI-RADS 0 (findings are unclear); BI-RADS 1 (no lesion, breast tissue appear normal); BI-RADS 2 (a lesion(s) is present but it is benign); BI-RADS 3 (lesion detected is probably benign); BI-RADS 4A (lesion is suspicious cancer and has 2-9% chances of being malignant); BI-RADS 4B (detected lesion has moderate chance of being cancerous 10– 49%); BI-RADS 4C (detected lesion has 50-94% chance of being cancerous); BI-RADS 5 (detected lesion has 95% or higher chance of being cancerous); BI-RADS 6 (lesion has been confirmed to be cancerous using biopsy) (77).

As a process in BreastScreen Australia (BSA), the Australian national population breast screening program, radiologists and breast physicians use the Royal Australian and New Zealand College of Radiologists (RANZCR) classification scale, where mammography findings are classified into one of five grades: Grade 1 (Normal); Grade 2 (Benign); Grade 3 (Indeterminate); Grade 4 (Suspicious of malignancy); Grade 5 (Highly suspicious of malignancy) (79). These gradings systems are used to determine if a woman should be called back for additional testing or pass through the normal biennial screening pathway. In the BI-

RADS system, lesions rated BI-RADS 3 undergo short interval follow-up. However, in the RANZCR system, lesions rated 3 are recalled for additional testing, which may involve additional DM views such as magnified views, DBT, ultrasound, and/or biopsy (79). Also, when findings from the current mammograms are unclear, additional imaging or comparison with prior imaging is often needed to clarify ambiguities. The diagnosis of breast cancer in current mammograms is made based on the identification of abnormal image features. Since breast cancer cells divide and grow rapidly over time, they induce changes in mammograms acquired at different screening rounds. Therefore, even without ambiguities, many screening programmes simultaneously display mammograms from the current and previous screening rounds to allow for comparison and identification of changes that may indicate the presence of breast cancer (80, 81).

Traditionally, screening mammograms are reported by radiologists. However, there is a global shortage of radiologists, and this issue has persisted to-date (82, 83). The shortage of breast radiologists and the ratio of breast radiologists to women vary across the world. For example, in Mexico, 260 breast radiologists have been identified as being involved in the reporting of screening mammograms, resulting in a ratio of one breast radiologists to 60,000 women eligible for mammography screening (84, 85). In the United Kingdom (UK), the workforce survey identified 407 full-time equivalent breast radiologists providing screening and diagnostic mammography interpretation services to 62 National Health Service Breast Screening Program units and approximately 35 million women (86). Data from the survey also showed that 25% of screening services had two or fewer breast radiologists, with 52% of these breast radiologists working less than full-time. While the survey identified only 1% annual increase in the number of breast radiologists, the national data on retirement rates indicated that 24% of breast radiologists would retire in 2025 (86). In Australia, the recent workforce survey showed that

the number of radiologists involved in the national breast screening program had continued to dwindle (83, 87). The decline in the number of Australian breast radiologists is further challenged by a decline in the average retirement age, from 71.1 years to 66.1 years (87). With the increasing aging population, the extension of the screening age to include women up to 74 years, declining radiologists specialising in breast imaging, and higher proportion of retirements, it is challenging to sustain screening programs radiology workload. These challenges will also negatively impact upon the World Health Organisation (WHO) target to reduce breast cancer deaths by 2.5% or save the lives of 2.5 million women worldwide by 2040 (88). Therefore, strategies are needed to increase and support the screening mammography interpretation workforce by understanding the mechanism by which correct interpretations are made and enhance strategies that support workforce extension.

It has been suggested that the shortages of breast radiologists can be mitigated by non-radiology professionals such as radiographers and nurse practitioners. These non-radiology professionals may undertake preliminary evaluation of mammograms so that those with suspicious features are evaluated by radiologists (89). Since the late 20th century, radiographers have emerged as the front runners to support radiologists in the reporting of X-ray images (90). With technological innovations, increasing radiological workloads, radiographer aspirations to report radiographs, the UK national health service supported and established radiographer reporting (90). Currently, advanced practice radiographers in the UK report screening and diagnostic mammograms independently or are paired with radiologists in a double reading system. Published data support the integration of UK radiographers into the reporting framework due to their comparable sensitivity with UK radiologists (91, 92). UK radiographers have also been shown to demonstrate better diagnostic performance than radiologists from developing countries (93). Data emerging from Australia and New Zealand also indicate that

radiographers have the potential to interpret screening mammograms but currently do not undertake mammogram reporting clinically (94-97). These pieces of evidence suggest that there is potential for radiographer reporting of mammograms. Breastscreen Australia policies of extending the screening age to 74 years, incorporating digital breast tomosynthesis in the screening setting, and implementing mobile screening mammography further increases the reporting workload in Australia. The absence of radiologists at mobile screening sites provides opportunities for Breastscreen services to harness the mammography interpretation expertise of radiographers. BreastScreen radiographers may assist in identifying women with suspicious lesions who may require extra views, additional imaging, or whose mammograms need to be triaged for immediate radiologists' reporting. Also, an increase in the number of mammograms acquired from previous and current rounds requires additional time for breast readers to review and assess these images, a task where radiographers can assist radiologists. However, there is a lack of studies on the impact that prior mammograms have on the performance of Australian radiographers in interpreting mammograms. Studies that have examined the impact of prior mammograms on diagnostic efficacy focused on breast radiologists (80, 98-111) even though advanced practice radiographers now participate in screening mammography interpretation. Therefore, studies are needed to establish the impact that prior mammograms have on the interpretation of mammograms by radiographers.

Prior mammograms and outcomes of mammography interpretation

As described previously, the digitisation of the screening process has allowed for professionals interpreting mammograms to access prior mammograms when interpreting current screening mammograms more easily. The human breast has a heterogenous appearance on mammograms due to a mixture of fibrous, glandular, and fatty tissues. Breast cancer lesions induce changes

to the normal appearance of the breast as displayed on mammograms (112, 113). When radiologists perceive the changes suggestive of breast cancer, the features of these changes are described and classified or graded according to the perceived level of malignancy as described above. Sometimes, the changes induced by breast cancer may be subtle and difficult to discern or may require comparison with previous mammograms before a diagnostic decision is made (104, 114). The comparison with previous mammograms is to assess if the changes identified in the current mammograms are present or significant differently from that visible in the previous mammogram. Therefore, prior mammograms can be considered a form of imaging history that can be used to support decision-making around the classification of changes identified in current screening mammograms or rule out malignancy in the absence of changes in the detected lesion. Thus, many screening programs provide options for radiologists to display either the current mammograms or both prior and current mammograms on the screen at the time of interpretation.

To-date, only 15 studies have examined the role prior mammograms play in the success of screening programs (80, 98-111). A detailed description of the characteristics of published studies and their findings and limitations is presented in Chapter 2 (64). Overall, published data on the impact of prior mammograms to diagnostic accuracy is limited. The few published studies focused on diagnostic performance metrics such as sensitivity and specificity (103-105, 107-109), with a few other studies examining recall rates (80, 101, 108), false positive rates (106, 107), and positive predictive value (103, 111). Available data shows that access to prior mammograms does not improve cancer detection rate and sensitivity, but specificity, recall rates, false positive rates, and positive predictive value improve when radiologists reviewed current mammograms together with prior mammograms (64). A limitation of the diagnostic performance metrics used to assess the impact of prior mammograms on the diagnostic efficacy

of screening programs is that they involve a classification task, where mammograms are classified as normal or abnormal. Given that women with abnormal mammograms may need to undergo biopsy to confirm that they have malignant lesions, it is important that performance metrics that assess radiologists' ability to locate the lesion and rate the level of malignancy are used to assess the impact of prior mammograms. However, such metrics including location receiver operating characteristic curve (LROC), lesion sensitivity, and Jacknife alternative receiver operating characteristic curve (JAFROC), which quantify observers' ability to detect, localise, and establish the level of malignancy of the lesions were rarely considered (115, 116). These limitations need to be addressed.

Another factor that influences radiologists' decision to report the presence of breast cancer in mammograms is the feature of the lesion identified. Breast cancer presents mammographically with an array of features, commonly described radiologically as lesion types. Five common lesion types have been described including architectural distortions, calcifications, stellate/spiculated mass, discrete mass, and non-specific/ asymmetric breast density (108). When interpreting screening mammograms, radiologists look out for mammographic features that may indicate the presence of any of these lesion types (see Figure 5 and 6). Histologically, these lesion types are associated with different breast cancer subtypes but may also represent benign conditions (117, 118). For example, stellate/spiculated masses may be seen in scirrhous and invasive cancer (117, 118). Also, discrete masses may be benign or malignant (119), and architectural distortions may represent cancer or radial scars (120-122). Since not all breast lesions are malignant, breast readers may ignore some lesions identified or classify the lesions according to their perceived risk of malignancy using appropriate lexicons. In the USA, the Breast Imaging Reporting and Data Systems is used to classify lesions into seven categories as described previously (78).

In Australia and New Zealand, the Royal Australian New Zealand College of Radiologists (RANZCR) classification system is used to classify mammographic findings into five categories: 1 (Normal); 2 (benign); 3 (Indeterminate); 4 (suspicious of malignancy); 5 (highly suspicious of malignancy) (123, 124). Screening programs rely on the risk stratification of detected lesions to recall women for assessment and follow-up. The risk of malignancy varies across lesion types, with spiculated/stellate lesions identified as the most predictive of invasive breast cancer (117, 118). Therefore, apart from the changes induced in mammograms by breast lesions, the lesion type identified may also play a critical role in breast readers' decisions to report the presence of cancer or recall women for additional testing. However, the impact of prior mammograms on the detection and risk classification of different lesion types is unknown. The only study that examined radiologists' diagnostic performance for each lesion type compared priors from the same vendor technologies versus different vendors and the radiologists who interpreted mammograms with priors were not always the same as those who interpreted mammograms without priors (108). Therefore, studies are needed to establish the influence of prior mammograms on reader performance in detecting and discriminating different types of breast lesions.

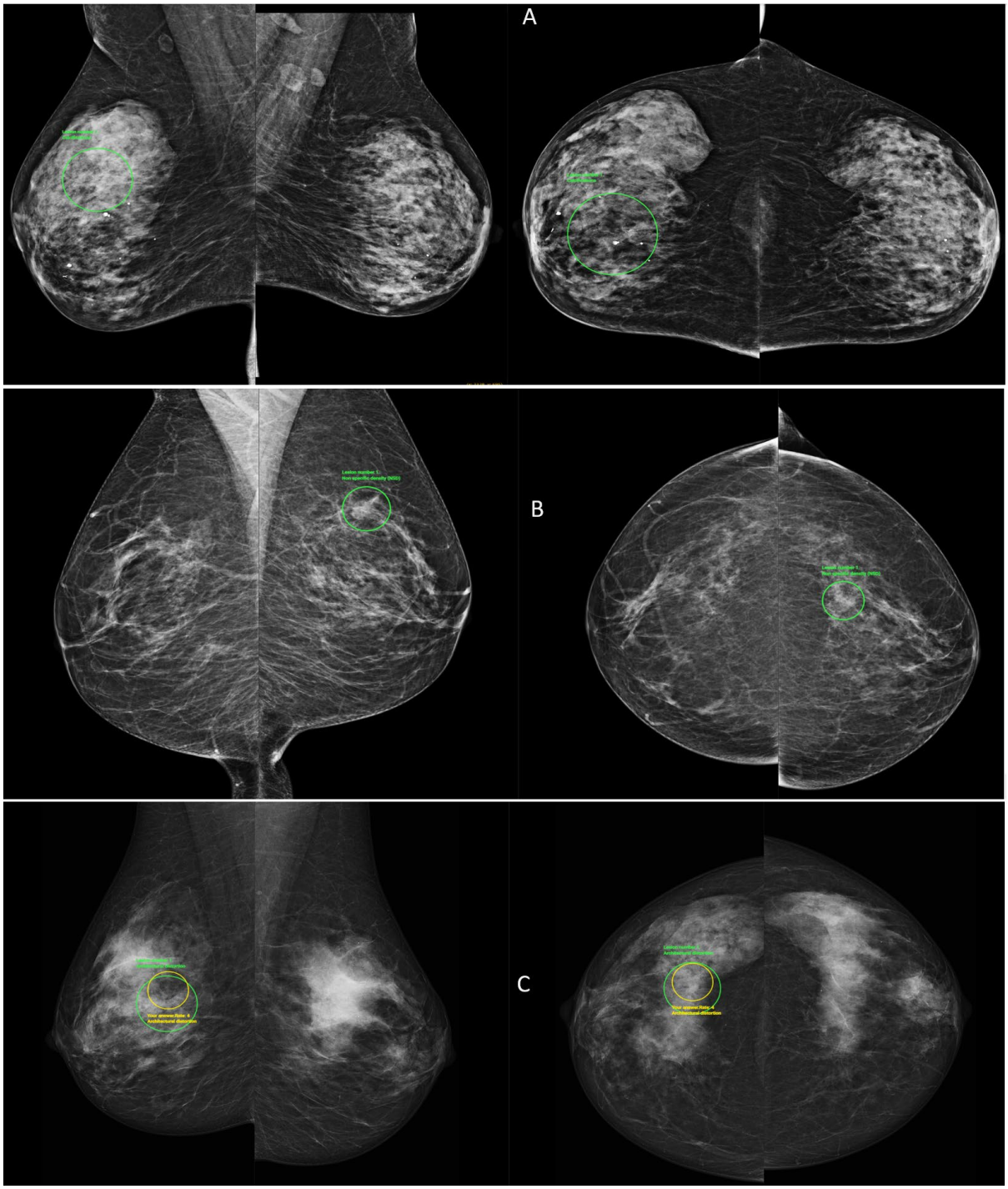


Figure 5: Mammograms showing the features of breast lesions in craniocaudal and mediolateral oblique views. A: Calcifications; (B) Non-specific density; (C) Architectural distortions. Green circles represent the true cancer locations and yellow circles represent markings of breast readers for suspected cancer locations during mammogram readings.

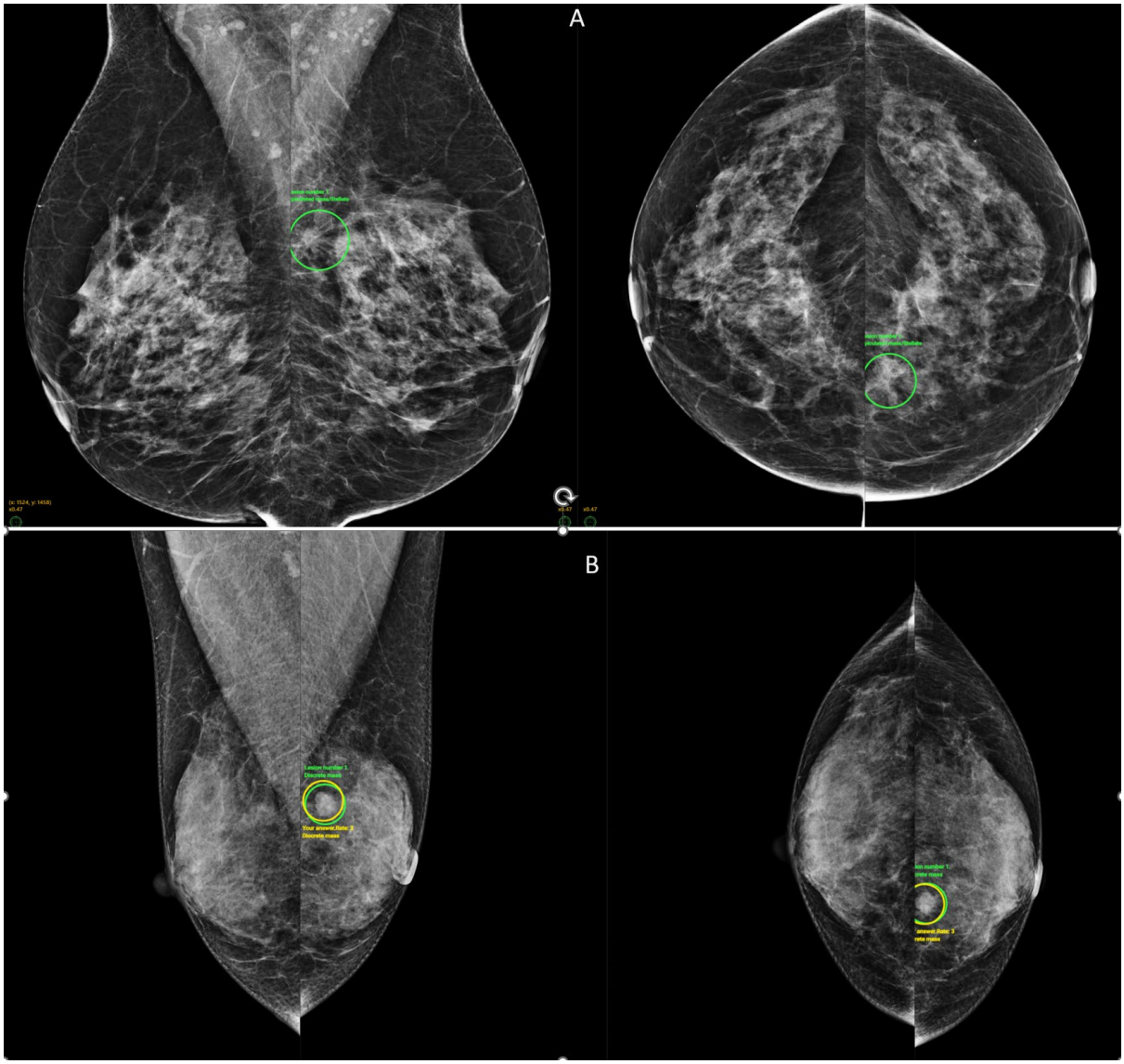


Figure 6: Mammograms showing the mammographic features of breast cancer masses. A: spiculated/stellate mass; B: discrete mass. Green circles represent the true cancer locations and yellow circles represent markings of breast readers for suspected cancer locations during mammogram readings.

Role of test-set interventions in assessing observer performance and the effect of technology and techniques on observer performance

Breast cancer screening programmes regularly undertake quality assurance and quality control to improve the effectiveness and diagnostic outcomes of the screening service. Also, when new techniques and technologies are introduced in clinical practice, performance assessments are often undertaken to ensure that these technologies and techniques are fit for purpose. Technical factors such as positioning and compression can affect the quality of mammograms (35); however, the BSA National Accreditation Standards (NAS) and the Mammography Quality Assurance Program (MQAP) ensure screening services produce consistent image standards. The National Safety and Quality Medical Imaging Standards (NSQMI) also provide quality frameworks for radiology in Australia. While there may be potential differences in technical parameters used for acquiring mammograms at different screening rounds, acceptance of the images for interpretation indicate that these images meet the acceptable quality standards for mammography. Therefore, differences in compression and positioning between different screening rounds may be minimal and have negligible impact on image interpretation. Many screening services undertake quality control and service performance through clinical audits, a process examining the delivery and outcomes of healthcare services against agreed standards (125). A challenge with clinical audits for diagnostic performance evaluation includes that it takes a long time to complete, and the results are provided to the screening service, not individual radiologists. Therefore, it is difficult for individuals involved in the image interpretation process to obtain feedback about their errors and seek corrective interventions to improve their diagnostic accuracy. Since the essence of audits is to improve the quality of the service, strategies that identified errors made and provide tailored individualised feedback are needed.

Over the years, test-set educational interventions have been implemented for continuous professional development (CPD) of radiologists. The first platform for mammography test set training and self-assessment platform was the PERsonal PerFORmance in Mammographic Screening (PERFORMS) implemented in the United Kingdom (126). In Australia, the BreastScreen Reader Assessment Strategy (BREAST) was also developed to support radiologists in identifying their errors and getting feedback on how to reduce error rates (127, 128). In the last 10 years, data generated from test set interventions such as PERFORMS and BREAST have shown that these platforms are robust resources for assessing the impact of technology on the diagnostic efficacy of breast screening services (126-128). Through these platforms, researchers have identified the characteristics of radiologists associated with diagnostic performance, and how technology and the characteristics of the screening population affect radiologists' performance. Knowledge gained from using the BREAST platform has been shown to improve the diagnostic performance of radiologists by about 34% (129). A recent study has also shown that test set interventions can be used as audit tools for quality improvement of screening services (130). There are arguments that data generated in a test setting may not necessarily reflect the clinical setting due to participants' behavioural adjustment from knowledge of being tested (131). However, it has been shown that radiologists' performance in a test set environment is not different from their performance in a clinical setting (130). It has also been shown that and that test-set performance can be used to predict clinical performance of radiologists (130). Therefore, test set platforms provide accessible and available avenues for assessing the impact that prior mammograms have on radiologists' performance.

Reader and patient factors affecting the interpretation of screening mammograms

The X-ray appearance of the breasts makes the interpretation of screening mammograms challenging (35, 132). The literature shows that 20–30% of cancers are missed in mammograms. Some of these missed cancers are visible in mammograms but were missed due the readers' inability to perceive and accurately report the cancer (35). The missed cancer rates vary between radiologists (35, 133), suggesting that limitations of readers interpreting the images may be responsible for errors in the interpretation of mammograms. Therefore, there has been growing interest in identifying the characteristics of radiologists associated with improved performance in screening mammography interpretation. A recent systematic review could not establish specific reader characteristics associated with expertise in the interpretation of mammograms, but some trends emerged (134). Radiologists with higher annual volume of mammograms read, and radiologists that were considered breast imaging specialists consistently showed higher sensitivity, cancer detection rate, and specificity. Radiologists who were older in age were also found to demonstrate lower false positive rates regardless of their reading volume.

The goal of screening programs is to increase cancer detection rate with high sensitivity and specificity while reducing the recall rate (30, 31). The systematic review described above (134) did not report whether the methodological differences between studies particularly the use of prior mammograms had any impact on the findings. Also, given that reading volume, specialisation, and age were found to be associated with key performance metrics (128, 134), it is important to establish the characteristics of readers who benefit from accessing prior mammograms when interpreting mammograms of women from the current screening round.

Identifying such reader characteristics may be useful for optimising the use of prior mammograms in screening programs.

It has also been suggested that technical factors affect the detection of breast cancer in mammograms. There are currently different mammographic systems in use across different screening programs (135, 136). These systems differ in detector technology, target/filter combinations, and image post-processing (137, 138). These factors may lead to differences in image quality across imaging systems and the perception of cancer in the images produced by these systems. Women participating in mammography screening may be screened using different systems across screening rounds. That is, the prior images of these women may have been acquired with mammography systems that are different from the system used for the current screening round. Whether the combination of priors and current mammograms acquired by different imaging systems influence reader performance in screening mammography interpretation is unclear. To-date, only one study has assessed whether pairing priors and current mammograms from the same or different vendor technologies has effect on reader performance (108). This study showed that prior and current mammograms from different vendor technologies improved cancer detection, while prior and current mammograms from the same vendor technology improved the correct interpretation of mammograms with no cancer. These findings need to be interrogated if we are to use prior mammograms from different vendor technologies to improve the diagnostic effectiveness of screening programs.

The characteristics of women participating in breast cancer screening have also been identified as possible factors influencing the accuracy of screening mammography interpretation (35). One of the most studied patient factors affecting the interpretation of screening mammograms is breast density. A major concern about breast density is its masking effect from

superimposing tissue, meaning that there is a decrease in the detection of breast cancer (139, 140). The sensitivity of mammography in women with almost fatty breasts with scattered fibroglandular tissue is as high as 90% but can be as low as 48% in women with extremely dense breasts (141, 142). Women with extremely dense breasts constitute about 60% of women recalled at screening even though they constitute 30 – 50% of women participating in screening (143, 144). Women with dense breasts are more likely to receive false positive outcomes due to artefacts in their mammograms caused by breast density (145, 146). Therefore, breast density affects the effectiveness of screening services from detection to recall through to assessment.

The premise of screening programs is to recall as fewer women as possible, hence it is important to examine if prior mammograms can be used to reduce the recall rate. There is limited data on the association between the availability of prior mammograms and diagnostic performance in dense and non-dense breasts. Only one study has assessed the impact of breast density on reader performance with and without prior mammograms reported higher performance values without prior mammograms with prior mammograms from different vendor technologies compared to with prior mammograms from the same vendor across different breast densities (108). Thus, research is needed to establish the role of prior mammograms in optimising breast cancer detection in women with dense breast.

Gaps and deficiencies in the literature

The literature demonstrates significant gaps in evidence regarding the value of prior mammograms on the performance of breast cancer screening services (80, 99, 100, 101, 105, 147). First, the few studies that examined the value of prior mammograms used different diagnostic performance metrics and there is limited data on each of these metrics. Second,

evidence for the association between the availability of prior mammograms and diagnostic accuracy is limited and unclear. Previous studies show that access to prior mammograms could reduce recall rates and increase specificity; however, this does not show an improvement in sensitivity and cancer detection rate. Third, slow growing cancerous lesions require an interval of two to five years to show subtle changes on mammograms, but there were short intervals between prior and current mammograms in published studies. These factors limit radiologists' ability to detect subtle changes related to breast cancer. Fourth, prior mammograms used in most of these studies were based on film-screen technology, which does not reflect current practice. Fifth, patient and technological factors that affect diagnostic accuracy of screening mammography such as breast density, mammographic features of breast cancer lesions, imaging technology used for acquiring the mammograms were not considered. While most of the published evidence were observer performance studies, almost all the studies were based older observer performance methodologies which do not consider the location and malignancy rating of the lesion. Furthermore, there were small numbers of radiologists who evaluated the mammograms and lack of information on their sampling strategy. Also, there is no data on the characteristics of radiologists who may benefit from accessing prior mammograms so that these can be used to improve the use of prior mammograms. In addition, radiographers have been increasingly undertaking independent interpretation of mammograms and have been identified to fill the global mammography interpretation workforce shortage. Understanding and comparing the findings from prior and current mammograms may provide radiographers baseline information to support decision-making when interpreting mammograms. However, the relevance of prior mammograms to breast radiographers has not been examined. Lastly, the diagnosis of breast cancer in mammograms relies on the identification of suspicious lesions in the breast parenchyma, and the imaging features of these lesions are used to stratify the lesions according to their risk of malignancy. However, there is a lack of data on the impact that prior

mammograms have on the detection and risk stratification of different lesion types. All these gaps must be addressed if we are to improve the performance of screening programs using prior mammograms.

Aims and Objectives of the thesis

The overarching aim of this thesis is to assess the relevance of prior mammograms to the diagnostic efficacy of breast cancer screening programs. To achieve this aim, the thesis addresses three strategic objectives:

- To assess radiologists' performance with prior digital mammograms compared to their performance without prior digital mammograms and the characteristics of radiologists associated with performance under these reading conditions.
- To assess the influence of prior mammograms on radiographers' performance in the interpretation of digital mammograms.
- To assess the influence of prior mammogram availability on the detection of different breast lesion types.

Thesis layout

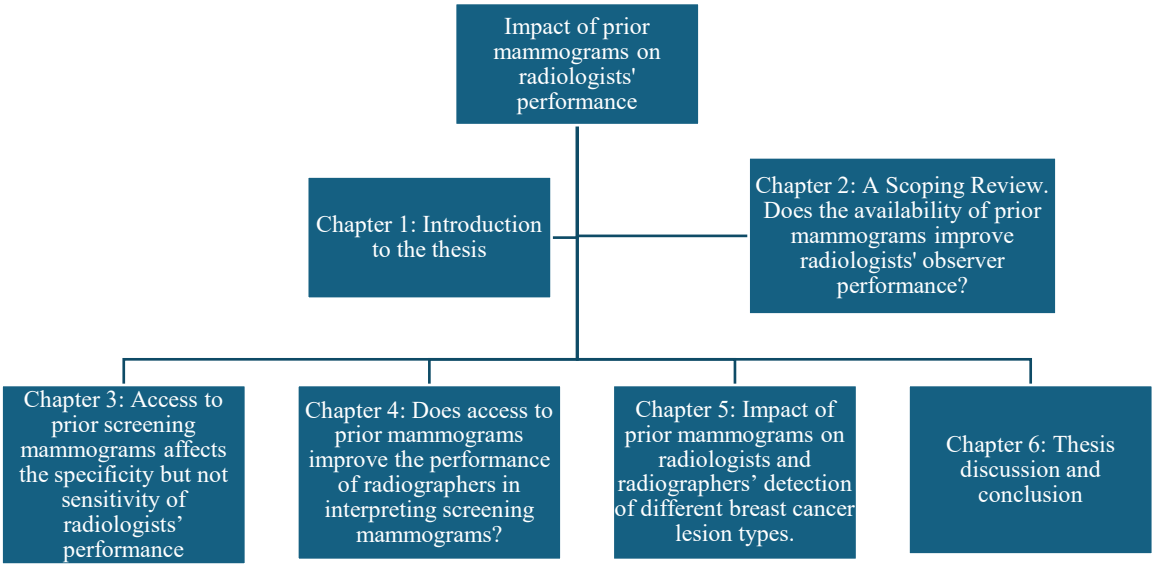


Figure 7: Layout of chapters in the thesis

References

1. WHO. Cancer Statistics. 2023. [Accessed: November 2024]; https://www.who.int/health-topics/cancer#tab=tab_1.

2. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer Journal for Clinicians. 2023;73(1):17-48.

3. International Agency for Research on Cancer (IARC). Monographs on the classification of carcinogenic harzards to humans. WHO report series, 2023 [Accessed: October 2024]; https://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-Humans.

4. Soffritti M, Minardi F, Maltoni C. Physical Carcinogens. In: Kufe DW, Pollock RE, Weichselbaum RR, et al., editors. Holland-Frei Cancer Medicine. 6th edition. Hamilton (ON): BC

Decker; 2003. Chapter 21. Available from: https://www.ncbi.nlm.nih.gov/books/NBK13884/5. Das S, Kundu M, Jena BC, Mandal M. Chapter 25 - Causes of cancer: physical, chemical, biological carcinogens, and viruses. In: Kundu SC, Reis RL, editors. Biomaterials for 3D Tumor Modeling: Elsevier; 2020. p. 607-41.

6. Schwartz, J.L. (2007). Physical Causes of Cancer. In The Cancer Handbook, M.R. Alison (Ed.). https://doi.org/10.1002/9780470025079.chap22.pub2

7. National Cancer Institute (NCI). Cancer-Causing Substances in the Environment. 2022.[Accessed October 2024]; https://www.cancer.gov/about-cancer/causes-prevention/risk/substances#:~:text=But%20others%20are%20the%20result,smoke%20and%20the%20sun's%20rays

8. Williams GM, Iatropoulos MJ, Weisburger JH. Chemical carcinogen mechanisms of action and implications for testing methodology. Experimental and Toxicologic Pathology. 1996;48(2-3):101-11.

9. Anand P, Kunnumakkara AB, Sundaram C, Harikumar KB, Tharakan ST, Lai OS, et al. Cancer is a preventable disease that requires major lifestyle changes. Pharmaceutical Research. 2008;25(9):2097-116.

10. Wu S, Zhu W, Thompson P, Hannun YA. Evaluating intrinsic and non-intrinsic cancer risk factors. Nature Communications. 2018;9(1):3490.

11. Kanavos P. The rising burden of cancer in the developing world. Annals of Oncology. 2006;17 Suppl 8:viii15-viii23.

12. Australian Institute of Health and Welfare. Cancer data in Australia. Cat. no. CAN 122. [Accessed May 2024]; https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia.

13. Azamjah N, Soltan-Zadeh Y, Zayeri F. Global Trend of Breast Cancer Mortality Rate: A 25-Year Study. Asian Pacific Journal of Cancer Prevention. 2019;20(7):2015-20.

14. Huang J, Chan PS, Lok V, Chen X, Ding H, Jin Y, et al. Global incidence and mortality of breast cancer: a trend analysis. Aging (Albany NY). 2021;13(4):5748-803.

15. Arnold M, Morgan E, Rumgay H, Mafra A, Singh D, Laversanne M, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. Breast. 2022;66:15-23.

16. Lei S, Zheng R, Zhang S, Wang S, Chen R, Sun K, et al. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. Cancer Communications (Lond). 2021;41(11):1183-94.

17. Australian Institute of Health and Welfare. Cancer data in Australia. Cat. no. CAN 122. [Accessed Januray 2025]; https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/overview18. Wojtyla C, Bertuccio P, Ciebiera M, La Vecchia C. Breast Cancer Mortality in the Americas and Australasia over the Period 1980-2017 with Predictions for 2025. Biology (Basel). 2021;10(8).

19. Carolyn T, Paul M, Jake P, John B, Jackie C, Sarah CD, et al. Breast cancer mortality in 500 000 women with early invasive breast cancer diagnosed in England, 1993-2015: population based observational cohort study. British Medical Journal. 2023;381:e074684.

20. Duffy SW, Tabár L, Yen AM, Dean PB, Smith RA, Jonsson H, et al. Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women. Cancer. 2020;126(13):2971-9.

21. Duffy S, Vulkan D, Cuckle H, Parmar D, Sheikh S, Smith R, et al. Annual mammographic screening to reduce breast cancer mortality in women from age 40 years: long-term follow-up of the UK Age RCT. Health Technology Assessment. 2020;24(55):1-24.

22. Burton R, Bell R. The global challenge of reducing breast cancer mortality. Oncologist. 2013;18(11):1200-2.

23. National Cancer Institute Surveillance, Epidemiology, and End Results (SEER). Cancer Stat Facts: Female Breast Cancer. 2021 [Accesed December, 2024]; https://seer.cancer.gov/statfacts/html/breast.html

24. Misganaw M, Zeleke H, Mulugeta H, Assefa B. Mortality rate and predictors among patients with breast cancer at a referral hospital in northwest Ethiopia: A retrospective follow-up study. PLoS One. 2023;18(1):e0279656.

25. Lim YX, Lim ZL, Ho PJ, Li J. Breast Cancer in Asia: Incidence, Mortality, Early Detection, Mammography Programs, and Risk-Based Screening Initiatives. Cancers (Basel). 2022;14(17).

26. Altobelli E, Rapacchietta L, Angeletti PM, Barbante L, Profeta FV, Fagnano R. Breast Cancer Screening Programmes across the WHO European Region: Differences among Countries Based on National Income Level. International Journal of Environmental Research Public Health. 2017;14(4).

27. Gianino MM, Lenzi J, Bonaudo M, Fantini MP, Siliquini R, Ricciardi W, et al. Organized screening programmes for breast and cervical cancer in 17 EU countries: trajectories of attendance rates. BioMed Central Public Health. 2018;18(1):1236.

28. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. CA: A Cancer Journal for Clinicians. 2019;69(3):184-210.

29. Schünemann HJ, Lerda D, Quinn C, Follmann M, Alonso-Coello P, Rossi PG, et al. Breast Cancer Screening and Diagnosis: A Synopsis of the European Breast Guidelines. Annals Internal Medicine. 2020;172(1):46-56.

30. Cancer Australia. Breast cancer in Australia statistics. [Accessed October, 2024]; https://www.canceraustralia.gov.au/cancer-types/breast-cancer/breast-cancer-australia-statistics#:~:text=In%202020%2C%20breast%20cancer%20was,34%20males%20and%203%2C110%20females).

31. McCarthy AM, Kontos D, Synnestvedt M, Tan KS, Heitjan DF, Schnall M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. Journal of the National Cancer Institute. 2014;106(11).

32. Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. Journal of the American Medical Association. 2015;314(15):1599-614.

33. Estrada SS. Review of the New American Cancer Society Guidelines for Breast Cancer Screening for Women at Average Risk. Journal Advanced Practitioner in Oncology 2016;7(5):563-6.

34. Yankaskas BC, Schell MJ, Bird RE, Desrochers DA. Reassessment of Breast Cancers Missed During Routine Screening Mammography. American Journal of Roentgenology. 2001;177(3):535-41.

35. Ekpo EU, Alakhras M, Brennan P. Errors in Mammography Cannot be Solved Through Technology Alone. Asian Pacific Journal of Cancer Prevention. 2018;19(2):291-301.

36. Mann RM, Athanasiou A, Baltzer PAT, Camps-Herrero J, Clauser P, Fallenberg EM, et al. Breast cancer screening in women with extremely dense breasts recommendations of the European Society of Breast Imaging (EUSOBI). European Radiology. 2022;32(6):4036-45.

37. Tubiana M. [Wilhelm Conrad Röntgen and the discovery of X-rays]. Bulletin De L Academie Nationale De Medecine. 1996;180(1):97-108.

38. Salomon A. Beitrage zur Pathologie und Klinic der Mammacarzinome [in German]. Archiv für Klinische Chirurgie 1913;101:573–668.

39. Vogel W. Die roentgendarstellung der mammatumoren. Archiv für Klinische Chirurgie. 1932;171:618-26.

40. Gershon-Cohen J, Colcher A. An evaluation of the roentgen diagnosis of early carcinoma of the breast. Journal of the American Medical Association. 1937;108(11):867-71.

41. Nicosia L, Gnocchi G, Gorini I, Venturini M, Fontana F, Pesapane F, et al. History of Mammography: Analysis of Breast Imaging Diagnostic Achievements over the Last Century. Healthcare (Basel). 2023;11(11).

42. Leborgne R. Diagnosis of tumors of the breast by simple roentgenography; calcifications in carcinomas. American Journal of Roentgenology Radium Therapy. 1951;65(1):1-11.

43. Leborgne RA, de Leborgne LC. The Breast in Roentgen Diagnosis. English Translation by Lucy Crocker de Leborgne: Montevideo; 1953.

44. Clark RL, Copeland MM, Egan RL, Gallager HS, Geller H, Lindsay JP, et al. Reproducibility of the technic of mammography (egan) for cancer of the breast. American Journal of Surgery. 1965;109:127-33.

45. Haus AG. Historical technical developments in mammography. Technology in Cancer Research and Treatment. 2002;1(2):119-26.

46. Kuzmiak CM, Pisano ED, Cole EB, Zeng D, Burns CB, Roberto C, et al. Comparison of full-field digital mammography to screen-film mammography with respect to contrast and spatial resolution in tissue equivalent breast phantoms. Medical Physics. 2005;32(10):3144-50.

47. Farber R, Houssami N, Wortley S, Jacklyn G, Marinovich ML, McGeechan K, et al. Impact of Full-Field Digital Mammography Versus Film-Screen Mammography in Population Screening: A Meta-Analysis. Journal of the National Cancer Institute. 2021;113(1):16-26.

48. Faridah Y. Digital versus screen film mammography: a clinical comparison. Biomedical Imaging and Intervention Journal. 2008;4(4):e31.

49. Gambaccini M, Baldelli P. Digital mammography: physical principles and future applications. Radiologia Medica. 2003;106(5-6):454-64; quiz 65-6.

50. Yaffe MJ. Basic Physics of Digital Mammography. In: Bick U, Diekmann F, editors. Digital Mammography. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 1-11.

51. Berns EA. CHAPTER 4 - Mammography and Digital Equipment. In: Bassett LW, Mahoney MC, Apple SK, D'Orsi CJ, editors. Breast Imaging. Philadelphia: W.B. Saunders; 2011. p. 79-98.

52. Svahn TM, Chakraborty DP, Ikeda D, Zackrisson S, Do Y, Mattsson S, et al. Breast tomosynthesis and digital mammography: a comparison of diagnostic accuracy. British Journal of Radiology. 2012;85(1019):e1074-82.

53. Schulz RA, Stein JA, Pelc NJ. How CT happened: the early development of medical computed tomography. Journal of Medical Imaging (Bellingham). 2021;8(5):052110.

54. Kalaf JM. Mammography: a history of success and scientific enthusiasm. Radiologica Brasileira. 2014;47(4):Vii-viii.

55. Dhamija E, Gulati M, Deo SVS, Gogia A, Hari S. Digital Breast Tomosynthesis: an Overview. Indian Journal of Surgical Oncology. 2021;12(2):315-29.

56. Sujlana PS, Mahesh M, Vedantham S, Harvey SC, Mullen LA, Woods RW. Digital breast tomosynthesis: Image acquisition principles and artifacts. Clinical Imaging. 2019;55:188-95.

57. Rodriguez-Ruiz A, Teuwen J, Vreemann S, Bouwman RW, van Engen RE, Karssemeijer N, et al. New reconstruction algorithm for digital breast tomosynthesis: better image quality for humans and computers. Acta Radiologica. 2018;59(9):1051-9.

58. Sechopoulos I. A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications. Medical Physics. 2013;40(1):014302.

59. Chong A, Weinstein SP, McDonald ES, Conant EF. Digital Breast Tomosynthesis: Concepts and Clinical Practice. Radiology. 2019;292(1):1-14.

60. Nelson JS, Wells JR, Baker JA, Samei E. How does c-view image quality compare with conventional 2D FFDM? Medical Physics. 2016;43(5):2538.

61. Conant EF. Clinical implementation of digital breast tomosynthesis. Radiologic Clinics of North America. 2014;52(3):499-518.

62. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Russell Sexton J. Implementation of Breast Tomosynthesis in a Routine Screening Practice: An Observational Study. American Journal of Roentgenology. 2013;200(6):1401-8.

63. Papathanasiou S, Walton LA, Thompson JD. A systematic review of viewing conditions and monitor specifications in mammography. Radiography. 2020;26(4):325-31.

64. Akwo JD, Trieu P, Lewis S. Does the availability of prior mammograms improve radiologists' observer performance?-a scoping review. British Journal of Radiology Open. 2023;5(1):20230038.

65. Nodine CF, Mello-Thoms C, Kundel HL, Weinstein SP. Time Course of Perception and Decision Making During Mammographic Interpretation. American Journal of Roentgenology. 2002;179(4):917-23.

66. Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive Error in Radiology. American Journal of Roentgenology. 2017;208(4):739-49.

67. Berlin L. Radiologic errors, past, present and future. Diagnosis (Berl). 2014;1(1):79-84.

68. Brady A, Laoide R, McCarthy P, McDermott R. Discrepancy and error in radiology: concepts, causes and consequences. Ulster Medical Journal. 2012;81(1):3-9.

69. Fleck MS, Samei E, Mitroff SR. Generalized "satisfaction of search": adverse influences on dual-target search accuracy. The Journal of Experimental Psychology: Applied. 2010;16(1):60-71.

70. Adamo SH, Gereke BJ, Shomstein S, Schmidt J. From "satisfaction of search" to "subsequent search misses": a review of multiple-target search errors across radiology and cognitive science. Cognitive Research: Principles and Implications. 2021;6(1):59.

71. Palazzetti V, Guidi F, Ottaviani L, Valeri G, Baldassarre S, Giuseppetti GM. Analysis of mammographic diagnostic errors in breast clinic. Radiologica Medica. 2016;121(11):828-33.

72. Pinto A, Brunese L. Spectrum of diagnostic errors in radiology. World Journal of Radiology. 2010;2(10):377-83.

73. Patel SH, Stanton CL, Miller SG, Patrie JT, Itri JN, Shepherd TM. Risk Factors for Perceptual-versus-Interpretative Errors in Diagnostic Neuroradiology. American Journal of Neuroradiology. 2019;40(8):1252-6.

74. Zhang L, Wen X, Li J-W, Jiang X, Yang X-F, Li M. Diagnostic error and bias in the department of radiology: a pictorial essay. Insights into Imaging. 2023;14(1):163.

75. Degnan AJ, Ghobadi EH, Hardy P, Krupinski E, Scali EP, Stratchko L, et al. Perceptual and Interpretive Error in Diagnostic Radiology-Causes and Potential Solutions. Academic Radiology. 2019;26(6):833-45.

76. RJ S. Benign Breast Disease in Women. Feingold KR AB, Blackman MR, et al., editor. South Dartmouth (MA): MDText.com; 2018.

77. Magny SJ, Shikhman R, Keppke AL. Breast Imaging Reporting and Data System. [Updated 2023 Aug 28]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-.Available from: https://www.ncbi.nlm.nih.gov/books/NBK459169/.

78. Ghaemian N, Haji Ghazi Tehrani N, Nabahati M. Accuracy of mammography and ultrasonography and their BI-RADS in detection of breast malignancy. Caspian Journal of Internal Medicine. 2021;12(4):573-9.

79. Hadadi I, Clarke J, Rae W, McEntee M, Vincent W, Ekpo E. Diagnostic Efficacy across Dense and Non-Dense Breasts during Digital Breast Tomosynthesis and Ultrasound Assessment for Recalled Women. Diagnostics (Basel). 2022;12(6).

80. Hakim CM, Catullo VJ, Chough DM, Ganott MA, Kelly AE, Shinde DD, et al. Effect of the Availability of Prior Full-Field Digital Mammography and Digital Breast Tomosynthesis Images on the Interpretation of Mammograms. Radiology. 2015;276(1):65-72.

81. Hayward JH, Ray KM, Wisner DJ, Kornak J, Lin W, Joe BN, et al. Improving Screening Mammography Outcomes Through Comparison With Multiple Prior Mammograms. American Journal of Roentgenology. 2016;207(4):918-24.

82. Rawson JV, Smetherman D, Rubin E. Short-Term Strategies for Augmenting the National Radiologist Workforce. American Journal of Roentgenology. 2024;222(6):e2430920.

83. Jeganathan S. The Growing Problem of Radiologist Shortages: Australia and New Zealand's Perspective. Korean Journal of Radiology. 2023;24(11):1043-5.

84. Torres-Mejía G, Smith RA, Carranza-Flores Mde L, Bogart A, Martínez-Matsushita L, Miglioretti DL, et al. Radiographers supporting radiologists in the interpretation of screening mammography: a viable strategy to meet the shortage in the number of radiologists. BioMed Central Cancer. 2015;15:410.

85. National Institute of Statistics and Demography. Demography and Society 2020 [Available from: https://en.www.inegi.org.mx/temas/estructura/.

86. Royal College of Radiologists. The Breast Imaging and Diagnostic Workforce in the United Kingdom: results of survey of NHS breast screening programme units and radiology departments. Acessed, December, 2024. Available from: https://www.rcr.ac.uk/our-services/all-our-publications/clinical-radiology-publications/the-breast-imaging-and-diagnostic-workforce-in-the-united-kingdom/

87. Royal Australian and New Zealand College of Radiologists. 2020 Clinical Radiology Workforce Census Report: Accessed, January 2025. Available from: https://www.ranzcr.com/college/document-library/2020-workforce-census-report-australia

88. WHO. Global breast cancer initiative implementation framework: assessing, strengthening and scaling up of services for the early detection and management of breast cancer: executive summary. [Accessed December, 2023]; https://www.who.int/publications/i/item/9789240067134.

89. Haiart DC, Henderson J. A comparison of interpretation of screening mammograms by a radiographer, a doctor and a radiologist: results and implications. British Journal of Clinical Practice. 1991;45(1):43-5.

90. Price RC. Radiographer reporting: origins, demise and revival of plain film reporting. Radiography. 2001;7(2):105-17.

91. Chen Y, James JJ, Michalopoulou E, Darker IT, Jenkins J. Performance of Radiologists and Radiographers in Double Reading Mammograms: The UK National Health Service Breast Screening Program. Radiology. 2023;306(1):102-9.

92. Pauli R, Hammond S, Cooke J, Ansell J. Comparison of radiographer/radiologist double film reading with single reading in breast cancer screening. Journal of Medical Screening. 1996;3(1):18-22.

93. Sarah L, Barron M, Dung Trieu P. Diagnostic performance of radiographers compared with an international radiology cohort from developing countries in breast cancer detection. Journal of Medical Imaging and Radiation Sciences. 2022;53(4, Supplement 1):S23-S4.

94. Debono JC, Poulos AE, Houssami N, Turner RM, Boyages J. Evaluation of radiographers' mammography screen-reading accuracy in Australia. Journal of Medical Radiation Sciences. 2015;62(1):15-22.

95. Holt JJ, Pollard K. Radiographers' Ability to Perceive and Classify Abnormalities on Mammographic Images - Results of a Pilot Project. Journal of Medical Radiation Sciences. 2010;57(2):8-14.

96. Moran S, Warren-Forward H. The diagnostic accuracy of radiographers assessing screening mammograms: A systematic review. Radiography. 2016;22(2):137-46.

97. Moran S, Warren-Forward H. Can Australian radiographers assess screening mammograms accurately? Biennial follow-up from a four year prospective study and lesion analysis. Radiography. 2016;22(3):e201-e6.

98. Burnside ES, Sickles EA, Sohlich RE, Dee KE. Differential Value of Comparison with Previous Examinations in Diagnostic Versus Screening Mammography. American Journal of Roentgenology. 2002;179(5):1173-7.

99. Callaway MP, Boggis CR, Astley SA, Hutt I. The influence of previous films on screening mammographic interpretation and detection of breast carcinoma. Clinical Radiology. 1997;52(7):527-9.

100. Frankel SD, Sickles EA, Curpen BN, Sollitto RA, Ominsky SH, Galvin HB. Initial versus subsequent screening mammography: comparison of findings and their prognostic significance. American Journal of Roentgenology. 1995;164(5):1107-9.

101. Hakim CM, Anello MI, Cohen CS, Ganott MA, Lu AH, Perrin RL, et al. Impact of and interaction between the availability of prior examinations and DBT on the interpretation of negative and benign mammograms. Academic Radioliogy. 2014;21(4):445-9.

102. Hayward JH, Ray KM, Wisner DJ, Kornak J, Lin W, Joe BN, et al. Improving Screening Mammography Outcomes Through Comparison With Multiple Prior Mammograms. American Journal of Roentgenology. 2016;207(4):918-24.

103. Kim WH, Chang JM, Koo HR, Seo M, Bae MS, Lee J, et al. Impact of prior mammograms on combined reading of digital mammography and digital breast tomosynthesis. Acta Radiologica. 2017;58(2):148-55.

104. Roelofs AA, Karssemeijer N, Wedekind N, Beck C, van Woudenberg S, Snoeren PR, et al. Importance of comparison of current and prior mammograms in breast cancer screening. Radiology. 2007;242(1):70-7.

105. Soh BP, Lee WB, McEntee MF, Kench PL, Reed WM, Heard R, et al. Mammography test sets: Reading location and prior images do not affect group performance. Clinical Radiology. 2014;69(4):397-402.

106. Taylor-Phillips S, Wallis MG, Duncan A, Gale AG. Use of prior mammograms in the transition to digital mammography: a performance and cost analysis. European Journal of Radiology. 2012;81(1):60-5.

107. Thurfjell MG, Vitak B, Azavedo E, Svane G, Thurfjell E. Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. Acta Radiologica. 2000;41(1):52-6.

108. Trieu PDY, Borecky N, Li T, Brennan PC, Barron ML, Lewis SJ. The Impact of Prior Mammograms on the Diagnostic Performance of Radiologists in Early Breast Cancer Detection: A Focus on Breast Density, Lesion Features and Vendors Using Wholly Digital Screening Cases. Cancers (Basel). 2023;15(4).

109. Varela C, Karssemeijer N, Hendriks JH, Holland R. Use of prior mammograms in the classification of benign and malignant masses. European Journal of Radiology. 2005;56(2):248-55.

110. Wilson TE, Nijhawan VK, Helvie MA. Normal mammograms and the practice of obtaining previous mammograms: usefulness and costs. Radiology. 1996;198(3):661-3.

111. Yankaskas BC, May RC, Matuszewski J, Bowling JM, Jarman MP, Schroeder BF. Effect of observing change from comparison mammograms on performance of screening mammography in a large community-based population. Radiology. 2011;261(3):762-70.

112. Sun SS, Zhang B, Zhao HM, Cao XC. Association between mammographic features and clinicopathological characteristics in invasive ductal carcinoma of breast cancer. Molecular and Clinical Oncology. 2014;2(4):623-9.

113. Sickles EA. Mammographic features of "early" breast cancer. American Journal of Roentgenology. 1984;143(3):461-4.

114. Roelofs AAJ, Karssemeijer N, Wedekind N, Beck C, Woudenberg Sv, Snoeren PR, et al. Importance of Comparison of Current and Prior Mammograms in Breast Cancer Screening. Radiology. 2007;242(1):70-7.

115. Chakraborty DP. Recent advances in observer performance methodology: jackknife free-response ROC (JAFROC). Radiation Protection Dosimetry. 2005;114(1-3):26-31.

116. Clarkson E. Estimation receiver operating characteristic curve and ideal observers for combined detection/estimation tasks. Journal of the Optical Society of America. 2007;24(12):B91-8.

117. Evans AJ, Pinder SE, James JJ, Ellis IO, Cornford E. Is mammographic spiculation an independent, good prognostic factor in screening-detected invasive breast cancer? American Journal of Roentgenology. 2006;187(5):1377-80.

118. Liu S, Wu XD, Xu WJ, Lin Q, Liu XJ, Li Y. Is There a Correlation between the Presence of a Spiculated Mass on Mammogram and Luminal A Subtype Breast Cancer? Korean Journal of Radiology. 2016;17(6):846-52.

119. Berment H, Becette V, Mohallem M, Ferreira F, Chérel P. Masses in mammography: What are the underlying anatomopathological lesions? Diagnostic and Interventional Imaging. 2014;95(2):124-33.

120. Choudhery S, Johnson MP, Larson NB, Anderson T. Malignant Outcomes of Architectural Distortion on Tomosynthesis: A Systematic Review and Meta-Analysis. American Journal of Roentgenology. 2021;217(2):295-303.

121. Shaheen R, Schimmelpenninck CA, Stoddart L, Raymond H, Slanetz PJ. Spectrum of Diseases Presenting as Architectural Distortion on Mammography: Multimodality Radiologic Imaging with Pathologic Correlation. Seminars in Ultrasound, CT and MRI. 2011;32(4):351-62.

122. Aslan O, Oktay A. Diagnostic accuracy of the breast MRI Kaiser score in suspected architectural distortions and its comparison with mammography. Scientific Reports. 2024;14(1):447.

123. Spak DA, Plaxco JS, Santiago L, Dryden MJ, Dogan BE. Breast Imaging Reporting and Data Systems (BI-RADS)® fifth edition: A summary of changes. Diagnostic and Interventional Imaging. 2017;98(3):179-90.

124. Royal Australian and New Zealand Colle of Radiology (RANZCR). Breast Imaging Grading Comparison and Lesion Classification lists: RANZCR, 2018. Accessed, April 2025. Available from: https://www.ranzcr.com/search/breast-imaging-grading-comparison-and-lesion-classification.

125. Limb C, Fowler A, Gundogan B, Koshy K, Agha R. How to conduct a clinical audit and quality improvement project. International Journal of Surgical Oncology. 2017;2(6):e24.

126. Gale A, Chen Y. A review of the PERFORMS scheme in breast screening. British Journal of Radiology. 2020;93(1112):20190908.

127. Suleiman ME, Rickard M, Brennan PC. Perfecting detection through education. Radiography (Lond). 2020;26 Supplement. 2:S49-s53.

128. Trieu PD, Tapia K, Frazer H, Lee W, Brennan P. Improvement of Cancer Detection on Mammograms via BREAST Test Sets. Academic Radiology. 2019;26(12):e341-e7.

129. Suleiman WI, Rawashdeh MA, Lewis SJ, McEntee MF, Lee W, Tapia K, et al. Impact of Breast Reader Assessment Strategy on mammographic radiologists' test reading performance. Journal of Medical Imaging and Radiation Oncology. 2016;60(3):352-8.

130. Qenam BA, Li T, Ekpo E, Frazer H, Brennan PC. Test-set training improves the detection rates of invasive cancer in screening mammography. Clinical Radiology. 2023;78(3):e260-e7.

131. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, et al. The "laboratory" effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. Radiology. 2008;249(1):47-53.

132. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? Insights into Imaging. 2017;8(1):171-82.

133. Lamb LR, Fonseca MM, Verma R, Seely JM. Missed Breast Cancer: Effects of Subconscious Bias and Lesion Characteristics. RadioGraphics. 2020;40(4):941-60.

134. Wong DJ, Gandomkar Z, Lewis S, Reed W, Suleiman M, Siviengphanom S, et al. Do Reader Characteristics Affect Diagnostic Efficacy in Screening Mammography? A Systematic Review. Clinical Breast Cancer. 2023;23(3):e56-e67.

135. Keavey E, Phelan N, O'Connell AM, Flanagan F, O'Doherty A, Larke A, et al. Comparison of the clinical performance of three digital mammography systems in a breast cancer screening programme. British Journal of Radiology. 2012;85(1016):1123-7.

136. Evans T, Burlton B, Devenish G, Stevens G, Lewis M, Gower Thomas K. A comparison of two digital mammography systems: are there any differences? Clinical Radiology. 2016;71(1):27-31.

137. Elshemey WM, Saif RA, Elfiky AA. Target-filter combination effects on breast tissue characterization using mammographic X-rays: A Monte Carlo simulation study. Journal of X-ray Science and Technology. 2022;30(4):823-34.

138. Gingold EL, Wu X, Barnes GT. Contrast and dose with Mo-Mo, Mo-Rh, and Rh-Rh target-filter combinations in mammography. Radiology. 1995;195(3):639-44.

139. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. New England Journal of Medicine. 2007;356(3):227-36.

140. Bell RJ. Mammographic density and breast cancer screening. Climacteric. 2020;23(5):460-5.

141. Freer PE. Mammographic Breast Density: Impact on Breast Cancer Risk and Implications for Screening. RadioGraphics. 2015;35(2):302-15.

142. Lynge E, Vejborg I, Andersen Z, von Euler-Chelpin M, Napolitano G. Mammographic Density and Screening Sensitivity, Breast Cancer Incidence and Associated Risk Factors in Danish Breast Cancer Screening. Journal of Clinical Medicine. 2019;8(11).

143. Pandya T, Liu Z, Dolan H, Hersch J, Brennan M, Houssami N, et al. Australian Women's Responses to Breast Density Information: A Content Analysis. International Journal of Environmental Research Public Health. 2023;20(2).

144. Hadadi I, Rae W, Clarke J, McEntee M, Ekpo E. Diagnostic Performance of Adjunctive Imaging Modalities Compared to Mammography Alone in Women with Non-Dense and Dense Breasts: A Systematic Review and Meta-Analysis. Clinical Breast Cancer. 2021;21(4):278-91.

145. Lehman CD, White E, Peacock S, Drucker MJ, Urban N. Effect of age and breast density on screening mammograms with false-positive findings. Americal Journal of Roentgenology. 1999;173(6):1651-5.

146. Nelson HD, O'Meara ES, Kerlikowske K, Balch S, Miglioretti D. Factors Associated With Rates of False-Positive and False-Negative Results From Digital Mammography Screening: An Analysis of Registry Data. Annals of Internal Medicine. 2016;164(4):226-35.

147. Sumkin JH, Holbert BL, Herrmann JS, Hakim CA, Ganott MA, Poller WR, et al. Optimal Reference Mammography: A Comparison of Mammograms Obtained 1 and 2 Years Before the Present Examination. American Journal of Roentgenology. 2003;180(2):343-6.

CHAPTER 2

Does the availability of prior mammograms improve radiologists' observer performance? -a scoping review

To ensure that the gaps and limitations in the literature are appropriately identified and addressed in this thesis, a scoping review of the literature was undertaken in Chapter 2. The review examined the impact that prior mammograms have on radiologists' performance. Using the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) extension for scoping reviews (ScR), published literature were extracted from MEDLINE, PubMed, Web of Science, ScienceDirect, and CINAHL databases. Google cross-searches were also conducted, and reference lists of eligible articles were examined to capture articles not identified on database search. Individual sources of evidence were critically appraised using the Standard for Reporting Diagnostic Accuracy Studies (STARD) guidelines. The search strategy yielded 15 eligible studies. Findings from these studies showed that access to prior mammograms reduce recall rates, false-positive rates, abnormal interpretation rates, and increases specificity without affecting sensitivity and cancer detection rate. However, most of the evidence were from studies involved either older screening technologies or observer performance evaluation methodologies that do not account for observer evaluation at the lesion level. The findings in Chapter 2 are published as Akwo JD, Trieu PD, Lewis SJ (2023). Does the availability of prior mammograms improve radiologists' observer performance? -a scoping review. BJR Open. 2023; 5(1): 20230038. 10.1259/bjro.20230038

BJR|Open https://doi.org/10.1259/bjro.20230038

Received: 01 May 2023 | Revised: 31 July 2023 | Accepted: 01 August 2023

Cite this article as:
Akwo JD, Trieu P, Lewis S. Does the availability of prior mammograms improve radiologists' observer performance?—a scoping review. *BJR Open* (2023) 10.1259/bjro.20230038.

REVIEW ARTICLE

Does the availability of prior mammograms improve radiologists' observer performance?—a scoping review

[1,2]JUDITH D. AKWO, BSc,, Hons, [1]PHUONG TRIEU, PhD and [1]SARAH LEWIS, PhD

[1]Medical Image Optimisation and Perception Group, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia
[2]Orange Radiology, Laboratories, and Research Centre, Calabar, Nigeria

Address correspondence to: Mrs Judith D. Akwo
E-mail: *jakw2899@uni.sydney.edu.au*

Objective: The objective of this review was to examine the impact of previous mammogram availability on radiologists' performance from screening populations and experimental studies.
Materials and Methods: A search of the literature was conducted using five databases: MEDLINE, PubMed, Web of Science, ScienceDirect, and CINAHL as well as Google and reference lists of articles. Keywords were combined with "AND" or "OR" or "WITH" and included "prior mammograms, diagnostic performance, initial images, diagnostic efficacy, subsequent images, previous imaging, and radiologist's performance". Studies that assessed the impact of previous mammogram availability on radiologists' performance were reviewed. The Standard for Reporting Diagnostic Accuracy guidelines was used to critically appraise individual sources of evidence.
Results: A total of 15 articles were reviewed. The sample of mammogram cases used across these studies varied from 36 to 1,208,051. Prior mammograms did not affect sensitivity [with priors: 62–86% (mean = 73.3%); without priors: 69.4–87.4% (mean = 75.8%)] and cancer detection rate, but increased specificity [with priors: 72–96% (mean = 87.5%); without priors: 63–87% (mean = 80.5%)] and reduced false-positive rates [with priors: 3.7 to 36% (mean = 19.9%); without priors 13.3–49% (mean = 31.4%)], recall rates [with priors: 3.8–57% (mean = 26.6%); without priors: [4.9%–67.5% (mean = 37.9%)], and abnormal interpretation rate decreased by 4% with priors. Evidence for the associations between the availability of prior mammograms and positive-predictive value, area under the curve (AUC) from the receiver operating characteristic curve (ROC) and localisation ROC AUC, and positive-predictive value of recall is limited and unclear.
Conclusion: Availability of prior mammograms reduces recall rates, false-positive rates, abnormal interpretation rates, and increases specificity without affecting sensitivity and cancer detection rate.

INTRODUCTION

Breast cancer is identified as the second highest cause of cancer-related deaths in females across the world.[1,2] The World Health Organisation's (WHO) global estimate that 7.8 million women live with breast cancer.[3] In 2020, there were 2.3 million new cases of breast cancer and 685,000 deaths from the disease.[3] The incidence and deaths from breast cancer vary across the world,[2,4] and in the last two decades, mortality from breast cancer has reduced considerably due to early detection primarily through organised population screening programmes and advances in treatment strategies.[2,5] Current data show that advances in breast cancer screening and treatment have reduced breast cancer mortality by 20–40%.[5,6] Full-field digital mammography (FFDM) is the standard imaging tool for breast cancer screening in Australia, the USA and many European countries, and has been partly credited with the success in improving survival rates from the disease.[2,5] For mammographic images to be useful in the early detection of breast cancer, they must be accurately interpreted by trained radiologists and in some countries by breast physicians and advanced practice radiographers. However, heterogeneity in the appearance of the breast and breast cancer features, and anatomical noise due to tissue superposition make accurate interpretation of FFDM images challenging.[7] These factors together with the inherent limitations of humans interpretating the images reduce the diagnostic performance of FFDM.[7,8]

Despite the introduction of digital breast tomosynthesis (DBT), ultrasound and MRI into the breast imaging pathway for screening and diagnosis, FFDM continues to be the front-line tool for breast cancer screening in many countries.[8,9] In

the last two decades, both technological and human-related strategies have been explored to improve the performance of FFDM in breast cancer screening.[8,10] These technological strategies include computer-assisted software packages such as computer-aided detection (CAD), artificial intelligence (AI), and machine learning (ML) algorithms to flag suspicious areas in mammograms or provide diagnostic decision-support.[8,10] Although these tools have shown promise for improving cancer detection, they are not widely used clinically, and the final diagnostic decision is still mostly made by radiologists. Therefore, strategies for improving radiologists' performance have been the focus of many research studies.[8,11–13] These strategies have included double independent reading of FFDM, pairing of radiologists with the technological algorithms, identification of radiologists' characteristics associated with performance, and giving radiologists access to the previous images of screened females.[8,14,15] These human-related strategies have been shown to impact differently on radiologist's performances and understanding which of these factors improves performance is crucial to improving the outcomes of breast cancer screening programs.

The detection of breast cancer in mammograms of asymptomatic females is made based on the identification of abnormal changes displayed on these mammograms.[16] With the digitisation of the screening process, it has become easier to retrieve and compare previous and current screening mammograms for abnormal breast changes. However, screening programs are also challenged by storage capacity due to the big data generated from current and previous screening rounds in the digital era and movement of females between screening services and public *vs* private clinic attendance for imaging. Consequently, several prospective and retrospective studies have examined the association between availability of previous mammograms and radiologists' performance, but these studies have generated mixed outcomes.[17–22] The reasons for these findings are not clear, and no review has been conducted to establish the relevance of prior mammograms to screening programs. Therefore, this review aims to examine the impact of previous mammograms on diagnostic performance in screening populations and experimental studies. Findings from the review should provide evidence for the relevant usability of prior imaging in clinical practice.

METHODS

Eligibility criteria

Articles were considered eligible if they examined the impact of prior mammograms (FFDM, film-screen (FS) mammograms) on the accuracy of mammography interpretation and were published in the English language. We also included articles that used DBT in conjunction with mammography, in recognising that some screening and diagnostic imaging pathways have transitioned to the dual use of FFDM and DBT, or as DBT as an updated technology and FFDM was used for the prior cases. Articles were also considered eligible if they compared the performance of readers with and without prior images, had at least one reader and a reference standard for assessing reader performance. Articles that calculated at least one of the following metrics, sensitivity, specificity, recall rates (RRs), false-positive rates (FPRs), cancer detection rate (CDR), negative-predictive value (NPV), positive-predictive value (PPV), PPV of recall, abnormal interpretation rates (AIR), area under the curve from receiver operating characteristics (ROC AUC), Region-of-interest (ROI) figure of merit (FOM), jacknife alternative free response curve (JAFROC), location receiver operating characteristic curve (LROC), PPV of recall, and biopsy recommendation rate (BRR), were considered eligible. No restriction was placed on publication date.

Information sources

A search of the literature was conducted using five databases: MEDLINE, PubMed, Web of Science, ScienceDirect, and CINAHL to identify articles that examined the impact of prior mammograms on radiologic image interpretation or diagnostic accuracy. Other information sources included Google, Google Scholar, and reference lists of published articles.

Search strategy

The following keywords were used to search for articles and were combined with "AND" or "OR" or "WITH": prior images AND diagnostic performance; initial images AND diagnostic efficacy; initial OR subsequent images OR previous imaging AND radiologist's performance; prior radiographs AND reader performance; prior mammograms AND diagnostic efficacy; prior mammograms AND screening outcomes; previous imaging OR initial film; comparison WITH priors; prior breast imaging AND screening performance. A Google cross-search was conducted and reference lists of identified articles were also manually searched to identify additional articles.

Study/source of evidence selection

After the articles were identified, their titles and abstracts were first screened by one reviewer. Full texts of articles that fulfilled the inclusion criteria were retrieved. A second reviewer independently reviewed the abstracts of these articles identified for relevance. Thereafter, a detailed assessment of the full text of all selected articles was performed against the inclusion criteria by two independent reviewers. Where there was a disagreement, this was resolved through discussion. Articles that did not fulfill these criteria were excluded as were reviews, case reports, and conference proceedings. The results of the search and evidence selection were then presented in a Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for scoping review (PRISMA-ScR) flow diagram.[23]

Data extraction/charting process

A data charting table was developed to guide data extraction from eligible articles. Each of the two reviewers, who were involved in the evidence selection process working independently, extracted the data from papers that fulfilled inclusion criteria based on this data charting table. The data extracted included study year, population of study/participants, study context and design, key findings relevant to the review question and adjustments. Any disagreements between the reviewers were also resolved via discussion, and when a consensus could not be reached, a third reviewer was contacted to act as an arbiter.

Data items

Data items are summarised in Table 1 and included CDR, sensitivity, specificity, RR, ROC AUC, JAFROC, FPR, ROI FOM, LROC, PPV, AIR, and BRR.

Table 1. Data items included in the review

	Data items	Definition of data items
1	Cancer detection rate	The number of cancers detected per 1000 women screened.
2	Sensitivity	The rate of true-positive (cancer) cases that were correctly identified.
3	Specificity	The rate of true-negative (normal) cases that were correctly identified.
4	Recall rate	The rate of females recalled for follow-up imaging due to suspicious of cancer.
5	AUC from ROC	A measure of performance in correctly classifying normal and abnormal cases
6	JAFROC FOM	A measure of performance in localising a lesion and simultaneously rating the level of malignancy *vs* identifying correct normal cases
7	FPR	The number of negative cases that were incorrectly reported as positive.
8	ROI FOM	The empirical probability that a cancer containing ROI is rated higher than a normal ROI)
9	PPV	The proportions of positive results that are true positive
10	AIR	The number of abnormal findings that require additional follow-up or the number of mammograms with abnormal final interpretation.
11	Biopsy rate	The proportion of females had been recalled at screening and had a biopsy test.
12	LROC	A performance measure that quantifies readers ability to detect and locate an abnormality on the mammogram.

AIR, abnormal interpretation rate; AUC, area under the curve; FOM, figure of merit; FPR, false-positive rate; JAFROC, jacknife alternative free-response curve; LROC, location ROC; PPV, positive-predictive value; ROC, receiver operating characteristic; ROI, region-of-interest.

Critical appraisal of individual sources of evidence

Critical appraisal was performed using the Standard for Reporting Diagnostic Accuracy (STARD) recommendations.[24] Items assessed included research questions or study aims and methodological criteria such as participants, reference standard, blinding of readers, index text, flow and timing (interval between the index tests and the reference standard). Other items included methods for calculating measures of accuracy and quantify uncertainty, data collection, sampling, training and expertise of readers, characteristics of the population studied, description of disease distribution and severity, and reproducibility estimates.

RESULTS

Selection of sources of evidence

Database search yielded 584,018 articles and 11 articles were identified through other sources, including Google and reference list of published articles. After duplicates were removed, 11,150 articles were available for screening. Of these 11,150 articles screened, 10,781 were excluded because they were not mammography reader performance studies. Out of the remaining 369 articles selected for abstract and full-text review, 354 studies were excluded for the following reasons: (1) they did not compare reader performance with and without prior images (n = 201), were based on computer-aided detection such as AI or ML (n = 99) or were animal studies (n = 32). Duplicates (n = 11), conference papers (n = 6), commentaries or letters to the editor (n = 3), and studies that assessed change in case management or radiologist's opinion but did not report any performance metric (n = 2) were excluded. A total of 15 articles fulfilled the eligibility criteria and were included in the review (Figure 1).

Characteristics of sources of evidence

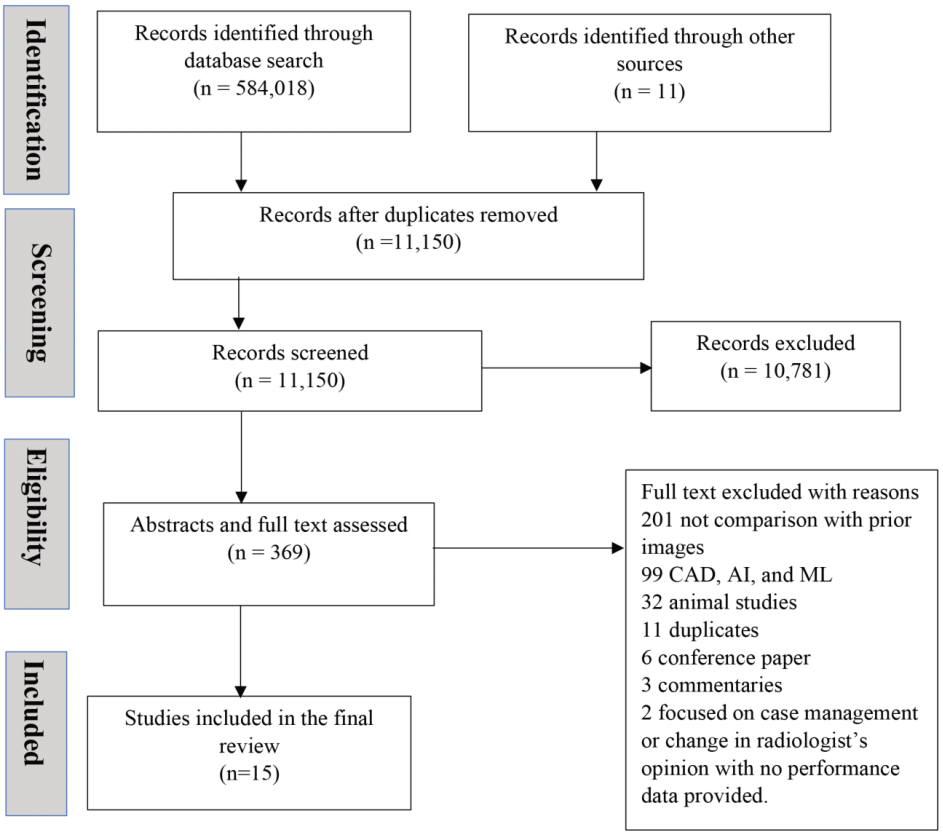
Table 1 summarises the characteristics of the studies identified. In terms of study location, there were 2 studies each conducted in the UK and Netherlands, 10 studies conducted in the USA, and 1 study each in Korea, Australia, and Denmark. 11 of these studies were observer performance studies examining test-sets[17,19–22,25–29] while four were retrospective analyses of mammographic databases, with three examining screening data whilst one also included a cohort of diagnostic mammograms. Eight studies were either based on FS mammography or a combination of FS and FFDM; four studies used FFDM alone, and three studies used FFDM and DBT.[19,20,25] The number of mammograms used in the studies reviewed ranged from 36 to 1,208,051 (Median: 160; Mean: 80,536.7). The sample of readers in 14 studies varied from 3 to 12, and 1 retrospective study evaluated test-set results of 612 radiologists.[30] In these 15 studies, the index test (the test being evaluated against a reference standard) was independently interpreted in 13 studies, and 2 studies did not provide any information about the index test.[31,32]

In comparing performance with and without prior images, four studies used ROC AUC,[17,20,25,29] four used RRs,[19,20,31,33] seven used sensitivity and specificity,[21,22,25,28,30] and four used CDR.[31,33,34] Two studies examined diagnostic accuracy. Outcomes metrics such as PPV of recall, number of cancers detected, abnormal interpretation rate, biopsy recommendation rate, and biopsy yield were each assessed in one study. In addition, JAFROC FOM, LROC, false-positive recall, FPR, and ROI FOM were each assessed in one study (Table 2).

Critical appraisal within sources of evidence

Appendix 1 summarises the strengths and weaknesses of the studies reviewed according to the STARD guidelines. 10 of the studies reviewed were prospective[17,19–22,25–29] and 5 were retrospective analyses.[18,30,31,33,34] All studies provided the aim, described the reference standard, blinded readers, and described the flow and timing of the index test. 10 studies described the

Figure 1. Prisma flow chart showing the selection of evidence. AI, artificial intelligence; CAD, computer-aided detection; ML, machine learning.



participants,[17,19–22,26–30] 3 studies provided limited participant information[31,33,34] and only 2 studies did not provide information about participants' demographics.[18] Almost all of the studies (14/15) provided information about the index test; however, the retrospective studies did not establish the reference standard prior to the index test. Only one study described how the readers were sampled[21] or assessed the reproducibility of the findings.[20] Information about the expertise of the reader were reported in 11 studies.[17,19–22,25–28,33,34] All studies provided information about the methods for calculating reader performance and uncertainty. The characteristics of the females whose mammograms were used (such as ethnicity, age) were clearly described in 10 studies, but this information was unclear in the remaining 5 studies. 13 of the studies described disease distribution and severity.

Overall, most of the studies were well conducted in terms of the reference standard, blinding of readers, index text, methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty. Weaknesses were noted in terms of reproducibility, description of the expertise of the readers, and the characteristics of the population studied. Only seven of the studies reported the interval between the prior and current mammograms, which ranged from 9 months to 4.5 years. Many of the studies (11/15) did not adjust for any confounding factor, two studies adjusted for breast density and age, and one study adjusted for reading location (Table 2).

Results of individual sources of evidence

Table 2 summarises the characteristics and results of studies reviewed. The results of studies comparing reader performance with and without prior images can be mapped into two themes: (1) diagnostic performance evaluation; (2) RRs.

Diagnostic performance evaluation

Of the three studies that assessed CDR, only two of these studies assessed CDR per 1000 women screened and the reported values ranged from 3.7 to 5.2% with prior mammograms and 5.5 to 7.1% without prior mammograms. One of the studies reported

Table 2. A summary of the characteristics of the studies identified

Author/Year	Study location	Study design	Sample size	Participants/ readers	Independent interpretation of index test	Outcomes/performance metrics			Key findings	Adjustments
						Outcome assessed	With priors	Without priors		
Callaway et al.1997[17]	UK	Observer performance	FS (n = 100): Normal: 88; Cancer:12	Radiologists (n = 8)	Yes	ROC AUC	0.78[a,d]	0.81[a,d]	Prior mammograms did not improve diagnostic accuracy	None
Hakim et al. 2014[19]	USA	Observer performance	DM+DBT (n = 36)	Radiologists (n = 8)	Yes	RRs	27%	64%	Prior mammograms reduced RRs	None
Kim et al. 2017[52]	Korea	Observer performance	DM+DBT n = 116 Cancers: 24	Radiologists (n = 3)	Yes	AUC	81%	82%	Adding prior mammograms did not significantly affect the AUC of DM +DBT; Prior mammograms significantly improved the specificity and PPV of DM +DBT; prior mammogram did not affect sensitivity of DM+DBT	None
						Sensitivity	69.4%	69.4%		
						Specificity	93.8%	85.9%		
						PPV	74.6%	56.2%		
Hayward et al. 2016[53]	USA	Retrospective	Prior = FS Current = DM n = 6,288	Radiologists, number not reported	Yes	RR	7.8%	16.6%	Multiple prior images reduced RR and increased CDR and PPV of recall	Age
						PPV of recall	5.6%	NR		
						CDR	Single prior (4.3/1000) Multiple prior (6.6/1000)	N/A		
Frankel et al.1995[18]	USA	Retrospective study	DM: n = 3386	Not reported	Yes	AIR	3%	7%	Substantially fewer abnormal screening interpretations were made when mammography had been performed previously and when the prior films were available for comparison.	None
						Number of cancers detected	41%	32%		

(Continued)

Table 2. (Continued)

Author/Year	Study location	Study design	Sample size	Participants/ readers	Independent interpretation of index test	Outcomes/performance metrics			Key findings	Adjustments
						Outcome assessed	With priors	Without priors		
Burnside et al. 2002[31]	USA	Retrospective study	FS: n = 48,281 screening n=:9825 Diagnostic: n = 38,456	Radiologists (numbers not reported	Not reported	RRs	Screening: 3.8%	4.9%	Screening: prior images reduced recall rates but not biopsy rate recommendation and cancer detection. Diagnostic: previous mammograms increased true-positive findings	None
						CDR	Screening: 5.2/1000 Diagnostic: 59/1000	5.5/1000 11/1000		
						Biopsy performed BRR	Screening 1.2% Diagnostic: 7.6% Diagnostic 9.4%	1.4% 3.0% 4.3%		
						Biopsy yield	Screening: 44% Diagnostic: 51%	40% 38%		
Sumkin et al. 2003[22]	USA	Observer performance study	FS: n = 128	Radiologists (n = 12)	Yes	Accuracy	72%	65%	Availability of prior images improved accuracy and specificity without affecting sensitivity.	None
						Sensitivity	71%	73%		
						Specificity	72%	63%		
Yankaskas et al. 2011[34]	USA	Retrospective study	Unclear if they are DM or FS: n = 1157 980	Radiologists (number not reported)	Yes	RR	6.9%	14.9%	Comparison mammograms: reduced CDR, RR, sensitivity, and PPV, but increased specificity.	Breast density
						Sensitivity	78.9%	87.4%		
						Specificity	93.5%	85.7%		
						CDR	3.7%	7.1%		
						PPV	5.4%	4.8%		
Hakim et al. 2015[20]	USA	Observer performance	DM+DBT 153 (50 cancers, 60 normal, 43 benign	Radiologists (n = 8)	Yes	RRs	Cancer: 81% Non-cancer: 33%	85% 50%	Prior images reduced RR in cancer and non-cancer images but decreased sensitivity. Prior images had no effect on ROC AUCs and specificity.	None
						AUC	83%	83%		
						Sensitivity	7% reduction	NR		
						Specificity	Reported p-values only	Reported p-values only		
Roelofs et al. 2007[26]	Netherlands	Observer performance	Digitised FS: n = 160	Radiologists (n = 12)	Yes	LROC/LLFS	26%	19%	Prior mammograms increased Lesion localisation.	None
Thurfjell et al. 2000[28]	Denmark	Observer performance	FS: n = 150 (SDC: 49; IC:12; normal:89	Radiologists (n = 3)	Yes	FPR	3.7%	13.3%	Prior images increased specificity, reduced FPR with no change in sensitivity	None
						Sensitivity	62%	69%		
						Specificity	96%	87%		

(Continued)

Table 2. (Continued)

Author/Year	Study location	Study design	Sample size	Participants/ readers	Independent interpretation of index test	Outcomes/performance metrics			Key findings	Adjustments
						Outcome assessed	With priors	Without priors		
Taylor-Phillips et al. 2012[27]	UK	Observer performance	DM & FS: n = 160 (cancer:66)	Radiologists (n = 4) Radiographers (n = 4)	Yes	JAFROC	87%	83%	Prior images improved JAFROC and reduced RR and FPR	None
						FPR	40%	59%		
						FPR RR JAFROC FOM difference	36% 43% 0.3%	49% 53% Baseline		
Varela et al. 2005[29]	Netherlands	Observer performance	FS: n = 198	Radiologists (n = 6)	Yes	AUC	79.6%	76%	Prior images improved performance	None
Soh et al. 2014[21]	Australia	Observer performance	DM: n = 200 (screening+diagnostic)	Radiologists (n = 10)	Yes	ROI FOMs	88%	85 %	Prior images had no effect on ROI FOMs, sensitivity, and specificity	Reading location
						Sensitivity	80%	76%		
						Specificity	91%	87%		
Trieu et al. 2023[30]	Australia	Retrospective study	Screening (n = 540; Cancer:179; Normal: 361	Radiologists (n = 612)	Yes	ROC AUC	0.782–0.820	0.814	AUC and Sensitivity without priors were higher, but specificity decreased without prior images. RRs were only lower if the priors were obtained with the same vendor technology as the current images	Breast density and technology
						Sensitivity	0.712–0.785	0.803		
						Specificity	0.771–0.787	0.749		
						RR	0.353–0.461	0.444		

AIR: abnormal interpretation rate; AUC: area under ROC; BRR: biopsy recommendation rate;CDR: cancer detection rate; FPR: false-positive rate;IC: interval cancer; JAFROC: jacknife free-response receiver operating characteristic; LLFS: lesion localised fraction score; LROC: location ROC; NR: values not reported;; ROC: receiver operating curve; RR: recall rate;SDC, screen-detected cancers.
[a]Figures estimated from ROC AUC curves.

that the availability of prior mammograms did not affect the CDR,[31] one reported an increase in CDR,[33] and one study reported a decrease in the CDR.[34] Overall, evidence for the association between the availability of prior mammograms and CDR is limited and unclear.

Among the five studies that compared performance using ROC AUC, three reported ROC values ranging from 79.6 to 83% (Mean: 80.16%) with prior mammogram availability and 76 to 83% (Mean: 80.68%) without prior images available. Four of the five studies showed that there is no difference in diagnostic accuracy between reading with and without prior mammograms,[17,20,25,29,30] and one study showed that when analysis was stratified by breast density, AUC values without prior images were significantly higher than AUC obtained when prior and current mammograms were acquired using the same vendor technology.[30] JAFROC and LROC, which rely on the reader's ability to localise the lesion within an acceptable radius, were each assessed in one study and were found to improve with the availability of prior mammograms.[26,27]

Six out of seven studies reported sensitivity values ranging from 62 to 86% (mean = 73.3%) with prior mammograms and 69.4 to 87.4% (mean = 75.8%) without prior mammograms, and one study reported a 7% reduction in sensitivity with prior mammograms without providing the sensitivity values obtained. Many of these studies (5/7) indicated that availability of prior mammograms did not have any effect on reader sensitivity.[21,22,25,28] The studies that used specificity reported values ranging from 72 to 96% (mean = 87.5%) with prior mammograms and 63 to 87% (mean = 80.5%) without prior mammograms. Most of the studies (5/7) indicate that availability of prior mammograms increased specificity.[22,25,28,30,34]

Reported FPRs ranged from 3.7 to 36% (mean = 19.9%) with prior mammograms and 13.3 to 49% (mean = 31.4%) without prior mammograms. Data from the studies show that prior mammograms reduced FPR.[27,28] The only study that assessed abnormal interpretation rate reported that the availability of prior mammograms reduced AIR by 4%.[18] The two studies that examined the impact of prior mammograms using PPV showed significant improvement.[25,34]

Recall rates

The reporting of RRs is included here as the rate of screening females recalled for follow-up imaging test due to suspicion of cancer requiring further imaging and biopsy is closely linked and reported as separate entities. While there is a relationship between specificity and RRs in screening mammography, RRs are also linked to sensitivity and TP identification and have been the discrete focus of some studies included in this review.

Five of the six studies that examined RRs showed that the availability of prior mammograms reduced RRs. RRs with prior mammograms ranged from 3.8 to 57% (mean = 26.6%), which was significantly less that that reported without prior mammograms [4.9% to 67.5% (mean = 37.9%)].[19,20,30,31,33] There was limited literature on the association between access to prior mammograms and PPV of recall.[33]

DISCUSSION

The findings suggest that the availability of prior mammograms improve specificity and PPV and reduce FPRs and RRs. In the studies identified through our search, access to prior mammograms did not improve reader sensitivity, CDR, and AUC values. Limited data were available to establish the effect of prior mammographic availability on PPV of recall.

The findings above can be explained by a few factors. Breast cancer presents with an array of features including masses, stellate lesions, non-specific densities, and architectural distortions.[16] While some of these features may be easy to identify, other features such as architectural distortions and non-specific densities may be subtle and associated with benign conditions or are indeterminate.[35] The detection and discrimination of these subtle and indeterminate lesions may be challenging to radiologists regardless of prior mammogram availability,[36] if lesions are not visible in prior mammograms and do not elicit significant changes in current mammograms.[36] Also, even if a lesion is apparent in the prior and current mammograms, but comparison with prior mammograms did not show significant changes in the lesion, this radiological feature may be dismissed. Therefore, comparison with prior mammograms may only result in the interpretation that a lesion is malignant when significant changes are observed in the lesion detected in current images. Thus, it is not surprising that diagnostic performance metrics such as sensitivity and ROC AUC, which depend on readers' ability to distinguish between normal and abnormal image features, and CDR, which depend on the ability of the readers to identify and classify lesions in mammograms, do not change in the presence or absence of prior mammograms.[21,22,25,28]

Conversely, in the absence of significant differences in the radiographic features of current and prior mammograms or changes in lesion appearance between these two sets of mammograms, the radiologists' interpretation that the female has no cancer signs may increase. Such interpretation differences may explain the increased specificity and lower false-positive and abnormal interpretation rates when prior mammograms are available. This in turn affects the recall rate, as improved specificity should lead to reduced recall rates if false-positives are improved simultaneously.

Although screening mammograms are performed on asymptomatic females, prior mammograms can be considered as a form of imaging history akin to written clinical notes about a person's medical history, which may influence the pre-test probability of breast cancer, particularly if these mammograms show features suggestive of cancer. Evidence in the literature shows that embedded imaging history has a positive influence on radiologists' decisions not to recommend further assessment by reducing the number of females recalled at screening.[19,20,31,33] In Australia, females recalled at screening undergo any of, or a combination of, mammography spot and compression views, ultrasound, and DBT assessment as well as biopsy.[37] Only 1–2% of females recalled at screening return a positive result after these series of testing, and as such, increased multiple testing can increase psychosocial harms and may deter some females from

rescreening.[38] The number of females recalled at screening is used as one of the most important indicators of the performance of a breast cancer screening program.[39] Interestingly, the literature reports that the availability of prior mammograms reduced RRs and improved the PPV of recall, suggesting that prior mammograms have a positive impact on recall decisions and the females recalled benefitted from these further assessments.

Whilst the findings of the review question the role of prior mammograms in improving cancer detection, particularly in small lesions which is one of the goals of screening mammography, it supports the use of prior mammograms to reduce the number of females recalled to assessment clinics. Low FPRs and low RRs that do not affect sensitivity are also important to a screening program because they improve workflow, reduce cost and harms such as patient anxiety, overdiagnosis, and overtreatment. The availability of prior mammograms reduces RRs and FPRs and increases specificity. Surprisingly, reading mammograms with the prior images did not improve sensitivity and CDRs. While these findings suggest that prior mammograms can be considered an important strategy to improve the specificity and RRs of screening programs and reducing radiologists' assessment workload, questions will remain about the high visual workload in reviewing prior images in the screening round.

Whilst the current review has identified 15 studies that assessed the impact of prior mammograms on diagnostic efficacy, it is important to note that most of these studies focused on different performance metrics. Sensitivity and specificity were examined in only seven studies, with other metrics considered in four or fewer studies. The differences in performance metrics across studies, wide variations in the study methodologies (retrospective analysis and experimental observer performance study), sample sizes, imaging technologies, and participant characteristics limits the comparison of results across the studies reviewed. For example, retrospective studies mostly did not report the number and experience of radiologists that interpreted the mammograms and almost all the studies prior to 2011 were based on the older FS technology, with a few of these combining FS, FFDM and DBT technologies. The Digital Mammographic Imaging Screening Trial (DMIST) which compared FS and FFDM has shown that FFDM performed better than FS mammography in females with dense breasts. Also, only one of the studies in this scoping review that assessed reader performance adjusted for the effect of breast density, but did not report the results of this adjustment; hence it could be impractical to compare results of studies that use a blended technology.[40]

Other important factors that should be considered when interpreting the results of published studies include adjustments for factors that confound reader performance such as breast density, characteristics of the population and lesions included in the data set, and characteristics of the radiologists that interpreted the images. For example, 12 of the studies did not adjust for any confounding factor, and only 2 studies accounted for breast density and age, which are established factors limiting cancer detection in FFDM. In addition, even though the distribution of cancer were reported in many studies, the characteristics of the lesions included in the data sets were not described as required by the STARD guidelines.[24] Furthermore, the small numbers of readers in the published studies and lack of information on their sampling strategy limit the generalisability of the results.

The methods of assessing performance across most of the studies reviewed focused on case-based analyses which is appropriate to a screening scenario. Only two studies performed lesion-level analyses using JAFROC and LROC,[26,27] with the results of these studies insufficient to draw definite conclusions about the impact of prior mammograms on lesion sensitivity performance, although we recognise that lesion sensitivity is not generally associated with the performance of screening programs. In addition, the imaging features of cancer, such as asymmetric densities, generally require prior mammograms for identification but these were not accounted for in statistical analysis in published studies. Finally, only 46% of the studies reported the intervals between the prior and current mammograms which demonstrated wide variation and, in some cases, very short intervals between prior and current mammograms (9 months to 3 years) in published studies.[19,20,22,25,27,34] Only one study included prior mammograms obtained 3–4.5 years before the current mammograms.[28] These short intervals and variations across studies may potentially bias the outcomes of published studies. This is because slow growing cancerous lesions generally require an interval of 2–5 years to show subtle changes on mammograms and shorter intervals may limit radiologists' ability to detect subtle changes related to breast cancer, noting that national programs such as BreastScreen Australia have a recommended interval of 24 months.

The limitations of the studies reviewed highlight the need for well-designed studies that consider current technologies, assess performance at case and lesion levels, and account for all the factors that confound radiological diagnosis. Such studies may provide a better understanding of the relevance of prior mammograms in breast cancer screening programs and how their role is validated in radiologists' education for breast cancer early detection.

This scoping review has a few limitations. Firstly, the literature search was conducted in the English language and only studies published in the English language were included in the review. Secondly, many retrospective studies that used the statistics from screening populations and did not determine the index test prior to the interpretation or provide information of the radiologists that interpreted the index test were included. To our knowledge, this scoping paper is the first to review the literature on the availability of prior mammograms and diagnostic efficacy. The outcomes of this review should inform future studies to address the deficiencies in the literature and provide informed evidence for the use or prior mammograms in the clinical setting. The review of prior images has somewhat been unchallenged in radiology, always anecdotally considered a positive task; however, the evidence could be clearer.

CONCLUSION

Evidence show that access to prior mammograms reduces RRs, FPRs, abnormal interpretation rates, and increased specificity,

however this does not show an improvement in sensitivity and the CDR. Future studies that consider reader performance at case and lesion levels and account for factors that confound radiological diagnosis are needed to confirm the true effect of prior mammograms on performance in the digital era.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021; **71**: 209–49. https://doi.org/10.3322/caac.21660
2. Huang J, Chan PS, Lok V, Chen X, Ding H, Jin Y, et al. Global incidence and mortality of breast cancer: a trend analysis. *Aging* 2021; **13**: 5748–5803. https://doi.org/10.18632/aging.202502
3. WHO. Breast cancer. WHO 2021.
4. Azamjah N, Soltan-Zadeh Y, Zayeri F. Global trend of breast cancer mortality rate: A 25-year study. *Asian Pac J Cancer Prev* 2019; **20**: 88661: 2015–20: . https://doi.org/10.31557/APJCP.2019.20.7.2015
5. Duffy S, Vulkan D, Cuckle H, Parmar D, Sheikh S, Smith R, et al. Annual Mammographic screening to reduce breast cancer mortality in women from age 40 years: long-term follow-up of the UK age RCT. *Health Technol Assess* 2020; **24**: 1–24. https://doi.org/10.3310/hta24550
6. Duffy SW, Tabár L, Yen AM-F, Dean PB, Smith RA, Jonsson H, et al. Mammography screening reduces rates of advanced and fatal breast cancers: results in 549,091 women. *Cancer* 2020; **126**: 2971–79. https://doi.org/10.1002/cncr.32859
7. Hooshmand S, Reed WM, Suleiman ME, Brennan PC. Screening Mammography: diagnostic efficacy—issues and considerations for the 2020S. *Radiat Prot Dosimetry* 2021; **197**: 54–62. https://doi.org/10.1093/rpd/ncab160
8. Ekpo EU, Alakhras M, Brennan P. Errors in Mammography cannot be solved through technology alone. *Asian Pac J Cancer Prev* 2018; **19**: 291–301. https://doi.org/10.22034/APJCP.2018.19.2.291
9. Shahan CL, Layne GP. Advances in breast imaging with current screening recommendations and controversies. *Obstet Gynecol Clin North Am* 2022; **49**: 1–33. https://doi.org/10.1016/j.ogc.2021.11.001
10. Nassif AB, Talib MA, Nasir Q, Afadar Y, Elgendy O. Breast cancer detection using artificial intelligence techniques: A systematic literature review. *Artif Intell Med* 2022; **127**: 102276. https://doi.org/10.1016/j.artmed.2022.102276
11. Geller BM, Bogart A, Carney PA, Sickles EA, Smith R, Monsees B, et al. Educational interventions to improve screening Mammography interpretation: a randomized controlled trial. *AJR Am J Roentgenol* 2014; **202**: W586–96. https://doi.org/10.2214/AJR.13.11147
12. Laming D, Warren R. Improving the detection of cancer in the screening of mammograms. *J Med Screen* 2000; **7**: 24–30. https://doi.org/10.1136/jms.7.1.24
13. Burnside ES, Park JM, Fine JP, Sisney GA. The use of batch reading to improve the performance of screening Mammography. *AJR Am J Roentgenol* 2005; **185**: 790–96. https://doi.org/10.2214/ajr.185.3.01850790
14. Jackson SL, Abraham L, Miglioretti DL, Buist DSM, Kerlikowske K, Onega T, et al. Patient and Radiologist characteristics associated with accuracy of two types of diagnostic mammograms. *AJR Am J Roentgenol* 2015; **205**: 456–63. https://doi.org/10.2214/AJR.14.13672
15. Hadadi I, Rae W, Clarke J, McEntee M, Ekpo E. Breast cancer detection across dense and non-dense breasts: markers of diagnostic confidence and efficacy. *Acta Radiol Open* 2022; **11**(1): 20584601211072279. https://doi.org/10.1177/20584601211072279
16. Park JS, Kim YS, Lee SE. Mammographic and Sonographic features of breast cancer in women before 30 years of age. *Curr Med Imaging* 2020; **16**: 1161–69. https://doi.org/10.2174/1573405615666191018154704
17. Callaway MP, Boggis CRM, Astley SA, Hutt I. The influence of previous films on screening Mammographic interpretation and detection of breast carcinoma. *Clin Radiol* 1997; **52**: 527–29. https://doi.org/10.1016/s0009-9260(97)80329-7
18. Frankel SD, Sickles EA, Curpen BN, Sollitto RA, Ominsky SH, Galvin HB. Initial versus subsequent screening Mammography: comparison of findings and their Prognostic significance. *AJR Am J Roentgenol* 1995; **164**: 1107–9. https://doi.org/10.2214/ajr.164.5.7717214
19. Hakim CM, Anello MI, Cohen CS, Ganott MA, Lu AH, Perrin RL, et al. Impact of and interaction between the availability of prior examinations and DBT on the interpretation of negative and benign mammograms. *Acad Radiol* 2014; **21**: 445–49. https://doi.org/10.1016/j.acra.2013.10.015
20. Hakim CM, Catullo VJ, Chough DM, Ganott MA, Kelly AE, Shinde DD, et al. Effect of the availability of prior full-field Digital Mammography and Digital breast Tomosynthesis images on the interpretation of mammograms. *Radiology* 2015; **276**: 65–72. https://doi.org/10.1148/radiol.15142009
21. Soh BP, Lee WB, McEntee MF, Kench PL, Reed WM, Heard R, et al. Mammography test SETS: reading location and prior images do not affect group performance. *Clin Radiol* 2014; **69**: 397–402. https://doi.org/10.1016/j.crad.2013.11.010
22. Sumkin JH, Holbert BL, Herrmann JS, Hakim CA, Ganott MA, Poller WR, et al. Optimal reference Mammography: A comparison of mammograms obtained 1 and 2 years before the present examination. *AJR Am J Roentgenol* 2003; **180**: 343–46. https://doi.org/10.2214/ajr.180.2.1800343
23. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping reviews (PRISMA-SCR): checklist and explanation. *Ann Intern Med* 2018; **169**: 467–73. https://doi.org/10.7326/M18-0850
24. Simel DL, Rennie D, Bossuyt PMM. The STARD statement for reporting diagnostic accuracy studies: application to the history and physical examination. *J Gen Intern Med* 2008; **23**: 768–74. https://doi.org/10.1007/s11606-008-0583-3
25. Kim WH, Chang JM, Koo HR, Seo M, Bae MS, Lee J, et al. Impact of prior mammograms on combined reading of Digital Mammography and Digital breast Tomosynthesis. *Acta Radiol* 2017; **58**: 148–55. https://doi.org/10.1177/0284185116647211
26. Roelofs AAJ, Karssemeijer N, Wedekind N, Beck C, van Woudenberg S, Snoeren PR, et al. Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology* 2007; **242**: 70–77. https://doi.org/10.1148/radiol.2421050684
27. Taylor-Phillips S, Wallis MG, Duncan A, Gale AG. Use of prior mammograms in the transition to Digital Mammography: a performance and cost analysis. *Eur J Radiol* 2012; **81**: 60–65. https://doi.org/10.1016/j.ejrad.2010.10.025

28. Thurfjell MG, Vitak B, Azavedo E, Svane G, Thurfjell E. Effect on sensitivity and specificity of Mammography screening with or without comparison of old mammograms. *Acta Radiol* 2000; **41**: 52–56.
29. Varela C, Karssemeijer N, Hendriks JHCL, Holland R. Use of prior mammograms in the classification of benign and malignant masses. *Eur J Radiol* 2005; **56**: 248–55. https://doi.org/10.1016/j.ejrad.2005.04.007
30. Trieu PDY, Borecky N, Li T, Brennan PC, Barron ML, Lewis SJ. The impact of prior mammograms on the diagnostic performance of Radiologists in early breast cancer detection: A focus on breast density, lesion features and vendors using wholly Digital screening cases. *Cancers (Basel)* 2023; **15**(4): 1339. https://doi.org/10.3390/cancers15041339
31. Burnside ES, Sickles EA, Sohlich RE, Dee KE. Differential value of comparison with previous examinations in diagnostic versus screening Mammography. *AJR Am J Roentgenol* 2002; **179**: 1173–77. https://doi.org/10.2214/ajr.179.5.1791173
32. Wilson TE, Nijhawan VK, Helvie MA. Normal mammograms and the practice of obtaining previous mammograms: usefulness and costs Radiology. *Radiology* 1996; **198**: 661–63. https://doi.org/10.1148/radiology.198.3.8628851
33. Hayward JH, Ray KM, Wisner DJ, Kornak J, Lin W, Joe BN, et al. Improving screening Mammography outcomes through comparison with multiple prior mammograms. *AJR Am J Roentgenol* 2016; **207**: 918–24. https://doi.org/10.2214/AJR.15.15917
34. Yankaskas BC, May RC, Matuszewski J, Bowling JM, Jarman MP, Schroeder BF. Effect of observing change from comparison mammograms on performance of screening Mammography in a large community-based population. *Radiology* 2011; **261**: 762–70. https://doi.org/10.1148/radiol.11110653
35. Gaur S, Dialani V, Slanetz PJ, Eisenberg RL. Architectural distortion of the breast. *AJR Am J Roentgenol* 2013; **201**: W662–70. https://doi.org/10.2214/AJR.12.10153
36. Jaglan P, Dass R, Duhan M. Breast cancer detection techniques: issues and challenges. *J Inst Eng India Ser B* 2019; **100**: 379–86. https://doi.org/10.1007/s40031-019-00391-2
37. Hadadi I, Clarke J, Rae W, McEntee M, Vincent W, Ekpo E. Diagnostic efficacy across dense and non-dense breasts during Digital breast Tomosynthesis and ultrasound assessment for recalled women. *Diagnostics (Basel)* 2022; **12**(6): 1477. https://doi.org/10.3390/diagnostics12061477
38. Sim MJH, Siva SP, Ramli IS, Fritschi L, Tresham J, Wylie EJ. Effect of false-positive screening mammograms on Rescreening in Western Australia. *Med J Aust* 2012; **196**: 693–95. https://doi.org/10.5694/mja11.10892
39. Australia C. National cancer control indicators: recall to assessment. Canberra 2022.
40. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of Digital versus film Mammography for breast-cancer screening. *N Engl J Med* 2005; **353**: 1773–83. https://doi.org/10.1056/NEJMoa052911

Review of Chapter 2 and introduction to chapter 3 "access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance"

The literature review presented in Chapter 2 highlighted important knowledge gaps around the impact that prior mammograms have on the outcome of screening mammography (1-5). For example, of the 16 studies that previously examined the impact that prior mammograms have on diagnostic performance, only seven used current screening technologies such as digital mammography and/or digital breast tomosynthesis (1, 2). Amongst these seven studies (2, 4-9), six used older observer performance evaluation technologies and assessed the impact that prior mammograms had on radiologists' performance at the case level. Such case level analyses do not completely reflect the reporting templates used clinically to describe the presence of cancer in mammograms. Clinically, when radiologists identify features that are suggestive of breast cancer, they describe the location of the lesion and the rate risk of malignancy of the lesion. The lesion description and rating are used by screening programs to make decisions around recall for additional assessment and biopsy recommendation (10). However, previous studies only examined how the availability of prior mammograms influenced the classification of mammograms as normal or abnormal without considering correct lesion localisation and the risk of malignancy in the lesions (1-9). The only study that examined the influence of prior mammograms on the lesion level was a retrospective analysis of observer performance data of radiologists (5). In this previous study the radiologists who interpreted the test sets with prior mammograms were not always the ones who interpreted the mammograms without priors (5). The lack of data on current technologies and diagnostic performance methodologies highlights the need for studies to examine the impact that prior mammograms have on the performance of screening programs in the digital era.

Published literature also shows that the personal and workload characteristics of breast readers such as experience, number of years reading mammograms, volume of mammograms read per year, and participation in diagnostic workup influence diagnostic performance of mammograms (11). The physical characteristics of the population screened, particularly breast density, is another factor that affects the efficacy of screening mammography (12). However, data presented in Chapter 2 shows that previous studies did not adjust for the confounding effects of breast density and reader characteristics. These methodological limitations needed to be addressed to verify the impact that prior mammograms have on reader performance.

Further to the limitations identified in Chapter 2, the work in Chapter 3 examined the impact that prior mammograms have on radiologists' performance in interpreting screening mammograms and the influence that radiologists' characteristics and breast density have on their subsequent performance. The participants in Chapter 3 included one breast physician and a trainee radiologist. The breast physician routinely reports screening mammograms for BSA with 10 years of mammography reading experience and between 151 to 200 mammograms read per week. The trainee radiologist has been working in breast imaging for four years and independently reading approximately 20 mammograms per week. Therefore, these participants were included because they routinely and independently interpret mammograms in Australia. The work is published in Clinical Radiology and is entitled "Access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance" (13). The work showed that reference to prior screening mammograms improved radiologists' specificity and reduced the false positive rates without affecting sensitivity, lesion sensitivity, ROC, and JAFROC. It also showed that the influence of prior mammograms on radiologists' performance was not affected by radiologists' personal and workload characteristics such as the number of years reading mammograms, hours spent reading mammograms per week, number of years

specialised as breast radiologist, and volume of mammograms read per week. Breast density also did not affect the influence that prior mammograms have on radiologists' performance in interpreting screening mammograms. The outcome of this paper provided evidence to support policies around the retention of mammograms for future reference. The findings also highlighted the need for the creation of a national population mammography database to support the interpretation of screening mammograms of women moving across states or between screening services.

References

1. Akwo JD, Trieu P, Lewis S. Does the availability of prior mammograms improve radiologists' observer performance?-a scoping review. British Journal of Radiology Open. 2023;5(1):20230038.

2. Hakim CM, Anello MI, Cohen CS, Ganott MA, Lu AH, Perrin RL, et al. Impact of and interaction between the availability of prior examinations and DBT on the interpretation of negative and benign mammograms. Academic Radiology. 2014;21(4):445-9.

3. Roelofs AAJ, Karssemeijer N, Wedekind N, Beck C, Woudenberg Sv, Snoeren PR, et al. Importance of Comparison of Current and Prior Mammograms in Breast Cancer Screening. Radiology. 2007;242(1):70-7.

4. Taylor-Phillips S, Wallis MG, Duncan A, Gale AG. Use of prior mammograms in the transition to digital mammography: a performance and cost analysis. European Journal of Radiology. 2012;81(1):60-5.

5. Trieu PDY, Borecky N, Li T, Brennan PC, Barron ML, Lewis SJ. The Impact of Prior Mammograms on the Diagnostic Performance of Radiologists in Early Breast Cancer Detection: A Focus on Breast Density, Lesion Features and Vendors Using Wholly Digital Screening Cases. Cancers (Basel). 2023;15(4).

6. Frankel SD, Sickles EA, Curpen BN, Sollitto RA, Ominsky SH, Galvin HB. Initial versus subsequent screening mammography: comparison of findings and their prognostic significance. American Journal of Roentgenology. 1995;164(5):1107-9.

7. Kim WH, Chang JM, Koo HR, Seo M, Bae MS, Lee J, et al. Impact of prior mammograms on combined reading of digital mammography and digital breast tomosynthesis. Acta Radiologica. 2017;58(2):148-55.

8. Soh BP, Lee WB, McEntee MF, Kench PL, Reed WM, Heard R, et al. Mammography test sets: Reading location and prior images do not affect group performance. Clinical Radiology. 2014;69(4):397-402.

9. Thurfjell MG, Vitak B, Azavedo E, Svane G, Thurfjell E. Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. Acta Radiologica. 2000;41(1):52-6.

10. RANZCR. 2020 Clinical Radiology Workforce Census Report. Accessed, January 2025. Available from: https://www.ranzcr.com/college/document-library/2020-workforce-census-report-australia.

11. Wong DJ, Gandomkar Z, Lewis S, Reed W, Suleiman M, Siviengphanom S, et al. Do Reader Characteristics Affect Diagnostic Efficacy in Screening Mammography? A Systematic Review. Clinical Breast Cancer. 2023;23(3):e56-e67.

12. Ekpo EU, Alakhras M, Brennan P. Errors in Mammography Cannot be Solved Through Technology Alone. Asian Pacific Journal of Cancer Prevention. 2018;19(2):291-301.

13. Akwo JD, Trieu PDY, Barron ML, Reynolds T, Lewis SJ. Access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance. Clinical Radiology. 2024; 79(12): e1549-e1556

CHAPTER 3

Access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance.

Chapter 3 is published as Akwo JD, Trieu PD, Robinson M, Reynolds T, Lewis SJ (2024). Access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance. *Clinical. Radiology.* 79(12): e1549-e1556. DOI: https://doi.org/10.1016/j.crad.2024.09.007

This published paper is the first of three original studies performed within the institutional ethics review board approval entitled "The influence of prior cases on radiology performance with screening mammograms (Human research and ethics committee approval number: 2023/101)".

Clinical Radiology 79 (2024) e1549–e1556

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance

J.D. Akwo [a,*], P. D. (Yun) Trieu [a], M.L. Barron [a], T. Reynolds [a], S.J. Lewis [a,b]

[a] Medical Image Optimisation and Perception Group, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia
[b] School of Health Sciences, Western Sydney University, Campbelltown, Australia

ARTICLE INFORMATION

Article history:
Received 25 July 2024
Received in revised form
5 September 2024
Accepted 12 September 2024

AIMS: To establish the impact that access to prior mammograms has on radiologists' performance and the influence of radiologists' characteristics and breast density on their subsequent performance.

METHODS: Eight participants independently interpreted 72 digital screening mammograms in two reading sessions using the Royal Australian and New Zealand College of Radiologist's classification. In the first reading session, participants were given access to current and prior mammograms. In the second reading session six months later, participants only had access to the current mammograms. Radiologists' specificity, sensitivity, lesion sensitivity, Receiver Operating Characteristic (ROC) curve, and Jacknife Alternative Free-response ROC (JAFROC) were calculated. A Paired T-test was used to compare readings with and without prior mammograms, and to assess if breast density influenced participants performance. Independent Sample T-test was used to compare performance across radiologists' characteristics. A relative risk analysis was conducted to assess the probability of false positives and false negatives when prior mammograms were available.

RESULTS: Access to prior mammograms improved specificity in dense and non-dense breasts ($p \leq 0.01$) and reduced false positives ($p = 0.01$) but had no effect on sensitivity ($p = 0.37$), lesion sensitivity ($p = 0.67$), ROC ($p = 0.16$), and JAFROC ($p = 0.24$). Prior mammogram also reduced the probability of false positives (RR = 0.38; 95%CI:0.26–0.57, $p<0.0001$) without affecting the false negative rate (RR = 1.14; 95%CI:0.88–1.49, $p = 0.30$). The impact of prior mammograms on performance was not influenced by breast density or radiologists' characteristics.

CONCLUSIONS: Access to prior mammograms improves radiologists' specificity and reduces false positives without affecting sensitivity and the false negative rate regardless of radiologists' characteristics and breast density.

* Guarantor and correspondent: J. D. Akwo, Medical Image Optimisation and Perception Group, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia.
E-mail address: jakw2899@uni.sydney.edu.au (J.D. Akwo).

https://doi.org/10.1016/j.crad.2024.09.007

Background

Female breast cancer is the second most common cancer and leading cause of deaths due to cancer in women worldwide.[1,2] Early detection is key to reducing deaths from the disease.[3] To facilitate early detection, Australia established a population-based screening program for women aged 50–74 years in 1991, called BreastScreen Australia (BSA).[4] Mammography is the standard imaging tool for breast cancer screening in Australia, and the images produced are interpreted by trained readers, normally radiologists. However, the interpretation of mammographic images is challenging.[5,6] About 20–30% cancers are missed in mammograms and radiologists demonstrate wide variability in their ability to detect cancer.[5]

The mammographic diagnosis of breast cancer is made based on the identification of features of abnormality on a woman's mammogram. Sometimes, these features are subtle and may be missed or not judged to be cancer by radiologists interpreting the mammograms.[5] Breast cancer contains rapidly proliferating cells, which causes the cancer lesions to grow over time and sometimes change the architecture of the breast tissues.[7,8] Therefore, previous mammograms may serve as a baseline to identify features or changes in current mammograms suggestive of breast cancer, which could help improve diagnostic decision-making[9,10] With the digitisation of the mammography screening process, it has become easier to retrieve and compare previous and current images of women undergoing screening for abnormal beast changes. However, screening programs are also challenged by storage capacity due to the big data generated from current and previous screening rounds in the digital era. In cases where previous mammograms are available, radiologists usually spend additional time reviewing and comparing these previous mammograms to the current screening exam, which could increase their workload, both operationally and visually.

Radiologists demonstrate variability in the interpretation of screening mammograms, and experience and case load are both documented to be predictors of observer performance.[11] It is possible that some radiologists more than others may benefit from looking at both prior and current mammograms together when interpreting mammograms, but this is poorly understood. Another factor affecting the interpretation of screening mammograms is the composition of a woman's breast, with the sensitivity of mammography documented to be lower in women with extremely dense breasts compared to those with fatty breast.[12,13] Women with dense breasts are more frequently recalled at screening and often receive false positive outcomes due to summation artefacts in their mammograms.[14,15]

Screening programs aim to keep recall rates low and is generally one of the major criteria for judging the performance of the program. It is possible that comparing current and previous mammograms, particularly of women with dense breasts may also help screening programs to reduce recall rates. Therefore, identifying the characteristics of radiologists and women who benefit from prior mammograms being reviewed with current mammograms could lead to informed strategies to improve screening outcomes. This study aims to establish the impact that access to prior mammograms has on accurate interpretation of screening mammograms and to establish the characteristics of radiologists and breast compositions that benefit from prior mammograms.

Methodology

This study received institutional review board approval to conduct an observer performance study involving breast radiologists and other designated screen readers for BSA interpreting different mammographic cases at two different time points: first with access to prior mammograms and second, without access to prior mammograms (Human Research Ethics Committee Number 2023/101).

Participant recruitment

A simple convenience sampling strategy was used to recruit readers at the Breast Imaging Group (BIG) conference and the Royal Australian and New Zealand College of Radiologists (RANZCR) annual scientific meeting in 2023. Eleven radiologists with varying levels of experience and workload characteristics were initially recruited for the study and 8 participants completed both reading sessions.

Mammogram test set

A total of 72 full-field digital mammograms were sampled from the Breastscreen REeader Assessment Strategy STrategy (BREAST) databank of cases, ensuring even distribution of women of breast densities (A-B-C-D), breast cancer types and characteristics (calcifications, discrete mass, architectural distortion, and non-specific density), and ages. Briefly, BREAST was established to provide breast screening readers the opportunity to assess their performance in mammography interpretation and receive feedback on false negative and false positive errors like PERFORMS (Personal Performance in Mammographic Screening) in the United Kingdom National Health Service Breast Screening Programme (NHSBSP).[16,17] BREAST is used to support continuous professional development and serves as a national quality training tool in BSA as PERFORMS is to the NHSBSP.[16,18] Therefore, BREAST mammography test-set cases are derived from the imaging libraries of BSA like the selection of mammograms for PERFORMS are derived from the NHSBSP.[16] The presence of cancer was confirmed by at least two expert radiologists and a positive biopsy result. The normal mammograms with no cancer were confirmed to be cancer-free using negative mammograms acquired 2–4 years later. All cases had a prior mammogram available through the BREAST databank, with the interval between the current and prior mammograms ranging from 2 to 4 years to ensure that slow growing cancers developed enough to be detected. The 72 cases contained prior and

current mammograms from different vendors (General Electric, Sectra, Hologic, Fujifilm and Siemens HealthCare).

Experimental design

The cases in the test sets were uploaded to the BREAST platform for viewing and interaction. The eight participants independently interpreted the digital screening mammograms in two reading sessions using the Royal Australian and New Zealand College of Radiologists' (RANZCR) classification: 1 (normal); 2 (benign); 3 (indeterminate); 4 (suspicious); 5 (highly suspicious). If the reader detected a lesion, they then used a mouse to click on the lesion and assign a rating. If a rating of 3 or higher was assigned, the reader was asked to state the mammographic feature of the lesion detected (Fig 1). A rating of 2 denoted that reader considered the detected lesion as benign. The first reading session involved mammograms acquired in the current screening round and their priors (mammograms acquired 2–4 years prior to the current screening round). The second reading session involved the interpretation of the same 72 cases, but the prior mammograms were hidden so that the readers only had access to the current mammograms of these women. The two reading sessions were conducted at least six months apart to reduce memory effect. All cases were read on 5MP monitors and ambient lighting was kept between 15–20 lux to conform with the RANZCR standard. Prior to the assessment, each participant completed a short survey, which enabled us to collect demographic data and practice-related characteristics such as age, gender, years of experience reading mammograms, number of mammograms read per week, whether they are working in screening services. No feedback was given to the readers about their performance after the first reading session.

Statistical analysis

At the end of each the reading session, the BREAST platform automatically calculates the performance of each participant using metrics such as sensitivity, specificity, false positive rates, receiver operating characteristic (ROC) Area Under the curve (AUC), Jack-knife alternative free-response receiver operating characteristic (JAFROC), and lesion sensitivity. We collected and compared the diagnostic performance metrics for the participants in the two reading scenarios (with versus without access to prior mammograms).

Shapiro-wilk test demonstrated that data were normally distributed ($P>0.05$ for all performance metrics). Therefore, a Paired T-Test, which compares the means of two assessments from the same participants was used to compare performance with and without prior mammograms. Independent sample T-test, which compares the means of two independent groups was used to compare the performance of groups of readers with different characteristics. The thresholds used for the grouping were based on the characteristics of the participants who completed the two readings and ensured that participants were evenly distributed between groups. To assess the impact of prior mammograms across breast densities, we grouped data into dense (BI-RADS C – D) and non-dense breasts (BI-RADS A – B) and compared participants' performance with and without prior mammograms in dense and non-dense breasts. A relative risk analysis (RR) analysis was performed to assess the probability of false positive and false negative outcomes when prior mammograms were available. For the RR analysis, sample size power calculation showed that 16 true positive and 16 true negative cases, and a cohort of five participants each reading with and without prior mammograms were sufficient to yield 80% power at

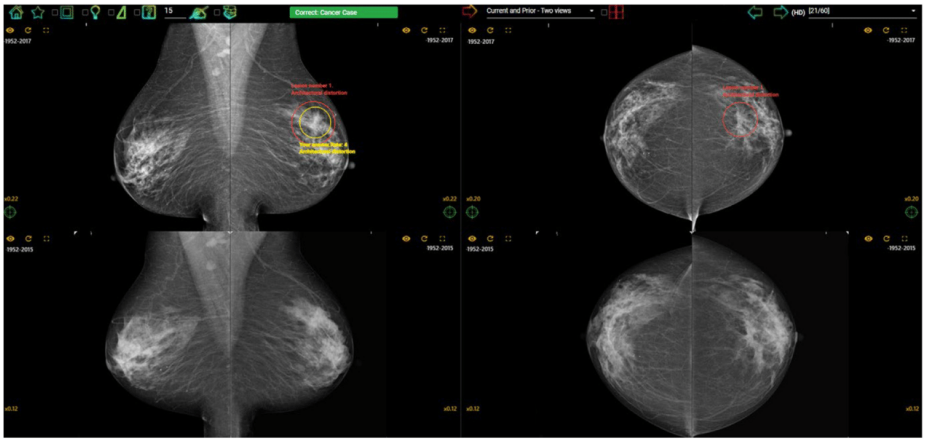


Figure 1 BREAST platform showing readers lesion marks. Yellow circle represents the reader's mark denoting the location of cancer and red circle represents the true location of the cancer.

Table 1
Participants' performance with compared to without prior mammograms.

Performance metric	With priors	Without priors	*p*-value
Specificity	91.1 (80–100)	76.3 (68–85)	0.009*
Sensitivity	67.5 (44–91)	71.3 (53–88)	0.37
Lesion sensitivity	64.1 (38–91)	66.9 (41–88)	0.67
ROC	80.7 (69.5–91.8)	77.6 (69.4–89.6)	0.16
JAFROC	78.1 (66.4–91.8)	74.7 (63.3–89.6)	0.24

ROC: Receiver Operating Characteristic curve; JAFROC: Jacknife Alternative Free-Response Receiver Operating Characteristic curve. Values in bracket represent the minimum and maximum scores.

95% confidence interval and <5% margin of error.[19] P-values ≤ 0.05 were considered statistically significant.

Results

A total of eight participants (six radiologists, one senior radiology trainee, and one breast physician) completed the two reading sessions. The years in their current roles varied from four to 35 years (mean = 20.3), with years of reading mammograms ranging from three to 35 years (Mean = 17.3). The number of hours spent reading mammograms per week varied from four to 30 hours, and the numbers of mammograms cases read per week ranged from 20 to 200.

Table 1 summarises the performance outcomes of the eight participants 'with prior mammograms' compared to 'without prior mammograms'. Only specificity was significantly higher when radiologists had access to prior mammograms compared to when prior mammograms were not available. True negatives were significantly higher with prior availability compared to without prior mammograms (mean = 36.4 vs. 30.5; p=0.01) and false positives were significantly lower with priors compared to without priors (mean = 3.6 vs 9.5; p = 0.01).

The probability of a false positive result was significantly lower when participants had access to prior mammograms (RR= 0.38; 95%CI: 0.26–0.57, p<0.0001). The probability of a false negative screening result was not significantly different between readings with and without prior mammograms (RR = 1.14; 95%CI: 0.88–1.49, p = 0.30).

When the analyses were stratified by breast density, participants' performance was not significantly different between dense and non-dense breast cases (p>0.05 for specificity, sensitivity, and lesion sensitivity). However, specificity was significantly higher when radiologists had access to prior mammograms in dense (p = 0.01) and non-dense (p = 0.01) compared to when prior mammograms were not available (Table 2). No differences were detected in sensitivity and lesion sensitivity when readings with priors were compared to those without prior mammograms.

Independent sample T-test showed that participants with >20 years of specialisation as breast radiologists or as breast physicians demonstrated significantly higher specificity (p=0.04) and true negatives (p=0.04), and lower false positives (p=0.04) when prior mammograms were available compared to those with ≤ 20 years of specialisation. No differences were observed between participants with different characteristics in other performance metrics (Table 3). No significant differences were observed between these groups of participants when prior mammograms were not available. There were also no differences between participants who spent ≤4 hours per week and > 4 hours per week reading mammograms in all performance metrics regardless of whether prior mammograms were available.

When prior mammograms were available, no significant differences were found between participants who have read mammograms for ≤ 10 years and those who had more than 10 years of mammography reading experience. Conversely, when prior mammograms were not available, participants with more than 10 years of mammography reading experience demonstrated better sensitivity, lesion sensitivity, ROC, JAFROC, and true positives (p≤0.05 for all). These group of participants also demonstrated fewer false negatives (p = 0.05). We did not find any significant differences between participants who read fewer cases (≤150) per week and those who reader more than 150 cases per week.

When the readings with and without prior mammograms were compared for participants employment characteristics, specificity emerged as the only performance metric that discriminated readings with versus without prior mammograms (Table 4). Specificity was significantly higher for participants with greater than 20 years of experience as radiologist or breast physician (p = 0.008), those who spent >4 hours/week (p = 0.05) or ≤4 hours/week (p = 0.02), have >10 years (p = 0.05) or ≤10 years (p = 0.01) of mammography reading experience, and those who read more than 150 mammograms per week (p = 0.04) when prior mammograms were available compared with when prior mammograms were not available.

Discussion

The findings demonstrate that access to mammograms from the previous screening round improves radiologists' ability to correctly identify women who have no signs of cancer on their images without affecting the detection and classification breast cancer lesions in women who have

Table 2
Comparison of participants' performance in dense and non-dense breasts with compared to without prior mammograms.

Performance	Dense breasts (BI-RADS C – D)			Non-dense breasts (BI-RADS A – B)		
	With priors	Without priors	*p*-value	With priors	Without priors	*p*-value
Specificity	92.5 (80–100)	76.9 (60–95)	0.01*	92.5 (80–100)	75.6 (60–90)	0.01*
Sensitivity	66.4 (50–93.8)	68.8 (50–87.5)	0.88	68.7 (37.5–93.8)	73.4 (56.3–87.5)	0.37
L. Sensitivity	62.5 (43.8–87.5)	64.8 (37.5–87.5)	0.66	65.6 (31.3–93.8)	68.8 (43.8–87.5)	0.62

L.Sensitivity: lesion sensitivity. Values in bracket represent the minimum and maximum scores.

Table 3
Comparison of the performance of participants with different characteristics across readings with and without prior mammograms.

Performance metric	With priors			Without priors		
	Years qualified as radiologist or breast physician/specialty					
	≤20 years (n=4)	>20 years (n=4)	*p*-value	≤20 years (n=4)	>20 years (n=4)	*p*-value
Specificity	86.5 (80–90)	95.8 (88–100)	0.04*	75.6 (68–85)	76.8 (68–85)	0.87
Sensitivity	71.8 (62–78)	63.3 (44–91)	0.45	68.6 (53–78)	73.6 (66–88)	0.54
L. Sensitivity	67.3 (56–72)	61 (38–91)	0.61	61.8 (41–75)	72 (62–88)	0.35
ROC	81.4 (75.5–79)	80.2 (69.5–91.8)	0.82	76.8 (69.4–82.8)	78.5 (70.3–89.6)	0.76
JAROC	77 (72.5–79)	79 (66.4–91.8)	0.75	71.2 (63.3–82.8)	77.6 (68.3–89.6)	0.39
	Hours spent reading mammograms per week					
	≤4 hours/week (n=4)	>4 hours/week (n=4)	p	≤4 hours/week (n=4)	>4 hours/week (n=4)	*p*-value
Specificity	91.5 (88–100)	90.6 (80–100)	0.89	75.8 (68–85)	76.8 (68–85)	0.87
Sensitivity	68.5 (62–78)	66.5 (44–91)	0.86	68.8 (53–78)	73.8 (66–88)	0.53
L. Sensitivity	64 (59–72)	64.3 (38–91)	0.98	61 (41–75)	72.8 (62–88)	0.27
ROC	81.5 (75.5–85.6	80 (65.5–91.8)	0.78	76.3 (69.4–80.9)	78.9 (70.3–89.6)	0.63
JAROC	77.3 (72.5–81.8)	78.9 (66.4–91.8	0.78	70.9 (63.3–79.3)	78.5 (68.3–89.6)	0.26
	Number of years reading mammograms					
	≤10 years (n=4)	>10 years (n=4)	p	≤10 years (n=4)	>10 years (n=4)	*p*-value
Specificity	90.3 (88–95)	91.6 (80–100)	0.82	74 (68–82)	77.6 (68–85)	0.57
Sensitivity	61.3 (44–91)	71.2 (59–72)	0.39	62.7 (53–88)	76.4 (66–78)	0.05*
L. Sensitivity	55.3 (38–91)	69.4 (56–72)	0.25	54.3 (41–88)	74.4 (62–75)	0.04*
ROC	76.7 (69.5–91.8)	83.1 (79.1–85.6)	0.22	72.2 (69.4–89.6)	80.9 (70.3–82.8)	0.04*
JAROC	73.6 (66.4–91.8)	80.7 (75–79.7)	0.20	67.6 (63.3–89.6)	78.9 (68.3–82.8)	0.05*
	Number of mammograms read per week					
	≤150 reads/week (n=3)	>150 reads/week; n=5	p	≤150 reads/week; n=3	>150 reads/week; n=5	p-value
Specificity	90.6 (88–100)	92 (95–100)	0.80	75 (68–85)	78.3 (68–85)	0.60
Sensitivity	68.2 (59–78)	66.3 (44–91)	0.87	74.6 (53–75)	65.7 (66–88)	0.26
L. Sensitivity	65.2 (56–72)	62.3 (38–91)	0.82	73.2 (41–72)	56.3 (62–88)	0.10
ROC	81.1 (75.5–85.2)	80.1 (69.5–91.8)	0.85	79.3 (69.4–80.9)	74.8 (70.3–89.6)	0.42
JAROC	78.1 (72.5–81.8)	78 (66.4–91.8)	0.99	77.7 (63.3–79.3)	69.6 (68.3–89.6)	0.24

L. Sensitivity: lesion sensitivity; ROC: receiver operating characteristics curve; JAFROC: Jacknife Alternative Free-response receiver operating characteristics curve; *significantly different. Values in bracket represent the minimum and maximum scores.

cancer. Data generated also show that the availability of prior mammograms for radiologists to view reduces the chances of false positive results for women undergoing screening. After adjustments were made for breast density and radiologists' characteristics such as years of specialisation, number of years reading mammograms, volume of mammograms read per week, and hours spent reading mammograms, the availability of prior mammograms consistently led to improvement in specificity and a reduction in false positive rate without affecting the false negative rate. It is important to note that while sensitivity and lesion sensitivity were lower with priors, they were not significantly different from the values obtained when readings occurred without priors. Also, despite the slightly lower sensitivity and lesion sensitivity with priors, ROC and JAFROC, which measures observers' abilities to detect the lesion and simultaneously rate the level of malignancy were higher when prior mammograms were available even though these were not significantly greater than chance.

These findings can be explained because breast cancer elicit perturbations in mammograms, which mimic normal breast parenchyma or benign changes.[7,20] However, unlike perturbations from normal breast parenchyma or benign lesions, the perturbations due to breast cancer gradually change overtime as the cancer grows or invades surrounding tissues.[7,8] Therefore, comparison with prior mammograms should help radiologists detect changes in mammographic features that are most discriminative of breast cancer. These findings are relevant because they suggest that viewing prior mammograms can improve the efficiency of screening programs by ensuring that women who have no cancer are correctly informed thereby reducing unnecessary recall and testing. However, due to population mobility across state lines, access to prior images is not always available to readers. The findings support screening guidelines and policies that recommend that mammograms of women undergoing screening should be retained for future reference.[21] Whilst archiving, cost, and retention issues are factors to consider when retaining prior mammograms,[22] the utility of prior mammograms in reducing false positives,[23,24] which mitigates psychosocial harms and cost of assessments for women wrongly recalled emphasise the importance of prior mammograms to screening programs.[25,26]

Our findings align with published evidence which shows that availability of prior mammograms to radiologists improves specificity and reduce false positives and recall rates without affecting cancer detection and sensitivity.[24,27–32] For example, experimental observer performance and retrospective studies have reported a 4–14.5%

Table 4
Comparison between readings with prior compared to without prior mammograms for participants of similar characteristics.

	Years qualified as radiologist or breast physician/specialty					
	With priors (>20 years; n=4)	Without priors (>20 years; n=4)	*p*-value	With priors (≤20 years; n=4)	Without priors (≤20 years; (n=4)	*p*-value
Specificity	95. (95–100)	76.8 (72–85)	0.008*	86.5 (80–90)	75.8 (68–85)	0.07
Sensitivity	63.3 (44–91)	73.8 (66–88)	0.38	71.8 (62–78)	68.8 (53–78)	0.66
L. Sensitivity	61 (38–91)	72 (62–88)	0.41	67.3 (56–72)	61.8 (41–75)	0.56
ROC	80 (70–92)	78.5 (73–90)	0.79	81.4 (76–86)	76.8 (69–83))	0.28
JAROC	79 (66–92)	77.6 (68–90)	0.84	77.1 (73–82)	71.7 (63–83)	0.11
	Hours spent reading mammograms per week					
	With priors (>4 hours/week; n=4)	Without priors (>4 hours/week; n=4)	*p*-value	With priors (≤4 hours/week; n=4)	Without priors (≤4 hours/week; n=4)	*p*-value
Specificity	90.8 (80–100)	76.8 (68–85)	0.05*	91.5 (88–100)	75.6 (68–85)	0.02*
Sensitivity	66.5 (44–91)	73. (66–88)8	0.54	68.5 (59–78)	68.8 (53–78)	0.97
L. Sensitivity	64.3 (38–91)	72.8 (62–88)	0.52	64 (56–72)	61 (41–75)	0.74
ROC	80 (70–92)	79 (70–90)	0.87	81.5 (76–86)	76.3 (69–81)	0.20
JAROC	78.9 (66–92)	78.5 (68–90)	0.95	77.3 (73–80)	70.9 (63–79)	0.18
	Number of years reading mammograms					
	With priors (>10 years; n=4)	Without priors (>10 years; n=4)	*p*-value	With priors (≤10 years; n=4)	Without priors (≤10 years; n=4)	*p*-value
Specificity	92.5 (80–100)	76.5 (68–85)	0.05*	89.8 (88–95)	76 (68–82)	0.01*
Sensitivity	66.3 (59–75)	73.5 (66–78)	0.19	68.8 (44–91)	69 (53–88)	0.98
L. Sensitivity	64 (56–72)	71 (62–75)	0.20	64.3 (38–91)	62.8 (56–88)	0.92
ROC	81 (79–86)	78.7 (70–83)	0.49	80.5 (70–92)	76.8 (69–90)	0.57
JAROC	78 (75–80)	76.2 (68–83)	0.61	78.1 (66–92)	73.1 (63–90)	0.55
	Number of mammograms read per week					
	With priors (>150 reads/week; n=5	Without priors (>150 reads/week; n=5	*p*-value	With priors ≤150 (reads/week; n=3	Without priors (≤150 reads/week; n=3	*p*-value
Specificity	91 (80–100)	73.7 (68–85)	0.04*	86.3 (88–100)	77.3 (82–85)	0.19
Sensitivity	68 (44–75)	73 (66–88)	0.40	68.2 (59–78)	70 (53–75)	0.89
L. Sensitivity	66.2 (38–91)	69.7 (62–88)	0.56	64.6 (56–72)	61.7 (41–72)	0.85
ROC	81.4 (70–92)	77.3 (70–90)	0.22	79 (76–85)	77.7 (69–81)	0.85
JAROC	78.9 (66–92)	74 (68–90)	0.13	77.1 (73–82)	73.1 (63–79)	0.64

L Sensitivity: lesion sensitivity; ROC: receiver operating characteristics curve; JAFROC: Jacknife Alternative Free-response receiver operating characteristics curve; *significantly different. Values in bracket represent the minimum and maximum scores.

improvement in specificity, no significant change in sensitivity,[24,27–31] and 8–13.5% reduction in false positive rates[23,24] when prior mammograms were available. This current study found a 14.8% improvement in specificity and a 6% reduction in the false positives rate with prior mammograms. Relative risk analysis demonstrated a 62% reduction in the absolute probability of false positives when prior mammograms were available. It is worth noting that assessment of sensitivity, specificity, and false positive rates involve classification tasks, which requires observers to state whether mammograms contain cancer without indicating the location of cancer in the mammogram. The JAFROC and lesion sensitivity methodologies used in our study capture observers correct lesion detection and the level of malignancy. The only previous study that has used these methodologies compared the performance of radiologists who read different test sets, some with and others without prior mammograms,[30] which may not completely capture the impact of prior mammograms on performance. This study is the first to examine the impact of prior mammograms where the same cohort of participants read the same mammograms with and without prior images, analysed performance with observer performance methodologies at the level of the lesion, and adjusted for factors that could influence their performance. Therefore, our findings account for both the detection and malignancy rating of cancer lesions and provide stronger evidence for the impact of prior examinations in identifying the normal cases better than cancer cases via screening mammograms.

The literature shows that factors such as the characteristics of women screened and radiologists who interpret these mammograms influence observer performance in interpreting screening mammograms.[11,13] Studies based BREAST and PERFORMS have demonstrated the impact of these platforms in identifying the influence that technology, patient, and radiologists' characteristics have on performance.[17,18,33] For example, analysis of data of over 400 PERFORMS participants showed that individual and practice-related characteristics were the most discriminative of high and low performing screen readers.[33] To account for these confounding factors, we adjusted for breast density and radiologists' characteristics such as years of experience, years reading mammograms, number of mammograms read per week, and hour spent reading mammograms using the demographic and practice-related information self-reported by participants in the BREAST platform. Regardless of the

adjustments that were made, access to prior mammograms consistently improved specificity without affecting sensitivity, lesion sensitivity, ROC, and JAFROC. When readings without prior mammograms were considered separately for each reader characteristic, only radiologists with greater than 10 years of experience were able to outperform their colleagues with fewer years of experience in terms of sensitivity, lesion sensitivity, ROC, and JAFROC (Table 3). These findings confirm that the improvement observed in specificity were not due to reader characteristics or the composition of the breast of women whose mammograms were used for the study as demonstrated by lack of difference in performance between dense and non-dense breasts. Rather, this study indicates that improved performance was due to the availability of prior mammograms as shown in Tables 2 and 4. These findings highlight the importance of prior mammograms to all radiologists regardless of their years of experience and workload-related characteristics, and women of all breast densities, and we consider this the first time such an effect has been shown.

A few limitations of the study must be acknowledged. The current study was based on an enriched mammography test-set with even distribution of breast densities. While this may not be representative of the clinical environment, the test-set design ensures that all cancer lesion types and breast densities were adequately represented, and accounts for the smaller number of images used in experimental test-set observer performance studies. It should also be noted that studies based on BREAST and PERFORMS have shown that test-set performance can predict screen readers' performance in a real-life clinical setting.[27,34,35] Therefore, our findings can be used to support the use of priors in screening programs and the establishment of a national reference mammography database for women participating in screen mammography. Secondly, only the assessments of eight participants who completed both test sets (readings with and without prior mammograms) were used in the analysis. However, this number is above the 75th percentile of the participant sample sizes used in similar previous studies.[24,27–32] Three participants did not respond to communication to complete the second test set and were therefore excluded from the study. The number of cases (n=72) could limit the generalisability of findings; however, it was important to develop an experimental protocol that could be completed in a reasonable time (2–3 hours per test set) so as not to fatigue the readers. Nonetheless, the sample of cases and readers in this study fall within the range recommended for phase II observer performance studies of medical imaging.[19,36]

Conclusion

Availability of prior mammograms to radiologists improves specificity and reduces false positive rates without reducing the detection and characterisation of breast cancer in mammograms. Therefore, the simultaneous display of prior and current mammograms may be a useful strategy to improve the efficiency of screening programs by reducing the number of women who are incorrectly called back to assessment clinics.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contribution

1. Guarantor of integrity of the entire study: Sarah J. Lewis.

2. Study concepts and design: Judith D. Akwo, Phuong Dung (Yun) Trieu, Melissa L. Barron, Tess Reynolds, Sarah J. Lewis.

3. Literature research: Judith D. Akwo.

4. Clinical studies: N/A.

5. Experimental studies/data analysis: Judith D. Akwo, Phuong Dung (Yun) Trieu, Melissa L. Barron, Tess Reynolds, Sarah J. Lewis.

6. Statistical analysis: Judith D. Akwo, Phuong Dung (Yun) Trieu.

7. Manuscript preparation: Judith D. Akwo.

8. Manuscript editing: Judith D. Akwo, Phuong Dung (Yun) Trieu, Melissa L. Barron, Tess Reynolds, Sarah J. Lewis.

Conflict of interest

The authors declare no conflict of interest.

References

1. Azamjah N, Soltan-Zadeh Y, Zayeri F. Global trend of breast cancer mortality rate: a 25-year study. *Asian Pac J Cancer Prev* 2019;**20**:2015–20.
2. Huang J, Chan PS, Lok V, *et al.* Global incidence and mortality of breast cancer: a trend analysis. *Aging (Albany NY)* 2021;**13**:5748–803.
3. Morrell S, Taylor R, Roder D, *et al.* Mammography screening and breast cancer mortality in Australia: an aggregate cohort study. *J Med Screen* 2012;**19**:26–34.
4. Olver IN, Roder D. History, development and future of cancer screening in Australia. *Public Health Res Pract* 2017;**27**.
5. Ekpo EU, Alakhras M, Brennan P. Errors in mammography cannot be solved through technology alone. *Asian Pac J Cancer Prev* 2018;**19**:291–301.
6. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? Insights into imaging 2017;**8**(1):171–82.
7. Stachs A, Stubert J, Reimer T, *et al.* Benign breast disease in women. *Dtsch Arztebl Int* 2019;**116**:565–74.
8. Cho N. Imaging features of breast cancer molecular subtypes: state of the art. *J Pathol Transl Med* 2021;**55**:16–25.
9. Hakim CM, Catullo VJ, Chough DM, *et al.* Effect of the availability of prior full-field digital mammography and digital breast tomosynthesis images on the interpretation of mammograms. *Radiology* 2015;**276**:65–72.
10. Hayward JH, Ray KM, Wisner DJ, *et al.* Improving screening mammography outcomes through comparison with multiple prior mammograms. *AJR Am J Roentgenol* 2016;**207**:918–24.
11. Wong DJ, Gandomkar Z, Lewis S, *et al.* Do reader characteristics affect diagnostic efficacy in screening mammography? A systematic review. *Clin Breast Cancer* 2023;**23**:e56–67.
12. Hadadi I, Rae W, Clarke J, *et al.* Diagnostic performance of adjunctive imaging modalities compared to mammography alone in women with

non-dense and dense breasts: a systematic review and meta-analysis. *Clin Breast Cancer* 2021;**21**:278–91.
13. Hadadi I, Rae W, Clarke J, *et al.* Breast cancer detection across dense and non-dense breasts: markers of diagnostic confidence and efficacy. *Acta Radiol Open* 2022;**11**:20584601211072279.
14. Hadadi I, Clarke J, Rae W, *et al.* Diagnostic efficacy across dense and non-dense breasts during digital breast tomosynthesis and ultrasound assessment for recalled women. *Diagnostics (Basel)* 2022;**12**.
15. Payne NR, Hickman SE, Black R, *et al.* Breast density effect on the sensitivity of digital screening mammography in a UK cohort. *Eur Radiol* 2024, https://doi.org/10.1007/s00330-024-10951-w.
16. Gale A, Chen Y. A review of the PERFORMS scheme in breast screening. *Br J Radiol* 2020;**93**:20190908.
17. Suleiman ME, Rickard M, Brennan PC. Perfecting detection through education. *Radiography (Lond)* 2020;**26**:S49–53.
18. 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**:352–8.
19. Wang X, Ji X. Sample size estimation in clinical research: from randomized controlled trials to observational studies. *Chest* 2020;**158**:S12–20.
20. Fraker JL, Clune CG, Sahni SK, *et al.* Prevalence, impact, and diagnostic challenges of benign breast disease: a narrative review. *Int J Womens Health* 2023;**15**:765–78.
21. Albus K. Record Keeping: Mammography. 2019 (Revised 12/12/2019) [cited 2024/09/07]; https://accreditationsupport.acr.org/support/solutions/articles/11000051613-record-keeping-mammography-revised-12-12-19-.
22. Mezrich JL, Siegel E. Who owns the image? Archiving and retention issues in the digital age. *J Am Coll Radiol* 2014;**11**:384–6.
23. Taylor-Phillips S, Wallis MG, Duncan A, *et al.* Use of prior mammograms in the transition to digital mammography: a performance and cost analysis. *Eur J Radiol* 2012;**81**:60–5.
24. Thurfjell MG, Vitak B, Azavedo E, *et al.* Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. *Acta Radiol* 2000;**41**:52–6.
25. Roelofs AAJ, Karssemeijer N, Wedekind N, *et al.* Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology* 2007;**242**:70–7.
26. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. *Ann Fam Med* 2013;**11**:106–15.
27. Soh BP, Lee WB, McEntee MF, *et al.* Mammography test sets: reading location and prior images do not affect group performance. *Clin Radiol* 2014;**69**:397–402.
28. Roelofs AA, Karssemeijer N, Wedekind N, *et al.* Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology* 2007;**242**:70–7.
29. Kim WH, Chang JM, Koo HR, *et al.* Impact of prior mammograms on combined reading of digital mammography and digital breast tomosynthesis. *Acta Radiologica* 2017;**58**:148–55.
30. Trieu PDY, Borecky N, Li T, *et al.* The impact of prior mammograms on the diagnostic performance of radiologists in early breast cancer detection: a focus on breast density, lesion features and vendors using wholly digital screening cases. *Cancers (Basel)* 2023;**15**.
31. Varela C, Karssemeijer N, Hendriks JH, *et al.* Use of prior mammograms in the classification of benign and malignant masses. *Eur J Radiol* 2005;**56**:248–55.
32. Akwo JD, Trieu P, Lewis S. Does the availability of prior mammograms improve radiologists' observer performance?-a scoping review. *BJR Open* 2023;**5**:20230038.
33. Scott HJ, Gale AG. What makes a good mammographer? Looking at the homogeneity of those groups who perform best and least well on the PERFORMS self-assessment scheme. *Breast Cancer Res* 2008;**10**:P74.
34. Qenam BA, Li T, Alshabibi A, *et al.* Test-set results can predict participants' development in breast-screen cancer detection: an observational cohort study. *Health Sci Rep* 2024;**7**:e2161.
35. Chen Y, James JJ, Cornford EJ, *et al.* The relationship between mammography readers' real-life performance and performance in a test set-based assessment scheme in a national breast screening program. *Radiol Imaging Cancer* 2020;**2**:e200016.
36. Obuchowski NA. How many observers are needed in clinical studies of medical imaging? *Am J Roentgenology* 2004;**182**:867–9.

Introduction to chapter 4 "Does access to prior mammograms improve the performance of radiographers in interpreting screening mammograms?"

The interpretation of screening mammograms is traditionally the role of breast radiologists. However, with the increase in the number of women participating in breast cancer screening and the higher volume of screening mammograms and assessment images acquired per day, the interpretation workload has increased over the last three decades (1, 2). The high workload of mammography interpretation is compounded by a shortage of breast radiologists. As described in Chapter 1, the global shortage of breast radiologists has persisted, with national workforce surveys in the UK and Australia highlighting breast radiologists' workforce gaps ranging from 30 – 35% (3). The workforce shortages appear to be getting worse with breast radiologists' retirement rates disproportionately higher than the growth rate. For example, the UK the national data on retirement rates indicated that 24% of breast radiologists would retire in 2025, but only 1% annual increase in the number of breast radiologists (3). In Australia, the breast radiologist's workforce involved in the national breast screening program has also been shown to consistently decline over the last five years (4, 5). These workforce shortages have reinforced the need for alternative solutions to mitigate screening mammography interpretation workload challenges.

It has been suggested that the breast radiologists' workload shortages can be filled by non-medical personnel (6, 7). The utilisation of non-medical personnel was first explored in the USA followed by the UK (6). Since radiographers are the first healthcare personnel to interact with women participating in screening and acquire the screening mammograms, the UK developed advance practice training programs for radiographers to undertake mammography interpretation in the 1990s (6). Since then, the National Health Service Breast Screening Program (NHSBSP) has utilised radiographers as breastscreen readers (6, 8, 9). The practice

of radiographers interpreting mammograms is gaining attention in the literature. However, the literature review presented in Chapter 2 shows that previous studies that examined the impact that prior mammograms have on the accuracy of mammography interpretation were based on radiologists and breast physicians, or do not distinguish radiography readers in the data (10). Given the differences between the training of radiographers and medical personnel, it is important to examine the impact that prior mammograms have on radiographers' interpretation of screening mammograms.

To address the gaps identified in the literature (10), the study presented in Chapter 4 examined for the first time the impact that prior mammograms have on Australian radiographers' interpretation of screening mammograms and the influence of radiographer characteristics on their subsequent performance. The study is published in the Journal, *Radiography* and is entitled "Does access to prior mammograms improve the performance of radiographers in interpreting screening mammograms?" (11). The test set used for this study only included digital mammograms. Digital breast tomosynthesis (DBT) images were not included because DBT is mostly used as an adjunct to digital mammography and for the assessment of women recalled at screening. Findings from the study show that the availability of prior mammograms to radiographers improved specificity, ROC, and JAFROC, and reduced the false positive rate without affecting sensitivity and lesion sensitivity. It also showed that the impact that prior mammograms have on performance is not affected by the number of years qualified as radiographer and the number of hours spent working in a mammography service. The outcome of the study highlighted the need for curricular innovations for radiographers that incorporate systematic analysis of prior mammograms when interpreting mammograms from the current screening round to reduce radiographers' false positive rates.

References

1. Smith-Bindman R, Miglioretti DL, Rosenberg R, Reid RJ, Taplin SH, Geller BM, et al. Physician Workload in Mammography. American Journal of Roentgenology. 2008;190(2):526-32.

2. Wing P, Langelier MH. Workforce Shortages in Breast Imaging: Impact on Mammography Utilization. American Journal of Roentgenology. 2009;192(2):370-8.

3. RCR. The Breast Imaging and Diagnostic Workforce in the United Kingdom: results of survey of NHS breast screening programme units and radiology departments. Available from: https://www.rcr.ac.uk/our-services/all-our-publications/clinical-radiology-publications/the-breast-imaging-and-diagnostic-workforce-in-the-united-kingdom/

4. Jeganathan S. The Growing Problem of Radiologist Shortages: Australia and New Zealand's Perspective. Korean Journal of Radiology. 2023;24(11):1043-5.

5. RANZCR. 2020 Clinical Radiology Workforce Census Report. Available from: https://www.ranzcr.com/college/document-library/2020-workforce-census-report-australia.

6. Culpan AM. Radiographer involvement in mammography image interpretation: A survey of United Kingdom practice. Radiography. 2016;22(4):306-12.

7. Bennett RL, Blanks RG, Moss SM. Workforce issues in breast imaging: radiographers as screen readers: Breast Cancer Research. 2006;8(Suppl 1):P31.

8. Chen Y, James JJ, Michalopoulou E, Darker IT, Jenkins J. Performance of Radiologists and Radiographers in Double Reading Mammograms: The UK National Health Service Breast Screening Program. Radiology. 2023;306(1):102-9.

9. Williams B, Brown PN. An evaluation of the current role of radiographer film readers in breast screening: Breast Cancer Research. 2002;4(Suppl 1):60.

10. Akwo JD, Trieu P, Lewis S. Does the availability of prior mammograms improve radiologists' observer performance?-a scoping review. British Journal of Radiology Open. 2023;5(1):20230038.

11. Akwo JD, Trieu PD, Barron ML, Reynolds T, Lewis SJ. Does access to prior mammograms improve the performance of radiographers in interpreting screening mammograms? Radiography. 2025;31(1):247-53.

CHAPTER 4

Does access to prior mammograms improve the performance of radiographers in interpreting screening mammograms?

Chapter 4 is published as Akwo JD, Trieu PD, Robinson M, Reynolds T, Lewis SJ (2024). Does access to prior mammograms improve the performance of radiographers in interpreting screening mammograms? *Radiography*; 31(1): 247-253. Doi: https://doi.org/10.1016/j.radi.2024.11.025.

This published paper is the second of the three original studies performed within the institutional ethics review board approval entitled "The influence of prior cases on radiology performance with screening mammograms (Human research and ethics committee approval number: 2023/101)".

Radiography 31 (2025) 247–253

Contents lists available at ScienceDirect

Radiography

journal homepage: www.elsevier.com/locate/radi

Does access to prior mammograms improve the performance of radiographers in interpreting screening mammograms?

J.D. Akwo [a, *], P.D. (Yun) Trieu [a], M.L. Barron [a], T. Reynolds [c], S.J. Lewis [a, b]

[a] Medical Image Optimisation and Perception Group, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia
[b] School of Health Sciences, Western Sydney University, Campbelltown, Australia
[c] Image X Institute, School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Camperdown, Australia

ARTICLE INFO

Article history:
Received 8 September 2024
Received in revised form
28 October 2024
Accepted 26 November 2024

Keywords:
Prior mammograms
Previous mammograms
Initial mammograms
Radiographer mammography interpretation

ABSTRACT

Introduction: The impact of previous screening mammograms on radiographers' performance in mammography interpretation is unknown. This study assesses the impact that previous screening mammograms has on radiographers' interpretation of mammograms.
Methods: Thirteen Australian radiographers working for the national breast screening service independently interpreted a mammography test-set containing mammograms of 28 women based on the Royal Australian and New Zealand College of Radiologists' classification. Twelve radiographers completed the "No prior test-set" (no previous mammograms available) while one radiographer completed the "Prior test-set" (most current screening mammograms with access to previous mammograms) in the first reading session. In the second reading session, 12 radiographers completed the "Prior test-set" and one radiographer completed the "No prior test-set". Their performance with and without previous mammograms were calculated and compared.
Results: The availability of prior mammograms significantly improved specificity [81(range:58–95) vs. 60(range:37–79); $p = 0.002$], ROC [91(range:80–99) vs. 82 (range:57–91); $p = 0.003$], and JAFROC 87(range:73–99) vs. 79 (range:52–91); $p = 0.01$]. Prior mammograms also significantly reduced false positives ($p = 0.002$). No differences were observed between readings with and without previous mammograms in terms of sensitivity ($p = 0.70$) and lesion sensitivity ($p = 0.82$). Years qualified as a radiographer did not modify the influence of previous mammograms on specificity, ROC, and false positives. Years specialised as breast radiographer slightly modified the influence of previous mammograms in radiographers with ≥25 years of experience but not those with <25 years of experience as breast radiographers.
Conclusions: The availability of previous screening mammograms improves radiographers' ability to discriminate between normal and abnormal mammograms and reduce the false positive rate without affecting the detection of breast cancer.
Implications for practice: The findings highlight the need for practices to store screening mammograms and for radiographers to actively refer to previous screening mammograms when interpreting mammograms from the current screening round. It also highlights the need for policies to establish a national accessible mammographic database platform for integrated clinics and to account for population mobility across states.

Introduction

Current health trends suggest that one in eight women in developed countries will be diagnosed with breast cancer in their lifetime.[1,2] Breast cancer accounts for one in three new cases of cancers in women, and one in six women affected with breast cancer will die from the disease.[1,3,4] Given the obvious health burden the World Health Organisation (WHO), through the Global Breast Cancer Initiative, seeks to achieve a 2.5 % reduction in breast cancer deaths by 2040, which will save the lives of an estimated 2.5 million women worldwide.[5] To achieve such a significant reduction in breast cancer deaths, a combination of different approaches including early detection and improvement in treatment strategies and patient care is required.[6,7] The literature shows lower mortality and higher survival rates for women whose cancer were detected

* Corresponding author.
E-mail address: jakw2899@uni.sydney.edu.au (J.D. Akwo).

https://doi.org/10.1016/j.radi.2024.11.025

early.[8,9] Thus, many countries have implemented population breast cancer screening programs and invite eligible women for mammography screening. As an example, BreastScreen Australia invites women ages 50–74 years for biennial imaging. Some countries provide optional screening to women in their forties and supplemental screening using magnetic resonance imaging (MRI) for women at higher risk of developing breast cancer.[10,11]

Early detection through screening mammography involves the interpretation of mammograms, usually by radiologists, with some other healthcare practitioners such as breast physicians and radiographers also interpreting mammograms. Evidence shows that the interpretation of screening mammograms is a challenging task, with about 20–30 % of breast cancers missed.[12,13] However these numbers vary depending on the characteristics of the women screened such as breast density, breast size, body mass index and the experience and expertise of the radiologist who interpreted the images.[12] A major challenge faced by screening programs around the world is the limited number of radiologists to interpret mammograms.[14–16] In the United Kingdom (UK), for example, the workforce interpreting screening mammograms includes radiographers as second readers and as autonomous advanced practitioners.[17,18] The number of breast radiologists in Australia is below the workforce needed to meet the target number of personnel interpreting screening mammograms.[14] Similar to programs in the UK and around the world, it has been suggested that Australian radiographers have the potential to fill the mammogram interpretation workforce gap.[19] Thus, radiographers could become an integral part of screening mammography reporting teams.

Mammography readers often review previous mammograms of women whose mammograms from the current screening round show features suggestive of breast cancer or demonstrate inconclusive findings. Artificial intelligence software packages such as Terapixel, recently introduced into the reporting framework to improve the efficacy of the interpretation process, also rely on prior images to resolve ambiguities or make diagnostic decisions about suspicious areas on current mammograms.[20] Many screening programs display previous mammograms of each woman alongside mammograms from the most recent screening round for comparison and identification of changes that may indicate breast cancer.[21] However, due to factors such as storage limitation, lost images and movement of women across states and countries, prior mammograms are not always available. While the practice of storing mammograms for future comparison has become well-established across screening programs, the impact of this practice on screening outcomes is poorly understood.

A literature review of previous studies that have examined the impact of prior mammograms on the diagnostic efficacy of screening programs found that past studies focus only on experienced radiologists and breast physicians.[21] Findings from these studies show that access to prior mammograms improved radiologists' specificity and positive predictive value while reducing both recall rate and false positive rate. Diagnostic performance metrics such as sensitivity, cancer detection rate, and accuracy were not affected by the availability of prior mammograms.[21]

Outside Australia, other healthcare practitioners are also employed as independent mammography readers. In the United Kingdom (UK), advanced practice radiographers have become increasingly involved in the autonomous interpretation of mammograms. Evidence shows that UK radiographers demonstrate strong diagnostic efficacy in interpreting both screening and diagnostic mammograms, either as single independent readers or as second readers in a double reading system.[22,23] Studies that have compared UK radiographers and radiologists in reporting screening mammograms in a double reading system have reported similar performance across all performance metrics[22] or sensitivity[23] and lower specificity for radiographers.[23,24] It has also been shown that UK radiographers outperformed radiologists from developing countries in all performance metrics.[25] These findings support the integration of radiographers into the screening mammography reporting system, particularly in countries with radiologists' shortage.

In Australia and New Zealand, radiographers' roles in mammography are usually limited to image acquisition and patient care, but there exists potential for reader status. A few studies have examined the potential of Australian radiographers to interpret mammograms,[26–29] the impact of breast density of radiographer performance in mammography interpretation[30] and the impact of targeted training packages to improve Australian radiographers' performance in the interpretation of screening mammograms.[31] These studies focused on case-level analysis, which requires radiographers to classify mammograms as normal or abnormal without identifying the location of the cancer in the mammograms. Findings from the studies demonstrate that Australian radiographers can participate in mammography interpretation but highlight a wide variability in their interpretation abilities and the need for interventions to improve performance. However, no study has examined the impact of prior mammograms on radiographers' performance in interpreting screening mammograms using observer performance methodologies that capture their performance at the lesion level.

In this new study, we hypothesise that access to previous mammograms will improve the diagnostic efficacy of radiographers and reduce false positive errors. Therefore, this study aims to assess the impact of providing access to the previous screening mammograms on the interpretation of mammograms from the current screening round by breast imaging radiographers.

Methods

An observer performance study methodology was employed and involved the interpretation of two sets of screening mammograms by the same observers. The study was approved by our Institutional Human Research Ethics Committee (Institutional Review Board Number: 2023/101).

Participant recruitment

Twenty-two radiographers from Australia and New Zealand who worked in their respective screening services (BreastScreen Australia and BreastScreen Aotearoa) were recruited to participate in the study. Participants were recruited at the 2024 Breast screen Australia conference held in Canberra between the 13th and 15th of March 2024. Participants were also recruited by email through the Breast Screen Australia units within the Local Health Districts (LHDs). All participants were provided with the participant information statement and those who consented to the study were recruited. However, only participants who completed the reading of the two test sets (with and without prior images) used for the study were included in the final analysis.

Mammogram test set

The test set used for the study contained 28 digital screening mammography cases (Cancer: n = 9; normal: n = 19) with prior images which were taken in the previous screening round. These mammograms were selected from the BreastScreen Reader Assessment STrategy (BREAST) image databank to ensure that mammograms were included from women of different breast densities (almost entirely fatty, scattered fibroglandular tissue, heterogeneously dense, and extremely dense), cancer

characteristics (calcifications, mass, architectural distortion, and non-specific density), and acquired using different vendor technologies (General Electric, Sectra, Hologic, Fujifilm and Siemens HealthCare) were evenly represented. Similar observer performance studies involving radiologists used between 36 and 200 mammograms and three to 12 radiologists.[21] Statistical analyses of clinical assessment in medical imaging also show that exploratory (Phase 1) studies require between two to three observers and 10–50 patients since the goal is to test observers' ability to distinguish images with and without disease.[32,33] Therefore, two sets of 28 mammograms were developed: "No Prior test-set" containing only the most current mammograms of these 28 women; "Prior test-set" containing mammograms from the most recent screening round and mammograms of the same cohort of women acquired 2–4 years prior (previous mammograms). The reference standard or absolute truth for each mammogram had been collected from radiology reports and pathological reports for cancer cases which were independently confirmed by three senior BreastScreen radiologists who had more than 25 years in interpreting the screening mammograms. The absence of cancer in mammograms classified as normal (cancer-free) was established based on negative screening mammograms of the same women acquired 2–4 years later. All prior mammograms used for the study were acquired 2–4 years prior to provide sufficient time for slow growing cancers to develop to a point where they can be detected.

Experimental design

Prior to starting the study, each participant was presented with an user interface on the BREAST platform (www.breastaustralia.com) to complete a short survey, where information such as their demographics including age, gender, years of experience taking mammograms, number of mammograms taken per week, and practice-related information was collected. Twelve radiographers completed the "No prior test-set" (no previous mammograms available) while one radiographer completed the "Prior test-set" (mammograms from the most recent screening round and the previous mammograms of these women) in the first reading session. Each participant independently reviewed the test-sets and reported their findings based on the Royal Australian and New Zealand College of Radiologist's Tabar classification: normal (RANZCR = 1); benign (RANZCR = 2); indeterminate or equivocal (RANZCR = 3); suspicious (RANZCR = 4); highly suspicious of malignancy (RANZCR = 5).[34] If no lesion or feature of abnormality was detected, the participant was asked to move to the next case in the test-set. However, if the participant detects a lesion, they were asked to mark the lesion and assign a rating from 2 to 5 according to the RANZCR rating scale using the online BREAST platform. If a rating of 3 or higher was assigned, the participants was presented with a drop-list of lesion types to select the lesion type that was detected (calcifications, mass, architectural distortion, and non-specific density). Ratings of 3, 4, 5 were used to establish a participant's level of confidence that the detected lesion was malignant. No feedback was provided to the participants about their scores or correct decisions after the first round of reading the test-set. The second reading session occurred after a wash-out period with 12 radiographers completing the "Prior test-set" and one radiographer completing the "No prior test-set". After the radiographers had completed both test sets, they were presented with their scores and could cycle back through the cases to receive case-based feedback.

Statistical analysis

The true positives, true negatives, false positives, false negatives, specificity, sensitivity, lesion sensitivity (ratio of the number of cancers correctly marked and the total number of cancer lesions in the test set), Receiver Operating Characteristics (ROC) Curve, which measures performance in discriminating between diseased and non-diseased cases, and Jacknife Alternative Free-Response Receiver Operating Characteristics (JAFROC) scores of each radiographer were extracted from the BREAST platform. JAFROC uses the confidence ratings assigned to lesions to calculate reader performance in correctly localising cancer lesions and rating their level of malignancy versus classifying images without cancer. Both Kolmogorov Smirnov and Shapiro-wilk tests indicated that the distributions of true negatives, false positives, specificity, ROC, and JAFROC of participants were parametric while true positives, false negatives, sensitivity, and lesion sensitivity were non-parametric. Therefore, a paired t-test was used to the radiographers' false positives, specificity, ROC, and JAFROC with versus without prior mammograms. Wilcoxon sign-ranked test was used to compare sensitivity and lesion sensitivity with versus without prior mammograms.

The radiographers self-reported the number of years qualified as radiographer, number of years specialised as breast radiographer, number hours spent working in a mammography service, and whether they interpret mammograms. For each of these characteristics, the radiographers were split into two groups: Years qualified as radiographer $\geq$30 versus <30 years; years specialised in breast imaging $\geq$25 years versus <25 years; number of hours spent working in mammography ($\geq$16 h versus <16 h). The thresholds used for the groupings of participants were based on the characteristics of radiographers who participated in the study to ensure even distribution between groups. A Mann-Witney-U test (for sensitivity and lesion sensitivity) and t-test (for specificity, ROC, JAFROC, and false positives) were used to assess if reader characteristics influenced performance in both readings with and without prior mammograms.

To test whether radiographers' characteristics influenced the impact of prior mammograms on performance, a paired sample t-test was used to compare the false positives, ROC, JAFROC, and specificity of each group with versus without prior mammograms. The Wilcoxon sign-ranked test was used to compare the sensitivity and lesion sensitivity of each group with versus without prior mammograms. A two-sided p value $\leq$ 0.05 was considered significant.

Results

Out of 22 radiographers who participated in the study, five radiographers completed only one test-set and three radiographers completed the test-set as a team. Therefore, the data of the five radiographers who completed only one test-set, and three radiographers completed the test-set as a team were excluded. Of the 14 remaining radiographers, one completed the readings sessions using a low-resolution monitor (<5 MP), which was below that recommended for mammography image interpretation. Therefore, 13 radiographers were included in the final analysis. Their years of experience as radiographers ranged from 8 to 44 (mean = 28.2) years, with the number of years specialised as breast radiographer ranging from 8 to 33 (mean = 21.4).

False positives ($p = 0.002$) were lower and True negatives ($p = 0.002$) were higher when readings occurred with prior mammograms. True positives ($p = 0.65$) and False negatives ($p = 0.65$) were not different between readings with versus without prior mammograms. Specificity ($p = 0.002$), ROC ($p = 0.003$), and JAFROC ($p = 0.01$) were significantly better when radiographer read with previous mammograms compared to readings without previous mammograms. No differences were observed in sensitivity ($p = 0.70$) and lesion sensitivity ($p = 0.82$) as shown in Table 1.

Table 1
Comparison of radiographers' performance with versus without prior mammograms (values in bracket represent the minimum and maximum scores of the radiographers for each performance metric).

Performance metric	With previous	Without previous	p-value
Specificity	81(58–95)	60(37–79)	0.002[a]
Sensitivity	96(78–100)	94(67–100)	0.70
Lesion sensitivity	89 (67–100)	88(56–100)	0.82
False positives	3.5(1–8)	7.5(4–12)	0.002[a]
ROC	91 (80–99)	82 (57–91)	0.003[a]
JAFROC	87 (73–99)	79 (52–91)	0.01[a]

ROC: Receiver Operating Characteristics Curve; JAFROC: Jacknife Alternative Free-Response Receiver Operating Characteristics.
[a] Statistically significant.

Table 2 summarises the results of the comparison between radiographers of different characteristics in readings with current mammograms alone and in readings with previous mammograms. There were no significant differences between radiographers in terms of the number of years qualified (≥30 years versus <30 years; $p > 0.05$), number of years specialised as breast radiographer (≥25 years versus <25 years; $p > 0.05$), and number of hours spent working in mammography ($p > 0.05$).

When the reading of each group was compared with and without previous mammograms, readings with previous mammograms consistently showed higher specificity, ROC, JAFROC, and lower false positives. In both radiographers with ≥30 years of experience and those with <30 years of experience, previous mammograms significantly improved specificity and reduced false positives ($p \leq 0.04$). The number of years qualified slightly modified the effect previous mammograms had on the ROC and JAFROC values of radiographers with <30 years' experience to below chance levels. The influence of previous mammograms on specificity, ROC, JAFROC, and false positives were also slightly attenuated to below chance levels in radiographers with ≥25 years of specialisation as breast radiographers but remained significant in those with <25 years of specialisation. Number of hours spent working in mammography did not modify the influence of previous mammograms on radiographer's performance (Table 3).

Discussion

The results show that the availability of previous mammograms to breast imaging radiographers improved their skills in correctly distinguishing women who have cancer from those who do not have cancer. The results also show that when radiographers had access to previous screening mammograms, their ability to correctly classify mammograms of women who do not have cancer improved without reducing their ability to detect breast cancer in women with cancer. Also, previous mammograms are helpful to radiographers in detecting cancer in mammograms and simultaneously establishing the level of malignancy and reducing the proportion of women whose mammograms are incorrectly classified as positive. The influence of previous mammograms on radiographers' mammogram interpretation performance remained noticeable after considering their years of experience as radiographers and the number of hours spent working in mammography. When radiographers with different years of experience, years of specialisation as breast radiographers, and hours spent working in a mammography service were compared in readings with or without previous mammograms alone, we did not find any difference. However, when the interpretations with versus without previous mammograms were compared for each of these groups of radiographers, prior mammograms continuously improved specificity and ROC and reduced false positives for participants with ≥30 years and <30 years qualified as radiographers, but the improvement in ROC for participants with less than 30 years' experience was not statistically significant. Prior mammograms also improved specificity, ROC, and JAFROC, and reduced false positives in

Table 2
Comparison of the performance of radiographers of different characteristics in readings with prior mammograms and without prior mammograms.

Performance metric	With previous mammograms			Without previous mammograms		
	Number of years qualified as radiographer					
	≥30 years (n = 7)	<30 years (n = 6)	p-value	≥30years (n = 7)	<30 years (n = 6)	p- value
Specificity	81(58–90)	82(63–89)	0.95	62 (37–74)	58(47–79)	0.56
Sensitivity	95 (78–100)	96(89–100)	1.00	97.3(89–100)	86.8(67–100)	0.95
L. sensitivity	89(78–100)	89(67–100)	1.00	87(56–100)	89(78–100)	1.00
False positives	2.8(78.1–97.7)	3.8(74–90.1)	0.39	6(65.5–87)	8(51.5–84.5)	0.24
ROC	91 (88–99)	91 (85–99)	0.94	81 (57–91)	83 (71–91)	0.73
JAFROC	87 (73–98)	87(74–99)	0.98	78 (52–86)	80 (66–91)	0.83
	Years of specialisation as breast radiographer					
	≥25 years (n = 6)	<25 years (n = 7)	p-value	≥25 years (n = 6)	<25 years (n = 7)	p-value
Specificity	80(58–90)	83(63–95)	0.69	62(47–79)	59 (37–74)	0.86
Sensitivity	94(78–100)	92(67–100)	1.00	93 (67–100)	95(78–100)	0.73
L. sensitivity	84(67–100)	91(67–100)	0.14	86 (56–100)	89(78–100)	0.53
False positives	3.8 (2–8)	3.2 (1–7)	0.66	7.3 (4–10)	7.7 (5–12)	0.80
ROC	90 (80–99)	92(85–99)	0.90	81 (57–91)	84(71–91)	0.19
JAFROC	86(73–98)	89 (74–99)	0.68	78(52–86)	78 (66–91)	0.62
	Number of hours spent working in a mammography service					
	≥16 h (n = 4)	<16 h (n = 9)	p-value	≥16 h (n = 4)	<16 h (n = 9)	p-value
Specificity	86(79–90)	80(58–95)	0.39	67(47–79)	57(37–74)	0.25
Sensitivity	95(89–100)	96(78–100)	0.82	89(78–100)	96(67–100)	0.15
L. sensitivity	86(67–100)	90(78–100)	0.60	86(78–100)	89(56–100)	0.50
False positives	2.8 (2–4)	3.9(1–8)	0.39	6.3 (4–10)	8 (5–12)	0.24
ROC	91(85–96)	91(80–99)	0.96	81(71–87)	83(57–91)	0.76
JAFROC	87(74–95)	88(73–99)	0.91	78(66–87)	79 (52–91)	0.92

ROC: Receiver Operating Characteristics Curve; JAFROC: Jacknife Alternative Free-Response Receiver Operating Characteristics; *: statistically significant; L. sensitivity: lesion sensitivity.

Table 3
Comparison of the performance of radiographers with similar characteristics with prior versus without prior mammograms.

	Number of years qualified as radiographer					
	With previous (≥30 years: n = 7)	Without previous (≥30 years: n = 7)	p-value	With previous (<30 years: n = 6)	Without previous (<30 years: n = 6)	p-value
Specificity	81(58–90)	62(47–79)	0.04[a]	82(63–95)	58(37–74)	0.03[a]
Sensitivity	95 (78–100)	94(67–100)	1.00	96(89–100)	95(78–100)	0.65
L. sensitivity	89(79–100)	87(56–100)	0.91	89(67–100)	89(78–100)	0.46
False positives	3.6(2–8)	7(4–10)	0.04[a]	3.5(1–7)	8(5–12)	0.03[a]
ROC	91(80–99)	81(57–91)	0.03[a]	91(85–99)	83(71–91)	0.08
JAFROC	87 (73–98)	78(52–86)	0.11	87 (74–99)	76(66–91)	0.06
	Years of specialisation as breast radiographer					
	With previous ≥25 years (n = 6)	Without previous ≥25 years (n = 6)	p-value	With previous <25 years (n = 7)	Without previous <25 years (n = 7)	p-value
Specificity	80 (58–90)	62 (47–79)	0.08	83 (63–95)	59 (37–74)	0.01[a]
Sensitivity	95 (78–100)	93 (67–100)	0.95	95 (89–100)	95 (78–100)	1.00
L. sensitivity	89 (79–100)	87(56–100)	0.73	91 (67–100)	89(78–100)	0.65
False positives	3.8 (2–8)	7.3 (4–10)	0.08	3.3 (1–7)	7.7 (5–12)	0.01[a]
ROC	90 (80–99)	81 (57–91)	0.06	92(85–99)	84(71–91)	0.04[a]
JAFROC	86(73–98)	78(52–86)	0.21	89 (74–99)	78 (66–91)	0.02[a]
	Number of hours spent working in a mammography service					
	With previous (≥16 h: n = 4)	Without previous (≥16 h: n = 4)	p-value	With previous (<16 h: n = 9)	Without previous (<16 h: n = 9)	p-value
Specificity	86(79–90)	67(47–79)	0.03[a]	80(58–95)	57(37–74)	0.01[a]
Sensitivity	95(89–100)	89 (78–100)	0.28	96 (78–100)	96 (67–100)	0.65
L. sensitivity	86(67–100)	86 (78–100)	0.59	90(78–100)	89 (56–100)	0.73
False positives	2.8 (2–4)	6.3 (4–10)	0.04[a]	3.9(1–8)	8(5–12)	0.02[a]
ROC	91(85–96)	81(71–87)	0.006[a]	91(80–99)	83 (57–91)	0.04[a]
JAFROC	87 (74–95)	78(66–87)	0.008[a]	88 (73–99)	79 (52–91)	0.08

ROC: Receiver Operating Characteristics Curve; JAFROC: Jacknife Alternative Free-Response Receiver Operating Characteristics.
[a] Statistically significant; L. sensitivity: lesion sensitivity.

participants with <25 years of specialisation as breast radiographers. Finally, prior mammograms improved specificity, ROC, and JAFROC, and reduced false positives in participants who spend ≥16 h and those who spend <16 h working in mammography service, although the improvement in JAFROC for participants spending <16 h working in mammography service did not reach statistical significance.

To achieve the goal of breast cancer screening programs of detecting breast cancer in its early stages, many women whose mammograms show suspicious, abnormal, or ambiguous features are called back for further assessment.[35] However, less than 10 % of these women are eventually confirmed to have cancer at assessment.[36] It has been suggested that strategies that reduce the proportion of women incorrectly recalled to assessment without reducing the detection of cancer may standardise and expand screening services.[37] Interestingly, reference to previous mammograms when interpreting mammograms from the current screening round improved specificity, ROC, JAFROC, and reduced false positives without affecting sensitivity. These outcomes may be due to the radiographers' ability to dismiss suspicious features when reference to previous mammograms show a lack of change or insignificant change within the suspicious area in keeping with the absence of abnormality or slow-growing benign lesions. Even though we did not directly examine recall rates, the findings suggest that comparison with previous mammograms can reduce the number of benign lesions recalled without reducing the positive predictive value of recall. Screening programs aim to increase cancer detection rate with high sensitivity and specificity while reducing the recall rate.[38] Reducing recall rate is particularly important since it has major clinical, economic, and psychosocial implications. These include increased workload to screening services and personnel from unnecessary additional testing, increased cost and radiation dose associated with assessment of the women recalled, and psychological harms such as fear and anxiety, which reduces screening re-attendance particularly for the 80–85 % of women who receive a false-positive diagnosis. Since screening programs aim to recall as fewer women as possible, the storage and reference to previous mammograms when interpreting mammograms from the current screening round may standardise and expand screening services by reducing the proportion of women incorrectly targeted for additional testing and associated costs.

Previous studies, which examined breast radiologists, show that comparison with previous mammograms improved specificity by 4–15 % and reduced the false positives by 8–14 % without reducing sensitivity and cancer detection rate.[21,39] The current study demonstrated a 22 % increase in specificity, 9 % improvement in radiographers' abilities to correctly distinguish between mammograms with and without cancer, and 53 % reduction in false positives when previous mammograms were available. Some women with breast cancer had two or more lesions and the JAFROC methodology accounted for the number of lesions detected by the radiographers. Therefore, the 8 % improvement observed in JAFROC may be a combination of increase in the number of lesions detected and rating of lesion malignancy. To the best of our knowledge, this is the first study to examine the impact of prior mammograms on breast radiographers' performance in mammography interpretation. An interesting finding was that the cohort of radiographers examined demonstrated very high sensitivity regardless of their experience as radiographers, years specialised as breast radiographers, and the number of hours spent working in mammography. Previous studies that have examined radiographers' performance in mammography interpretation consistently reported comparable sensitivity with radiologists,[22,40] but a high false positive rate for radiographers ranging 35 %–49 %,[40,41] which is a concern for integrating radiographers into the interpretation process.[40,41] We found a mean false positive rate of 40 % when readings occurred

without prior mammograms. Interestingly, when the radiographers had access to previous mammograms, the false positive rate dropped to 18 %, suggesting that providing previous mammograms to radiographers is a useful strategy to reduce their false positive rate. The findings also have implications for advanced practice mammography training programs. It suggests that the incorporation of systematic approaches for evaluating and comparing current and prior mammograms into the training curriculum may be a useful strategy for improving radiographers' mammography interpretation performance.

Experience and workload characteristics of screen-readers have been shown to impact upon radiologists' performance in the interpretation of screening mammograms.[38,42] To confirm that the effect observed when readings occurred with previous mammograms were not due to the experiences of the radiographers, we adjusted the analysis based on the years of practice as radiographers, years of experience as breast radiographers, and number of hours spent working in mammography. In all adjustments, specificity, ROC, and JAFROC were consistently higher, and false positives decreased for all readings with previous mammograms, and only failed to reach significance in radiographers with ≥25 years of specialisation as breast radiographers. Also, sensitivity and lesion sensitivity remained unchanged at below chance levels (see Table 3). These results suggest that the experience and time spent by the radiographer in mammography did not change the impact of previous mammograms on diagnostic performance. To the best of our knowledge, this is the first time that the impact that radiographers' practice experience has on diagnostic performance in mammography interpretation has been examined. The similar performance across all the radiographers suggests that breast imaging radiographers can serve as adjuncts to breast radiologists, particularly where there is breast radiologists' shortage. The findings also provide preliminary evidence to explore the use of breast radiographers for identifying women whose mammograms may need to be triaged for immediate interpretation or those who may require additional imaging such as spot view or digital breast tomosynthesis to provide sufficient information to assist in the diagnostic decision-making process. Triaging and additional imaging has become particularly important with the expansion of screening services to include mobile screening mammography.

We acknowledge that our study has a few limitations. First, participants examined in this work were self-selected breast imaging radiographers. The radiographers did not undertake official interpretation training to read the mammograms; however, most had many years of experience (8–44 years) working in mammography, and this experience could have been helpful to the radiographers in identifying the abnormal/normal mammograms. This selection bias may limit the generalisation of the findings to all radiographers. Secondly, the study was an exploratory baseline study, reflective of our smaller test-set (mammograms of 28 women) and reader numbers (13 radiographers), who completed the readings with and without prior mammograms. The small number of mammograms may limit the generalisability of the findings; however, it was necessary to develop an experimental protocol that mitigates radiographers' fatigue in completing the test-sets. It is important to note that the sample of mammograms in our test-sets is within the 10–50 recommended for exploratory phase of diagnostic performance evaluation,[32,33] and that the sample of readers in the current study is higher than previously reported in similar studies involving radiologists.[21,43–48] Therefore, the numbers of mammograms and readers are sufficient to establish the influence of previous mammograms on radiographers' performance in screening mammography interpretation.

Conclusion

Access to previous screening mammograms improves radiographers' ability to discriminate between normal and abnormal mammograms, detect and establish the malignancy level in abnormal mammograms, and reduce the false positive rates without affecting the detection of breast cancer. The influence of prior mammograms on radiographers' performance is not affected by the number of hours spend working in a breast imaging service, and the number of years qualified as radiographer or specialised as breast radiographer. The findings highlight the potential for integrating breast imaging radiographers into mammography interpretation framework, particularly in countries with radiologists' shortages. The data also emphasise the need for practices to store screening mammograms for future reference, and for policies to establish a national accessible mammographic database platform for one-stop clinics and to account for population mobility across states.

Conflict of interest statement

None.

Acknowledgement

The authors would like to express appreciation to the BreastScreen Reader Assessment Strategy (BREAST) program at the University of Sydney for access and use of the data. BREAST is funded/supported by the Commonwealth Government Department of Health and Aged Care, with their images provided by BreastScreen Australia and the Cancer Institute New South Wales (CINSW).

References

1. American Cancer Society. *Cancer statistics centre*. American Cancer Society; 2024. Available at: https://cancerstatisticscenter.cancer.org/#/. [Accessed 23 July 2024].
2. Arnold M, Morgan E, Rumgay H, Mafra A, Singh D, Laversanne M, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *Breast* 2022;**66**:15–23.
3. WHO. *Cancer statistics*. 2023. https://www.who.int/health-topics/cancer#tab=tab_1. [Accessed 23 July 2024].
4. Cancer Australia. *Breast cancer in Australia statistics*. 2024. Available at: https://www.canceraustralia.gov.au/cancer-types/breast-cancer/statistics. [Accessed 23 July 2024].
5. WHO. *Global breast cancer initiative implementation framework: assessing, strengthening and scaling up of services for the early detection and management of breast cancer: executive summary*. 2023.
6. American Cancer Society. *Survival rates for breast cancer*. 2024. Available at: https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html. [Accessed 23 July 2024].
7. Lei S, Zheng R, Zhang S, Wang S, Chen R, Sun K, et al. Global patterns of breast cancer incidence and mortality: a population-based cancer registry data analysis from 2000 to 2020. *Cancer Commun* 2021;**41**:1183–94.
8. Kalaf JM. Mammography: a history of success and scientific enthusiasm. *Radiol Bras* 2014;**47** [Vii-viii].
9. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA A Cancer J Clin* 2023;**73**:17–48.
10. Schünemann HJ, Lerda D, Quinn C, Follmann M, Alonso-Coello P, Rossi PG, et al. Breast cancer screening and diagnosis: a synopsis of the European breast guidelines. *Ann Intern Med* 2020;**172**:46–56.
11. Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast cancer screening for women at average risk: 2015 guideline update from the American cancer society. *JAMA* 2015;**314**:1599–614.
12. Ekpo EU, Alakhras M, Brennan P. Errors in mammography cannot be solved through technology alone. *Asian Pac J Cancer Prev APJCP* 2018;**19**:291–301.
13. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? *Insights Imaging* 2017;**8**:171–82.
14. Jeganathan S. The growing problem of radiologist shortages: Australia and New Zealand's perspective. *Korean J Radiol* 2023;**24**:1043–5.

15. Nakajima Y, Yamada K, Imamura K, Kobayashi K. Radiologist supply and workload: international comparison–Working Group of Japanese College of Radiology. *Radiat Med* 2008;**26**:455–65.
16. American College of Radiologists (ACR). *The radiology labor shortage*. ACR Bulletin; 2022. Available at: https://www.acr.org/Practice-Management-Quality-Informatics/ACR-Bulletin/Articles/March-2022/The-Radiology-Labor-Shortage. [Accessed 23 July 2024].
17. Moran S, Warren-Forward H. The diagnostic accuracy of radiographers assessing screening mammograms: a systematic review. *Radiography* 2016;**22**:137–46.
18. Tanaka R, Takamori M, Uchiyama Y, Nishikawa RM, Shiraishi J. Using breast radiographers' reports as a second opinion for radiologists' readings of microcalcifications in digital mammography. *Br J Radiol* 2015;**88**:20140565.
19. Moran S, Warren-Forward H. Can Australian radiographers assess screening mammograms accurately? Biennial follow-up from a four year prospective study and lesion analysis. *Radiography* 2016;**22**:e201–6.
20. Tuvemba N. *Therapixel releases a new version of mammoscreen that laverages prior mammograms for better accuracy*. Paris: Therapixel; 2022.
21. Akwo JD, Trieu P, Lewis S. Does the availability of prior mammograms improve radiologists' observer performance?-a scoping review. *BJR Open* 2023;**5**:20230038.
22. Chen Y, James JJ, Michalopoulou E, Darker IT, Jenkins J. Performance of radiologists and radiographers in double reading mammograms: the UK national health service breast screening program. *Radiology* 2023;**306**:102–9.
23. Pauli R, Hammond S, Cooke J, Ansell J. Comparison of radiographer/radiologist double film reading with single reading in breast cancer screening. *J Med Screen* 1996;**3**:18–22.
24. van den Biggelaar FJHM, Nelemans PJ, Flobbe K. Performance of radiographers in mammogram interpretation: a systematic review. *Breast* 2008;**17**:85–90.
25. Sarah L, Barron M, Dung Trieu P. Diagnostic performance of radiographers compared with an international radiology cohort from developing countries in breast cancer detection. *J Med Imag Radiat Sci* 2022;**53**:S23–4.
26. Moran S, Warren-Forward H. Can Australian radiographers assess screening mammograms accurately? Biennial follow-up from a four year prospective study and lesion analysis. *Radiography* 2016;**22**:e201–6. London, England 1995.
27. Moran S, Warren-Forward H. A retrospective study of the performance of radiographers in interpreting screening mammograms. *Radiography* 2011;**17**:126–31.
28. Debono JC, Poulos AE, Houssami N, Turner RM, Boyages J. Evaluation of radiographers' mammography screen-reading accuracy in Australia. *J Medical Radiation Sci* 2015;**62**:15–22.
29. Holt JJ, Pollard K. Radiographers' ability to perceive and classify abnormalities on mammographic images - results of a pilot project. *J Medical Radiation Sci* 2010;**57**:8–14.
30. Alakhras M, Al-Mousa DS, Alqadi AK, Sabaneh HA, Karasneh RM, Spuur KM. The influence of breast density and key demographics of radiographers on mammography reporting performance – a pilot study. *J Medical Radiation Sci* 2022;**69**:30–6.
31. Moran S, Warren-Forward H. Development of a training package to increase the performance of radiographers in assessing screening mammograms. *J Medical Radiation Sci* 2011;**58**:10–3.
32. Obuchowski NA. How many observers are needed in clinical studies of medical imaging? *Am J Roentgenol* 2004;**182**:867–9.
33. Wang X, Ji X. Sample size estimation in clinical Research: from randomized controlled trials to observational studies. *Chest* 2020;**158**:S12–20.
34. RANZCR. Breast imaging grading comparison and lesion classification lists. In: *The royal Australian and New Zealand College of Radiologists®*; 2018.
35. Lee CS, Parise C, Burleson J, Seidenwurm D. Assessing the recall rate for screening mammography: comparing the medicare hospital compare dataset with the national mammography database. *Am J Roentgenol* 2018;**211**:127–32.
36. Elmore JG, Armstrong K, Lehman CD, Fletcher SW. Screening for breast cancer. *JAMA* 2005;**293**:1245–56.
37. Rauscher GH, Murphy AM, Qiu Q, Dolecek TA, Tossas K, Liu Y, et al. The "sweet spot" revisited: optimal recall rates for cancer detection with 2D and 3D digital screening mammography in the metro Chicago breast cancer registry. *AJR Am J Roentgenol* 2021;**216**:894–902.
38. Wong DJ, Gandomkar Z, Lewis S, Reed W, Suleiman M, Siviengphanom S, et al. Do reader characteristics affect diagnostic efficacy in screening mammography? A systematic review. *Clin Breast Cancer* 2023;**23**:e56–67.
39. Akwo JD, Trieu PDY, Barron ML, Reynolds T, Lewis SJ. Access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance. *Clin Radiol* 2024;**12**:e1549–56.
40. Torres-Mejía G, Smith RA, Carranza-Flores Mde L, Bogart A, Martínez-Matsushita L, Miglioretti DL, et al. Radiographers supporting radiologists in the interpretation of screening mammography: a viable strategy to meet the shortage in the number of radiologists. *BMC Cancer* 2015;**15**:410.
41. van den Biggelaar FJNP, Flobbe K. Performance of radiographers in mammogram interpretation: a systematic review. *Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews [Internet]* 2008. Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews [Internet]: Centre for Reviews and Dissemination (UK); 1995. Available from: https://www.ncbi.nlm.nih.gov/books/NBK75605/.
42. Hadadi I, Rae W, Clarke J, McEntee M, Ekpo E. Breast cancer detection across dense and non-dense breasts: markers of diagnostic confidence and efficacy. *Acta Radiol Open* 2022;**11**:20584601211072279.
43. Soh BP, Lee WB, McEntee MF, Kench PL, Reed WM, Heard R, et al. Mammography test sets: reading location and prior images do not affect group performance. *Clin Radiol* 2014;**69**:397–402.
44. Roelofs AA, Karssemeijer N, Wedekind N, Beck C, van Woudenberg S, Snoeren PR, et al. Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology* 2007;**242**:70–7.
45. Kim WH, Chang JM, Koo HR, Seo M, Bae MS, Lee J, et al. Impact of prior mammograms on combined reading of digital mammography and digital breast tomosynthesis. *Acta Radiol* 2017;**58**:148–55.
46. Thurfjell MG, Vitak B, Azavedo E, Svane G, Thurfjell E. Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. *Acta Radiol* 2000;**41**:52–6.
47. Trieu PDY, Borecky N, Li T, Brennan PC, Barron ML, Lewis SJ. The impact of prior mammograms on the diagnostic performance of radiologists in early breast cancer detection: a focus on breast density, lesion features and vendors using wholly digital screening cases. *Cancers* 2023;**15**.
48. Varela C, Karssemeijer N, Hendriks JH, Holland R. Use of prior mammograms in the classification of benign and malignant masses. *Eur J Radiol* 2005;**56**:248–55.

Introduction to Chapter 5 "Impact of prior mammograms on radiologists and radiographers' detection of different breast cancer lesion types."

Breast radiologists, physicians, and advanced practice reporting radiographers use features identified on mammograms to classify mammograms as normal or abnormal (1). Although the breast parenchyma shows a heterogenous appearance on mammograms, specific features are used to establish the presence of abnormality, but these abnormal features may be benign or malignant (1, 2). Radiologically, features such as architectural distortions, calcifications, spiculations, stellate appearances, discrete mass, and non-specific density raise the suspicion of malignancy (1, 2). These features are often referred to as mammographic lesion types and demonstrated differences in their risk of malignancy. For example, the risk of malignancy in stellate lesions varies from 75–90% (3, 4) compared to 10% for discrete masses and non-specific densities, and 10–50% for architectural distortions (5-8). To account for the risk of malignancy in detected lesions, mammography reporting systems such BI-RADS and the RANZCR classification systems were developed. These systems allow BreastScreen readers to highlight the presence of a lesion and describe the type of lesion identified and their perceived risk of malignancy. It has also been shown that longitudinal changes in the appearance of breast lesions can predict the risk of malignancy (9, 10). For example, changes in discrete masses and non-specific densities increase the risk of malignancy by only 10% whereas change in stellate lesions may increase the risk of malignancy by more than 50% (9). Therefore, changes in lesion features can be used to improve diagnostic efficacy of screening mammography, and prior mammograms may provide opportunities for breast readers to identify changes in lesion features over time.

The results presented in Chapters 3 and 4 focused on the impact prior mammograms have reader performance and investigated if personal and workload characteristics of the readers, and breast density, influenced the impact that prior mammograms have on performance. However, the impact that prior mammograms have on the identification of the different lesion types was not explored. While the findings in Chapters 3 and 4 pointed to an upward trend in JAFROC values (lesion detection and risk classification) when prior mammograms were available, it was not clear if a particular lesion type was responsible for the improvement in JAFROC. In addition, the literature review in Chapter 2 demonstrated a lack of studies focusing on lesion-level and lesion type analysis to elucidate the impact that prior mammograms may have on the detection of different lesion types. Only one study adjusted the analysis for lesion types; however, the radiologists who interpreted the mammograms with priors were not always the same as those who interpreted mammograms without priors (2). Therefore, there is limited understanding of the impact that prior mammograms have on the detection of different lesion types. The scarcity of scientific evidence on the impact that prior mammograms have on the detection of different lesion types and the higher JAFROC values recorded with priors in Chapters 3 and 4 highlighted the need for lesion type analyses.

To address these knowledge gaps, Chapter 5 examined the impact that prior mammograms have on the detection of different lesion types. The outcome of this chapter has been published in the Journal of Medical Radiation Sciences and is entitled “Impact of prior mammograms on radiologists and radiographers’ detection of different breast cancer lesion types”. In this paper, observer performance metrics such as specificity, ROC, JAFROC, and false positives were not considered because these metrics incorporate cases without cancer (normal cancer-free cases) in their analysis. Also, these performance metrics have been reported in Chapters 3 and 4 (25, 26). Therefore, the study in Chapter 5 focused on the impact that prior mammograms have on

the detection of different lesion types, not the interpretation of negative cases. Briefly, the findings indicate that reference to prior mammograms improves radiologists' detection of stellate/spiculated lesions without affecting the detection of other lesion types. Prior mammograms did not impact upon radiographers' abilities to detect any lesion type but improved their performance in discriminating between normal cancer-free mammograms and those with cancer. The outcome of the analysis provided important insight into the impact that prior mammograms have on the radiological detection of changes in stellate/spiculated lesions which are most discriminative of malignancy.

References

1. Trieu PDY, Lewis SJ, Li T, Ho K, Tapia KA, Brennan PC. Reader characteristics and mammogram features associated with breast imaging reporting scores. British Journal of Radiology. 2020;93(1114):20200363.

2. Trieu PDY, Borecky N, Li T, Brennan PC, Barron ML, Lewis SJ. The Impact of Prior Mammograms on the Diagnostic Performance of Radiologists in Early Breast Cancer Detection: A Focus on Breast Density, Lesion Features and Vendors Using Wholly Digital Screening Cases. Cancers (Basel). 2023;15(4).

3. Evans AJ, Pinder SE, James JJ, Ellis IO, Cornford E. Is mammographic spiculation an independent, good prognostic factor in screening-detected invasive breast cancer? American Journal of Roentgenology. 2006;187(5):1377-80.

4. Liu S, Wu XD, Xu WJ, Lin Q, Liu XJ, Li Y. Is There a Correlation between the Presence of a Spiculated Mass on Mammogram and Luminal A Subtype Breast Cancer? Korean Journal of Radiology. 2016;17(6):846-52.

5. Berment H, Becette V, Mohallem M, Ferreira F, Chérel P. Masses in mammography: What are the underlying anatomopathological lesions? Diagnostic and Interventional Imaging. 2014;95(2):124-33.

6. Choudhery S, Johnson MP, Larson NB, Anderson T. Malignant Outcomes of Architectural Distortion on Tomosynthesis: A Systematic Review and Meta-Analysis. American Journal of Roentgenology. 2021;217(2):295-303.

7. Shaheen R, Schimmelpenninck CA, Stoddart L, Raymond H, Slanetz PJ. Spectrum of Diseases Presenting as Architectural Distortion on Mammography: Multimodality Radiologic Imaging with Pathologic Correlation. Seminars in Ultrasound, CT and MRI. 2011;32(4):351-62.

8. Aslan O, Oktay A. Diagnostic accuracy of the breast MRI Kaiser score in suspected architectural distortions and its comparison with mammography. Scientific Reports. 2024;14(1):447.

9. Farshid G, Walker A, Battersby G, Sullivan T, Gill PG, Pieterse S, et al. Predictors of malignancy in screen-detected breast masses with indeterminate/equivocal (grade 3) imaging features. The Breast. 2011;20(1):56-61.

10. Sturesdotter L, Sandsveden M, Johnson K, Larsson AM, Zackrisson S, Sartor H. Mammographic tumour appearance is related to clinicopathological factors and surrogate molecular breast cancer subtype. Scientific Reports. 2020;10(1):20814.

CHAPTER 5

Impact of prior mammograms on radiologists and radiographers' detection of different breast cancer lesion types

Chapter 5 is published as **"**Akwo JD, Trieu PD, Barron M, Reynolds T, Lewis SJ (2025). Impact of prior mammograms on radiologists and radiographers' detection of different breast cancer lesion types" in the Journal of Medical Radiation Sciences. 2025; 0:1–7

This published paper is the last of the three original studies performed within the institutional ethics review board approval entitled "The influence of prior cases on radiology performance with screening mammograms (Human research and ethics committee approval number: 2023/101)".

Journal of Medical Radiation Sciences

Journal of Medical Radiation Sciences

ORIGINAL ARTICLE OPEN ACCESS

Impact of Prior Mammograms on Radiologists and Radiographers' Detection of Different Breast Cancer Lesion Types

Judith D. Akwo[1] | Phuong Dung (Yun) Trieu[1] | Melissa L. Barron[1] | Tess Reynolds[2] | Sarah J. Lewis[1,3]

[1]Medical Image Optimisation and Perception Group, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia | [2]Image X Institute, School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Camperdown, Australia | [3]School of Health Sciences, Western Sydney University, Campbelltown, Australia

Correspondence: Judith D. Akwo (jakw2899@uni.sydney.edu.au)

Received: 29 January 2025 | **Revised:** 21 May 2025 | **Accepted:** 15 July 2025

Funding: The authors received no specific funding for this work.

Keywords: breast cancer | mammographic features | observer performance | prior mammograms | radiographers' performance | radiologists' performance

ABSTRACT

Introduction: Mammographic interpretation relies on the detection of suspicious abnormal changes, and reference to prior mammograms may support the detection of these changes. However, the influence of prior mammograms on the detection of different breast lesions is unclear. This paper assesses the influence of prior mammograms on the detection of different lesion types in mammograms.

Methods: Mammographic test sets comprising different lesion types were interpreted in two stages. In Stage 1, eight radiologists interpreted current mammograms of 72 women (normal: $n = 40$; cancer: $n = 32$) with and without access to the prior mammograms. In Stage 2, 13 radiographers interpreted another test set containing 28 mammograms (normal: $n = 19$; cancer: $n = 9$) with and without access to the prior mammograms. Radiologists and radiographers' performances in detecting different lesion types with and without prior mammograms were compared separately using a paired t-test.

Results: In Stage 1, reference to prior mammograms did not improve sensitivity for architectural distortions ($p = 0.48$), calcifications ($p = 0.85$), discrete masses ($p = 0.45$), and non-specific density ($p = 0.22$). However, prior mammograms improved the detection of spiculated/stellate lesions ($p = 0.05$) and diagnostic accuracy for architectural distortions ($p = 0.006$) and discrete/spiculated/stellate masses ($p = 0.01$). Prior mammograms had no impact on lesion sensitivity ($p > 0.05$). In Stage 2, no differences were observed in sensitivity and lesion sensitivity when compared to without prior mammograms for all lesion types ($p > 0.05$), but prior mammograms improved overall diagnostic accuracy ($p \leq 0.002$).

Conclusions: Prior mammograms improve the detection of spiculated/stellate lesions but have no impact on the detection of other lesion types when measuring radiologists' and radiographers' performance.

1 | Introduction

Breast cancer remains the most prevalent cancer in women worldwide, and the incidence of the disease continues to increase annually [1]. Globally, more than three million new cases of breast cancer are predicted to be diagnosed annually by 2040 [2]. Current estimates show that one in eight women will be diagnosed with breast cancer in their lifetime, and one in 40 affected women will die from the disease [3]. In the last three decades, deaths from breast cancer have reduced by 44% [4], and this has been attributed to early detection through population-based screening and advances in treatment of

breast cancer [4]. Estimates from the United States and Australia show that screening mammography and improvement in breast cancer treatment have reduced breast cancer deaths per 100,000 women from 64 to 27 in the United States and 74 to 37 in Australia—a decrease of 37 per 100,000 women [5, 6].

Estimates show that 47% of the reduction in breast cancer deaths are due to early detection and treatment of Stages 1, 2, and 3 breast cancers [6]. Screening mammography aims to detect early-stage breast cancers and relies on the identification of features of the disease. Across the world, radiologists and other healthcare practitioners involved in mammography interpretation use terms such as discrete mass, spiculated mass, stellate lesion, calcification, non-specific density, and architectural distortion to describe mammographic features identified in mammograms and to report the presence of cancer [7]. These mammographic features have been shown to be correlated with histopathology features of breast cancer but may also indicate benign conditions [8, 9]. Therefore, when breast lesions are identified, they are classified according to their risk of malignancy using different lexicons such as the Breast Imaging Reporting and Data Systems (BI-RADS) in the United States and Royal Australian and New Zealand College of Radiologists (RANZCR) classification in Australia and New Zealand [10, 11]. The initial identification and risk classification of these lesions inform screening programs' recall to assessment clinics and follow-up.

Mammographic images are traditionally interpreted by breast radiologists. However, breast radiologists' workforce shortages have reinforced the need for breast physician and radiographer reporting. In the United Kingdom (UK) National Health Service Breast Screening Programme (NHS BSP), advanced practice radiographers undertake independent interpretation of mammograms or act as second independent double readers [12, 13]. In Australia, radiographers have demonstrated the potential to report mammograms at a level that supports their integration into the reporting workforce [14–17] but have difficulty in detecting certain lesion types [14]. Previous studies based on breast radiologists have explored prior mammograms for improving the diagnostic efficacy of mammography [18–21]. The rationale behind these studies is that comparing current mammograms to mammograms acquired in the previous screening rounds may facilitate the identification of parenchymal changes associated with malignancy [22, 23]. However, these studies were mostly based on older technologies and observer performance evaluation methodologies that do not consider performance at the lesion level, and only evaluated radiologists even though advanced practice radiographers, particularly in the UK NHS BSP, now participate in mammography interpretation [18].

To close these knowledge gaps, our group examined the impact that prior mammograms have on radiologists [24] and radiographers' [25] performance using current technologies and observer performance evaluation methodologies. The findings from these two studies showed that access to prior mammograms improves specificity without affecting sensitivity and lesion sensitivity, and these findings were not confounded by the experience and work-related characteristics of radiologists and radiographers [24, 25]. However, in both studies, we observed that access to prior mammograms led to a similar trend in JAFROC, which examines performance in detecting lesions and simultaneously establishing malignancy risk [24, 25]. These findings indicated that lesion features may impact upon their risk classification and question the influence that prior mammograms have on the detection and risk classification of different lesion types. The test sets used in our previous studies included different lesion types but did not examine the impact that prior mammograms have on the detection of each lesion type [24, 25]. A recent review by our group also showed a lack of studies on the impact that prior mammograms have on radiologists' or radiographers' detection of different lesion types [18]. The only study that adjusted radiologists' performance for each lesion type was a retrospective analysis of observer performance data, and the radiologists who interpreted with prior mammograms were not always the same as those who interpreted mammograms without priors, which does not fully account for the impact that radiologists' experience has on performance [21]. Since the risk of malignancy varies across lesion types [9, 26, 27] and these risk ratings are used for reporting the presence of breast cancer in the BI-RADS and RANZCR lexicons [10, 11], it has become increasingly important to address the knowledge gaps around the impact that prior mammograms have on performance. Therefore, this study aimed to assess the influence of prior mammograms on the detection of different breast lesion types via mammograms by radiologists and radiographers. The outcome of the work has the potential to support the design of training and continuous professional development programmes around the systematic analysis of prior mammograms for radiologists and radiographers.

2 | Methods

This observer performance study approved by the Human Research Ethics Committee of the University of Sydney Human Research Ethics Committee (Number 2023/101) involved the interpretation of screening mammograms of women in two phases. Participants' recruitment and the interpretation protocol have been reported previously [24, 25].

2.1 | Mammogram Test Set

Mammograms comprising of different breast lesion types/mammographic features were used to design separate test sets for radiologist and radiographers. For each participant group, two test sets were developed: 'prior test set' and 'no prior test set'. The design and description of these tests and the vendor technologies used for acquiring the mammograms used for the test sets are described in our previous studies [24, 25]. The sample sizes for the test sets were based on the requirements for exploratory phase (radiographer test set) and challenge phase (radiologists test set) observer performance evaluation [28, 29]. The reference standard for the normal mammograms (mammograms of women with no cancer) was established independently by at least two expert radiologists and negative mammograms acquired 2 to 4 years later. Biopsy was used as the reference standard for the cancer cases (mammograms containing breast cancer) and expert radiologists who also established the lesion-type descriptions. The characteristics of the test sets are summarised in Table 1. All mammograms were selected from the BreastScreen Reader Assessment

TABLE 1 | Characteristics of mammographic the test sets.

Test set characteristics	Radiologists test set	Radiographer test set
Number of cases	72	28
Cancer cases	32	9
Normal cases	40	19
Dense breasts	36	18
Non-dense breasts	36	10
Architectural distortions	8	0
Calcifications	8	1
Non-specific density	7	4
Stellate/spiculated lesions	5	2
Discrete mass	4	2

STrategy (BREAST) image databank and uploaded to the BREAST platform, a cloud-based platform technology that allows readers to interpret the mammograms and receive feedback on their performance [30–32].

2.2 | Experimental Design

The study was conducted in two stages. In the first stage, eight participants interpreted the radiologist's test set containing 72 mammograms (normal: $n=40$; cancer: $n=32$) in two reading sessions as described in our previous studies [24, 25]. Each reader independently interpreted the mammograms in both reading sessions based on the RANZCR breast imaging lesion classification using the BREAST platform. If a lesion was detected, the reader was asked to provide a rating from 2 to 5. A lesion rating of 2 was considered benign and non-actionable. If a reader marked a lesion and assigned ratings of 3, 4, or 5, a list of checkboxes for different lesion-type descriptors (spiculated/stellate lesion, discrete mass, calcification, non-specific density, or architectural distortion) was presented for the reader to check the box corresponding to the lesion type detected. A detailed experimental design, including viewing conditions, is reported elsewhere [24, 25].

2.3 | Statistical Analysis

Once each reader completes a test set, the BREAST software records the reader's true positives (TP), true negatives (TN), false negatives (FN), and false positives (FP) scores. The software uses these values to automatically calculate each reader's sensitivity, specificity, lesion sensitivity, receiver operating characteristic (ROC), area under the curve (AUC) and jackknife alternative free-response receiver operating characteristic (JAFROC). For this study, we extracted the raw data of each reader and established their TP, TN, FN and FP for each lesion type in both reading sessions: spiculated lesion, discrete mass, calcification, non-specific density, architectural distortion or stellate lesion. The sensitivity and diagnostic accuracy analysis for each lesion type were then performed using the MedCalc software package (Version 23.0.9) [33]. Since discrete and spiculated/stellate lesion types present as masses in mammograms, these were also combined and analysed. If a reader's mark falls within the lesion, the BREAST software records a value to indicate a correct lesion localisation. Lesion sensitivity or localisation for each lesion type was calculated by dividing the number of that lesion type correctly marked by the total number of that lesion type in the dataset. Specificity, ROC and JAFROC outputs were not analysed since the focus of the study was on the impact of prior mammograms on the detection of different lesion types and because they have been reported elsewhere [24, 25]. Kolmogorov Smirnov test showed that data for each lesion type in both the radiologists/breast physicians and radiographers' datasets were normally distributed. Therefore, the performance of the readers with and without prior mammograms for each lesion type were compared using a paired t-test. $p \leq 0.05$ was considered as statistically significant.

3 | Results

The eight participants in Stage 1 routinely interpret mammograms in clinical practice. The mean number of years in their current roles was 20.3 ± 11.6 years, and the mean number of years of reading mammograms was 17.3 ± 10.9 years. These readers spent between four and 30 h interpreting mammograms and interpret 20 to 200 mammograms per week.

Table 2 shows the mean performance and the minimum and maximum scores of the participants in each lesion type when reading occurred with and without prior mammograms in Stage 1. Overall, there were no differences in sensitivity when reading occurred with compared to without prior mammograms for architectural distortions ($p=0.48$), calcifications ($p=0.85$), discrete masses ($p=0.45$) and non-specific density ($p=0.22$). However, prior mammograms marginally improved the classification of spiculated and stellate lesions ($p=0.05$). No differences were observed in lesion sensitivity for all lesion types ($p \geq 0.16$). Overall diagnostic accuracy did not differ between reading with and without prior mammograms when calcifications, discrete masses, spiculated and stellate lesions and non-specific density lesions were examined independently ($p \geq 0.16$). However, the availability of prior mammograms significantly improved diagnostic accuracy for architectural distortions ($p=0.006$).

Thirteen radiographers with mean years of experience as radiographers of 28 ± 11 years, and 21 ± 8.4 years specialised as breast radiographers completed the reading in Stage 2. All the radiographers correctly classified all the discrete masses and non-specific density lesions when prior mammograms were available. Since only one calcification was available in the dataset, it was excluded from the final analysis. Overall, there were no differences in sensitivity and lesion sensitivity when readings occurred with compared to without prior mammograms for all lesion types ($p > 0.05$ for all). However, overall diagnostic accuracy was significantly higher for all lesion

TABLE 2 | Radiologists' performance in lesion type detection with prior and no prior mammograms.

Performance metric	With priors	Without priors	*p*
Architectural distortions ($n=8$)			
Sensitivity	58 (25–75)	62 (50–75)	0.48
Lesion sensitivity	48 (13–75)	53 (38–75)	0.49
Accuracy	86 (79–92)	74 (67–83)	0.006*
Calcifications ($n=8$)			
Sensitivity	80 (63–100)	79 (63–100)	0.85
Lesion sensitivity	79 (63–100)	78 (63–100)	0.16
Accuracy	86 (58–100)	77 (67–85)	0.95
Discrete mass ($n=4$)			
Sensitivity	72 (22–100)	78 (75–100)	0.45
Lesion sensitivity	68 (25–100)	69 (25–100)	0.94
Accuracy	85 (57–100)	76 (68–89)	0.18
Non-specific density ($n=7$)			
Sensitivity	52 (14–100)	60 (42–85)	0.22
Lesion sensitivity	48 (14–100)	57 (42–85)	0.40
Accuracy	81 (59–89)	76 (66–85)	0.29
Spiculated mass/stellate ($n=5$)			
Sensitivity	75 (40–100)	61 (20–100)	0.05*
Lesion sensitivity	77 (40–100)	77 (20–100)	1.0
Accuracy	86 (58–100)	76 (67–87)	0.17
Discrete mass + spiculated mass/stellate masses ($n=9$)			
Sensitivity	74 (44–100)	78 (44–100)	0.66
Lesion sensitivity	72 (44–100)	74 (22–100)	0.89
Accuracy	87 (58–100)	76 (67–87)	0.01*

Note: Values in brackets represent the minimum and maximum scores of the readers per lesion type.
*Significant difference.

TABLE 3 | Radiographers' performance in lesion type detection with prior and no prior mammograms.

Performance metric	With priors	Without priors	*p*
Discrete mass ($n=2$)			
Sensitivity	100 (100–100)	94 (50–100)	0.35
Lesion sensitivity	100 (50–100)	88 (50–100)	0.17
Accuracy	85 (62–95)	66 (43–81)	0.002*
Non-specific density ($n=4$)			
Sensitivity	100 (100–100)	94 (75–100)	0.17
Lesion sensitivity	90 (75–100)	84 (50–100)	0.45
Accuracy	86 (65–96)	69 (48–83)	0.001*
Spiculated/stellate mass ($n=2$)			
Sensitivity	75 (0–100)	75 (0–100)	1.00
Lesion sensitivity	69 (0–100)	75 (0–100)	0.59
Accuracy	83 (62–95)	64 (48–76)	< 0.001*
Discrete mass + spiculated mass/stellate masses ($n=4$)			
Sensitivity	92 (50–100)	90 (50–100)	0.76
Lesion sensitivity	87 (50–100)	87 (25–100)	1.0
Accuracy	86 (62–95)	66 (48–83)	0.001*

Note: Values in brackets represent the minimum and maximum scores of the readers per lesion type.
*Significant difference.

types ($p \leq 0.002$) when prior mammograms were available to the radiographers. The mean scores of all readers, and minimum and maximum scores for each performance metric are shown in Table 3.

4 | Discussion

The findings demonstrate that prior mammograms improve radiologists' detection of spiculated/stellate lesions without impacting upon the detection of discrete masses, calcifications, non-specific densities and architectural distortions. The results also show that prior mammograms did not have an influence on the ability of radiographers to detect different breast cancer lesion types. Diagnostic accuracy, which accounts for the reader's ability to classify both normal and abnormal mammograms, was consistently higher when prior mammograms were available to radiographers. Although the mean accuracy of radiologists for all lesion types was consistently higher when prior mammograms were available, this improvement only reached statistical significance in architectural distortions.

The correct detection and discrimination of different lesion types on mammograms are important because they influence both the false-positive recall and false-negative rate, which have both clinical and economic implications for screening and assessment [34]. Breast imaging readers have a mental picture of cancer features [35], and when these cancer features are detected, the lesion may likely be reported as cancer regardless of whether prior mammograms are available. This may explain the similar sensitivity and lesion sensitivity between readings with and without prior mammograms. However, different lesion types can also indicate different benign and malignant conditions [36–38], but some lesion types are more predictive of cancer than others [8, 37, 38]. For example, the risk of malignancy

in discrete masses and non-specific densities is 10% compared to 10%–50% in architectural distortions, and 75%–90% in spiculated/stellate lesions [37–39]. Our findings suggest that not all lesion types visible in mammograms are reported as malignant, even when comparison with prior mammograms suggests subtle changes in current mammograms. The discriminatory powers of lesion types may explain the lower radiologists' sensitivity and lesion sensitivity in both architectural distortions and non-specific density lesions overall (Table 2), but higher sensitivity for spiculated/stellate lesions when prior mammograms were available. This is further supported by the results obtained when discrete masses were combined with spiculated/stellate masses, where the impact of prior mammograms on the detection of spiculated/stellate masses was attenuated.

Although the results of sensitivity analysis indicate that prior mammograms have no impact on the detection of most lesion types, the higher mean diagnostic accuracy scores with priors suggests that the availability of prior mammograms improves the discrimination between malignant and benign/normal features in mammograms. These findings highlight the influence of prior mammograms in clinical decision-making around recall for further assessment and the management and follow-up of patients with suspicious cancer lesions. Also, despite our use of prior mammograms obtained 2–4 years prior to allow changes in slow-growing cancers to be visible, these changes did not improve the detection rate except for spiculated/stellate lesions. The high risk of malignancy in spiculated/stellate lesions may explain the improvement in radiologists' sensitivity in this lesion type. However, this raises concerns for radiographers who demonstrated similar sensitivity, when priors were available despite this lesion type being most predictive of breast cancer and most recalled by radiologists [39, 40]. These findings highlight the need for radiographer intervention that incorporates strategies to thoroughly analyse and compare prior and current mammograms to identify spiculated/stellate lesions and changes in this lesion type that indicate malignancy.

Previous studies that have explored the mammographic detection of different breast lesion types focussed on radiologists' sensitivity [41] or inter-radiologist concordance in classifying breast lesion types [7, 42] and positive predictive value of different lesion types [42]. The studies that assessed the influence of prior mammograms on radiologists' performance mostly report improvement in specificity and reduction in the false-positive rate with no impact on sensitivity and cancer detection rate [18, 43, 44]. However, none of these studies examined performance according to lesion types. The only study that attempted to assess the influence of prior mammograms on radiologists' detection of different lesion types compared prior mammograms from different vendor technologies and showed that using prior mammograms from a different vendor improved cancer detection, particularly architectural distortions compared to priors from the same vendor [21]. The authors also reported that prior mammograms from the same vendor improved the classification of normal mammograms. However, the radiologists who interpreted the mammograms with priors were not always the same radiologists who interpreted the mammograms without priors [21], and differences in the expertise of the radiologists could have influenced the results. Our test sets combined mammograms from both the same and different vendors, and the same cohort of participants read mammograms of the same women with and without priors. Therefore, our methodology considered both the impact of vendor technology and reader expertise on the influence of prior mammograms on performance in detecting different lesion types. The lack of changes in sensitivity and lesion sensitivity, but higher mean accuracy scores for readings with prior mammograms may be due to the readers performance in classifying normal mammograms when prior mammograms were available. This is supported by findings of previous studies that consistently reported higher true negatives, lower false positives and higher specificity when prior mammograms were available [18, 21, 44–48]. The results of our lesion-type analyses provide further evidence that the lack of improvement in sensitivity and cancer detection rate with prior mammograms reported in previous studies [18, 21, 44–48] may not be due to the imaging features of breast cancer lesions.

Our study has some limitations that should be considered when interpreting the findings. First, we employed a laboratory-based experimental study design involving cancer-enriched mammography test sets and self-selected radiologists and radiographers. Given that the study was a lesion-type analysis, it was necessary for different lesion types to be adequately represented in our test sets. Second, the samples of mammograms used in the test sets, particularly the radiographer test sets, were relatively small, which may limit the generalisation of the findings. Since Australian and New Zealand radiographers do not interpret mammograms clinically, it was necessary to undertake their performance evaluation using a sample of mammograms within the range (10–50 images) scientifically recommended for exploratory baseline performance evaluation studies [28, 29]. Also, the study involved significant radiologists and radiographers' time commitment at two different time points, which underscored the need to develop test sets that have scientific merit but also mitigate the effect that fatigue may have on their performance. Thirdly, the sample of participants (radiologist/breast physicians: $n=8$ and radiographer: $n=13$) who completed the test sets may not represent the population of breast imaging readers. However, these numbers are within the participant sample sizes used in studies ($n=3$ to 12 readers) that have investigated the impact of prior mammograms on reader performance, and many were highly experienced [18, 21, 44–48]. Although the laboratory nature of our study and the limitations highlighted may impact upon the generalisability of the findings, it should be remembered that our methodology followed the recommendations for clinical studies of observer performance in medical imaging [28, 29]. Furthermore, evidence indicates that the findings from a laboratory setting can be generalised to a clinical setting [45, 49, 50]. Therefore, the results of our study provide baseline evidence for the impact of priors in the detection of different lesion types.

5 | Conclusions

The access to prior mammograms marginally improved the detection of spiculated/stellate lesions by radiologists but did not have an impact on the detection of other lesion types. Similarly, access to prior mammograms did not improve radiographers' ability to detect different lesion types. In this laboratory study, the higher mean diagnostic accuracy when prior mammograms

were available may be due to readers' ability to correctly classify normal images rather than detect cancer, which is an important capability of screening programmes. The importance of access to prior screening images remains strong when all elements are considered, including reducing recall rates and public confidence in the balance between sensitivity and specificity of mammographic imaging.

Ethics Statement

The study was approved by the University of Sydney Human Research Ethics Committee (Approval Number: 2023/101).

Acknowledgement

Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australian University Librarians.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data used for this paper are available in the BreastScreen Reader Assessment Strategy (BREAST) database and can be made available with approval of the data custodian and the University of Sydney Human Research Ethics Committee.

References

1. A. N. Giaquinto, H. Sung, L. A. Newman, et al., "Breast Cancer Statistics 2024," *CA: A Cancer Journal for Clinicians* 74, no. 6 (2024): 477–495.

2. The Lancet Oncology, "Curbing the Climb in Cancer Incidence," *Lancet Oncology* 25, no. 5 (2024): 529.

3. NBCF, *Breast Cancer Facts & Stats* (National Breast Cancer Foundation INC, 2024), https://www.nationalbreastcancer.org/breast-cancer-facts/.

4. ACS, *Breast Cancer Facts & Figures 2024–2025* (American Cancer Society INC, 2024), https://www.cancer.org/content/dam/cancer-org/cancer-control/en/booklets-flyers/breast-cancer-facts-and-figures-2425.pdf.

5. AIHW, *BreastScreen Australia Monitoring Report 2024* (Australian Institute of Health and Welfare, 2024).

6. J. L. Caswell-Jin, L. P. Sun, D. Munoz, et al., "Analysis of Breast Cancer Mortality in the US—1975 to 2019," *Journal of the American Medical Association* 331, no. 3 (2024): 233–241.

7. A. Y. Lee, D. J. Wisner, S. Aminololama-Shakeri, et al., "Inter-Reader Variability in the Use of BI-RADS Descriptors for Suspicious Findings on Diagnostic Mammography: A Multi-Institution Study of 10 Academic Radiologists," *Academic Radiology* 24, no. 1 (2017): 60–66.

8. A. J. Evans, S. E. Pinder, J. J. James, I. O. Ellis, and E. Cornford, "Is Mammographic Spiculation an Independent, Good Prognostic Factor in Screening-Detected Invasive Breast Cancer?," *AJR. American Journal of Roentgenology* 187, no. 5 (2006): 1377–1380.

9. S. Liu, X. D. Wu, W. J. Xu, Q. Lin, X. J. Liu, and Y. Li, "Is There a Correlation Between the Presence of a Spiculated Mass on Mammogram and Luminal A Subtype Breast Cancer?," *Korean Journal of Radiology* 17, no. 6 (2016): 846–852.

10. D. A. Spak, J. S. Plaxco, L. Santiago, M. J. Dryden, and B. E. Dogan, "BI-RADS() Fifth Edition: A Summary of Changes," *Diagnostic and Interventional Imaging* 98, no. 3 (2017): 179–190.

11. RANZCR, "Breast Imaging Grading Comparison and Lesion Classification Lists: The Royal Australian and New Zealand College of Radiologists" (2018), https://www.ranzcr.com/search/breast-imaging-grading-comparison-and-lesion-classification.

12. Y. Chen, J. J. James, E. Michalopoulou, I. T. Darker, and J. Jenkins, "Performance of Radiologists and Radiographers in Double Reading Mammograms: The UK National Health Service Breast Screening Program," *Radiology* 306, no. 1 (2023): 102–109.

13. R. Pauli, S. Hammond, J. Cooke, and J. Ansell, "Comparison of Radiographer/Radiologist Double Film Reading With Single Reading in Breast Cancer Screening," *Journal of Medical Screening* 3, no. 1 (1996): 18–22.

14. S. Moran and H. Warren-Forward, "Can Australian Radiographers Assess Screening Mammograms Accurately? Biennial Follow-Up From a Four Year Prospective Study and Lesion Analysis," *Radiography* 22, no. 3 (2016): e201–e206.

15. S. Moran and H. Warren-Forward, "A Retrospective Study of the Performance of Radiographers in Interpreting Screening Mammograms," *Radiography* 17, no. 2 (2011): 126–131.

16. J. C. Debono, A. E. Poulos, N. Houssami, R. M. Turner, and J. Boyages, "Evaluation of Radiographers' Mammography Screen-Reading Accuracy in Australia," *Journal of Medical Radiation Sciences* 62, no. 1 (2015): 15–22.

17. J. J. Holt and K. Pollard, "Radiographers' Ability to Perceive and Classify Abnormalities on Mammographic Images - Results of a Pilot Project," *Radiographer* 57, no. 2 (2010): 8–14.

18. J. D. Akwo, P. Trieu, and S. Lewis, "Does the Availability of Prior Mammograms Improve Radiologists' Observer Performance? A Scoping Review," *BJR Open* 5, no. 1 (2023): 20230038.

19. M. P. Callaway, C. R. Boggis, S. A. Astley, and I. Hutt, "The Influence of Previous Films on Screening Mammographic Interpretation and Detection of Breast Carcinoma," *Clinical Radiology* 52, no. 7 (1997): 527–529.

20. T. Evans, B. Burlton, G. Devenish, G. Stevens, M. Lewis, and K. Gower Thomas, "A Comparison of Two Digital Mammography Systems: Are There Any Differences?," *Clinical Radiology* 71, no. 1 (2016): 27–31.

21. P. D. Y. Trieu, N. Borecky, T. Li, P. C. Brennan, M. L. Barron, and S. J. Lewis, "The Impact of Prior Mammograms on the Diagnostic Performance of Radiologists in Early Breast Cancer Detection: A Focus on Breast Density, Lesion Features and Vendors Using Wholly Digital Screening Cases," *Cancers (Basel)* 15, no. 4 (2023): 1339.

22. A. Stachs, J. Stubert, T. Reimer, and S. Hartmann, "Benign Breast Disease in Women," *Deutsches Ärzteblatt International* 116, no. 33–34 (2019): 565–574.

23. N. Cho, "Imaging Features of Breast Cancer Molecular Subtypes: State of the Art," *Journal of Pathology and Translational Medicine* 55, no. 1 (2021): 16–25.

24. J. D. Akwo, P. D. Y. Trieu, M. L. Barron, T. Reynolds, and S. J. Lewis, "Access to Prior Screening Mammograms Affects the Specificity but Not Sensitivity of Radiologists' Performance," *Clinical Radiology* 79 (2024): e1549–e1556.

25. J. D. Akwo, P. D. Trieu, M. L. Barron, T. Reynolds, and S. J. Lewis, "Does Access to Prior Mammograms Improve the Performance of Radiographers in Interpreting Screening Mammograms?," *Radiography* 31, no. 1 (2025): 247–253.

26. L. Sturesdotter, M. Sandsveden, K. Johnson, A.-M. Larsson, S. Zackrisson, and H. Sartor, "Mammographic Tumour Appearance Is Related to Clinicopathological Factors and Surrogate Molecular Breast Cancer Subtype," *Scientific Reports* 10, no. 1 (2020): 20814.

27. C.-Y. Chang, S.-J. Kuo, H.-K. Wu, Y.-L. Huang, and D.-R. Chen, "Stellate Masses and Histologic Grades in Breast Cancer," *Ultrasound in Medicine & Biology* 40, no. 5 (2014): 904–916.

28. N. A. Obuchowski, "How Many Observers Are Needed in Clinical Studies of Medical Imaging?," *American Journal of Roentgenology* 182, no. 4 (2004): 867–869.

29. X. Wang and X. Ji, "Sample Size Estimation in Clinical Research: From Randomized Controlled Trials to Observational Studies," *Chest* 158, no. 1s (2020): S12–S20.

30. S. J. Lewis, N. Borecky, T. Li, M. L. Barron, P. Brennan, and P. D. Y. Trieu, "Radiologist Self-Training: A Study of Cancer Detection When Reading Mammograms at Work Clinics or Workshops," *Journal of Cancer Education* 38, no. 2 (2023): 571–577.

31. P. D. Y. Trieu, D. Abu Awwad, M. L. Barron, and S. J. Lewis, "Optimizing Radiological Education: The Role of Learning Spacing via Test Sets in Enhancing Diagnostic Proficiency in Breast Screening Readers," *Academic Radiology* 31, no. 10 (2024): 4261–4271.

32. P. D. Y. Trieu, C. R. Mello-Thoms, M. L. Barron, and S. J. Lewis, "Look How Far We Have Come: BREAST Cancer Detection Education on the International Stage," *Frontiers in Oncology* 12 (2022): 1023714.

33. MedCalcSoftware L, "Diagnostic Test Evaluation Calculator" (2024), https://www.medcalc.org/calc/diagnostic_test.php.

34. E. U. Ekpo, M. Alakhras, and P. Brennan, "Errors in Mammography Cannot Be Solved Through Technology Alone," *Asian Pacific Journal of Cancer Prevention* 19, no. 2 (2018): 291–301.

35. S. Waite, A. Grigorian, R. G. Alexander, et al., "Analysis of Perceptual Expertise in Radiology – Current Knowledge and a New Perspective," *Frontiers in Human Neuroscience* 13 (2019): 213, https://doi.org/10.3389/fnhum.2019.00213.

36. M. G. Thurfjell, A. Lindgren, and E. Thurfjell, "Nonpalpable Breast Cancer: Mammographic Appearance as Predictor of Histologic Type," *Radiology* 222, no. 1 (2002): 165–170.

37. J. S. Mitnick, M. F. Vazquez, M. N. Harris, and D. F. Roses, "Differentiation of Radial Scar From Scirrhous Carcinoma of the Breast: Mammographic-Pathologic Correlation," *Radiology* 173, no. 3 (1989): 697–700.

38. T. Retson, P. Kocher, D. Anaam, S. Fazeli, M. Eghtedari, and H. Ojeda-Fournier, "Do Not Miss Architectural Distortion," *Radiographics* 44, no. 9 (2024): e240024.

39. P. Cherel, V. Becette, and C. Hagay, "Stellate Images: Anatomic and Radiologic Correlations," *European Journal of Radiology* 54, no. 1 (2005): 37–54.

40. N. Mohd Norsuddin, C. Mello-Thoms, W. Reed, M. Rickard, and S. Lewis, "An Investigation Into the Mammographic Appearances of Missed Breast Cancers When Recall Rates Are Reduced," *British Journal of Radiology* 90, no. 1076 (2017): 20170048.

41. W. I. Suleiman, M. F. McEntee, S. J. Lewis, et al., "In the Digital Era, Architectural Distortion Remains a Challenging Radiological Task," *Clinical Radiology* 71, no. 1 (2016): e35–e40.

42. E. Lazarus, M. B. Mainiero, B. Schepps, S. L. Koelliker, and L. S. Livingston, "BI-RADS Lexicon for US and Mammography: Interobserver Variability and Positive Predictive Value," *Radiology* 239, no. 2 (2006): 385–391.

43. J. H. Hayward, K. M. Ray, D. J. Wisner, et al., "Improving Screening Mammography Outcomes Through Comparison With Multiple Prior Mammograms," *American Journal of Roentgenology* 207, no. 4 (2016): 918–924.

44. A. A. J. Roelofs, N. Karssemeijer, N. Wedekind, et al., "Importance of Comparison of Current and Prior Mammograms in Breast Cancer Screening," *Radiology* 242, no. 1 (2007): 70–77.

45. B. P. Soh, W. B. Lee, M. F. McEntee, et al., "Mammography Test Sets: Reading Location and Prior Images Do Not Affect Group Performance," *Clinical Radiology* 69, no. 4 (2014): 397–402.

46. W. H. Kim, J. M. Chang, H. R. Koo, et al., "Impact of Prior Mammograms on Combined Reading of Digital Mammography and Digital Breast Tomosynthesis," *Acta Radiologica* 58, no. 2 (2017): 148–155.

47. M. G. Thurfjell, B. Vitak, E. Azavedo, G. Svane, and E. Thurfjell, "Effect on Sensitivity and Specificity of Mammography Screening With or Without Comparison of Old Mammograms," *Acta Radiologica* 41, no. 1 (2000): 52–56.

48. C. Varela, N. Karssemeijer, J. H. Hendriks, and R. Holland, "Use of Prior Mammograms in the Classification of Benign and Malignant Masses," *European Journal of Radiology* 56, no. 2 (2005): 248–255.

49. B. A. Qenam, T. Li, A. Alshabibi, H. Frazer, E. Ekpo, and P. Brennan, "Test-Set Results Can Predict Participants' Development in Breast-Screen Cancer Detection: An Observational Cohort Study," *Health Science Reports* 7, no. 6 (2024): e2161.

50. Y. Chen, J. J. James, E. J. Cornford, and J. Jenkins, "The Relationship Between Mammography Readers' Real-Life Performance and Performance in a Test Set-Based Assessment Scheme in a National Breast Screening Program," *Radiology: Imaging Cancer* 2, no. 5 (2020): e200016.

CHAPTER 6

DISCUSSION AND CONCLUSIONS

Breast cancer screening programs are established to detect early-stage breast cancer where treatment is most likely to be successful (1, 2). Screening programs aim to increase cancer detection rate with high sensitivity and specificity, and lower recall rate (3-5). Maintaining a balance between the detection of breast cancer via mammography and the number of women incorrectly recalled for additional testing is crucial to the success of screening programs (6). However, approximately 30% of cancers are missed in mammograms and about 80-85% of women whose mammograms show features suspicious of breast cancer and recalled for further testing turn out to be false positives (7-9).

Despite the technological advances in breast imaging and human interventions, the proportion of cancers missed, and the high number of women incorrectly recalled for further testing remains an issue for screening programs (7-9). Mammographic interpretation relies on the identification of suspicious abnormal changes in the breast parenchyma and surrounding tissues (10-12). It has been suggested that comparing mammograms from two or more screening rounds may allow these suspicious changes to be detected more easily (13-18). Consequently, there are recommendations for screening services to store mammograms for future reference, and the practice of displaying previous and current screening mammograms has become well-established in programs such as BreastScreen Australia (19). However, due to challenges with image data storage and retrieval as well as population mobility, breast imaging readers might not always have access to previous mammograms.

This thesis examined the influence that access to prior mammograms has on the interpretation of screening mammograms and the detection and classification of different types of breast lesions. First the literature was examined to establish current evidence around the influence of prior mammograms on radiologists' observer performance and to identify gaps that need to be addressed to establish the role of prior mammograms to screening programs (13). While data from the literature review suggest that prior mammograms improved specificity without affecting sensitivity and cancer detection rate, the data were sparse and based on older screening technologies and observer performance methodologies. These older methodologies do not consider the influence of prior mammograms at the level of the lesion (13). Also, radiographers are increasingly undertaking autonomous interpretation of mammograms, but the influence of prior mammograms on their performance has not been examined. Furthermore, while the diagnosis of breast cancer is based on the detection and risk classification of lesions in mammograms, the impact of prior mammograms on the detection and classification different lesion types has not been established (13). This thesis closed these gaps in three pieces of work establishing the impact that reference to prior mammograms has on:

- radiologists' performance at both the lesion and case levels, and the influence that radiologists' characteristics and breast density have on their subsequent performance (Chapter 3)
- radiographers' performance in interpreting screening mammograms (Chapter 4)
- the detection of different breast lesion types in screening mammograms by radiologists and radiographers (Chapter 5)

Evidence generated from these pieces of work demonstrate that: 1) access to prior mammograms improves radiologists' ability to correctly classify mammograms with no cancer features without affecting their interpretation of mammograms with breast cancer (20). Prior

mammograms also reduced their likelihood of making false positive errors. The impact of prior mammograms on radiologists' diagnostic efficacy was not influenced by a woman's breast density or the experience and practice characteristics of the radiologist interpreting the mammograms as shown in Chapter 3 (20); 2) reference to prior mammograms improves the ability of breast imaging radiographers to distinguish between normal mammograms from those with breast cancer and establish the level of malignancy of detected lesions. Prior mammograms also reduces the false positive rate irrespective of radiographers' years of experience and the number of hours spent working in mammography as shown in Chapter 4 (21); 3) access to prior mammograms improves radiologists' ability to detect spiculated/stellate lesions without affecting their ability to detect discrete masses, calcifications, non-specific densities, and architectural distortions. Prior mammograms do not influence the ability of radiographers to detect different breast cancer lesion types but improve their overall diagnostic performance in distinguishing normal and cancer containing mammograms as reported in Chapter 5. Insights from these studies and their implications and relevance are discussed below.

FINDINGS FROM CHAPTER 3 "ACCESS TO PRIOR SCREENING MAMMOGRAMS AFFECTS THE SPECIFICITY BUT NOT SENSITIVITY OF RADIOLOGISTS' PERFORMANCE".

The findings in Chapter 3 show that reference to prior mammograms when interpreting mammograms from the current do not influence the radiological detection of breast cancer. Rather, prior mammograms improve the interpretation of normal mammograms that do not show radiological signs of breast cancer. This is illustrated by the results of the different diagnostic performance metrics. For example, sensitivity, which represents the ratio of the number of mammograms with cancers that were correctly classified, and the total number of

cancers present in the dataset were similar when readings occurred with versus without prior mammograms. Also, lesion sensitivity values, which is the ratio of the number cancer lesions correctly marked by radiologists and the total number of lesions in the dataset, and JAFROC which accounts for the number of lesions detected and their risk classification were statistically similar across readings with and without prior mammograms. On the other hand, results of the performance evaluation metrics that focus on normal cancer-free mammograms showed that radiologists were more accurate in classifying mammograms with no cancer features with high specificity when prior mammograms were available. Consequently, reference to prior mammograms reduced the proportion of cancer-free mammograms that were incorrectly reported as having cancer and the likelihood of radiologists' making false positive interpretations.

Radiologists are trained to identify features of disease in radiological images (22). In the breast cancer screening setting, their major task is to detect breast cancer (23). Many breast radiologists have a mental schema of cancer features and actively look out for these features when interpreting mammograms (22, 24). When features that are discriminative of breast cancer are detected, radiologists are most likely to detect and classify these features according to their perceived risk of malignancy irrespective of whether prior mammograms are available (24). The risk classification is used by screening programs to recall women whose mammograms show suspicious cancer features and receive a high-risk rating for additional testing or diagnostic work-up. The distinctiveness of cancer features in mammograms and the ability of radiologists to detect these features may explain why no significant differences were observed between readings with and without prior mammograms in terms of false negatives, sensitivity, and lesion sensitivity. While only a small percentage of mammograms of women participating in screening mammography contain breast cancer (1), a significant number of

mammograms are incorrectly determined to contain cancer features due to the heterogeneity of the breast parenchyma, and lead to many women being incorrectly recalled for further testing. For example, in Australia, approximately 190,000 of the 1.9 million women participating in breast cancer screening are recalled due to suspicious findings in their mammograms (1, 25). Women recalled at screening undergo additional testing using different imaging modalities such as ultrasound, digital breast tomosynthesis, contrast-enhanced mammography depending on the routine in the screening service (26), but less than 10% of these women are eventually confirmed to have breast cancer (1, 4, 27, 28). These statistics suggest that reducing false positive recalls is equally as important as improving the cancer detection rate.

Interestingly, data from the study in Chapter 3 revealed that access to prior mammograms reduces radiologists' false positive recall as demonstrated by improvement in specificity and reduction in the false positive rate when prior mammograms were available. It should be noted that some of the normal mammograms in the test set used in Chapter 3 contained benign changes, which were considered malignant when priors were not available, but were reported as normal (no cancer) when radiologists had access to prior mammograms. These findings suggest that prior mammograms not only improved the classification of normal images but also helped radiologists to distinguish between benign and malignant changes in mammograms. The impact of prior mammograms on the correct interpretation of normal mammograms may be explained by factors such as the nature of changes induced by benign and malignant lesions. Benign lesions are usually slow-growing and less likely to significantly disrupt the breast parenchyma unlike relatively faster growing cancer lesions which have the potential to cause pronounced changes in the breast parenchyma and invade surrounding tissues (29, 30). Therefore, benign lesions detected on mammograms from the current screening round are more likely to be correctly classified as benign when comparison with prior mammograms show no

significant differences in the features of the lesions. The ability of breast screen readers to combine information from current and prior mammograms may be the reason why screening recall rates are lower at subsequent screens than at first screen. Current data show that the proportion of women recalled for assessment following a first screen in the Australian national breast screening program continues to increase whereas the recall to assessment has decrease at subsequent screens (1, 31). For example, recent data from the Breastscreen Australia monitoring reports show that in 2021, the recall to assessment for women aged 50 – 74 years was 11.1% in the first screening round, but this dropped to 4% for those participating in the subsequent screening rounds (1, 31). Such important decreases and the findings in Chapter 3 highlight the relevance of prior mammograms to Breastscreen readers and screening programs.

The findings in chapter 3 are in concordance with previous studies that have examined the influence of prior mammograms of radiologists' diagnostic performance in terms of sensitivity and specificity (32-37). Chapter 2, which summarises evidence from previous studies found that access to prior mammograms does not improve the sensitivity and cancer detection rate of radiologists; however, most of the studies included in the review reported improvement in specificity and reduction in the false positive rate and recall rates when radiologists referred to prior mammograms (32-37). Findings from the review in Chapter 2 show that on the average, specificity is 8% higher and the false positive rate is 14% lower when radiologists interpret mammograms with priors compared to when reading without prior mammograms. Evidence from the review in Chapter 2 aligns with the results in Chapter 3, which showed that access to prior mammograms led to a15% improvement in specificity and 2.6 times lower false positive rate when prior mammograms were available to radiologists. Chapter 2 also showed that reference to prior mammograms reduced the recall rates by 15%. while recall rates were not directly examined in Chapter 3, the improvement observed in specificity and reduced false

positive rate suggest that prior mammograms influence the recall rate by reducing false positives, which are a major concern for screening programs.

Although some of the findings in Chapter 3 have been previously reported in other studies, there were important gaps and methodological limitations, and some of the data reported in previous studies did not represent the reality of current practice. When breast cancer is suspected in mammograms, the suspicious regions of the breast are biopsied, and histopathology analysis is used as a reference standard to assess the diagnostic performance of imaging tools (26). The outcome of histopathology depends on the representativeness of biopsy samples taken from the lesion (38, 39), and this highlights the need for radiologists to correctly localise the lesion in mammograms. The evidence reviewed in Chapter 2 were from studies using older observer performance methodologies or retrospectively analysed data incorporating obsolete screen-film technology and required radiologists to classify mammograms as normal or abnormal, without stating where in the mammogram the cancer is located. These gaps and limitations were addressed in Chapter 3, where mammograms from currently available digital systems were used, the same radiologists interpreted mammograms of the same women with and without priors, and lesion sensitivity and JAFROC analyses were incorporated into the evaluation of the impact of prior mammograms on radiologists' performance. The lesion sensitivity and JAFROC methodologies in Chapter 3 accounted for the impact that prior mammograms have on radiologists' ability to localise the cancer lesions and classify them according to their risk of malignancy in keeping with the clinical guidelines for reporting breast lesions. Also, for the first time, the work in Chapter 3 examined the risk of false positives with and without prior mammograms and reported a 62% reduced risk of false positive interpretations when prior mammograms were available. Therefore, evidence reported in this thesis considered the influence of prior mammograms on radiologists' performance both

at the case level (sensitivity, specificity, and ROC) and lesion level (lesion sensitivity and JAFROC).

One variable that influences the interpretation of mammograms is the composition of a woman's breast. As described previously in Chapter 1, radiologists have difficulty in detecting breast cancer in mammograms of women with extremely dense breasts compared to those with fatty breast tissue (40, 41). Also, women with dense breasts account for 60% of screen recalls and are most likely to be false positive recalls (42, 43). Therefore, the assessment of the influence of prior mammograms on the efficacy of radiologists' interpretations should consider the impact that breast density has on the interpretation process. Prior to the study in Chapter 3, only one study (44) examined whether mammographic breast density influenced the results obtained when readings with prior mammograms were compared to those without prior mammograms. However, this study (44) was a retrospective analysis of mammograms read with versus without prior mammograms, with most of the radiologists who read the mammograms with priors different from those who read without prior mammograms. Therefore, the outcome of the study did not account for the impact that radiologists' expertise has on the results of this comparison. To ensure that the results presented in Chapter 3 were not influenced by mammographic breast density, radiologists' performance in dense (heterogeneously dense and extremely dense) were compared their performance (almost entirely fatty and scattered fibroglandular tissue) in non-dense breasts. The outcome of this comparison showed that radiologists' sensitivity, specificity, and lesion sensitivity were similar between dense and non-dense breasts. To further confirm these findings, radiologists' performance with versus without prior mammograms were compared separately in dense and non-dense breasts. These adjustments did not change the impact of prior mammograms on radiologists' performance. Specificity remained consistently higher when prior mammograms

were available, and sensitivity and lesion sensitivity remained unchanged in both dense and non-dense breasts.

Another factor that determines the accuracy of mammogram interpretation is the expertise of the radiologist who interprets the mammograms (3, 7, 45). Data from previously published studies show that the accuracy of mammography interpretation vary across radiologists (45, 46). The findings of these studies indicate that radiologists are the most important piece to improving early cancer detection and reducing false positive error rates. Since the beginning of the 20th century, there has been significant research focusing on the identification of the personal and practice-related characteristics of radiologists that improve diagnostic efficacy in the mammography screening and diagnostic settings (3, 7, 45-48). Radiologists' parameters widely investigated include specialisation in breast imaging, radiologists' experience often quantified by the number of years reading mammograms, and the volume of mammograms read per year. Evidence from studies that have examined these factors demonstrate mixed results (3); however, the volume of mammograms read per year appears to be a better discriminator of performance in mammography interpretation. Therefore, many countries use reading volume criterion to certify or renew the certification of radiologists for independent mammography interpretation (45, 49, 50). Thus, it is necessary to adjust for the confounding effects of these factors when assessing the impact of an intervention on the accuracy of screening mammography interpretation. To further test whether the impact of prior mammograms identified in Chapter 3 was influenced by the characteristics of the radiologists, the analysis was adjusted for the number of years qualified as breast radiologist, the number of hours spent reading mammograms per week, the number of years reading mammograms, and the number of mammograms read per week. The results from these adjustments confirmed that the improvements observed in specificity and reduction in the false positive rate were not due

to the characteristics of the radiologists. Rather, the higher specificity and true negative rates, and lower false positives were due to the availability of prior mammograms to the radiologists.

These findings around the influence of prior mammograms are supported by the results presented in the Tables 3 and 4 of Chapter 3. In Table 3, when the readings with prior mammograms of radiologists of different characteristics were compared, breast radiologists with >20 years of specialisation demonstrated similar sensitivity, lesion sensitivity, ROC, and JAFROC with those with ≤20 years specialist experience; only specificity was higher for those specialised for more than 20 years when prior mammograms were available. When other radiologists' characteristics (reading volume and number of years reading mammograms) were examined independently, no significant differences were observed. When these radiologists were compared in readings without prior mammograms, no differences were found in terms of number of years specialised as breast radiologist, the number of hours spent reading mammograms per week, and the number of mammograms read per. While sensitivity, lesion sensitivity, ROC, and JAFROC were significantly higher in radiologists with greater than 10 years of experience, no difference was observed in specificity. Conversely, when the readings with prior mammograms were compared to readings without prior mammograms for radiologists with similar characteristics, specificity remained consistently higher when prior mammograms were available regardless of the adjustment made even though it did not reach statistical significance when years of specialisation and number of mammograms read per week were considered. Other performance metrics remained unchanged despite the multiple adjustments that were made to the analysis. These results confirmed that radiologists' personal and workload did not confound the effect that prior mammograms had on performance.

Together, the findings reported in Chapter 3 confirm that access to prior mammograms improve radiologists' ability to correctly classify mammograms that do not show features of breast cancer without affecting their ability to detect cancer in mammograms. The data also confirm that prior mammograms reduced the risk of false positive screening mammography interpretations. The work presented in Chapter 3 considered for the first time, the potential factors that may influence the impact of prior mammograms on radiological interpretation of screening mammograms and used novel observer performance evaluation methodologies that accounted for radiologists' performance at both the case and lesion levels. These considerations addressed the gaps in previous studies that examined the influence of prior mammograms on radiologists' performance. The findings that the impact of prior mammograms on the interpretation process is not influenced by the experience and characteristics of the radiologists interpreting the mammograms provide strong evidence for screening programs to store mammograms for future reference. These findings suggest that all radiologists benefit from having access to prior mammograms when interpreting mammograms from the current screening round. The findings also highlight the importance of prior mammograms to women of different breast densities and support the practice of screening programs of storing screening mammograms for future reference.

FINDINGS FROM CHAPTER 4 "DOES ACCESS TO PRIOR MAMMOGRAMS IMPROVE THE PERFORMANCE OF RADIOGRAPHERS IN INTERPRETING SCREENING MAMMOGRAMS?"

The work in Chapter 3 focused on radiologists who are the traditional custodians of radiological image interpretation. However, as described in Chapter 1, the interpretation of mammograms faces several workforce challenges. First the current workforce is less than that required to

sustain screening service, particularly with half of breast radiologists working part time (51, 52). Second, the breast radiologists' retirement rate is 23% higher than the growth rate, which has led to a decline in the workforce across the UK and Australia (52-54). Third, screening mammograms uptake continues to increase with ageing population, and the extension of the screening age to 74 years and large volumes of images from screening and assessment of radiologists further challenge the interpretation of mammograms (1).

To address the breast radiologist's workforce shortage, advanced practice radiographers, particularly in the United Kingdom have been commissioned to interpret both screening and diagnostic mammograms independently or in pairs with radiologists (55, 56). Given the differences in training between radiologists and radiographers, it was necessary to examine the impact that prior mammograms have on the performance of radiographers. Therefore, the work presented in Chapter 4 examined for the first time the influence that prior mammograms have on radiographers' interpretation of mammograms.

The results presented in Chapter 4 revealed that prior mammograms improved specificity, ROC, and JAFROC, and reduced the false positive rate without affecting sensitivity and lesion sensitivity. These findings imply that when radiographers have access to prior mammograms, they are better able to discriminate between mammograms which show features of breast cancer from normal mammograms of women who do not have breast cancer. The results also indicate that prior mammograms improve radiographers' performance in detecting breast cancer and classifying lesions according to their risk of malignancy and reducing the proportion of mammograms are incorrectly reported as showing cancer features.

Previous studies that have examined radiographers' performance in the interpretation of mammograms have shown promising results (55, 56). Studies based on the UK's setting report that advanced practice radiographers interpret screening and diagnostic mammograms at the level comparable to UK radiologists (55, 56). These findings have supported the integration of radiographers into the national health service breast screening program (NHSBSP) as autonomous interpreters of screening and diagnostic mammograms. Studies across other countries including Mexico and Australia report that radiographers demonstrate optimal performance, but wide variabilities in the interpretation of mammograms (57-61). For example, in Australia and New Zealand, radiographer performance analyses reported sensitivity values ranging from 55% to 100% and specificity ranging from 63% to 98% (57-60). However, pooled estimates of the results from these studies demonstrate sensitivity and specificity values below that reported for Australian radiologists (57-60), highlighting the need for interventions to improve the mammography interpretation performance of these radiographers.

While evidence from these pieces of research has demonstrated the potential for radiographer participation in mammography interpretation, it has also highlighted the need for strategies to optimise their performance. However, no study has examined the impact that prior mammograms have on radiographer performance. It should be noted that radiographers do not routine interpret mammograms in Australia and New Zealand. The study presented in Chapter 4 involved Australian and New Zealand radiologists working in their respective breast screening services. Despite these radiographers not reporting mammograms clinically, they demonstrated strong abilities to detect cancer in mammograms, with a mean sensitivity of 95%, and these abilities did not change regardless of whether they had access to the prior mammograms of these women. The high sensitivity values recorded with and without prior mammograms may be due to their experience and knowledge gained through interacting with

mammograms of women with and without cancer features over the course of their breast imaging practice. These findings suggest that experiential learning may have played an important role in equipping these radiographers to interpret mammograms.

The findings also provide evidence to support the incorporation of breast imaging radiographers into Breastscreen Australia as screen readers. Alternatively, with the expansion of screening services to include mobile screening mammography in Australia, breast imaging radiographers can provide preliminary comments when acquired mammograms show suspicious cancer features. Even though sensitivity and lesions sensitivity were not significantly different between readings with and without priors in Chapter 3, JAFROC values were significantly higher when prior mammograms were available. These findings indicate that radiographers can detect more cancer lesions and more accurately establish the risk of malignancy of these lesions when prior mammograms were available. Therefore, it is important for screening services to retain mammograms for future reference and for breast imaging radiographers to actively refer to these images when undertaking preliminary review of mammograms or considering the need for additional views to support diagnostic decision-making. Breast radiographers' ability to identify suspicious features of breast cancer at the time of screening and comment on these features will improve screening service efficiency by ensuring that women with suspicious lesions are offered immediate additional views or supplemental imaging at the screening sites. This will also facilitate the triaging of images with suspicious findings by radiographers for immediate reporting by radiologists and assist in expediting clinical decision-making around biopsy recommendations and early treatment of pathologically confirmed breast cancers in line with the goal of breast cancer screening programs.

As much as it is important for breast radiographers to be able to identify features that are suggestive of breast cancer, it is equally important for them to correctly identify mammograms of women who do not have cancer. Some screening services in Australia allow breast radiographers to acquire additional mammographic views for women whose screening mammograms are deemed to show suspicious cancer features. Therefore, correctly classifying mammograms with no cancer will reduce unnecessary additional views and radiation dose to women do not have breast cancer. Interestingly finding in Chapter 4 revealed access to prior mammograms improved radiographers' abilities to discriminate mammograms without breast cancer from those with cancer as shown by a significantly higher ROC values when prior mammograms were available. It should be noted that the higher ROC values with prior mammograms were driven by high specificity, which describes the ability of the radiographers to correctly classify mammograms without breast cancer. Specificity was approximately 21% lesser when radiographers read the mammograms without access to their priors. As discussed previously, women whose mammograms are deemed to show features suggestive of breast cancer are recalled to assessment clinics for additional testing. Screening services use the recall rate to assess the performance of the screening program and aim to detect as many cancers as possible while reducing the number of normal cancer-free women who are called back to assessment clinics (1, 25). The lower the specificity, the more the number of women who are incorrectly called back for additional testing; this incorrect recall to assessment does not benefit both women participating in breast cancer screening programs and the screening services. Interestingly the findings in Chapter 4 concur with the results presented in Chapter 3, where specificity was significantly higher in readings with prior mammograms and further confirm the impact that prior mammograms have a significant impact on the effectiveness of screening services. These results indicate that prior mammograms have a similar impact on

mammography interpretation rate regardless of whether mammograms are interpreted by radiologists or radiographers.

Another common finding across studies involving radiographer reporting of mammograms has been high false positive rates ranging from 35 – 49% (61, 62). The false positive rate describes the proportion of normal cancer-free mammograms incorrectly classified as having breast cancer. High false positive rates for radiographers reporting mammograms have been cited as a major issue limiting the use of radiographers as Breastscreen readers in many countries (61, 62). This is because have false positive rates have negative implications for women participating in screening, the screening services, and the government. For example, false positive screening results lead to long-term psychosocial harms (63, 64) and increase women's physical and economic burdens (65-67); these negative outcomes cause some women to avoid future screening or delay their subsequent screening examinations (67) even though they may be at higher-than-average risk of developing cancer in the future (68) and being diagnosed of late stage breast cancer(67). Women who received a false positive result would have undergone additional testing including spot mammographic views, ultrasound, digital breast tomosynthesis (DBT), contrast-enhanced mammography (CEM), and/or magnetic resonance imaging (MRI) as part of their diagnostic assessment before being told that they are cancer-free (26). These additional tests further increase radiation dose to women and lead to lost time without benefits (26). Additional testing further increase pressure on already outstretch screening services' resources including equipment and personnel and shifts additional cost to Medicare (69). These negative outcomes justify the concerns around integrating radiographers as screening mammography interpreter. The study in Chapter 4 highlighted for the first time, the impact that prior mammograms have on the false positive rate of radiographers when interpreting mammograms. The findings showed that prior mammograms reduced the false

positive rate from 40% to 18%, which is more than half the false positive rate when readings occurred without prior mammograms. Breast radiologists who routinely interpret mammograms demonstrate wide differences in their false positive rates, which can be as high as 24% (70, 71), suggesting that solutions that lower the false positive rates for radiographers to 18% may alleviate the concerns around employing radiographers as screen readers. If the findings in Chapter 4 can be confirmed in studies with larger samples of mammograms and radiographers, it may support the development of educational resources and continuous professional development activities for radiographers to reduce their false positive mammography interpretation rate.

Given that the radiographers that were examined in did not interpret mammograms clinically, the work in Chapter 4 examined if their work-related experiences influenced the results obtained when readings with prior mammograms were compared to these without prior mammograms. Using the self-reported information provided by the radiographers who participated in the study, three work-related characteristics were established: years qualified as radiographer (≥30 versus <30 years); years specialised in as breast radiographer (≥25 years versus <25 years); number of hours spent working in mammography (≥16 hours versus <16 hours). When the radiographers were compared in readings with prior mammograms based on these work-related characteristics, no statistically significant differences were observed. Similarly, no differences were found when they were compared in readings without prior mammograms. However, when the performance of radiographers with the same work-related characteristics were compared between readings with versus without prior mammograms, significant trends emerged. The radiographers' performance when prior mammograms were available was consistently better than when prior mammograms were not available, particularly in terms of specificity, false positives, ROC, and JAFROC (see Table 3 in Chapter 4). The

work presented in Chapter 4 was also the first to report whether the work-related experience or characteristics of radiographers affect their performance in mammography interpretation. These findings indicate that the impact of prior mammograms on radiographers' abilities to interpret mammograms is not influenced by their work-related experience characteristics. It also suggests that all radiographers benefit from looking at prior mammograms when interpreting mammograms from the current screening round. The findings in Chapter 4 around the lack of influence of radiographer experience on the impact of prior mammograms also align with the results presented in Chapter 3 regarding radiologists. Collectively, these pieces of research indicate that access to prior mammograms is an independent factor that enhances the efficacy of screening mammography interpretation through improvement in the classification of normal cancer-free mammograms.

With the Australian breast radiologists' workforce shortage expected to get worse in the next five years (52, 53), it has become increasingly important to train Australian radiographers in mammography interpretation. Published studies that have examined the mammography interpretation skills of Australian radiographers who have not received formal training in mammography interpretation show that formal screening mammography training may address their limitations (57-60). Targeted training packages for Australian radiographers has only been effective in improving their sensitivity without impacting specificity and false positive rates (72). Interestingly, the findings in Chapter 4, which considered radiographers' performance at both the case and lesion level show that reference to prior mammograms addresses the specificity and false positive concerns for using radiographers as screen readers. Therefore, training interventions that incorporate systematic analysis of prior mammograms may upskill radiographers to report mammograms so that the breast radiologists' shortage gaps can be filled and the benefits of radiographer Breastscreen-reading accrued in Australia.

FINDINGS FROM CHAPTER 5 "IMPACT OF PRIOR MAMMOGRAMS ON RADIOLOGISTS AND RADIOGRAPHERS' DETECTION OF DIFFERENT BREAST CANCER LESION TYPES".

The pieces of work in Chapters 3 and 4 provided evidence on the influence of prior mammograms on the performance of radiologists and radiographers in the interpretation of mammograms. As described in these two Chapters, reader performance in mammography interpretation is assessed based on their abilities to distinguish between normal cancer-free mammograms and those with breast cancer, as well as their abilities to detect and classify breast lesions according to their risk of malignancy. Radiologically, breast cancer lesions display different features in mammograms (73-75). These lesion features are used by radiologists to classify lesions into different types by different breast reporting lexicons including the BI-RADS® and RANZCR classification systems (76, 77). Since the risk of malignancy vary across lesion types, not all lesions visible in mammograms are considered cancerous by breast imaging readers. The results presented in Chapters 3 and 4 show that access to prior mammograms did not improve sensitivity and lesion sensitivity for both radiologists and radiographers. However, JAFROC values were better when readings occurred with prior mammograms available even though it did not reach statistical significance for radiologists (20, 21). JAFROC examines observer performance in detecting multiple lesions and simultaneously establishing malignancy risk. Thus, the findings in Chapters 3 and 4 highlighted the potential impact that prior mammograms may have on the detection and risk classification of different lesion types. The results presented in Chapters 3 and 4 were the performances of radiologists and radiographers respectively for all lesion types combined without considering the impact that different lesion types had on the results. Chapter 1 and the evidence in presented

in Chapter 2 also highlighted the knowledge gaps around the impact that prior mammograms have on the detection of different lesion types, which needed to be addressed.

Further to the knowledge gaps identified in the literature as described previously in the thesis, the work in chapter 5 provided new knowledge to close these gaps. The results demonstrated that the availability of prior mammograms to radiologists improved their ability to detect and classify spiculated/stellate lesions without limiting their abilities to detect and classify other lesion types including discrete masses, calcifications, non-specific densities, and architectural distortions. When the mammography interpretations by radiographers were considered separately, the availability of prior mammograms did not influence on the radiographers' abilities to detect and classify different lesion types. However, their overall diagnostic accuracy, a metric that considers their abilities to distinguish mammograms with cancer from those without cancer consistently improved when prior mammograms were available. The diagnostic accuracy of radiologists also improved reaching statistical significance in architectural distortions. As described in Chapters 1 and 5, the classification of lesions into radiological types has become a common norm across all breast imaging reporting lexicons due to their differences in the risk of malignancy. The literature demonstrates that different radiological lesion types have different histological correlations and cancer predictive abilities (73-75). For example, the risk of malignancy in architectural distortions vary from 10–50% due to an array of other benign conditions, which may be responsible for such mammographic appearance, from complex sclerosing lesion, radial scars, and fat necrosis (78). Discrete masses and non-specific densities are both associated with a wider range of causative factors, from benign to malignant (79). Changes in discrete masses and non-specific densities have been shown to increase the risk of malignancy by only 10% (79). Therefore, these lesion types are most likely to be reported as benign or indeterminate (80). The work in Chapter 5 included

prior mammograms acquired 2 – 4 years prior to enable sufficient time for cancer lesions to induce changes in current mammograms. Despite the interval between the prior and current mammograms in the test set, no significant differences were observed in the performance of radiologists in terms of sensitivity and lesion sensitivity, particularly for calcification, discrete masses, non-specific density, and architectural distortions. These findings may be due to the low risk of malignancy in these lesion types or the inability of the radiologists and radiographers to identify changes in the lesion over time that indicate malignancy.

Evidence from published studies show that when cancer cells interact with normal adipose tissue, these interactions cause changes in the breast parenchyma which appear spiculated/stellate in mammograms (81, 82). Spiculated/stellate lesions have a high propensity for malignant transformation (75). Evidence shows that stellate appearance is the most typical feature of breast cancer in mammograms and constitutes between 75–85% of palpable malignant breast lesions, with a positive predictive value between 75% –90% (83). A previous study reported that radiologists easily detected stellate lesions and were most likely to recall women whose mammograms show stellate lesions for assessment compared to discrete masses, calcifications, non-specific densities, and architectural distortions, even at lower recall rates (80). Therefore, when spiculated/stellate lesions or changes in this lesion type are detected, radiologists are mostly likely to assign a higher risk rating to facilitate histopathology analysis (83). Histopathology analyses also show that stellate mammographic appearances are highly predictive of invasive breast cancer (73-75). Approximately 71% of stellate lesions are luminal Type A molecular subtype of breast cancer, which is associated with good prognostic outcome compared to other lesion types (73, 74). Therefore, since stellate lesions are less likely to be false positive and have a good prognostic outcome, changes in the lesions following comparison with prior mammograms may increase the suspicion of malignancy. The

previously established high risk of malignancy in spiculated/stellate lesions and the results presented in Chapter 5, suggest that prior mammograms provide opportunities for Breastscreen readers to identify changes in spiculated/stellate lesions associated with higher-than-average risk of malignancy. These pieces of evidence may explain why access to prior mammograms improved radiologists' detection and classification of spiculated/stellate lesions. It is also possible that in addition to the high risk of malignancy in stellate lesions, perceptual mechanisms may underpin the results in Chapter 5. Therefore, eye-tracking studies are needed to provide further information to establish whether the interrogation of the different lesion types differ with and without prior mammograms and explain the other reasons for the findings in Chapter 5.

Published research regarding radiologists' performance in detecting different lesion types focused on metrics such as sensitivity (47) or inter-reader agreement (84, 85) and on the likelihood of malignancy for each lesion type (84). Prior to the study in Chapter 5, only one study had explored the impact that prior mammograms have on radiologists' ability to detect different lesion types (44); however, the study had important limitations as discussed in Chapter 5, which may impact upon the translation of these findings. For example, the study was a retrospective analysis of radiologists' performance results, and the radiologists who interpreted the mammograms with priors were not always the same as the ones who interpreted mammograms without prior mammograms. The literature shows most errors in the interpretation of mammograms arise from the intrinsic limitations of radiologist and that radiologists' factors such as volume of mammograms read per year and years of reading mammograms affect reader performance (3, 7). Therefore, the previous study that had examined the impact of prior mammograms on the detection of different lesion types did not account for the radiologists' factors that could influence interpretation outcomes (44). The

work in Chapter 5 performed the first robust analysis of radiologists' performance comparing the results of the same cohort of radiologists in detecting each lesion type with and without access to prior mammograms of the same cohort of women. It also reported for the first time the impact that prior mammograms have on radiographers in detecting each lesion type when prior mammograms were available compared to when they were not available. Therefore, the work reported in Chapter 5 addressed the limitations in the published literature. It also provided new knowledge, which highlighted for the first time that the lack of improvement in radiologists' sensitivity and cancer detection rate reported previous studies (13, 15, 17, 18, 36, 44, 86) when reading with prior mammograms may be due to the mixture of different lesion types in these studies.

The findings in Chapter 5 suggest that the correct classification of mammographic features into different lesion types may impact upon the performance of screening services in different ways. First, when suspicious areas are identified in mammograms of women, these women are called back for additional testing using different imaging modalities as discussed previously (26). The literature shows that the diagnostic efficacy of these imaging assessment tools varies across different lesion types (26). For example, while DBT has been shown to have a sensitivity of up to 100% in detecting calcifications, its specificity for calcifications can be as low as 0% (87). Also, while ultrasound underestimates the risk of malignancy in breast calcifications, its specificity can be as high as 72% (87). These findings suggest that the detection and correct classification of lesion types can be used to select the choice of imaging assessment modality or the combination of imaging modalities that improve the detection of breast cancer. Such strategies will lead to a higher number of cancers being detected and treated early to reduce deaths from the disease. Secondly, another goal of screening services is to reduce the number of women whose mammograms show features suggestive of cancer who turn out to be false

positives following further testing (7-9). It is well established that approximately 80–85% of suspicious features on mammograms are normal or benign (1, 4, 27, 28). Since the risk of malignancy vary across different lesion types, the detection and correct classification of lesion types can be used to establish the risk of malignancy and reduce the chances of benign lesions being incorrectly recalled for assessment. Because the analysis in Chapter 5 focused on lesion analysis, specificity, which measures the influence that prior mammograms have on the reporting of the normal and benign cases was not assessed; however, the specificity results in Chapters 3 and 4, and the outcome of the diagnostic accuracy analysis in Chapter 5 support the impact that prior mammograms have radiologists' abilities to discriminate normal and benign features from cancer features. Prior mammograms also improved radiographers' abilities to discriminate normal and benign mammographic features from features due to malignant discrete masses, nonspecific density, and spiculated/stellate mases. The ability of prior mammograms to improve the classification of normal mammograms and those with benign changes support their usefulness in reducing the number of lesions incorrectly recalled to assessment clinics. Such reduction will reduce screening services workload of conducting and reporting assessment images, the cost of additional testing for women recalled, and the number of women exposed to the psychosocial harms associated with false positive recalls. Therefore, even though the initial costs of archiving mammograms may be high, the long-term benefits accrued from storing and retrieving mammograms may surpass these initial costs.

The results pertaining to radiographers in Chapter 5 also highlight the role that prior mammograms may play in supporting Breastscreen radiographers involved in providing mobile screening services. In Australia where the studies reported in this thesis were conducted, specialised trucks or vans customised with portable digital mammography machines are used to provide mobile screening mammography to communities. Such mobile screening ensures

that screening services are accessible to women in rural areas, underserved communities, working women, and to reduce long distance travel to screening sites. With the absence of radiologists at these screening sites, the initial decision-making around additional views due to suspicious findings rests solely on breast radiographers. Interestingly, the higher diagnostic accuracy when prior mammograms were available suggest that providing radiographers access to prior mammograms may improve their decision-making around additional views. It should be noted that when prior mammograms were available, radiographers detected more discrete masses and non-specific density lesions, but fewer spiculated/stellate lesions, which are more predictive of breast cancer. These findings contrast slightly with those of radiologists, who detected significantly more spiculated/stellate lesions when prior mammograms were available.

While the differences in the detection of different lesion type by radiographers did not reach statistical significance, the findings do highlight the need for solutions to improve radiographers' use of prior mammograms as a benchmark to identify the most discriminant malignant changes in mammograms from the current screening round. Such a solution will ensure that women whose mammograms show spiculated/stellate lesions are effectively identified so that additional views can be taken for radiological assessment.

LIMITATIONS OF THE THESIS

The research contained in this thesis have common limitations which must be considered when interpreting the findings of this thesis. These limitations are related to the designs of the studies including the laboratory nature, participant recruitment, and mammographic case sample size.

However, the rationale for the designs is well-grounded in the scientific literature as described below.

The work reported in Chapters 3, 4, and 5 employed a laboratory-based experimental study design involving the interpretation of mammographic test sets under controlled conditions. There are concerns that laboratory-based experimental study designs are limited by selection bias arising from participant recruitment, information bias or inconsistent data collection, lack of adjustment for potential confounders, and the laboratory effect (88-91). These challenges are considered to limit the generalisability and external validity of the findings (88-91). The studies reported in this thesis utilised a convenience sampling methodology to recruit radiologists and radiographers, resulting in cohorts of self-selected radiologists and radiographers, which could introduce selection bias. The study in Chapter 3 was a phase II observer performance study of medical imaging. The analysis of clinical assessment of clinical performance of medical imaging requires a sample of five to 10 observers. The sample of radiologists who participated in previous studies that have assessed the impact of prior mammograms on radiologists' performance used between 3 to 12 radiologists (13, 15, 17, 18, 36, 44, 86). In Chapter 3 of this thesis, 11 radiologists were recruited, and the eight radiologists completed the reading of the two test sets and were included in the analysis. Three radiologists were excluded from the analysis because they completed only one test set and did not respond to reminders to complete the second test set. Therefore, the sample of radiologists in Chapter 3 meets the requirements for phase II observer studies and sits in the top quartile of the participant sample sizes used in similar studies. While a simple convenient sampling strategy was employed, the characteristics of the radiologists who participated in the study represent the population of radiologists reporting mammograms clinically for BSA.

The study in chapter 4 was a phase I (exploratory phase) observer performance study, which assessed for the first time the impact that prior mammograms have on radiographer performance. Phase I assessment of performance of medical imaging is necessary when a new protocol is first evaluated in a cohort of human subject and usually requires two or three participants (92). The study in Chapter 4 involved 13 breast radiographers, which is significantly higher than the numbers needed for exploratory baseline studies. Australian and New Zealand radiographers do not report mammograms clinically, which could limit the generalisation of their findings to advanced practice radiographers who report mammograms independently or as second readers in a double reading system. Therefore, the use of radiographers who do not interpret mammograms routinely could be considered as a limitation of the study. However, Australian radiographers have been shown to demonstrate optimum performance in mammography interpretation (57, 59, 60, 72, 93, 94), the radiographers examined in this thesis have an average of 28 years of experience, which is sufficient for developing knowledge in mammography interpretation as demonstrated by their which was within the range reported for advanced practice reporting radiographers (55, 56).

The studies in this thesis also involved image interpretation in two readings sections, which could introduce memory bias. The concerns related to memory and information bias, or inconsistent data collection was mitigated in Chapters 3, 4, 5 using the BREAST platform. Regarding memory bias, it has been shown that radiologists' image perception memory declines within weeks (95, 96). The second reading of mammograms in pieces of work reported in this thesis were undertaken after a washout period, which were up to six months – a washout period that was sufficient to mitigate memory bias. Therefore, memory bias is less likely to influence the results presented in this thesis.

Regarding concerns related to information bias, the use of the BREAST platform ensured consistency in the data collection from the prior and current mammographic test sets used for these studies. In addition, the participants involved in each of the studies reported the same set of prior and current mammograms belonging to the same set of women. Furthermore, only data of participants who completed the reading of the prior and current test sets using calibrated clinical grade primary monitors (5 mega-pixel) and in controlled ambient lighting conditions recommended clinically were included in the analysis. The standardisation of monitors and ambient lighting conditions ensured consistency in the reading conditions for all participants.

Lastly, the same survey tool was used to collect information related to the characteristics of the radiologists and radiographers who participated in the studies reported in Chapters 3, 4, and 5. The studies reported in this thesis accounted for potential confounders that influence the performance of mammography in breast cancer detection. For example, the work in Chapter 3 adjusted for breast density, which is a major factor that influence the sensitivity of mammography. The work in Chapters 3 adjusted for the characteristics of the radiologists that have been shown to confound reader performance including years qualified as radiologist or breast physician/specialty, hours spent reading mammograms per week, number of years reading mammograms, and the number of mammograms read per week. In Chapter 4 involving radiographers, adjustments were made for the number of years qualified as radiographer, years of specialisation as breast radiographer, and the number of hours spent working in a mammography service. These adjustments confirmed that breast density and the characteristics of the participants did not confound the impact that prior mammograms have on participants' performance. Since the analysis reported in Chapter 5 were based on the work in Chapters 3 and 4, no further adjustments were needed. Finally, regarding the laboratory nature of the study, an earlier study reported that radiologists' performance in the clinical setting is better than their

performance in a laboratory setting (88). The authors attributed their findings to the laboratory effect due to participants' knowledge of being tested (88). However, more recent studies show that radiologists' laboratory-based test set performance reflects the performance of these radiologists in clinical settings (36, 97, 98). Also, imaging referral guidelines such as the iRefer and Radiation protection 118 support the use of evidence from observer performance of diagnostic evaluation studies to inform clinical practice (99, 100). Thus, the external validity of the findings reported in this thesis are unlikely to be affected by the laboratory effect. Therefore, the findings can be used to support practice and policies around the use of prior mammograms by screening services.

Another potential limitation of the thesis is related to the design of the test sets and their sample sizes. Clinically, approximately one in 87 or more screening mammograms may contain breast cancer depending on the prevalence of the disease in the population (1). However, the pieces of research reported in Chapters 3, 4, and 5 were based on highly enriched test sets. In Chapter 3 there were 32 cancer cases in 72 mammograms. In chapter 4, there were nine cancers in 28 mammograms, and these test sets were combined in Chapter 5. It can be argued that the distributions of cancer cases in the test sets do not reflect the clinical setting; however, it was necessary for mammograms of different breast densities, breast lesion types, and from women of different ages to be adequately and evenly represented in the test sets. In addition, the designs of the test sets were to conform with the requirements for clinical assessment of clinical performance of medical imaging which require sufficient diseased cases (92). These requirements are to allow for robust comparisons, assess false positives and false negatives to establish how well observers perform when prior mammograms are available, and to assess performance on difficult cases and different lesion types (92). Regarding the sample sizes in Chapters 3, 4, and 5, it is important to note that they reflect practical constraints, including

limited availability of eligible participants and time/resource limitations. However, the small numbers conform to the sample sizes required for evaluation of observer performance in medical imaging and those used in previous studies that had assessed the influence of prior mammograms on radiologists' performance. For example, Chapter 3 was a phase II study involving radiologists which requires 50 – 100 mammograms, and Chapter 4 was a phase I study involving radiographers and requires10–50 mammograms (92, 101). Secondly, prior to the work reported in this thesis, the influence of prior mammograms was only tested on radiologists, and the sample size used in Chapter 3 is within the top quartile of the sample sizes used in similar observer performance studies reviewed in Chapter 2 of this thesis. Thirdly, assessing the impact of prior mammograms on performance require image interpretation at two different time points and involve significant time commitment. Therefore, it was necessary to develop test sets following scientific recommendation that can be completed in a reasonable time without overwhelming the participants or negatively impacting upon the results of the studies. Despite these limitations, the data provided valuable results consistent with findings from similar studies discussed in Chapter 2.

THESIS IMPLICATIONS

The pieces of evidence presented in this thesis collectively have four major implications:

- ***Implications for screening programs and women participating in mammography screening:*** Chapters 3 and 4 consistently demonstrate that reference to mammograms from the previous screening rounds improve the ability of radiologists and radiographers to correctly identify women whose mammograms do not contain breast cancer and reduce the number of women incorrectly classified as having breast cancer.

Chapter 5, which focused on lesion analysis also indicate that the ability of radiologists and radiographers to discriminate different lesion types from normal mammograms or benign lesions was mostly better when references were made to prior mammograms. These findings have implications for the recall rates, cost of screening services, and psychosocial harms and screening re-attendance for women participating in screening mammography. As discussed previously in this chapter, the recall rate is key criterion for assessing the performance of mammography screening services. Although the goal of screening is to detect as many cancers as possible in the early stages, screening services also aim to reduce the number of false positive recalls. This is because lowering the recall rate, reduces the number of cancer-free women recalled to breast assessment clinics, and consequently the human resources, workload, and costs associated with additional testing of these women (65). False positive recalls have been reported to cause long-term psychosocial harms such as anxiety, negative impact on behaviour, sleep, sexuality, attractiveness, and mental health (64, 102, 103). It has also been shown that false positive screening recalls significantly reduce the screening re-attendance rate (104-107). Therefore, the higher specificity and reduced false positive rate indicate that the storage, retrieval, and reference to prior mammograms is relevant for improving the efficiency of screening programs and the experiences of women participating in screening through lowering the false positive recall rate and associated costs and harms.

- ***Implications for a national accessible mammography database to support diagnostic decision-making.*** The additional analysis in Chapter 3 highlighted that radiologists' experience and workload characteristics do not influence the impact that prior mammograms have on their performance in interpreting mammograms. Prior mammograms consistently improved specificity and reduced false positive rates, and

the influence of prior mammograms was not altered by the number of years qualified as radiologist or breast physician/specialty, hours spent reading mammograms per week, number of years reading mammograms, and the number of mammograms read per week for radiologists. In Chapter 4, prior mammograms consistently improved specificity, ROC, and JAFROC, and reduced the false positive rate. Like with radiologists, the number of years qualified as radiographer, years of specialisation as breast radiographer, and the number of hours spent working in a mammography service did not alter the impact of prior mammograms on radiographers' performance. When the analyses in Chapter 3 were adjusted for breast density, the availability of prior mammograms consistently led to an increase in specificity in both dense and non-dense breasts. These findings imply that breast radiologists and breast radiographers of all experience and workload characteristics benefit from referring to prior mammograms when interpreting mammograms from the current screening round. They also indicate that women of all breast densities benefit from their mammograms being interpreted with the mammograms from their previous and current screening round reviewed simultaneously. These findings collectively emphasise the need for the establishment of a national accessible mammography database so that Breastscreen readers can have access and refer to prior mammograms of women moving across states.

- ***Implications for curricular innovations for radiographers participating in mammography interpretation.*** As discussed previously, the major factor limiting the implementation of radiographer reporting in Breastscreen programs is their high false positive rate. In Chapter 4 of this thesis, reference to prior mammograms reduced the false positive rate by 22% and improved the ROC and JAFROC values of radiographers. In Chapter 5 reference to prior mammograms improved radiographers'

diagnostic accuracy. The findings around ROC and accuracy suggest that when prior mammograms are made available, radiographers perform better in distinguishing cancer containing mammograms from normal mammograms and those with benign lesions. The findings around JAFROC suggest that when radiographers refer to prior mammograms during interpretation, they detect more cancer lesions in women with multiple lesions and more correctly establish the risk of malignancy in these lesions. However, in Chapter 5, fewer spiculated/stellate lesions were detected when prior mammograms were available suggesting that radiographers were unable to identify spiculated/stellate lesions or changes in this lesion type that indicate malignancy. Identifying and correctly classifying all cancer lesions is important for targeting biopsy and treatment planning. Thus, the findings in Chapters 4 and 5 pertaining to radiographers have important implications for radiographer training and continuous professional development in mammography interpretation. These findings suggest that curricular innovations that incorporate strategies to systematically evaluate and compare prior and current mammograms may mitigate radiographers' false positive errors and improve the detection of the lesion type that is most predictive of malignancy. Such curricular innovations may strategically position radiographers to undertake mammography interpretation with higher accuracy.

- ***Implications of lesion type detection and risk classification.*** The results in Chapters 3 and 4 demonstrate that prior mammograms do not have any impact on sensitivity; however, when the analysis was nested on lesion types in Chapter 5, prior mammograms improved the detection and classification of spiculated/stellate lesions by radiologists. The availability of prior mammograms mostly improved the detection of other lesion types even though the results did not reach statistical significance. The

implications of detecting and correctly classifying spiculated/stellate lesions have been discussed above (see discussion pertaining to Chapter 5). Briefly, while 75–90% of spiculated/stellate lesions turn out to be malignant, only 10% of discrete masses and non-specific density lesions are cancerous, highlighting why prior mammograms only significantly improved the detection of spiculated/stellate lesions. The findings in Chapter 5 provide insight into how comparison of mammograms from the previous screening rounds can improve the detection and risk classification of breast lesions that are most discriminative of malignancy. The findings emphasise the importance of reference to prior mammograms when mammographic features in mammograms from the current screening round point to spiculated/stellate lesions, as changes in the lesion most likely indicate the presence of malignancy.

FUTURE DIRECTIONS

The thesis has provided evidence on the impact that prior mammograms have on the interpretation of screening mammograms based on the current screening technology and observer performance evaluation methodologies incorporating both case- and lesion-level analyses. Further larger studies are required to confirm these findings. For example, the study in Chapter 3 was a phase II (challenge phase) evaluation of observer performance, which compared diagnostic performance at case and lesion levels, and adjusted for both radiologists and patient characteristics (breast density). Future studies should conduct multi-observer phase III (advanced phase) evaluation of the impact of prior mammograms on radiologists' performance using hundreds of patients sampled prospectively. The study in Chapter 4 was a baseline exploratory study (phase I study), which tested radiographer diagnostic abilities with versus without prior mammograms. Future studies exploring the influence of prior mammograms on radiographers' performance should be sufficiently powered to meet the

requirements of phase II or III observer performance evaluation studies. In Chapter 5, reference to prior mammograms to radiologists significantly impacted upon the detection and classification of spiculated/stellate lesions, but not other lesion types; however, the perceptual mechanisms underpinning the results presented in this thesis were not examined. Therefore, future studies should incorporate eye-tracking into the assessment of the impact of prior mammograms on observer performance. Such eye-tracking data may elucidate the perceptual mechanisms responsible for the results presented in this thesis.

CONCLUSIONS

The storage, retrieval and simultaneous presentation of prior and current mammograms for reporting improves specificity and reduces false positive rates of both radiologists and radiographers without reducing their abilities to detect and characterise breast cancer in mammograms. In addition, reference to prior mammograms improves radiologists' abilities to detect and classify spiculated/stellate lesions but have no impact on the detection of other lesion types. Reference to prior mammograms also improve radiographers' abilities to detect multiple breast lesions in mammograms and simultaneously establish their risk of malignancy. The impact of prior mammograms on the interpretation of screening mammograms is not affected by women's breast density and radiologists and radiographers experience or workload characteristics such as the number of years qualified, specialisation in breast imaging, and reading volume. These pieces of evidence highlight the potential for prior mammograms to improve the efficiency of screening programs by reducing false positive recall, and the need for policies to establish a national accessible mammographic database platform to account for population mobility across states.

REFERENCES

1. Australian Institute of Health and Welfare. BreastScreen Australia monitoring report 2024. In: AIHW, editor. Canberra: Australian Institute of Health and Welfare; 2024. p. 136.

2. National Breast Cancer Foundation (NBCF). Breast Cancer Facts & Stats Dallas TX: NBCF INC; 2024 [Available from: https://www.nationalbreastcancer.org/breast-cancer-facts/.

3. Wong DJ, Gandomkar Z, Lewis S, Reed W, Suleiman M, Siviengphanom S, et al. Do Reader Characteristics Affect Diagnostic Efficacy in Screening Mammography? A Systematic Review. Clinical Breast Cancer. 2023;23(3):e56-e67.

4. Grabler P, Sighoko D, Wang L, Allgood K, Ansell D. Recall and Cancer Detection Rates for Screening Mammography: Finding the Sweet Spot. American Journal of Roentgenology. 2017;208(1):208-13.

5. Rauscher GH, Murphy AM, Qiu Q, Dolecek TA, Tossas K, Liu Y, et al. The "Sweet Spot" Revisited: Optimal Recall Rates for Cancer Detection With 2D and 3D Digital Screening Mammography in the Metro Chicago Breast Cancer Registry. American Journal of Roentgenology. 2021;216(4):894-902.

6. Committee on New Approaches to Early Detection and Diagnosis of Breast Cancer. Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis. In: Joy JE PE, Petitti DB, editor. Improving Breast Cancer Screening Services: Washington (DC): National Academies Press (US); 2005.

7. Ekpo EU, Alakhras M, Brennan P. Errors in Mammography Cannot be Solved Through Technology Alone. Asian Pacific Journal of Cancer Prevention. 2018;19(2):291-301.

8. Wadhwa A, Sullivan JR, Gonyo MB. Missed Breast Cancer: What Can We Learn? Current Problems in Diagnostic Radiology. 2016;45(6):402-19.

9. Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Annals of Internal Medicine. 2011;155(8):481-92.

10. Tan PS, Ali MA, Eriksson M, Hall P, Humphreys K, Czene K. Mammography features for early markers of aggressive breast cancer subtypes and tumor characteristics: A population-based cohort study. International Journal of Cancer. 2021;148(6):1351-9.

11. Sun SS, Zhang B, Zhao HM, Cao XC. Association between mammographic features and clinicopathological characteristics in invasive ductal carcinoma of breast cancer. Molecular and Clinical Oncology. 2014;2(4):623-9.

12. Pulappadi VP, Dhamija E, Baby A, Mathur S, Pandey S, Gogia A, et al. Imaging Features of Breast Cancer Subtypes on Mammography and Ultrasonography: an Analysis of 479 Patients. Indian Journal of Surgical Oncology. 2022;13(4):931-8.

13. Akwo JD, Trieu P, Lewis S. Does the availability of prior mammograms improve radiologists' observer performance?-a scoping review. British Journal of Radiology Open. 2023;5(1):20230038.

14. Callaway MP, Boggis CR, Astley SA, Hutt I. The influence of previous films on screening mammographic interpretation and detection of breast carcinoma. Clinical Radiology. 1997;52(7):527-9.

15. Roelofs AA, Karssemeijer N, Wedekind N, Beck C, van Woudenberg S, Snoeren PR, et al. Importance of comparison of current and prior mammograms in breast cancer screening. Radiology. 2007;242(1):70-7.

16. Roelofs AAJ, Karssemeijer N, Wedekind N, Beck C, Woudenberg Sv, Snoeren PR, et al. Importance of Comparison of Current and Prior Mammograms in Breast Cancer Screening. Radiology. 2007;242(1):70-7.

17. Thurfjell MG, Vitak B, Azavedo E, Svane G, Thurfjell E. Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. Acta Radiolgica. 2000;41(1):52-6.

18. Varela C, Karssemeijer N, Hendriks JH, Holland R. Use of prior mammograms in the classification of benign and malignant masses. European Journal of Radiology. 2005;56(2):248-55.

19. Albus K. Record Keeping: Mammography Reston VA: American College of Radiologists; 2019 [updated (Revised 12/12/2019); cited 2024. Available from:

https://accreditationsupport.acr.org/support/solutions/articles/11000051613-record-keeping-mammography-revised-12-12-19-.

20. Akwo JD, Trieu PDY, Barron ML, Reynolds T, Lewis SJ. Access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance. Clinical Radiology. 2024.

21. Akwo JD, Trieu PD, Barron ML, Reynolds T, Lewis SJ. Does access to prior mammograms improve the performance of radiographers in interpreting screening mammograms? Radiography. 2025;31(1):247-53.

22. Waite S, Grigorian A, Alexander RG, Macknik SL, Carrasco M, Heeger DJ, et al. Analysis of Perceptual Expertise in Radiology - Current Knowledge and a New Perspective. Frontiers in Human Neuroscience. 2019;13:213.

23. Harvey JA. The Evolving Role of Breast Radiologists. Journal of Breast Imaging. 2020;2(1):1.

24. Alexander RG, Waite S, Macknik SL, Martinez-Conde S. What do radiologists look for? Advances and limitations of perceptual learning in radiologic search. Journal of Vision. 2020;20(10):17.

25. Mohd Norsuddin N, Reed W, Mello-Thoms C, Lewis SJ. Understanding recall rates in screening mammography: A conceptual framework review of the literature. Radiography. 2015;21(4):334-41.

26. Akwo J, Hadadi I, Ekpo E. Diagnostic Efficacy of Five Different Imaging Modalities in the Assessment of Women Recalled at Breast Screening-A Systematic Review and Meta-Analysis. Cancers (Basel). 2024;16(20).

27. Pace LE. False-Positive Results of Mammography Screening in the Era of Digital Breast Tomosynthesis. Journal of American Medical Association Network Open. 2022;5(3):e222445-e.

28. Yankaskas BC, Cleveland RJ, Schell MJ, Kozar R. Association of Recall Rates with Sensitivity and Positive Predictive Values of Screening Mammography. American Journal of Roentgenology. 2001;177(3):543-9.

29. Stachs A, Stubert J, Reimer T, Hartmann S. Benign Breast Disease in Women. Deutsches Ärzteblatt International. 2019;116(33-34):565-74.

30. Posso M, Alcántara R, Vázquez I, Comerma L, Baré M, Louro J, et al. Mammographic features of benign breast lesions and risk of subsequent breast cancer in women attending breast cancer screening. European Radiology. 2022;32(1):621-9.

31. Cancer Australia. National Cancer Control Indicators: Recall to assessment. 2024. [Accessed December, 2024]; https://ncci.canceraustralia.gov.au/screening/abnormal-breast-screen-assessment/recall-assessment

32. Callaway MP, Boggis CRM, Astley SA, Hutt I. The influence of previous films on screening mammographic interpretation and detection of breast carcinoma. Clinical Radiology. 1997;52(7):527-9.

33. Frankel SD, Sickles EA, Curpen BN, Sollitto RA, Ominsky SH, Galvin HB. Initial versus subsequent screening mammography: comparison of findings and their prognostic significance. American Journal of Roentgenology. 1995;164(5):1107-9.

34. Hakim CM, Anello MI, Cohen CS, Ganott MA, Lu AH, Perrin RL, et al. Impact of and interaction between the availability of prior examinations and DBT on the interpretation of negative and benign mammograms. Academic Radiology. 2014;21(4):445-9.

35. Hakim CM, Catullo VJ, Chough DM, Ganott MA, Kelly AE, Shinde DD, et al. Effect of the Availability of Prior Full-Field Digital Mammography and Digital Breast Tomosynthesis Images on the Interpretation of Mammograms. Radiology. 2015;276(1):65-72.

36. Soh BP, Lee WB, McEntee MF, Kench PL, Reed WM, Heard R, et al. Mammography test sets: Reading location and prior images do not affect group performance. Clinical Radiology. 2014;69(4):397-402.

37. Sumkin JH, Holbert BL, Herrmann JS, Hakim CA, Ganott MA, Poller WR, et al. Optimal Reference Mammography: A Comparison of Mammograms Obtained 1 and 2 Years Before the Present Examination. American Journal of Roentgenology. 2003;180(2):343-6.

38. Gallegos LL, Gilchrist A, Spain L, Stanislaw S, Hill SM, Primus V, et al. A protocol for representative sampling of solid tumors to improve the accuracy of sequencing results. Structured Transparent Accessible Reproducible Protocols. 2021;2(3):100624.

39. Hedenström P, Marschall HU, Nilsson B, Demir A, Lindkvist B, Nilsson O, et al. High clinical impact and diagnostic accuracy of EUS-guided biopsy sampling of subepithelial lesions: a prospective, comparative study. Surgical Endoscopy. 2018;32(3):1304-13.

40. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. New England Journal of Medicine. 2007;356(3):227-36.

41. Bell RJ. Mammographic density and breast cancer screening. Climacteric. 2020;23(5):460-5.

42. Lehman CD, White E, Peacock S, Drucker MJ, Urban N. Effect of age and breast density on screening mammograms with false-positive findings. American Journal of Roentgenology. 1999;173(6):1651-5.

43. Nelson HD, O'Meara ES, Kerlikowske K, Balch S, Miglioretti D. Factors Associated With Rates of False-Positive and False-Negative Results From Digital Mammography Screening: An Analysis of Registry Data. Annals of Internal Medicine. 2016;164(4):226-35.

44. Trieu PDY, Borecky N, Li T, Brennan PC, Barron ML, Lewis SJ. The Impact of Prior Mammograms on the Diagnostic Performance of Radiologists in Early Breast Cancer Detection: A Focus on Breast Density, Lesion Features and Vendors Using Wholly Digital Screening Cases. Cancers (Basel). 2023;15(4).

45. Suleiman WI, Lewis SJ, Georgian-Smith D, Evanoff MG, McEntee MF. Number of mammography cases read per year is a strong predictor of sensitivity. Journal of Medical Imaging (Bellingham). 2014;1(1):015503.

46. Jackson SL, Abraham L, Miglioretti DL, Buist DS, Kerlikowske K, Onega T, et al. Patient and Radiologist Characteristics Associated With Accuracy of Two Types of Diagnostic Mammograms. American Journal of Roentgenology. 2015;205(2):456-63.

47. Suleiman WI, McEntee MF, Lewis SJ, Rawashdeh MA, Georgian-Smith D, Heard R, et al. In the digital era, architectural distortion remains a challenging radiological task. Clinical Radiology. 2016;71(1):e35-40.

48. Suleiman WI, Rawashdeh MA, Lewis SJ, McEntee MF, Lee W, Tapia K, et al. Impact of Breast Reader Assessment Strategy on mammographic radiologists' test reading performance. Journal of Medical Imaging and Radiation Oncology. 2016;60(3):352-8.

49. American College of Radiology. Interpreting Physician: Mammography (Revised 05-22-2024) Reston, VA 20191: American College of Radiologists; 2024 [updated 22/05/2024; accessed December, 2024];https://accreditationsupport.acr.org/support/solutions/articles/11000049778-interpreting-physician-mammography-revised-05-22-2024-

50. BreastScreen Australia National Accreditation Standards. Canberra: BreastScreen; 2022.

51. Rawson JV, Smetherman D, Rubin E. Short-Term Strategies for Augmenting the National Radiologist Workforce. American Journal of Roentgenology. 2024;222(6):e2430920.

52. Jeganathan S. The Growing Problem of Radiologist Shortages: Australia and New Zealand's Perspective. Korean Journal of Radiology. 2023;24(11):1043-5.

53. Royal Australian and New Zealand College of Radiology. 2020 Clinical Radiology Workforce Census Report: Australia. 2022. [Accessed 21/10/2023]; https://www.ranzcr.com/whats-on/news-media/ranzcr-releases-2020-australian-clinical-radiology-workforce-census-1

54. Royal College of Radiology. The Breast Imaging and Diagnostic Workforce in the United Kingdom: results of survey of NHS breast screening programme units and radiology departments. 2020.[Accessed, January, 2025]; https://www.ewin.nhs.uk/sites/default/files/bfcr162_bsbr_survey.pdf

55. Chen Y, James JJ, Michalopoulou E, Darker IT, Jenkins J. Performance of Radiologists and Radiographers in Double Reading Mammograms: The UK National Health Service Breast Screening Program. Radiology. 2023;306(1):102-9.

56. Pauli R, Hammond S, Cooke J, Ansell J. Comparison of Radiographer/Radiologist Double Film Reading with Single Reading in Breast Cancer Screening. Journal of Medical Screening. 1996;3(1):18-22.

57. Debono JC, Poulos AE, Houssami N, Boyages J. Evaluation of radiographers' mammography screen-reading accuracy in Australia. Journal of Medical Radiation Sciences. 2015;62(1):15-22.

58. Holt JJ, Pollard K. Radiographers' ability to perceive and classify abnormalities on mammographic Images -a Pilot Project. Journal of Medical Radiation Sciences. 2010;57(2):8-14.

59. Moran S, Warren-Forward H. The diagnostic accuracy of radiographers assessing screening mammograms: A systematic review. Radiography. 2016;22(2):137-46.

60. Moran S, Warren-Forward H. Can Australian radiographers assess screening mammograms accurately? Biennial follow-up from a four year prospective study and lesion analysis. Radiography. 2016;22(3):e201-e6.

61. Torres-Mejía G, Smith RA, Carranza-Flores Mde L, Bogart A, Martínez-Matsushita L, Miglioretti DL, et al. Radiographers supporting radiologists in the interpretation of screening mammography: a viable strategy to meet the shortage in the number of radiologists. BioMed Central Cancer. 2015;15:410.

62. van den Biggelaar FJ NP, Flobbe K. . Performance of radiographers in mammogram interpretation: a systematic review. Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews [Internet]. Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews [Internet]: Centre for Reviews and Dissemination (UK); 2008.

63. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Annals of Internal Medicine. 2007;146(7):502-10.

64. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. Annals of Family Medicine. 2013;11(2):106-15.

65. Soon J, Houssami N, Clemson M, Lockie D, Farber R, Barratt A, et al. The financial implications of investigating false-positive and true-positive mammograms in a national breast cancer screening program. Australian Health Review. 2023;47(2):159-64.

66. Ong MS, Mandl KD. National expenditure for false-positive mammograms and breast cancer overdiagnoses estimated at $4 billion a year. Health Aff (Millwood). 2015;34(4):576-83.

67. Dabbous FM, Dolecek TA, Berbaum ML, Friedewald SM, Summerfelt WT, Hoskins K, et al. Impact of a False-Positive Screening Mammogram on Subsequent Screening Behavior and Stage at Breast Cancer Diagnosis. Cancer Epidemiology Biomarkers and Prevention. 2017;26(3):397-403.

68. Mao X, He W, Humphreys K, Eriksson M, Holowko N, Yang H, et al. Breast Cancer Incidence After a False-Positive Mammography Result. Journal of American Medical Association Oncology. 2024;10(1):63-70.

69. Chubak J, Boudreau DM, Fishman PA, Elmore JG. Cost of breast-related care in the year following false positive screening mammograms. Medical Care. 2010;48(9):815-20.

70. Tan A, Freeman DH, Jr., Goodwin JS, Freeman JL. Variation in false-positive rates of mammography reading among 1067 radiologists: a population-based assessment. Breast Cancer Research and Treatment. 2006;100(3):309-18.

71. Elmore JG, Miglioretti DL, Reisch LM, Barton MB, Kreuter W, Christiansen CL, et al. Screening mammograms by community radiologists: variability in false-positive rates. Journal of the National Cancer Institute. 2002;94(18):1373-80.

72. Moran S, Warren-Forward H. Development of a training package to increase the performance of radiographers in assessing screening mammograms. Journal of Medical Radiation Sciences. 2011;58(2):10-3.

73. Liu S, Wu XD, Xu WJ, Lin Q, Liu XJ, Li Y. Is There a Correlation between the Presence of a Spiculated Mass on Mammogram and Luminal A Subtype Breast Cancer? Korean Journal of Radiology. 2016;17(6):846-52.

74. Sturesdotter L, Sandsveden M, Johnson K, Larsson A-M, Zackrisson S, Sartor H. Mammographic tumour appearance is related to clinicopathological factors and surrogate molecular breast cancer subtype. Scientific Reports. 2020;10(1):20814.

75. Chang C-Y, Kuo S-J, Wu H-K, Huang Y-L, Chen D-R. Stellate Masses and Histologic Grades in Breast Cancer. Ultrasound in Medicine & Biology. 2014;40(5):904-16.

76. Spak DA, Plaxco JS, Santiago L, Dryden MJ, Dogan BE. BI-RADS(®) fifth edition: A summary of changes. Diagnostic and Interventional Imaging. 2017;98(3):179-90.

77. RANZCR. Breast Imaging Grading Comparison and Lesion Classification lists: The Royal Australian and New Zealand College of Radiologists; 2018 [Available from: Breast%20Grading%20Comparison%20and%20Lesion%20Classification%20FINAL%pdf.

78. Bahl M, Baker JA, Kinsey EN, Ghate SV. Architectural Distortion on Mammography: Correlation With Pathologic Outcomes and Predictors of Malignancy. American Journal of Roentgenology. 2015;205(6):1339-45.

79. Farshid G, Walker A, Battersby G, Sullivan T, Gill PG, Pieterse S, et al. Predictors of malignancy in screen-detected breast masses with indeterminate/equivocal (grade 3) imaging features. The Breast. 2011;20(1):56-61.

80. Mohd Norsuddin N, Mello-Thoms C, Reed W, Rickard M, Lewis S. An investigation into the mammographic appearances of missed breast cancers when recall rates are reduced. British Journal of Radiology. 2017;90(1076):20170048.

81. Moriuchi H, Yamaguchi J, Hayashi H, Ohtani H, Shimokawa I, Abiru H, et al. Cancer Cell Interaction with Adipose Tissue: Correlation with the Finding of Spiculation at Mammography. Radiology. 2016;279(1):56-64.

82. Sartor H, Sturesdotter L, Larsson A-M, Rosendahl AH, Zackrisson S. Mammographic features differ with body composition in women with breast cancer. European Radiology. 2025;35(1):151-9.

83. Cherel P, Becette V, Hagay C. Stellate images: anatomic and radiologic correlations. European Journal of Radiology. 2005;54(1):37-54.

84. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. Radiology. 2006;239(2):385-91.

85. Lee AY, Wisner DJ, Aminololama-Shakeri S, Arasu VA, Feig SA, Hargreaves J, et al. Inter-reader Variability in the Use of BI-RADS Descriptors for Suspicious Findings on Diagnostic Mammography: A Multi-institution Study of 10 Academic Radiologists. Academic Radiology. 2017;24(1):60-6.

86. Kim WH, Chang JM, Koo HR, Seo M, Bae MS, Lee J, et al. Impact of prior mammograms on combined reading of digital mammography and digital breast tomosynthesis. Acta Radiologica. 2017;58(2):148-55.

87. Hadadi I, Clarke J, Rae W, McEntee M, Vincent W, Ekpo E. Diagnostic Efficacy across Dense and Non-Dense Breasts during Digital Breast Tomosynthesis and Ultrasound Assessment for Recalled Women. Diagnostics (Basel). 2022;12(6).

88. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, et al. The "laboratory" effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. Radiology. 2008;249(1):47-53.

89. Popescu L. Value and Limitations of Observer Models. 2018. p. 300-4.

90. Gur D, Rockette HE, Maitz GS, King JL, Klym AH, Bandos AI. Variability in Observer Performance Studies: Experimental Observations1. Academic Radiology. 2005;12(12):1527-33.

91. Haygood TM, Smith S, Sun J. Memory bias in observer-performance literature. Journal of Medical Imaging. 2018;5(3):031412.

92. Obuchowski NA. How Many Observers Are Needed in Clinical Studies of Medical Imaging? American Journal of Roentgenology. 2004;182(4):867-9.

93. Alakhras M, Al-Mousa DS, Alqadi AK, Sabaneh HA, Karasneh RM, Spuur KM. The influence of breast density and key demographics of radiographers on mammography reporting performance – a pilot study. Journal of Medical Radiation Sciences. 2022;69(1):30-6.

94. Moran S, Warren-Forward H. A retrospective study of the performance of radiographers in interpreting screening mammograms. Radiography. 2011;17(2):126-31.

95. Evans KK, Marom EM, Godoy MC, Palacio D, Sagebiel T, Cuellar SB, et al. Radiologists remember mountains better than radiographs, or do they? Journal of Medical Imaging (Bellingham). 2016;3(1):011005.

96. Rust NC, Mehrpour V. Understanding Image Memorability. Trends Cognition Sciences. 2020;24(7):557-68.

97. Qenam BA, Li T, Alshabibi A, Frazer H, Ekpo E, Brennan P. Test-set results can predict participants' development in breast-screen cancer detection: An observational cohort study. Health Science Reports. 2024;7(6):e2161.

98. Chen Y, James JJ, Cornford EJ, Jenkins J. The Relationship between Mammography Readers' Real-Life Performance and Performance in a Test Set-based Assessment Scheme in a National Breast Screening Program. Radiology: Imaging Cancer. 2020;2(5):e200016.

99. Remedios D, France B, Alexander M. Making the best value of clinical radiology: iRefer Guidelines, 8th edition. Clinical Radiology. 2017;72(9):705-7.

100. European Commission: Directorate-General for E, Chateil J, Cavanagh P, Ashford N, Remedios D, Bezzi M, et al. Referral guidelines for medical imaging – Availability and use in the European Union: Publications Office; 2014.

101. Wang X, Ji X. Sample Size Estimation in Clinical Research: From Randomized Controlled Trials to Observational Studies. Chest. 2020;158(1s):S12-s20.

102. Heleno B, Siersma VD, Brodersen J. Diagnostic invasiveness and psychosocial consequences of false-positive mammography. Annals of Family Medicine. 2015;13(3):242-9.

103. Gram EG, Siersma V, Brodersen JB. Long-term psychosocial consequences of false-positive screening mammography: a cohort study with follow-up of 12-14 years in Denmark. British Medical Journal Open. 2023;13(4):e072188.

104. Setz-Pels W, Duijm LE, Coebergh JW, Rutten M, Nederend J, Voogd AC. Re-attendance after false-positive screening mammography: a population-based study in the Netherlands. British Journal of Cancer. 2013;109(8):2044-50.

105. Squillace L, Pizzi L, Rallo F, Bazzani C, Saguatti G, Mezzetti F. Subsequent attendance in a breast cancer screening program after a false-positive result in the Local Health Authority of Bologna (Italy). Scientific Reports. 2021;11(1):8530.

106. Voogd AC, Molnar Z, Nederend J, Schipper R-J, Strobbe LJA, Duijm LEM. Predictors of re-attendance at biennial screening mammography following a false positive referral: A study among women in the south of the Netherlands. The Breast. 2024;74:103702.

107. Fitzpatrick P, Fleming P, O'Neill S, Kiernan D, Mooney T. False-positive mammographic screening: factors influencing re-attendance over a decade of screening. Journal of Medical Screening. 2011;18(1):30-3.

Appendix 1: Human research ethics committee approval

THE UNIVERSITY OF SYDNEY

Research Integrity & Ethics Administration
HUMAN RESEARCH ETHICS COMMITTEE

Wednesday, 22 March 2023

Prof Sarah Lewis
Clinical Imaging; Faculty of Medicine and Health
Email: sarah.lewis@sydney.edu.au

Dear Sarah,

The University of Sydney Human Research Ethics Committee (HREC) has considered your application.

I am pleased to inform you that after consideration of your response, your project has been approved.

Details of the approval are as follows:

Project No.:	**2023/101**
Project Title:	**The influence of prior cases on radiology performance with screening mammograms**
Authorised Personnel:	**Lewis Sarah; Akwo Judith; Trieu Phuong Dung (Yun);**
Approval Period:	**22/03/2023** to **22/03/2027**
First Annual Report Due:	**22/03/2024**

Documents Approved:

Date Uploaded	Version Number	Document Name
16/03/2023		Email invitation
16/03/2023		PIS

The Committee approved the revised application in the absence of ethical objections and on the basis of satisfactory scientific merit.

Please note for future submissions: Researchers need to answer all questions in full on each application so that the reviewing committee has the information they need to independently assess each submission. Submissions need to stand alone in completeness for more efficient review. Where the application links to a previous submission to the HREC, researchers should clearly provide this information in their application (including protocol number) in order to expedite review.

Condition/s of Approval

- Research must be conducted according to the approved proposal.
- An annual progress report must be submitted to the Ethics Office on or before the anniversary of approval and on completion of the project.
- You must report as soon as practicable anything that might warrant review of ethical approval of the project including:
  - Serious or unexpected adverse events (which should be reported within 72 hours).
  - Unforeseen events that might affect continued ethical acceptability of the project.
- Any changes to the proposal must be approved prior to their implementation (except where an amendment is undertaken to eliminate *immediate* risk to participants).
- Personnel working on this project must be sufficiently qualified by education, training and experience for their role, or adequately supervised. Changes to personnel must be reported and approved.

Research Integrity & Ethics Administration
Research Portfolio
Level 3, F23 Administration Building
The University of Sydney
NSW 2006 Australia

T +61 2 9036 9161
E human.ethics@sydney.edu.au
W sydney.edu.au/ethics

ABN 15 211 513 464
CRICOS 00026A

- Personnel must disclose any actual or potential conflicts of interest, including any financial or other interest or affiliation, as relevant to this project.
- Data and primary materials must be retained and stored in accordance with the relevant legislation and University guidelines.
- Ethics approval is dependent upon ongoing compliance of the research with the *National Statement on Ethical Conduct in Human Research*, the *Australian Code for the Responsible Conduct of Research*, applicable legal requirements, and with University policies, procedures and governance requirements.
- The Ethics Office may conduct audits on approved projects.
- The Chief Investigator has ultimate responsibility for the conduct of the research and is responsible for ensuring all others involved will conduct the research in accordance with the above.

This letter constitutes ethical approval only.

Please contact the Ethics Office should you require further information or clarification.

Sincerely,

Acting Chair, Health Review Committee (Low Risk)

The University of Sydney of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2018) and the NHMRC's Australian Code for the Responsible Conduct of Research (2018)

Appendix II: Recruitment advertisement for radiologists

Research participants needed!

Help the University of Sydney researchers investigate the influence of prior mammograms on diagnostic performance. We are asking for your help because you are currently a radiologist or breast physician reporting mammographic images. The goal of the project is to investigate whether the availability of prior screening mammograms improve radiologists' ability to interpret screening mammograms.

For this study, you will be asked to complete a questionnaire about your demographics and interpret 75 mammograms with and without prior mammograms. There will be two reading sessions which will be 3-4 months apart. The first reading session will require you to interpret 75 images with looking at the prior mammograms of these women. The second reading session will require you to interpret another set of 75 mammograms and you will be given access to the prior mammograms of these women to support the interpretation process. We anticipate that each of the reading session will take about 2 hours and 30 minutes of your time. Participation in the study is completely voluntary. The results of your participation will be anonymised and used for research only. Data protection and confidentiality will be ensured by storing the data collected in the University of Sydney Research Data Store. We may also submit the information from this project to a public database for research information, so that other researchers can access it and use it in their projects, but no information will be linked to you."

To get signed up, please register your interest by contacting the Chief Investigator, **Prof. Sarah Lewis** (**sarah.lewis@sydney.edu.au**) or PhD student, **Judith Akwo** (jakw2899@uni.sydney.edu.au)

We look forward to working with you.

Sincerely

Sarah Lewis

Email: sarah.lewis@sydney.edu.au

Appendix III: Recruitment advertisement for radiographers

Research participants needed!

Help the University of Sydney researchers investigate the influence of prior mammograms on diagnostic performance. We are asking for your help because you are currently a BreastScreen radiographer or junior breast radiologist. The goal of the project is to investigate whether the availability of prior screening mammograms improve radiographers and junior breast radiologists' ability to interpret screening mammograms.

For this study, you will be asked to complete a questionnaire about your demographics and interpret two sets of mammograms with and without prior mammograms. There will be two reading sessions which will be 3-4 months apart and you will be required to interpret 32 images at each reading session. We anticipate that each of the reading session will take about 2 hours of your time. Participation in the study is completely voluntary. The results of your participation will be anonymised and used for research only. Data protection and confidentiality will be ensured by storing the data collected in the University of Sydney Research Data Store. We may also submit the information from this project to a public database for research information, so that other researchers can access it and use it in their projects, but no information will be linked to you."

To get signed up, please register your interest by contacting the Chief Investigator, **Prof. Sarah Lewis** (**sarah.lewis@sydney.edu.au**) or PhD student, **Judith Akwo** (jakw2899@uni.sydney.edu.au)

We look forward to working with you.

Sincerely

Sarah Lewis

Email: sarah.lewis@sydney.edu.au

Appendix IV: Participant information statement for radiologists

THE UNIVERSITY OF SYDNEY

Discipline of Medical Imaging Science
School of Health Sciences
Faculty of Medicine and Health

Prof. Sarah Lewis
Associate Dean Research Performance

Susan Wakil Health Building (D18)
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: M:
Facsimile:
Email: sarah.lewis@sydney.edu.au
Web: http://www.sydney.edu.au/

Influence of prior cases on radiologists' performance with screening mammograms

PARTICIPANT INFORMATION STATEMENT

(1) What is this study about?

You are being invited to take part in a research project that will investigate the influence of prior mammograms on diagnostic performance. The purpose of this study is to explore the impact of access to prior mammograms on the accurate interpretation by BreastScreen readers, and to identify the characteristics of cases and readers that will benefit from accessing previous mammograms. You have been asked to participate because you are a radiologist or breast physician reporting mammographic images and have previously read the same test-set containing prior mammograms. Please read the following information carefully, and feel free to ask any question(s) about anything you do not understand or anything you would like to know more about.

Participation in this research study is voluntary.

By giving your consent to take part in this study you are telling us that you:
- ✓ Understand what you have read.
- ✓ Agree to take part in the research study as outlined below.
- ✓ Agree to the use of your personal information as described.

You will be given a copy of this Participant Information Statement to keep.

(2) Who is running the study?

The study is being carried out by the following researchers:
- Prof. Sarah Lewis, Associate Dean, Research Performance, The University of Sydney
- Dr. Phuong D (Yun) Trieu, Research Fellow, The University of Sydney
- Ms Judith Akwo, PhD Student, The University of Sydney.

STUDENT DECLARATION
Ms Akwo is conducting this study as the basis for the degree of Doctor of Philosophy at The University of Sydney. This will take place under the supervision of Professor Sarah Lewis.

(3) What will the study involve for me?

As a participant in this study, you will be asked to interpret 30 digital breast tomosynthesis (DBT) images. You have previously read the same test-set containing prior mammograms using the BreastScreen Reader Assessment Strategy (BREAST) platform. For this study, you will be asked to

interpret the same set of DBT images without being able to access the prior mammograms of these cases. .

We will also ask you to provide your demographic and practice-related information to help us understand the characteristics of readers who benefit from looking at prior mammograms when reading DBT images from the current screening round (this is done through creating an account with BREAST). You are welcome to access the BREAST platform from your workplace during or outside your working hours in accordance with your department protocols. Alternatively, you are welcome to come to the Susan Wakil Health Building at the University of Sydney to use our facilities. After creating a BREAST login profile, or reactivating your existing profile, you will be asked to give consent to collect your responses to the cases and update your demographics. This consent option will occur prior to you providing any data points.

(4) How much of my time will the study take?

We anticipate that the reading will take approximately 1 hour of your time.

(5) Who can take part in the study?

Only radiologists and breast physicians who interpret screening mammograms can participate in the study. This is to ensure that the results of the study are considered representative of the professional community when considering the impact of the study results.

(6) Do I have to be in the study? Can I withdraw from the study once I've started?

Being in this study is completely voluntary. Your decision whether to participate will not affect your current or future relationship with the researchers, BREAST or anyone else at the University of Sydney. If you decide to take part in the study and then change your mind later, you are free to withdraw at any time up until the results are aggregated and published. After this time, we will not be able to remove your data points. You can withdraw from the study by contacting the Chief Investigator, Prof. Sarah Lewis or PhD student, Judith Akwo. There are no consequences for withdrawing from the study.

(7) Are there any risks or costs associated with being in the study?

Aside from giving up your time, we do not expect that there will be any risks or costs associated with taking part in this study.

(8) Are there any benefits associated with being in the study?

We cannot guarantee that you will receive any direct benefits from being in the study. You are able to claim CPD points from RANZCR for the BREAST research activities and your certificate will be added to your BREAST profile page once you have completed the study.

(9) What will happen to information about me that is collected during the study?

For this study, we will collect demographic information such as age, gender, and practice-related information such as years since qualification, number of images read per week, and completion of a breast fellowship as well as your performance data in reading the cases (such as Sensitivity, Specificity etc). You are able to see your own performance for the set of images once you complete all cases. . By providing your consent, you are agreeing to us collecting personal information about you for the purposes of this research study. The BREAST platform has a clear consent option for you to approve, indicating that you are comfortable for us to collect your data points. Your information will only be used for the purposes outlined in this Participant Information Statement, unless you consent otherwise.

Your information will be stored securely, and your identity/information will be kept strictly confidential, except as required by law. Study findings may be published, but you will not be individually identifiable in these publications.

.

(10) Can I tell other people about the study?

Yes, you are welcome to tell other people about the study.

(11) What if I would like further information about the study?

When you have read this information and would like to know more about the study, please feel free to contact **Prof Sarah Lewis (sarah.lewis@sydney.edu.au)** or **Judith Akwo (jakw2899@uni.sydney.edu.au)**

(12) Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. This feedback will be in the form of a lay summary of the findings sent to your nominated email account or through the BREAST newsletter.

(13) What if I have a complaint or any concerns about the study?

Research involving humans in Australia is reviewed by an independent group of people called a Human Research Ethics Committee (HREC). The ethical aspects of this study have been approved by the HREC of the University of Sydney. As part of this process, we have agreed to carry out the study according to the *National Statement on Ethical Conduct in Human Research (2007)*. This statement has been developed to protect people who agree to take part in research studies.

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the university using the details outlined below. Please quote the study title and protocol number.

The Manager, Ethics Administration, University of Sydney:

- **Telephone:** +61 2 8627 8176
- **Email:** human.ethics@sydney.edu.au
- **Fax:** +61 2 8627 8177 (Facsimile)

☐ I confirm that I have read and understood the information in this Participant Information Statement
☐ By completing this study, I give consent for my responses to be used in the research as detailed in the Participant Information Statement.

This information sheet is for you to keep

Appendix V: Participant information statement for radiographers

THE UNIVERSITY OF SYDNEY

Discipline of Medical Imaging Science
School of Health Sciences
Faculty of Medicine and Health

Prof. Sarah Lewis
Associate Dean Research Performance

Susan Wakil Health Building (D18)
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: M:
Facsimile:
Email: sarah.lewis@sydney.edu.au
Web: http://www.sydney.edu.au/

Influence of prior cases on radiologists' performance with screening mammograms

PARTICIPANT INFORMATION STATEMENT

(1) What is this study about?

You are being invited to take part in a research project that will investigate the influence of prior mammograms on diagnostic performance. The purpose of this study is to explore the impact of access to prior mammograms on the accurate interpretation by junior radiologists/ BreastScreen readers, and to identify the characteristics of cases and readers that will benefit from accessing previous mammograms. You have been asked to participate because you are a radiologist, breast physician, radiology trainee reporting mammographic images, or radiographer. Please read the following information carefully, and feel free to ask any question(s) about anything you do not understand or anything you would like to know more about.

Participation in this research study is voluntary.

By giving your consent to take part in this study you are telling us that you:
- ✓ Understand what you have read.
- ✓ Agree to take part in the research study as outlined below.
- ✓ Agree to the use of your personal information as described.

You will be given a copy of this Participant Information Statement to keep.

(2) Who is running the study?

The study is being carried out by the following researchers:
- Prof. Sarah Lewis, Associate Dean, Research Performance, The University of Sydney
- Dr Tess Reynolds, Deputy Director, Image X Institute, The University of Sydney
- Dr. Yun Trieu, Senior Lecturer, The University of Sydney
- Ms Judith Akwo, PhD Student, The University of Sydney.

STUDENT DECLARATION
Ms Akwo is conducting this study as the basis for the degree of Doctor of Philosophy at The University of Sydney. This will take place under the supervision of Professor Sarah Lewis.

(3) What will the study involve for me?

As a participant in this study, you will be asked to interpret 28 mammograms using the BreastScreen Reader Assessment Strategy (BREAST) platform. The study will require you to interpret 28 mammograms without being able to access the prior mammograms of these cases.

We will also ask you to provide your demographic and practice-related information to help us understand the characteristics of readers who benefit from looking at prior mammograms when reading mammograms from the current screening round (this is done through creating an account with BREAST). You are welcome to access the BREAST platform from your workplace during or outside your working hours in accordance with your department protocols. Alternatively, you are welcome to come to the Susan Wakil Health Building at the University of Sydney to use our facilities. After creating a BREAST login profile, or reactivating your existing profile, you will be asked to give consent to collect your responses to the cases and update your demographics. This consent option will occur prior to you providing any data points.

(4) How much of my time will the study take?

We anticipate that the reading session will take approximately 1 hour of your time.

(5) Who can take part in the study?

Only junior radiologists, breast physicians who interpret screening mammograms, or radiographers can participate in the study. This is to ensure that the results of the study are representative of the professional community when considering the impact of the study results.

(6) Do I have to be in the study? Can I withdraw from the study once I've started?

Being in this study is completely voluntary. Your decision whether to participate will not affect your current or future relationship with the researchers, BREAST or anyone else at the University of Sydney. If you decide to take part in the study and then change your mind later, you are free to withdraw at any time up until the results are aggregated and published. After this time, we will not be able to remove your data points. You can withdraw from the study by contacting the Chief Investigator, Prof. Sarah Lewis or PhD student, Judith Akwo. There are no consequences for withdrawing from the study.

(7) Are there any risks or costs associated with being in the study?

Aside from giving up your time, we do not expect that there will be any risks or costs associated with taking part in this study.

(8) Are there any benefits associated with being in the study?

We cannot guarantee that you will receive any direct benefits from being in the study. You are able to claim CPD points from RANZCR for the BREAST research activities and your certificate will be added to your BREAST profile page once you have completed the study.

(9) What will happen to information about me that is collected during the study?

For this study, we will collect demographic information such as age, gender, and practice-related information such as years since qualification, number of images read per week, and completion of a breast fellowship as well as your performance data in reading the cases (such as Sensitivity, Specificity etc). You are able to see your own performance for the set of images once you complete all 28 cases. By providing your consent, you are agreeing to us collecting personal information about you for the purposes of this research study. The BREAST platform has a clear consent option for you to approve, indicating that you are comfortable for us to collect your data points. Your information will only be

used for the purposes outlined in this Participant Information Statement, unless you consent otherwise.

Your information will be stored securely, and your identity/information will be kept strictly confidential, except as required by law. Study findings may be published, but you will not be individually identifiable in these publications.

.

(10) Can I tell other people about the study?

Yes, you are welcome to tell other people about the study.

(11) What if I would like further information about the study?

When you have read this information and would like to know more about the study, please feel free to contact **Prof Sarah Lewis (sarah.lewis@sydney.edu.au)** or **Judith Akwo (jakw2899@uni.sydney.edu.au)**

(12) Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. This feedback will be in the form of a lay summary of the findings sent to your nominated email account or through the BREAST newsletter.

(13) What if I have a complaint or any concerns about the study?

Research involving humans in Australia is reviewed by an independent group of people called a Human Research Ethics Committee (HREC). The ethical aspects of this study have been approved by the HREC of the University of Sydney. As part of this process, we have agreed to carry out the study according to the *National Statement on Ethical Conduct in Human Research (2007)*. This statement has been developed to protect people who agree to take part in research studies.

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the university using the details outlined below. Please quote the study title and protocol number.

The Manager, Ethics Administration, University of Sydney:

- **Telephone:** +61 2 8627 8176
- **Email:** human.ethics@sydney.edu.au
- **Fax:** +61 2 8627 8177 (Facsimile)

☐ I confirm that I have read and understood the information in this Participant Information Statement
☐ By completing this study, I give consent for my responses to be used in the research as detailed in the Participant Information Statement.

This information sheet is for you to keep

Appendix VI: Participant consent form

THE UNIVERSITY OF SYDNEY

PARTICIPANT CONSENT FORM

I, ...[PRINT NAME], give consent to my participation in the research project:

Influence of prior cases on radiologists' performance with screening mammograms

In giving my consent I acknowledge that:

1. The procedures required for the project and the time involved have been explained to me, including any inconvenience and any questions I have about the project have been answered to my satisfaction.

2. I have read the Participant Information Statement and have been given the opportunity to discuss the information and my involvement in the project with the researcher/s.

3. I understand that being in this study is completely voluntary – I am not under any obligation to consent.

4. I understand that my involvement is strictly confidential. I understand that any research data gathered from the results of the study may be published, presented in conferences, or disseminated in a student thesis; however, no information about me will be used in any way that is identifiable.

5. I understand that I can withdraw from the study at any time before and during data collection, without affecting my relationship with the researcher(s) or the University of Sydney now or in the future.

6. I understand that I can stop the study at any time if I do not wish to continue, and the information provided will be erased and not be included in the study.

7. I consent to participate in this study:

.....
Signature

.....
Please PRINT name

.....
Date