THE EFFECT OF IMAGE QUALITY ON RECALL RATES
IN A BREASTSCREENING PROGRAM

JENNIFER D. THOMPSON

Thesis submitted for the degree of
Masters of Research (Medical Radiation Science)

The University of Newcastle, Australia

February 2009
DECLARATION

The thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

Signed: ____________________________
ACKNOWLEDGMENTS

Firstly, I wish to express my sincere gratitude to my principal supervisor, Associate Professor Helen Warren-Forward for encouraging me to undertake this project and instructing me in the art of scientific investigation. I thank her for her guidance, tolerance, constructive criticism and humour. She has been instrumental in helping me achieve my biggest learning curve and I cannot thank her enough.

I would also like to thank my co-supervisor Dr. Jinnan Gao for his statistical and academic assistance and encouragement during my study. His input and understanding were invaluable.

My sincere thanks go to the staff (both former and current) and clients at Hunter Region and Wyong Shire BreastScreen – to the counsellors, data collection staff, doctors, film handlers, management, radiographers, receptionists and recruitment. I would like to specifically thank Dr. Michelle Bennett, Rhonda Brown, Jenny Forbes, Professor John Forbes, Dr. Fran Jones, Dr. Jim Lai, David MacMillan, Sheila Moran, Val Richardson, Emma Sessions, Peter Troke, Julie Wilson and Jennifer Woodhead.

I am deeply indebted to Melissa Murphy for her constructive comments during proof reading of this thesis.

To my family and friends who have offered me support, understanding, patience and love, I am honoured by your presence in my life. I send you all my deepest appreciation.

Finally, my friends Andy Eklund and Wendy Moroney deserve a special mention. Without their support, friendship, humour and reality checks, my life and this thesis would just be a dream.
“The point at which the learner becomes satisfied with or indifferent to his attained level of competence defines the ultimate level of ability for that person. An attitude towards work that includes continual self-criticism, progressive problem solving and continual re-investment in improvement is the description of a profession, where expertise, accountability, autonomy and authority are interrelated.” (1)
# TABLE OF CONTENTS

| LIST OF FIGURES | IX |
| LIST OF TABLES | XI |
| LIST OF TABLES | XI |
| ABBREVIATIONS | XII |
| ABSTRACT | XIV |

## 1 CHAPTER ONE: INTRODUCTION

1.1 OVERVIEW

1.2 THESIS OUTLINE

1.3 STUDY BACKGROUND

1.3.1 Breast Cancer

1.3.2 Mammography

1.3.3 Image Quality

1.3.4 Rationale For Study

1.4 RESEARCH AIM

1.5 RESEARCH QUESTIONS

1.6 RESEARCH OBJECTIVES

1.7 SCOPE / LIMITATIONS OF RESEARCH

1.8 SIGNIFICANCE OF THE STUDY

## 2 CHAPTER TWO: LITERATURE REVIEW

2.1 INTRODUCTION

2.2 BREAST ANATOMY

2.2.1 Breast Anatomy

2.2.2 Breast Density

2.2.3 Breast Regions

2.3 BREAST CANCER

2.3.1 Breast Cancer Definition

2.3.2 Incidence / Mortality of Breast Cancer

2.3.3 Breast Cancer Staging

2.3.4 Breast Cancer Risk Factors

2.4 BREAST CANCER DETECTION

2.4.1 Clinical Examination and Self-Examination

2.4.2 Imaging Modalities
2.5 SCREENING MAMMOGRAPHY ................................................................. 31
  2.5.1 History ............................................................................................ 31
  2.5.2 Advantages .................................................................................... 32
  2.5.3 Disadvantages ............................................................................... 32
  2.5.4 Accuracy ....................................................................................... 35
  2.5.5 Mammography Screening Guidelines ........................................... 38
2.6 FILM / SCREEN MAMMOGRAPHY ......................................................... 39
  2.6.1 History And Development ............................................................... 39
  2.6.2 Obtaining the Mammographic Image .............................................. 45
  2.6.3 Breast Anatomy Appearance ....................................................... 49
  2.6.4 Technical Assessment ................................................................... 50
  2.6.5 Screen Reader Review .................................................................. 67
  2.6.6 Reasons for Recall ........................................................................ 72
2.7 SUMMARY ............................................................................................ 79
3 CHAPTER THREE: METHODS ...................................................................... 80
  3.1 INTRODUCTION .................................................................................. 81
  3.2 ETHICAL ISSUES .............................................................................. 82
  3.3 RESOURCES AND MATERIALS .......................................................... 82
    3.3.1 Expert Panel ................................................................................ 82
    3.3.2 Sets of Images .......................................................................... 82
  3.4 PILOT STUDY (PHASE A) ................................................................. 84
    3.4.1 Selection of sets of images ....................................................... 84
    3.4.2 Retrieval, Data Collection and display of sets of images .......... 86
    3.4.3 Assessment of images ............................................................. 88
    3.4.4 Consensus Agreement ............................................................. 89
    3.4.5 Analysis of results ................................................................. 90
  3.5 MAIN COMPONENT (PHASE B) ....................................................... 91
    3.5.1 Selection of sets of images ....................................................... 92
    3.5.2 Retrieval, Data Collection and display of sets of images .......... 92
    3.5.3 Assessment of images ............................................................. 92
    3.5.4 Consensus Assessment ............................................................. 92
    3.5.5 Analysis of results ................................................................. 92
    3.5.6 Intra-observer Consistency ..................................................... 93
  3.6 CONTROL COMPONENT (PHASE C) ................................................. 93
    3.6.1 Selection of sets of images ....................................................... 93
    3.6.2 Retrieval, Data Collection and display of sets of images .......... 93
    3.6.3 Assessment of images ............................................................. 93
    3.6.4 Consensus Assessment ............................................................. 93
    3.6.5 Analysis of results ................................................................. 94
6 CHAPTER SIX: CONCLUSION

6.1 CONCLUSION

6.1.1 What is the level of consistency between panel members in reviewing image quality?...

6.1.2 Can the level of agreement be improved with training?.................................

6.1.3 What percentage of images is determined to be inadequate?........................

6.1.4 Public versus private sites of production for inadequate images..................

6.1.5 What are the common reasons for inadequate images?.................................

6.1.6 What factors influence the recall versus non-recall of clients in screening
mammography? .................................................................

6.2 FURTHER RECOMMENDATIONS

7 REFERENCES

APPENDIX ONE: ETHICAL APPROVAL

APPENDIX TWO: STATISTICAL ANALYSIS TESTS

A2.1 PAIRED T TEST

A2.2 WILCOXON MATCHED PAIRS TEST

A2.3 CONTINGENCY TABLES (CHI-SQUARE, RELATIVE RISK AND ODDS RATIO)

A2.3.1 Chi-square ($\chi^2$) and Fisher Exact test..............................

A2.3.2 Relative Risk..............................................................................

A2.3.3 Odds Ratio..............................................................................
LIST OF FIGURES

FIGURE 1-1: HUNTER BREASTSCREEN CATCHMENT AREA ................................................................. 9
FIGURE 1-2: NEW SOUTH WALES WITH HBS HIGHLIGHTED ...................................................... 10
FIGURE 2-1: TANGENTIAL AND SAGITTAL VIEW OF BREAST AND ASSOCIATED CHEST WALL ...... 14
FIGURE 2-2: BREAST SHOWING LYMPH NODES AND LYMPH VESSELS ........................................ 15
FIGURE 2-3: DESCRIPTIVE BREAST TERMINOLOGY ....................................................................... 16
FIGURE 2-4: FREQUENCY OF BREAST CANCERS BY LOCATION IN THE BREAST FROM A SERIES OF 961 CONSECUTIVE, HISTOLOGICALLY CONFIRMED CASES FROM THE FALUN CENTRAL HOSPITAL ... 17
FIGURE 2-5: PART OF THE BREAST SHOWING ENLARGED TDLU .................................................. 18
FIGURE 2-6: BREAST DUCT SHOWING IN SITU AND INVASIVE CARCINOMA ................................. 18
FIGURE 2-7: BREAST CANCER INCIDENCE AND MORTALITY WORLD-WIDE .................................. 21
FIGURE 2-9: PROCESS OF SCREENING ............................................................................................. 26
FIGURE 2-10: ULTRASOUND OF (A) SIMPLE CYST AND (B) CARCINOMA (24) ............................... 28
FIGURE 2-11: (A) MAMMOGRAPHY UNIT, (B) SCHEMATIC DRAWING ........................................... 44
FIGURE 2-12: STANDARD MAMMOGRAPHY POSITIONS .................................................................. 46
FIGURE 2-13: COMPRESSION SIGN AT HBS .................................................................................... 47
FIGURE 2-14: CLIENT QUESTIONNAIRE FOR HUNTER BREASTSCREEN ........................................ 48
FIGURE 2-15: A 45° OBLIQUE MAMMOGRAPHY PROJECTION DEMONSTRATING ANATOMICAL STRUCTURES THAT ARE VISUALIZED WITH MAMMOGRAPHY ................................................................. 49
FIGURE 2-16: PATTERNS OF BREAST DENSITY PER BI-RADS, WOLFE & TABAR ............................. 51
FIGURE 2-17: IMAGE QUALITY PARAMETERS .................................................................................. 56
FIGURE 2-18: X-RAY TUBE DETAIL ............................................................................................... 58
FIGURE 2-19: SCHEMATIC DIAGRAM OF AN AUTOMATIC EXPOSURE CONTROL CIRCUIT DESIGN ... 58
FIGURE 2-20: (A) COMPRESSION PLATE WITH AEC CHAMBERS (B) SHOWING CORRECT POSITION ON BREAST TISSUE ............................................................................................................. 59
FIGURE 2-21: LINEAR ATTENUATION COMPARING TISSUE TYPE AND CARCINOMA ....................... 60
FIGURE 2-22: MOLYBDENUM SPECTRUM ...................................................................................... 62
FIGURE 2-23: MOLYBDENUM AND TUNGSTEN SPECTRA .............................................................. 62
FIGURE 2-24: SCHEMATIC DIAGRAM OF PROJECTION OF ACTUAL AND EFFECTIVE FOCAL SPOT .... 64
FIGURE 2-25: SCHEMATIC OF MAMMOGRAPHY EQUIPMENT SHOWING COMPRESSED BREAST TISSUE ... 64
FIGURE 2-26: DESCRIPTORS OF SHAPES AND MARGINS OF MAMMOGRAPHIC LESIONS .............. 73
FIGURE 2-27: CIRCUMSCRIBED LESIONS ON CC MAMMOGRAMS – (A) BENIGN CYST AND (B) MALIGNANT CIRCUMSCRIBED Lesion ................................................................. 74
FIGURE 2-28: DESCRIPTORS AND IMAGE OF TYPICALLY BENIGN MICRO-CALCIFICATIONS ............ 75
FIGURE 2-29: DESCRIPTORS OF INDETERMINATE / MALIGNANT MICRO-CALCIFICATIONS AND IMAGE OF MALIGNANT MICRO-CALCIFICATIONS ................................................................. 75
FIGURE 2-30: CORE SPECIMEN RADIOGRAPH WITH CALCIFICATIONS ........................................... 76
LIST OF TABLES

TABLE 2-1: BREAST DENSITY DESCRIPTIONS .................................................................16
TABLE 2-2: 2002 BREAST CANCER INCIDENCE AND MORTALITY RATES ................20
TABLE 2-3: STAGING OF BREAST CANCER .................................................................23
TABLE 2-4: RISK CLASSIFICATION DERIVED BY WOLFE ........................................24
TABLE 2-5: NATIONAL ACCREDITATION STANDARDS GUIDELINES .........................41
TABLE 2-6: BREASTSCREEN AUSTRALIA STATISTICS .................................................42
TABLE 2-7: ADVANCES IN MAMMOGRAPHY .................................................................43
TABLE 2-8: PGMI EVALUATION OF CLINICAL IMAGE QUALITY (NAS, 2001) ..........54
TABLE 2-9: DETECTION OF SMALL CANCER DETECTION RATE ON OPTICAL DENSITY .................................57
TABLE 2-10: DENSITY & LINEAR ATTENUATION COEFFICIENT - MAMMARY TISSUE & LESION TYPE ...61
TABLE 2-11: CONDENSED VERSION OF EQUIPMENT RELATING TO IMAGE QUALITY AND MAMMOGRAPHY ..............................................................66
TABLE 2-12: COMPARISON OF BI-RADS AND HUNTER BREASTSCREEN CATEGORIES OF MAMMOGRAPHIC BREAST ASSESSMENT / INTERPRETATION .................................67
TABLE 3-1: DATA COLLECTION SHEET A .................................................................87
TABLE 3-2: DATA COLLECTION SHEET B .................................................................88
TABLE 3-3: MODIFIED PGMI SHEET .................................................................89
TABLE 3-4: KAPPA STATISTIC AGREEMENT VALUES .................................................91
TABLE 3-5: DATA COLLECTION SHEET PHASE C ..................................................95
TABLE 4-1: PANEL MEMBERS DEMOGRAPHICS ......................................................99
TABLE 4-2: BREAKDOWN FOR IMAGE SELECTION PHASE A ..................................100
TABLE 4-3: PHASE B: REASONS FOR RECALL .......................................................101
TABLE 4-4: ASSESSMENT BY PANEL MEMBERS .....................................................102
TABLE 4-5: PHASE A: KAPPA STATISTICS PGMI ...................................................105
TABLE 4-6: PHASE A: KAPPA STATISTICS A/I .......................................................106
TABLE 4-7: PHASE A: COMPARISON FOR KAPPA PGMI & KAPPA A/I ................106
TABLE 4-8: PHASE A: ASSESSMENT BY PANEL MEMBERS ...................................107
TABLE 4-9: PHASE B: KAPPA STATISTICS PGMI ...................................................111
TABLE 4-10: PHASE B: KAPPA STATISTICS A/I .....................................................113
TABLE 4-11: PHASE B: COMPARISON FOR KAPPA PGMI & KAPPA A/I ...............113
TABLE 4-12: PHASE C: KAPPA STATISTICS PGMI ................................................116
TABLE 4-13: PHASE C: KAPPA STATISTICS A/I .....................................................117
TABLE 4-14: PHASE C: COMPARISON FOR KAPPA PGMI & KAPPA A/I ...............118
TABLE 4-15: KAPPA OF INTRA-OBSERVER CONSISTENCY BASED ON FOUR CATEGORIES ................................................122
TABLE 4-16: ALL PHASES: INADEQUATE IMAGES BY PUBLIC / PRIVATE SITES ....124
TABLE 4-17: ALL PHASES: REASONS FOR INADEQUATE IMAGES .........................126
TABLE 4-18: CANCER DETECTION VERSUS ADEQUATE / INADEQUATE ...............127
TABLE 4-19: CASE CONTROL DATA ......................................................................129
ABBREVIATIONS

ACR - American College of Radiography
ACS - American Cancer Society
AEC - Automatic Exposure Control Device
BI-RADS - Breast Imaging-Reporting and Data System
BRCA - Breast Cancer Susceptibility Gene
BSA - BreastScreen Australia
BSE - Breast Self Exam
CAD - Computer-Aided Detection
CC - Cranio-Caudal Mammography View
CI - Confidence Interval
DCIS - Ductal Carcinoma In Situ
DES - Diethylstilbestrol
DMIST - Digital Mammographic Imaging Screening Trial
EAR - Excellent, Adequate, Repeat
ER - Estrogen Receptor
FDA - Food and Drug Administration
FNA - Fine Needle Aspiration
HBS - Hunter Region & Wyong Shire BreastScreen (Hunter BreastScreen)
HER2 - Human Epidermal Growth Factor Receptor 2
NHSBSP - National Health Service Breast Screening Programme
HRT - Hormone Replacement Therapy
kVp - Kilovoltage Peak
mAs - Milliampere-Seconds
MLO - Medio-Lateral Oblique Mammography View
MRI - Magnetic Resonance Imaging
NAS - National Accreditation Standards
NCI - National Cancer Institute
OD - Optical Density
OR - Odds Ratio
PET - Positron Emission Tomography
PGMI - Perfect, Good, Moderate, Inadequate
PIAA - Physician Insurers Association of America
PNL - Posterior Nipple Line
QA - Quality Assurance
RR - Relative Risk
SID - Source Image Distance
TDLU - Terminal Duct Lobular Unit
TNM - Tumour (T), Lymph Node Involvement (N) and Any Distant Metastases (M)
UK - United Kingdom
USA - United States of America
ABSTRACT

Introduction: Between 6-10% of women attending breast screening are recalled to investigate an unclear area on the mammogram. Image quality is known to affect image interpretation and it has been suggested that the number of recalls could be reduced with improved image quality.

Aim: This study aimed to investigate the effect image quality has on recall rates, to assess reader consistency using the PGMI classification system and to establish factors leading to recall.

Materials and Methods: A six member panel assessed 904 sets of images (698 recalled; 206 non-recalled) through a BreastScreening Program during three separate phases (pilot, main and non-recall). The pilot study was conducted without additional training in PGMI. Levels of agreement and Kappa statistics were calculated to assess intra- and inter-consistency. The percentage of and reasons for inadequate images was calculated; while a case-control study was conducted to establish factors increasing the likelihood of a client being recalled.

Results: The level of agreement between panel members significantly increased from the pilot to the main study (45.5% to 57.7%) before decreasing slightly for the non-recall (57.7% to 52.2%). Overall, 3.3% of the 904 sets of images were classed as inadequate; the most common PGMI reason was exposure (31%); the left MLO was considered the most common inadequate projection (30%), with more privately produced (66%) images considered inadequate compared to public images (34%). Inadequate image quality did not hinder the cancer detection rates. The case-control component demonstrated current and previous HRT use, increased breast density, better image quality and images being taken at a public site all contributed to a client being recalled.

Conclusion: The results of this study demonstrated that inadequate image quality was not a major factor leading to recall; although twice the number of recalled images were considered inadequate compared to the non-recalled images. The use of the PGMI classification system is highly subjective, with low levels of agreement amongst users. The use of HRT, breast density, imaging site and image quality all contribute to a client being recalled.
CHAPTER ONE: INTRODUCTION
1.1 OVERVIEW
This thesis presents the results of a retrospective review of 904 sets of mammograms from clients x-rayed through NSW Hunter Region & Wyong Shire BreastScreen (Hunter BreastScreen – HBS) during 2002 and 2003. This review was conducted to assess image quality, clinical outcomes for clients screened and for those clients returning for extra images and reasons for their return.

1.2 THESIS OUTLINE
This thesis is presented in six chapters. A brief overview of the content of each chapter is given below.

Chapter 1: This chapter introduces the reader to the background and rationale of this research before leading to the research aims. The objectives and individual research questions required to achieve the research aim are introduced. Finally, the limitations, scope and significance of the research being undertaken are discussed.

Chapter 2: This chapter deals with the literature review starting with an overview of world wide breast cancer statistics, breast anatomy and detection of breast cancer, in particular the modality of mammography. The principles and current research in the area of mammography recalls and image quality are further explored.

Chapter 3: This chapter describes the methodology of the study and includes panel selection, client selection, study design, data processing and analyses.

Chapter 4: The results of the study are reported in this chapter. This chapter begins with the presentation of inter and intra-observer consistency, reasons for and numbers of inadequate images, a case controlled regression to determine patterns of recall and assessment of whether inadequate images influence cancer detection.

Chapter 5: The discussion chapter correlates how the current information and research findings interrelate and diverge. This discussion includes how consistent panel members were in using an established grading system for mammograms, the relationship between image quality and mammographer performance in the public and private sector, reasons for recall, reasons for any images considered inadequate and comparison of client factors for recalled and non-recalled episodes.

Chapter 6: The final chapter deals with the conclusion and recommendation for future studies.
1.3 STUDY BACKGROUND

1.3.1 BREAST CANCER

Breast cancer is a significant health problem in the industrialised western world and the most common form of cancer among women. It is estimated that each year the disease is diagnosed in over one million women worldwide and is the cause of death in over 410,000 women (2). In Australia, it is the second most common cancer after non-melanoma skin cancer (3, 4).

The incidence is increasing in both industrialized and developing countries. The incidence in the United States has increased steadily by about 1-2% per annum since 1960 (5). The majority of other western industrialized countries have similar incidence rates. For example, the Australian incidence rose from 80 cases per 100,000 females in 1983 to 117.3 cases per 100,000 females in 2002 – an average increase of 2 % per annum (4, 6).

Overall, the incidence and prevalence of breast cancer increases with increasing age. In both the United States and Australia, over 70% of new cases are diagnosed in women aged 50 years or older (3, 6, 7). Numerous randomized trials have shown that early detection with mammography increases treatment options, the chance of successful treatment and survival (7, 8). When breast cancer is confined to the breast, the 5 year survival rate is close to 96% (7). As there is no way of preventing breast cancer, the focus in reducing the deaths from the disease has been on finding breast cancer at a localized stage (while still confined to the breast).

1.3.1.1 Detection of Breast Cancer

Breast changes can be detected by the women themselves (through breast self-examination), palpation by their physician or by breast imaging modalities such as diagnostic or screening mammography, ultrasound, magnetic resonance imaging and positron emission tomography. All these modalities will be discussed in Chapter 2 with emphasis on screening mammography.

1.3.2 MAMMOGRAPHY

Mammography is a type of x-ray imaging that produces detailed images of the breast by using low energy x-rays and compression of the breast. While mammograms are considered “uncomfortable”, they are non-invasive, relatively inexpensive and have a reasonable sensitivity (68-92%) that increases with age, when performed correctly (9-14). Mammographic procedures are dependent on variables such as knowledge and skill of the operator, positioning technique employed, amount of compression the client can tolerate, the type, size and position of breast cancer, the radiographic appearance
of the tissue and knowledge of the interpreting physician (15-19). The mammographers’ task is to apply their knowledge of breast anatomy to properly position the client to view as much tissue with as little discomfort as possible (15).

Mammograms can be performed in a diagnostic or screening setting. A diagnostic mammogram is performed on clients who have a referral from their physician. These clients may have a breast concern, a family history of breast cancer, be any age or gender. Screening is used as a mass tool to check for unexpected breast cancer in asymptomatic women. These women are able to make their own appointments in countries with centralized screening programs. Although BreastScreen Australia is specifically aimed at women 50-69 years (the target group) without symptoms, women 40 to 50 and 70 years and older may attend for screening (20). The standard mammogram consists of two views of each breast – the cranio-caudal (CC’s) are views whereby the breast is compressed from the superior to inferior aspect and the medial-lateral-oblique (MLO’s) whereby the breast is compressed from the medial to lateral aspect.

The challenge for mammography is to display subtle differences in tissues of similar density while retaining detail and reducing dose (21, 22). Screening does not provide a diagnosis; the challenge is to perceive the occasional abnormal lesions (23, 24). The differential diagnosis of the lesions perceived at screening requires an appropriate workup with a multimodality approach (23, 24) generally 2 to 3 weeks after the initial mammogram.

The goal of any screening program is to obtain high cancer detection rates while avoiding unnecessary diagnostic evaluation resulting in false positive results, which are costly and associated with ongoing psychological morbidity (25, 26).

1.3.2.1 Pathways for diagnosis within screening mammography

This section provides a very brief overview to the imaging pathways in screening mammography. A detailed description is provided in chapter 2.

The images are acquired and reviewed for technical quality by the mammographer. If not technically adequate, the mammograms are retaken. This process is referred to as a technical fault and is usually done before the films are reviewed by the screen readers (generally radiologists or trained breast physicians). There is a multitude of criteria to assess if quality mammograms are being produced. The method of assessing mammograms produced within BreastScreen Australia is the established grading system of PGMI (Perfect, Good, Moderate, Inadequate) (discussed in section 2.6.4.1.2).
Once technically adequate, the mammograms are reviewed by 2 (and sometimes 3) screen readers to determine whether there is evidence of any abnormality. If the images appear clear, the clients are invited to be re-screened in 2 years. If there is an area that needs clarification, these women are asked to return for a recall.

1.3.2.2 Recall and Diagnosis

At recall, further imaging is performed such as spot or magnified mammographic views and possibly ultrasound. If further imaging confirms an abnormality, the woman may undergo fine-needle aspiration, core biopsy or even open biopsy to determine the nature of the abnormality. The final diagnosis is then obtained by histo-pathological methods.

1.3.2.3 Advantages

Mammography’s primary advantage is the ability to detect a tumour less than 1 cm in size up to 2-3 years before it is clinically palpable (27) and possibly before it has spread. This early intervention improves the chance of survival as axillary lymph node metastases and tumour size are the two most important prognostic features of breast cancer (28-30). In the last 15 to 20 years, most developed countries have introduced screening mammography programs, as screening has been shown to reduce mortality by 30% (8, 31-33). BreastScreen Australia, a national screening program, was introduced in the early 1990’s. The Australian breast cancer mortality rate has declined by an average of 2% per annum from 1990 to 2004 – from 31 to 23.4 deaths per 100,000 females (4).

1.3.2.4 Disadvantages

A screening mammogram is considered positive if the screen readers’ interpretation was indeterminate or aroused a suspicion of cancer, or if there was a recommendation for non-routine follow-up, including physical examination, diagnostic mammogram within 12 months, ultrasound exam or a biopsy. The current practice of mammography generates high levels of recalls with false positive results (recalls where further imaging clarified the “lesion” and no further investigation was required). It is estimated that approximately 90% of the follow-up investigations (recalls) do not result in breast cancer diagnosis (34).

Studies have estimated that between 6% and 10% of all clients attending breast screening are recalled (34-37). A meta-analysis of world wide recall rates reveal that the USA has the highest and the Netherlands the lowest recall rates (15% versus 1.4%) (38, 39). In Australia, the National Accreditation Standards (NAS) contains strict
guidelines in relation to the number of women who should be recalled. Based on the results of overseas screening programs, the NAS of BreastScreen Australia indicate that less than 10% of women who attend for their first screen should be recalled for assessment and that less than 5% of women who attend for their second or subsequent screen should be recalled for assessment (40).

There are always costs and this is especially true of false positive screens. After the initial mammogram, there are the real financial costs of resources which include the follow-up costs of evaluating these mammograms. There are also important psychological adverse effects women may suffer such as the emotional toll of false positive mammograms (34, 37, 41-43). Encouraging studies indicate that the experience of a false positive mammogram does not deter women from re-attending subsequent screening in the future (42, 44-48). However, contradicting studies indicate that women who experience a false positive mammogram were less likely to re-attend for a subsequent screening (49, 50) and if the client had undergone a benign biopsy, 35% less likely (49). As a positive, although recalled women experience high levels of affective-cognitive distress, these effects are not long lasting (37, 41, 51). It is understandable that women attending a recall would experience higher levels of breast cancer worries, fears and beliefs than non-recalled women, but studies have shown that these effects are not relevant at 2 months following notification of a clear result (51).

The key to surviving breast cancer is early detection and treatment. The present gold standard for detection is screening mammography; although some costs for clients of screening include fear of compression, fear of being recalled, fear of breast cancer, and the increased diagnosis of ductal carcinoma in situ (DCIS) - which has an uncertain natural course and its detection has increased since screening commenced.

### 1.3.3 Image Quality

Satisfactory viewing of the mammogram is dependent upon adequate levels of image quality as this is known to affect image interpretation and can be influenced by a variety of factors, such as positioning technique, compression of the breast, exposure and equipment factors. For a mammographer, the technical quality of mammograms can be divided into two categories: those associated with the physics of the x-ray imaging process and those associated with performance of the examination by the mammographer (52). This later category is influenced by the client in terms of mobility, breast density and breast composition.
Increased breast tissue density is associated with a high recall rate compared to fatty tissue (7-10% versus 2-4%) (36, 53-55) and may increase the risk of getting breast cancer by 4-6 times (56-58). Only age and mutations in the BRCA 1 and 2 genes carry more risk than dense breast tissue (59). The use of hormone replacement therapy (HRT) can increase breast tissue density by arresting and often reversing the process of involution (24, 60, 61). Increased breast density may lead to less tolerance of compression due to breast tenderness (62) and more difficulty in observing subtle differences in tissue due to less attenuation, higher scatter and reduced contrast (21, 58, 63-65).

Poor quality mammograms may fail to achieve the mortality reductions and have adverse consequences such as missed cancers, increased false positives, increased costs and anxiety and discomfort for the women who undergo unnecessary additional procedures (66-68). One study conducted by Taplin et al demonstrated that incorrect breast positioning was the primary reason for not finding interval detected invasive breast cancers, after adjusting for age, film date and breast density (68). Image quality affects image interpretation and if this false negative study by Taplin et al concluded that the number of “missed cancers” might be reduced by better positioning technique, then maybe the number of false positive recalls could be reduced in the same way.

Poor image quality can also increase the number of pseudo lesions due to tissue superimposition. Superimpositions simulating a tumour mass have been reported as the most common cause for recall (69). Possible causes of superimpositions include tissue overlapping or inadequate compression.

1.3.4 RATIONALE FOR STUDY

While the numbers of recalls within BreastScreen Australia are generally low and within NAS guidelines (fewer than 10% recalled after first screen and less than 5% recalled after subsequent screens) (20); screen readers and mammographers at Hunter BreastScreen perceived that some recalls might have been avoided if the screening mammograms had been of better image quality. This study was conducted to review a series of mammograms from women screened including both recalled and non-recalled clients. This review included assessment of the mammographic image quality by a panel of readers, reasons and outcomes of the recall and any substandard images.
1.4 RESEARCH AIM
The primary aim of this research is to assess if inadequate image quality of screening mammograms contributes to false positive results among asymptomatic clients attending BreastScreen New South Wales Hunter Region & Wyong Shire.

The secondary aims of the research are to:

- identify factors attributing to mammographic images deemed inadequate by a panel of experts using the PGMI evaluation system
- investigate the intrinsic factors contributing to why breast screening clients are asked to return for further assessment after their initial mammogram

1.5 RESEARCH QUESTIONS
The research questions for this study are:

- What is the level of consistency between panel members in reviewing image quality of screening mammograms using the PGMI evaluation system?
- Can the level of agreement between panel members using the PGMI evaluation system be improved with training?
- What percentage of mammographic images is determined to be inadequate using the PGMI evaluation system?
- What are the common reasons for inadequate images among asymptomatic clients attending BreastScreen New South Wales Hunter Region & Wyong Shire?
- What factors influence the recall versus non-recall of asymptomatic clients attending BreastScreen New South Wales Hunter Region & Wyong Shire?

1.6 RESEARCH OBJECTIVES
In order to address the research aims and to answer the research questions, a series of objectives are set. The primary objectives of this study are:

- Conduct full literature review
- Develop research methodology
- Obtain ethical approval
- Identify and recruit participants
- Gather resources (mammograms and data)
- Conduct pilot study
• Collect, collate, enter and analyse data for pilot study
• Make necessary changes to research methodology
• Conduct main study
• Collect, collate, enter and analyse data for main study
• Write final report

1.7 SCOPE / LIMITATIONS OF RESEARCH
This research was limited to BreastScreen NSW Hunter Region and Wyong Shire (HBS) during the time period of 2002-2003 (Figure 1-1). This area extends as far north as Taree, west as far as Murrurundi / Merriwa and south as far as Wyong and incorporates approximately 83,570 potential clients in the target age group, approximately 53.4% have been screened through HBS. Figure 1-2 demonstrates the position of HBS in New South Wales, Australia.

Figure 1-1: Hunter BreastScreen Catchment Area (BreastScreen NSW Hunter Region & Wyong Shire; 2000)
Although limited to one BreastScreening Program, this research did incorporate sets of mammograms taken through both the public and private sector (incorporating four fixed sites and three mobile screening units) with images produced by approximately thirty mammographers of varying experience and capability.

This retrospective study reviewed sets of mammograms from 2002 to 2003. This research had proposed a prospective study during 2004, but this component of the study was cancelled due to circumstances beyond the author’s control.

The sets of mammograms were assessed using the established grading system of PGMI (Perfect, Good, Moderate, Inadequate). Although all the panel members were aware of the system, not all were familiar with its use. Plus, there may be discrepancies between routine reading and reading with PGMI. Firstly, the expert panel were aware of the nature of this research; therefore, may have scrutinized the films in a different manner for the study as opposed to the level in routine screening and while experienced, all readers have their own prejudices which became apparent during the research.

There is the possibility that there is a difference between calling a film inadequate and actually asking that client to return for a technical fault. Examples include:

- if the woman has fatty breast tissue with a very clear MLO (no abnormalities detected) but slight unsharpness on the CC film
• if the woman is disabled or has a sore shoulder and coverage is inadequate but even if brought back, the mammographer will obtain no better quality images.

These clients may not be asked to return for a technical fault (even if their films were considered inadequate by the grading system of PGMI). The PGMI is an academic and subjective rating, whereas the decision to fault a film technically is a combination of technical plus other factors, including client limitations and breast appearance.

Mammographers were used to assess the sets of mammograms as well as other screen readers. This may inadvertently bias the grading of the sets as the mammographers use this system routinely to assess mammograms while the screen readers generally tend to focus on whether the images contain any abnormalities.

Lastly, both the small number of inadequate images found during the study and the small number of clients evaluated for the conditional logistical regression analysis (N = 220) was another shortcoming in this study.

1.8 SIGNIFICANCE OF THE STUDY
The significance of any research in breast imaging should be to improve services for the client. This research tries to address this in several ways including:

- Reiterating for mammographers an awareness of the technical quality for the images they produce
- Identifying problem areas for inadequate imaging in mammography
- Determining which factors can and cannot be changed in terms of client recalls and possibly including information on which clients may be more likely to be recalled
- Trying to validate an established grading system for assessment of images

This research aims to investigate factors attributing to recall. Although screening mammography is the gold standard for detecting breast cancer before it becomes clinically apparent, poor quality mammograms may fail to achieve the mortality reductions and have adverse consequences such as missed cancers, increased false positives, increased costs and anxiety and discomfort for clients (66-68). While less than 10% of all clients screened are recalled, some factors may influence recall rates such as dense breast tissue, the use of hormone replacement therapy and family history of breast cancer. This research is, in essence, a microcosm of BreastScreen Australia.
2 CHAPTER TWO: LITERATURE REVIEW
2.1 INTRODUCTION
A comprehensive search was conducted using electronic databases. Databases included EMBASE, Medline and Meditext. Reference lists from potentially relevant papers were examined for additional papers. Information of non-changing aspects in mammography (anatomy, historical, positioning) were obtained from the University of Newcastle, Hunter BreastScreen and personal libraries. Although men may get breast cancer (1% to 2%) (70), this literature review is specifically targeted at women.

2.2 BREAST ANATOMY
2.2.1 BREAST ANATOMY
The breast is situated between the superficial and deep layers of the superficial fascia to the anterior, and to some extent the lateral aspects of the chest wall. Each breast extends from the level of the 2nd rib to the level of the 6th intercostal cartilage, and from the side of the sternum to the mid-axillary line. Breast tissue extending into the axilla is known as the tail of Spence, generally called the “tail” (71).

The breasts are cone shaped, with the base of the posterior surface of the breast overlying the pectoralis major and the serratus anterior muscle (Figure 2-1). The breast is composed of glandular tissue (breast parenchyma), together with fibrous connective tissue and fat (breast stroma) for support. The breast ends in a nipple where the lactiferous ducts open. The nipple may be erect, flattened or inverted (a congenital abnormality different to retraction). The areola is the pigmented area surrounding the nipple and may appear wrinkled or roughened. The breast is firmly attached to the skin by suspensory ligaments called Cooper’s ligaments. These ligaments determine the firmness of the breast (71).

The female mammary glands (breasts) are milk producing organs that provide nourishment to offspring (25). During embryonic development, there is growth and differentiation of the breasts. Paired glands develop along the milk lines, which extend between the future axilla to the future inguinal region (25).

There is variation among individual females in size and shape of the breasts. Individually, the size, weight and composition of the breast varies with age and stage of development i.e. the breasts may be small prior to puberty, enlarging with development, increasing in size during pregnancy and lactation, then atrophying with old age.
The glandular system comprises 15 to 20 lobes which are divided into lobules (the basic structural units of the breasts) (Figure 2-2). The lobules contain the glandular elements (or acini). Ductules are also known as acini in the lactating breast. Each lobule consists of several acini, draining ducts and interlobular or connective tissue. The lobule and its duct form a basic histo-pathological unit of the breast, which is designated as the terminal ductal lobular unit (TDLU), from which most benign and malignant lesions of the breast arise (19, 61, 71). The lobes and lobules are all part of the breast parenchyma and participate in hormonal changes. The lobules tend to decrease in size and number with increasing age and particularly after pregnancy (the start of involution). The process of Involution involves regression of the glandular and connective tissue by replacement of fat in the breast.

The primary route of lymphatic drainage of the breast is through the axillary lymph node groups. Lymph vessels drain laterally into axillary lymph nodes (75%) and medially into the internal mammary chain of lymph nodes (25%) (25). There are between 12 to 30 axillary nodes. The lymphatic drainage of the breast has diagnostic and therapeutic implications as tumours can spread through the lymphatic vessels (61).
Positive axillary nodes are a major risk factor for later systemic disease (24, 25, 29, 61, 72).

The female breast undergoes extensive development under the influence of hormonal changes until about the age of 20; by 40 the breast has begun to atrophy. Changes in the breast occur during menstruation, pregnancy, lactation and menopause. At menopause, the breast involutes and the fibro-glandular tissue is replaced with fat and connective tissue (25, 71).

![Breast Diagram](image)

**Figure 2-2: Breast Showing Lymph Nodes and Lymph Vessels (73)**

### 2.2.2 Breast Density

There is a wide range of normal breast size, shape and breast tissue composition (25). Breast composition can range from almost all fat to extremely fibro-glandular (dense) tissue and the composition usually relates to the individual sensitivity of the mammogram (24). Various descriptors of breast density include Wolfe’s classification and the American College of Radiologists categories of the Breast Image and Reporting Data System (BI-RADS) (Table 2-1). These descriptors will be compared with Tabar’s pattern of mammographic breast appearance (section 2.6.3).

In the past, assessing breast density has been a subjective evaluation; today breast density can be assessed more objectively. One method is to use a computer assisted method (scanner). The scanner automatically calibrates after turning on the machine. The density of breast tissue is measured to produce a digitized form of the mammogram. Thus the quality of the scanned images is assumed to remain constant for all mammogram films and this assessment work method has been validated to be accurate for assessment and reproduction (74).
## Table 2-1: Breast Density Descriptions (75, 76)

<table>
<thead>
<tr>
<th>Breast Density Descriptors</th>
<th>BI-RADS</th>
<th>Wolfe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Breast composed entirely of fat</td>
<td>N1 – Parenchyma entirely composed of fat</td>
<td>“normal” breast tissue</td>
</tr>
<tr>
<td>2. Scattered fibro-glandular densities that could obscure a lesion at mammography</td>
<td>P1 – Chiefly fat with up to 25% nodular densities</td>
<td>“mixed” breast tissue</td>
</tr>
<tr>
<td>3. Heterogeneously dense (which may lower the sensitivity of mammography)</td>
<td>P2 – Over 25% nodular densities</td>
<td>“dense” breast tissue</td>
</tr>
<tr>
<td>4. Extremely dense (which lowers the sensitivity of mammography)</td>
<td>DY – No ductal densities visible</td>
<td>“very dense” breast tissue</td>
</tr>
</tbody>
</table>

### 2.2.3 Breast Regions

Areas of the breast should always be described in a consistent manner. Each breast is divided into four quadrants: the upper outer quadrant (UOQ), upper inner quadrant (UIQ), lower outer quadrant (LOQ), and lower inner quadrant (LIQ) (17-19) and can also be viewed as the face of a clock (Figure 2-3). A breast lesion in the right UOQ at 10 o’clock for example, would correspond to the same area on the left UOQ at 2 o’clock.

![Figure 2-3: Descriptive Breast Terminology (19)](image)
Cancer of the breast affects the left breast slightly more than the right (77) with the left breast about 10% larger than the right in approximately 80% of women (19). The UOQ of the breast (tail) contains more glandular tissue than the remainder of the breast (25). Approximately 50% of cancers occur in the UOQ, 15% in the UIQ, 11% in the LOQ, 6% in the LIQ and 18% in the nipple (19, 24, 77). This percentage varies slightly as demonstrated in a series of cases in Falun, Sweden (24) (Figure 2-4).

Figure 2-4: Frequency of Breast Cancers by Location in the Breast from a Series of 961 Consecutive, Histologically Confirmed Cases from the Falun Central Hospital (24)

2.3 BREAST CANCER

2.3.1 BREAST CANCER DEFINITION

Breast cancers are derived from the epithelial cells that line the terminal duct lobular unit (19, 25, 71, 78) (Figure 2-5). Cancer cells that remain within the basement membrane of the TDLU and its draining duct are classed as in situ, non-invasive or ductal.

In situ cancer has remained intraductal and can remain so for years before penetrating the basement membrane to become infiltrating (or invasive). Ductal carcinomas account for about 85% of breast cancers (79); while it is estimated that approximately 15-30% of all screen detected mammography cancers are ductal carcinoma in situ (DCIS) (32, 61, 80).
An invasive or infiltrating breast cancer occurs when there are cancer cells outside the basement membrane of the ducts and lobules into the surrounding adjacent normal tissue. Both in situ and invasive cancers have characteristic patterns by which they can be classified (82) (Figure 2-6).

After excision, breast lesions can be classified histo-pathologically as benign or malignant. Some of these lesions include, but are not limited to:

- Benign lesions such as cysts, epithelial hyperplasia, atypical hyperplasia, sclerosing adenosis, radial scars, duct ectasia, fibroadenomas, phyllodes tumour and papillomas (25, 71, 83).
- Ductal in situ carcinomas (examples solid, cribiform, papillary and comedo), Paget’s disease, intracystic papillary and lobular carcinoma in situ (25, 71, 83).
- Invasive carcinomas such as ductal, lobular, tubular, papillary, medullary, mucinous (colloid) (25, 71, 83).
2.3.2 Incidence / Mortality of Breast Cancer

Breast cancer is a significant health problem in the industrialised western world and the most common form of cancer among women in North America and almost all of Europe. In Australia, it is the second most common cancer after non-melanoma skin cancer (3, 4) and the most common cause of cancer related deaths in women (84). It is estimated that each year the disease is diagnosed in over one million women worldwide and is the cause of death in over 410,000 women (2).

After continuously increasing for more than two decades, breast cancer incidence rates in women in the United States decreased by 3.5% per year from 2001 to 2004 (85). In 2008, the National Cancer Institute (NCI) estimates that 182,460 women in the United States will be diagnosed with and 40,480 women will die of cancer of the breast (86, 87).

The majority of other western industrialized countries have similar incidence trends. For example, while Australian incidence increased on average 2% per annum from 80 cases per 100,000 females in 1983 to 117.3 cases per 100,000 females in 2002; age standardized incidence of breast cancer is projected to remain at this level until 2011 (4, 6, 88). The incidence and mortality for some countries is shown in Table 2-2.

Overall, the incidence and prevalence of breast cancer increases with increasing age. In both the United States and Australia, over 70% of new cases were diagnosed in women over 50 years of age (4, 7, 70). The incidence of breast cancer is negligible for women in their teens and 20's. Under the age of 30 years, due to the low prevalence of disease, the radiation risk of screening mammography exceeds any benefit in detection (89). This is the primary reason for not performing screening mammography in this age group. Other reasons include the fact that tissue is more sensitive to radiation and mammograms will be more difficult to interpret due to the higher density of breast tissue (58, 90).

Although the increase in incidence with age is seen elsewhere in the world, the absolute rates of incidence are lower for each age group in Japan. Japanese women are four to five times less likely to develop breast cancer than a woman in the United States (91-94) Figure 2-7). This advantage is lost for Japanese immigrants to the United States within 1-2 generations (93).
Table 2-2: 2002 Breast Cancer Incidence and Mortality Rates (88)

<table>
<thead>
<tr>
<th>Population</th>
<th>Incidence Numbers</th>
<th>Incidence Crude</th>
<th>Incidence ASR(W)</th>
<th>Mortality Numbers</th>
<th>Mortality Crude</th>
<th>Mortality ASR(W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>World</td>
<td>1,151,298</td>
<td>37.4</td>
<td>37.4</td>
<td>410,712</td>
<td>13.3</td>
<td>13.2</td>
</tr>
<tr>
<td>More developed</td>
<td>636,128</td>
<td>103.7</td>
<td>67.8</td>
<td>189,765</td>
<td>30.9</td>
<td>18.1</td>
</tr>
<tr>
<td>Less developed</td>
<td>514,072</td>
<td>20.9</td>
<td>23.8</td>
<td>220,648</td>
<td>9.0</td>
<td>10.3</td>
</tr>
<tr>
<td>Australia</td>
<td>11,176</td>
<td>114.1</td>
<td>83.2</td>
<td>2,667</td>
<td>27.2</td>
<td>18.4</td>
</tr>
<tr>
<td>Canada</td>
<td>19,540</td>
<td>124.0</td>
<td>84.3</td>
<td>5,305</td>
<td>33.7</td>
<td>21.1</td>
</tr>
<tr>
<td>New Zealand</td>
<td>2,330</td>
<td>120.0</td>
<td>91.9</td>
<td>670</td>
<td>34.5</td>
<td>24.5</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>40,928</td>
<td>135.5</td>
<td>87.2</td>
<td>13,303</td>
<td>44.0</td>
<td>24.3</td>
</tr>
<tr>
<td>United States of America</td>
<td>209,995</td>
<td>143.8</td>
<td>101.1</td>
<td>42,913</td>
<td>29.4</td>
<td>19.0</td>
</tr>
<tr>
<td>Central &amp; Eastern Europe</td>
<td>100,262</td>
<td>63.4</td>
<td>42.6</td>
<td>45,310</td>
<td>28.7</td>
<td>17.9</td>
</tr>
<tr>
<td>Northern Europe</td>
<td>62,425</td>
<td>128.8</td>
<td>82.5</td>
<td>19,789</td>
<td>40.8</td>
<td>22.6</td>
</tr>
<tr>
<td>South-Eastern Asia</td>
<td>58,495</td>
<td>21.8</td>
<td>25.5</td>
<td>26,818</td>
<td>10.0</td>
<td>11.8</td>
</tr>
<tr>
<td>Southern Europe</td>
<td>72,458</td>
<td>97.8</td>
<td>62.4</td>
<td>24,617</td>
<td>33.2</td>
<td>18.1</td>
</tr>
<tr>
<td>Western Europe</td>
<td>125,604</td>
<td>134.3</td>
<td>84.6</td>
<td>39,297</td>
<td>42.0</td>
<td>22.3</td>
</tr>
</tbody>
</table>

Notes
1. Cancer number and rates are estimates for the middle of 2002, from the most recent data available, generally 3-5 years earlier.
2. Rates are expressed per 100,000 populations and age-standardised to the year 2002 Standard Population of the corresponding country and of the World Stand Population (ASR(W)).

Source: GLOBOCAN 2002, IARC, 2005

Epidemiological studies indicate that the advantage by Japanese women may be lessening, with the incidence of breast cancer in Japan doubling between 1960 and 1985 (95). The change in prevalence may reflect the increasing adoption of Western lifestyles in Japan over the last 50 years.

The incidence and mortality rates are more useful for comparison between two different populations or for comparisons within a population over time. Figure 2-8 shows the changes in incidence and mortality from breast cancer in Australia. The incidence rate is a function of the underlying time of disease trends, as well as social contribution of screening / diagnostic mammography and availability of services, while the morality rate is a function of incidence rates over time and the accompanying trends in survival. In Australia, the increase in incidence in the 50-69 year age group in the early 1990’s is thought to be attributable to the introduction of the National BreastScreening Program, as this is the age group targeted. The breast cancer mortality rate has declined by an average of 2% per annum from 1990 to 2004 – from 31 deaths per 100,000 females to 23.4 deaths per 100,000 females (4).
2.3.3 BREAST CANCER STAGING

Breast cancers are classified by the extent of the disease or stage. The stage is based on the size of the tumour and whether the cancer has spread. Axillary lymph node metastases and tumour size are the two most important prognostic features of breast cancer (28-30).

The international system for classifying cancer by tumour size and location, regional lymph node involvement and distant metastases (the TNM staging system) has served oncologists for more than 50 years and was developed by the American Joint Committee on Cancer (AJCC) and the Union Internationale Contre le Cancer (UICC). The system classifies by tumour (T), lymph node involvement (N) and any distant metastases (M). This classifies the stage, site, size of the tumour, extent of local advancement, regional node status and the presence of distant metastases. The stages for breast cancer are from 0 – IV, with local stage being confined to the breast, regional spreading to the surrounding tissue or lymph nodes and distant metastasising (spread) to distant organs (25, 86) (Table 2-3).

There was debate at the 2006 European Breast Cancer Conference concerning whether TNM is still relevant in breast cancer compared with more relevant knowledge about the role of estrogen and progesterone positive and negative markers (ER and PR), and HER-2 gene status (96). As no consensus was reached, a breast pathology
report in 2007, contains not only client identifiers (side, any further anatomical detail and procedure) and lesion classification by tumour (size, type, grade, margin status), lymph node involvement and any metastases (lymphatic / vascular invasion), but also predictive markers for tailored therapy such as estrogen receptor (ER + / -) and HER-2 status (97).

**Incidence of breast cancer in women, Australia, 1989–2003**

<table>
<thead>
<tr>
<th>Year</th>
<th>New cases per 100,000 women</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>70+ years</td>
</tr>
<tr>
<td>1990</td>
<td>50–69 years</td>
</tr>
<tr>
<td>1991</td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>Approximate commencement of the BreastScreen Australia Program</td>
</tr>
<tr>
<td>1993</td>
<td>All ages</td>
</tr>
<tr>
<td>1994</td>
<td>&lt; 50 years</td>
</tr>
<tr>
<td>1995</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td></td>
</tr>
</tbody>
</table>


**Mortality from breast cancer, females, Australia, 1990–2004**

<table>
<thead>
<tr>
<th>Year</th>
<th>Deaths per 100,000 women</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>70+ years</td>
</tr>
<tr>
<td>1991</td>
<td>50–69 years</td>
</tr>
<tr>
<td>1992</td>
<td>Approximate commencement of the BreastScreen Australia Program</td>
</tr>
<tr>
<td>1993</td>
<td>All ages</td>
</tr>
<tr>
<td>1994</td>
<td>&lt; 50 years</td>
</tr>
<tr>
<td>1995</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
</tr>
</tbody>
</table>

Source: AIHW National Mortality Database.

**Figure 2-8: Breast Cancer Changes in Incidence (1989-2003) and Mortality (1990-2004) in Australia (84)**
Table 2-3: Staging of Breast Cancer (adapted from (98))

<table>
<thead>
<tr>
<th>Stage</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 0</td>
<td>Carcinoma in situ</td>
</tr>
<tr>
<td>Stage I (local)</td>
<td>Higher the stage, the more extensive the disease: larger tumour size, and / or spread of cancer to lymph nodes and / or organs near primary tumour</td>
</tr>
<tr>
<td>Stage II (limited regional)</td>
<td></td>
</tr>
<tr>
<td>Stage III (extensive regional)</td>
<td></td>
</tr>
<tr>
<td>Stage IV (distant metastases)</td>
<td>Metastases to other organs</td>
</tr>
</tbody>
</table>

2.3.4 Breast Cancer Risk Factors

There are no known reasons why one woman develops breast cancer and another does not. There are, however, risk factors that make the development more likely. Advancing age is the most common risk factor (86). Breast cancer is extremely rare through the second decade of life, and only about 0.3% of breast cancers occur in women under 30. The incidence begins to increase rapidly around 35 and continues throughout life, at a more gradual pace in postmenopausal women. More than 70% of breast cancers occur in women over 50 years of age (4, 7, 70), with the latest statistics reporting that 97% of breast cancer deaths occur in women aged 40 and older (7, 99).

Unavoidable personal and family history risks include having a previous breast carcinoma, a first degree relative (mother, sister or daughter) who has had breast cancer, certain breast changes such as atypical hyperplasia, gene changes (such as the presence of BRCA1 and BRCA2), early menarche or late menopause, being a Caucasian woman or having dense breast tissue (56-59, 86, 100).

The significance of breast density as an indicator of risk for the development of malignancy has been a controversial issue in breast imaging. Wolfe assessed the relationship between breast density and the risk of malignancy after reviewing a series of mammograms produced during 1967-1971 (Table 2-4). This study concluded that women with radiographically dense breast pattern (DY) had a 22 times greater risk of developing breast cancer than women with almost all fat (N1) (101). Boyd et al showed that compared with women with density in less than 10% of the mammogram, women with density in 75% or more had a 4.7 times increased risk of breast cancer (95% CI 3.0-7.4) (56).
Table 2-4: Risk Classification Derived by Wolfe (76)

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Number of Women</th>
<th>%</th>
<th>Number of Breast Cancers</th>
<th>Incidence %</th>
<th>Risk of Breast Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1 (Fatty)</td>
<td>2190</td>
<td>41 %</td>
<td>3</td>
<td>0.1 %</td>
<td>Minimal</td>
</tr>
<tr>
<td>P1 (Mixed)</td>
<td>1355</td>
<td>26 %</td>
<td>6</td>
<td>0.4 %</td>
<td>Moderate</td>
</tr>
<tr>
<td>P2 (Dense)</td>
<td>1375</td>
<td>26 %</td>
<td>2.3</td>
<td>1.7 %</td>
<td>Significant</td>
</tr>
<tr>
<td>DY (Very Dense)</td>
<td>364</td>
<td>7 %</td>
<td>8</td>
<td>2.2 %</td>
<td>Highest</td>
</tr>
<tr>
<td>Totals</td>
<td>5284</td>
<td>100 %</td>
<td>40</td>
<td>0.8 %</td>
<td></td>
</tr>
</tbody>
</table>

Other risks include having a first child later in life, radiation therapy to the chest, taking DES (diethylstilbestrol) during pregnancy, being overweight or obese after menopause, lacking physical activity and possibly alcohol consumption (86, 102). The National Cancer Institute advises, however, that most women with known risk factors do not get breast cancer. In fact, except for getting older, most women with breast cancer have no clear risk factors (86). As there is no way yet of preventing breast cancer, the focus in reducing the deaths from the disease has been on the early detection of breast cancer.

2.4 BREAST CANCER DETECTION

2.4.1 CLINICAL EXAMINATION AND SELF-EXAMINATION

Early breast cancer usually produces no signs when it is still small and treatable. Early detection and diagnosis are therefore the key to surviving breast cancer. It is vital for women to follow guidelines to assist in finding breast cancers before symptoms develop. The American Cancer Society (ACS) and National Cancer Institute (NCI) advocate an annual clinical examination by a health care provider (7, 86). Also, women who conscientiously practice breast self-exam (BSE) improve their chance of detecting early breast cancer by noticing subtle differences that may lead them to consult a doctor more quickly. In 2003, the American Cancer Society (ACS) dropped its recommendation of BSE for women. The reasoning was that it is more important for women to have awareness of their breasts than a structured plan for checking. If a woman chooses to practice BSE, the ACS advises receiving instructions from a qualified health care provider (7).
2.4.2 IMAGING MODALITIES

Breast imaging includes all imaging methods used to detect and diagnose diseases of the breast. Some of the modalities currently used in breast imaging include film-screen mammography, digital mammography, ultrasound, magnetic resonance imaging and positron emission tomography.

The goal of any screening program is to obtain high cancer detection, while avoiding unnecessary diagnostic evaluation following false positive recall results, which are costly and associated with ongoing psychological morbidity (25, 26). The standards for optimal breast cancer screening are constantly evolving due to technological improvements and the experience of the health care team. At present, the gold standard for detection of breast cancer is screening mammography.

Numerous randomized trials have shown that early detection with mammography increases treatment options, the chance of successful treatment and survival (7, 8). When confined to the breast, the five year survival rate is close to 96% (7).

2.4.2.1 Mammography

Mammography visualizes the internal structures of the breast using low energy x-rays. A radiographer specializing in mammography (a mammographer) performs the examination. During the procedure, generally two or three standard views are taken of each compressed breast. A screen reader reviews each set of mammograms, any suspicious area(s) are noted and additional imaging is used to assess these area(s). Mammograms may be performed in a diagnostic or screening setting.

2.4.2.1.1 Screening and Diagnostic Mammography

While equipment and positioning remain constant whether performed in a diagnostic or screening setting, the clientele and progression in procedures differ.

Diagnostic mammography is used to evaluate clients who present with either breast symptoms (such as lump, pain, discharge, nipple retraction), have previously been diagnosed with breast cancer or have a strong family history of breast cancer. Clients must be referred by their physician and include women, girls and males. Diagnostic mammography is the work-up of clinically or radiologically detected areas of concern. These clients may require additional specialized views, use of other modalities (such as ultrasound / MRI) or other procedures (such as fine needle or core biopsies), which are generally performed at their initial visit.
Screening mammography is used to check for unexpected breast cancer in asymptomatic women. Women are able to make their own appointments in countries with centralized screening programs. BreastScreen Australia is specifically aimed at women without any breast symptoms who are between 50-69 years (the target group), although women 40 to 50 and 70 years and older may attend for screening (20). The goal of screening mammography is to detect breast cancers when they are too small to palpable. The process of screening is illustrated in (Figure 2-9).

The challenge of screening is to perceive the occasional abnormal lesions (23, 24). The differential diagnosis of the lesions perceived at screening requires an appropriate work-up (recall) with a multi-modality approach (24) generally 2 to 3 weeks after the initial mammogram. Determining whether an area is benign or malignant is complicated, as these disease processes often imitate one another. At these times a biopsy is necessary. If an area is not palpable, needle cytology is obtained using ultrasound guidance. Core biopsies, involving the removal of intact pieces of tissue and analysed for histo-pathological evaluation, are attained under both ultrasound and stereo-tactic guidance. The advantages of obtaining a core are two-fold; if the core is benign, no surgery is required and if malignant, more information is available from pathology to plan treatment (103). The procedure is also less traumatic than open biopsy for a client.

![Figure 2-9: Process of Screening](image)
2.4.2.1.2 Digital Mammography

While digital and film screen mammography use similar techniques to produce an image of the breast; digital separates image acquisition and display (104). This is achieved by using an image receptor to take an electronic image of the breast and store it on a computer. There are a number of methods how a digital image can be acquired. These include point scanning, line scanning, multiline scanning and area detectors including storage phosphor systems, charge-coupled devices, amorphous-silicon photodiode transistor array and amorphous selenium as a detector). The US Food and Drug Administration (FDA) approved the first digital mammography system for clinical use in January 2000 (86).

The spatial resolution of digital mammography is 8 to 10 line pair per mm (lp/mm) and was originally thought to limit the visibility of calcifications. Although film screen mammography can obtain 20 lp/mm, even 5 lp/mm with digital is considered adequate as digital is able to distinguish structures of similar contrast and has the ability to enhance the image once processed (21, 105).

Digital mammography has advantages over film screen mammography. Since the images are stored and retrieved electronically, consultation with distance specialists is possible. The images can be adjusted for subtle differences and there have been reports of a reduction in the number of recalls (86, 106).

A large-scale clinical trial, the Digital Mammographic Imaging Screening Trial (DMIST), reported no significant difference between digital and film screen detection for breast cancer in the screening population. The study did reveal that digital mammography was more sensitive (78% versus 51%) in women with more dense breast tissue, those under 50 years old and pre-menopausal women (104). In a later editorial, it was reported that women older than 50 years of age, postmenopausal, and those with less dense breasts had more cancers identified with film-screen mammography, though this was not statistically significant (107).

The major disadvantage of digital mammography is the cost. According to Kodak Australia, the published prices of digital machines are approximately double that of film screen at present ($118,000.00 versus $67,000.00) (Joanne Walker, Customer Relationship Manager, Kodak Health Group, personal email communication, January 2008). Other limitations of digital mammography include incompatibility of all systems, images that are more time consuming and difficult for screen readers to review with comparison to film screen mammography (61, 108, 109).
2.4.2.1.3 Computer-aided Detection

Another advantage of digital mammography is computer-aided detection (CAD). This technology involves the use of computers to bring suspicious areas on a mammogram to the attention of the screen reader after the initial review of the mammogram by the screen reader and is presently being used after in some clinical settings in the United States of America. The objective of CAD is to improve sensitivity of mammography screening which currently comes at the cost of low specificity (110). While several studies have shown that CAD use increases detection and sensitivity by up to 20% (111, 112); it was also reported to increase recall rates (110, 113, 114). Another study demonstrated that CAD use neither increased detection or recall rates (115).

There are varying sensitivity rates for marking cancers based on lesion type with the use of CAD. The use of CAD increases the detection of early stage tumours (110, 113, 114, 116). With the exception of masses in dense breast tissue, breast density does not impact overall CAD detection (117). Sensitivity for malignant micro-calcifications ranged from 86%-99%, with a drop to 57% sensitivity for malignant amorphous micro-calcifications; while sensitivity for masses ranged between 43%-85% (118). Although CAD is effective in finding calcifications and masses, studies have reported that it failed in the detection of architectural distortion (110, 119).

2.4.2.2 Ultrasound

Ultrasound of the breast uses high-frequency sound waves (7 to 10 MHz) to obtain an image of the tissues and internal structures. While not suitable as a screening tool, ultrasound has become a valuable tool in breast imaging for evaluating any lumps not visualized on a mammogram, as ultrasound can accurately visualize breast cancers as small as 3 mm (103). Ultrasound is also very useful in distinguishing between cystic and solid lesions (Figure 2-10) and for guidance in interventional procedures.

Figure 2-10: Ultrasound of (A) Simple Cyst and (B) Carcinoma (24)
An ultrasound evaluation is generally not helpful in diagnosing micro-calcifications, although on some of the newer high resolution equipment micro-calcifications may be seen, especially if located within a mass (103, 118, 120). Unlike mammography, ultrasound is not limited by breast density.

Kopans suggests that ultrasound studies are highly operator dependent with an unknown sensitivity (61); however, performance statistics in the United States for ultrasound sensitivity have demonstrated a 96% sensitivity for infiltrating ductal carcinoma (95% CI range 95%-97%) and 88% sensitivity for infiltrating lobular carcinoma (95% CI 82%-94%) (118). The American College of Radiology Imaging Network National Breast Ultrasound Trial (ACRIN 6666 - a triple blind evaluation of clinical breast examination, mammography and ultrasound) may provide more information on the sensitivity and specificity of ultrasound (61). The first round of screening in the ACRIN 6666 has consisted of more than 2,800 high risk participants (median age 55 years old and more than half of the participants having a personal history of breast cancer); ultrasound alone has shown an overall sensitivity of 50% (95% CI range 33.8%-66.2%) and an overall specificity of 91.8% (95% CI range 90.7%-92.8%) (121). Mammography, plus ultrasound in these high risk women has increased overall sensitivity to 77.5% (95% CI range 61.6%-89.2%) with a slight decrease in overall specificity to 89.4% (95% CI range 0.88.1%-90.6%) (121).

Breast diseases are not always diagnostically specific. During ultrasound of the breast, benign and malignant processes are similar and sometimes, identical. Distinguishing a benign from a malignant disease is not always possible without further intervention. Ultrasound can be used both as an evaluation tool and in diagnostic procedures, such as cyst aspiration, fine needle aspiration (FNA) or ultrasound guided core biopsies. Cyst aspirations (withdrawing the fluid from a cyst) are performed if the cyst is causing the client tenderness. FNA is the removal of fluid / tissue with a small needle for an evaluation under a microscope for cytology; whereas core is the removal of a core of tissue with a larger bore needle for evaluation in histopathology (similar to a stereotactic core biopsy).

2.4.2.3 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) creates detailed images inside the body through the use of a static magnetic field, radio frequency and a computer. These images can show the difference between normal and diseased tissue. During MRI of the breast, the client lies prone, with the breast hanging through a coil that transmit and receive the radio signal (86). The most important advantage of MRI is that contrast enhancement
makes lesions stand out from background tissues (unlike mammography and ultrasound, where the signals generated by a cancer may be similar to the background tissue, making the lesion difficult to perceive or even invisible on those tests).

Since most cancers are enhanced, MRI is quite sensitive to the presence of breast cancer. A study reported in 2004 comparing MRI, mammography, ultrasound and clinical breast examination for 236 high risk women over the age of 25 years old related sensitivity of 77% for MRI with a specificity of 95.4%; the corresponding figures were 36% and 99.8% for mammography, 33% and 95% for ultrasound and 9.1% and 99.3% for clinical examination (122). Other analysis of studies have shown high sensitivity for MRI (between 85%-100%), but suggest it lacks specificity (between 37%-97%) (118, 123, 124)

Breast MRI is not used for routine breast cancer screening, but has been reported to be valuable for clients with dense glandular tissue, which is usually associated with low mammographic sensitivity (18, 103, 125), such as young women at high risk for breast cancer. Another advantage is its use after a diagnosis of breast cancer to determine the extent of the tumour and any contra-lateral tumours (123, 126).

MRI is also considered useful for classifying suspicious lesions, some micro-calcifications and implants (for leaks or ruptures) (18, 125). MRI can visualize invasive carcinoma (123), lobular and occult cancers not visualized on mammography and ultrasound (126). Breast MRI can detect 2 – 3 mm enhancing lesions that do not correlate with either mammographic or ultrasonic findings (125).

In 2005-2006, the American Cancer Society stated that while MRI located more cancers in women who were genetically predisposed to breast cancer, it was unknown whether this difference was great enough to save additional lives (7). While MRI is not available at all facilities, its use is increasing. There have been concerns expressed; however, that some facilities performing MRI lack the ability to perform biopsies (127). Breast MRI has not yet been standardized due to the continual and rapid advances in this modality (128).

2.4.2.4 Positron Emission Tomography

The positron emission tomography (PET) scan creates computerized images of chemical changes that take place in tissue. The patient is given an injection consisting of a combination of glucose and radioactive material, usually F-18 fluro-2-deoxyglucose (FDG). The radioactive glucose can help in locating a tumour, as cancer cells absorb glucose faster than normal tissues in the body. Presently, PET is helpful in evaluating and staging recurrent disease.
Positron emission tomography breast scans may play a role in determining whether a breast mass is cancerous. A review on current literature for PET trials have shown high sensitivity of 85%-90%+ and specificity 85%-100% for axillary staging; while the detection of nodal or distant metastases sensitivity ranged from 78%-100% and specificity ranged from 20%-100% (129). However, for the detection of small cancers (less than 1 cm in size) sensitivity was 57% and for DCIS, the sensitivity was 25% (86, 118, 129). For this reason, PET would be unlikely as a screening tool for breast cancer.

2.4.2.5 Tomosynthesis

An emerging 3-D technology in breast imaging is tomosynthesis. This is comparable to a tomogram of a compressed breast with images acquired at a number of different angles. These images can be also be reconstructed into 3D images of the breasts (like CT images) either individually or in a dynamic cine mode. There are several advantages over traditional mammography such as reducing or eliminating the problems of tissue overlap and structure noise (130) and the possibility of reducing compression. These advantages may lead to fewer recalls, fewer biopsies (as more tissue is seen and not overlapping) and less pain.

2.5 SCREENING MAMMOGRAPHY

2.5.1 HISTORY

Breast screening became widely available in most developed countries in the mid 1980’s. Randomized breast trials were conducted between the 1960’s and the 1980’s; these included the Health Insurance Plan (HIP) in New York (1963-1969), Swedish Two County Trials (1978-1985), Edinburgh (1979-1988) and the Canadian Trials (1980) (71). These randomized, controlled trials demonstrated that the natural history of breast cancer can be interrupted, the individual can be cured or death from breast cancer delayed, if the malignancies are detected and treated before they have metastasized.

2.5.1.1 Australia BreastScreen

Australia began a National Program for free mammography screening in 1991, known as BreastScreen Australia (131). This program actively recruits women aged 50 – 69 and also accept 40 to 50 and over 70 year olds for screening. One of the goals set by the National Accreditation Standards for BreastScreen Australia Quality Improvement Program includes:

“To achieve agreed performance outcomes which minimize recall rates, retake films, invasive procedures, ‘false negatives’, and ‘false positives’, and maximize the number of cancers detected, particularly the number of small cancers.” (40).
The site where this current research is being undertaken, Hunter Region and Wyong Shire (HBS), began as a pilot program for mammography screening in 1989 (131).

2.5.2 ADVANTAGES
The reduction in breast cancer mortality due to screening has been reported to be between 18 – 46% (8, 31-33, 132); with most studies concluding there is good evidence that screening reduces breast cancer mortality among women aged 50-69 by about 25%. Some studies suggest there is little or no evidence of effectiveness among women aged 40-49 years (25, 133); while some authors recommend tailored guidelines for individual patients (134, 135)

Mammography can detect breast cancer as early as 2-3 years before a lump can be felt during manual examination (27). For the most accurate detection, a quality mammogram must be performed on dedicated mammography equipment by a competent, dedicated, qualified mammographer and interpreted by a competent, dedicated, qualified screen reader. Evidence from randomized trials suggests that the quality of the mammograms can affect cancer detection rates, stage at detection and interval cancer rates (66, 136).

Breast screening reduces mortality by detecting small cancers for treatment as tumour size and axillary lymph node metastases are the two most important prognostic features of breast cancer (28-30). A study conducted to assess the likelihood of systemic disease for over 550 screen detected invasive cancers reported low rates of nodal involvement and vascular invasion for both Grade 1 invasive tumours less than 20 mm in size and Grades II and III invasive tumours less than 10 mm in size (29).

2.5.3 DISADVANTAGES
There are several technical disadvantages of mammography. The most common disadvantages for clients include fear of compression, fear of being recalled and the diagnosis of breast cancer (51). Screening has also increased the diagnosis of DCIS, a breast cancer whose progression and risk is unknown (86, 137).

2.5.3.1 Fear of Compression
The fear of compression is a deterrent for some women to avoid mammography. Compression of the breast while uncomfortable, if properly performed should rarely be excruciatingly painful. One of the skills of a competent mammographer is not only to establish rapport with a potentially tense, frightened or embarrassed client, but to listen to their concerns and reposition if the client states there is too much discomfort (16, 17, 19, 77). There have been conflicting studies done on whether the pain of compression deters clients from returning for subsequent screening. Most studies indicate that while
many clients find the procedure painful, they are willing to return for another mammogram (46, 138-140). A study by Kornguth et al reported less than 15% of women complained of intense pain. The two variables that were shown to consistently predict a current painful mammographic experience included those who reported a higher level of pain at the time of their last mammogram and those who had more dense breast tissue (141).

Several strategies have been trialled in an attempt to decrease the discomfort felt by clients. These have included relaxation techniques such as listening to relaxing music (142); techniques such as decreasing the compression force while maintaining image quality (143); using breast cushions between compression paddle, bucky and breast (144, 145); and giving clients control over the compression (146). After reviewing current literature, the Cochrane Breast Cancer Group reported the only intervention suggesting a significant reduction in discomfort was that of client controlled compression (p = .0003) (146, 147). The randomly controlled study using 109 women involved producing one client and one mammographer controlled projection. Most images (93.5%) demonstrated adequate compression, with client controlled compression producing an image comparable to that produced by the mammographers, as long as the mammographer applied the first compression (146).

The current standard at Hunter BreastScreen is to give the client verbal control over the degree of compression. The mammographer asks the client to state if the compression is too painful while she is compressing the breast. Although all mammographers may react differently to this scenario, the response of the mammographer should be to release the compression paddle and re-position to see if a more comfortable position can be attained.

2.5.3.2 Fear of Being Recalled

Women are recalled when the screen reader detects an abnormality on the initial mammogram. Studies have estimated that between 6% and 10% of all women attending breast screening are recalled (34-37). A meta-analysis of world wide recall rates reveal that the USA has the highest and the Netherlands the lowest recall rates (15% versus 1.4%) (38, 39). All “abnormal” mammograms should be diagnostically followed-up, although most abnormalities will not be cancer (86). While recalls may occur in up to 10% of screening mammography examinations, it may be reassuring to clients to state that approximately 90% of recalled women do not have cancer (34).
Women who were recalled, yet given a “benign” outcome at recall (false positive recalls), reported more anxiety and concern compared to women whose original mammograms were reported as normal (34, 37, 41-43). A systematic review conducted by Brett et al assessing 54 papers from 13 countries reported a significant difference in terms of distress depending on the type of examinations performed at recall and the outcome (148). There are short term levels of anxiety / depression for clients who are diagnosed not to have cancer (37, 46, 149). The psychological distress, however, continues if clients are diagnosed with cancer (46) or if undergo FNA, surgical biopsy or 6 month follow-up (34, 50).

Encouraging studies indicate that the experience of a recall does not deter women from attending subsequent screenings (42, 44-48, 150); although other studies indicate women who are recalled are 18% less likely to attend for a subsequent screening (49, 50) and 35% less likely to attend if they had undergone a benign biopsy (49).

Research findings have reported that after false-positive recalls, there were higher numbers of patient-initiated visits to clinicians for both breast-related and non-breast related issues, including mental health services (34). This concern was noted in 10% of recalled clients by clinicians, though the numbers are probably underestimated. Women with false positive recalls initiated 3 times as many breast related visits as women with normal mammograms (34). On a positive note, however, recalled women tend to practice breast self exam (BSE) more frequently (42, 48).

As a positive, while recalled women experience high levels of affective-cognitive distress, these effects are not long lasting (37, 41, 51). Affective-cognitive distress includes any emotional concerns leading to symptoms of distress, worry, fear or feelings of vulnerability. While it is understandable that women attending a recall would experience higher levels of breast cancer worries, fears and beliefs than non-recalled women, studies have shown that these affects are not relevant 2 months following notification of a clear result (51, 151, 152).

In order to minimize anxiety in recalled women a significant amount of research has focused on ways to reduce the time between women receiving “abnormal” results and their appointment time (37, 48, 51, 153). One possible solution would be to minimize the time between receiving the recall notice and the recall appointment or informing women about the results of their mammograms immediately following the mammogram (48, 51). Several studies have undertaken immediate interpretation and work-ups (recalls) for clients (48, 153, 154). The advantages for clients include immediate results, reduced stress and convenience of recall in a single visit; however the disadvantages for the radiology practices include difficulty in scheduling, disruption of
work flow, longer waiting times for clients, overall inefficiency, decreased cost effectiveness and the inability for double interpretation (48, 153, 154).

2.5.3.3 DCIS

Over the past 30 years, mammography has detected a higher proportion of small tissue abnormalities called ductal carcinomas in situ (DCIS), abnormal cells of the milk ducts of the breast. Generally, the cells have not invaded the surrounding tissue. As some DCIS may become invasive they are usually removed surgically utilizing breast conservation techniques (80, 86). As the progression of DCIS is unknown, some consider this a limitation, as the presence of DCIS leads to intervention. Most DCIS is not clinically evident; however, the use of mammography has increased the diagnosis of this early lesion. In the 1970’s, DCIS represented fewer than 5% of the cancers diagnosed; it now represents 20-30% of cancers detected in a screening program. Mammography should not be blamed for detecting this early cancer; rather there is a need to determine the best method of treatment.

2.5.4 Accuracy

Specificity and sensitivity are two methods of predicting accuracy of a screening test. They are characteristics of the test itself and not related to the population being screened. Centralized mammography screening programs (Sweden, UK, Australia) emphasise high specificity and high sensitivity (39).

2.5.4.1 Specificity

The specificity is defined as $p(TN)$ the probability of a negative (normal) interpretation when no abnormality exists in the patient. It is known as the true-negative fraction (22).

$$p(TN) = \frac{\text{number of patients correctly identified as not having disease}}{\text{total number of patients without the disease}} = \frac{TN}{TN + FP}$$

\textbf{Equation 2-1: Specificity}

In mammography, specificity relates to the ability to understand anatomy and changes over time in a breast and to determine whether these changes are benign or malignant. As breast cancer has a low prevalence in the screening population, specificity rates may be misleading. For example, if a 1000 asymptomatic women are screened and reported as normal sight unseen, the specificity would be 99.9% as only 2 to 4 cancers
are detectable in 1000 asymptomatic women; even though all cancers were missed, the specificity is high (61). Therefore, sensitivity is the more useful measure.

To detect more true positives, a test must have a high sensitivity if the specificity is below 95%. As most screening tests do not meet this high standard of specificity, the screening must absorb the cost of false positives; in the case of screening mammography, this relates to the number of recalls. Carney et al showed that specificity significantly decreases as breast density increases and therefore additional imaging may be required (53). In practical terms, this means that for 1000 women with dense breast tissue, 100 will be recalled; while for a 1000 women with fatty tissue only 35 would be recalled (53).

Some studies have indicated a slight decrease in specificity with HRT use, although this was not clinically significant (155, 156). It does, however, indicate that different HRT regimes and durations may affect breast parenchyma differently.

2.5.4.2 Sensitivity

The sensitivity of an imaging procedure is defined as $p(\text{TP})$, the probability of detecting an abnormality when it is present in the patient by examining the images provided by the procedure. It is also known as the true positive fraction (22).

$$p(\text{TP}) = \frac{\text{number of patients correctly identified as having disease}}{\text{total number of patients with the disease}} = \frac{\text{TP}}{\text{TP + FN}}$$

Equation 2-2: Sensitivity

Sensitivity depends on the nature of the breast lesion, radiographic density, general composition of the breast tissue, use of HRT, location of the abnormality within the breast, technical quality of mammogram and the screen reader’s expertise in interpreting the imaging appearances (157). Mammography sensitivity is reported to be between 68–92% (9, 10, 12-14), with double reading of screening mammography increasing sensitivity by 15% (158, 159).

2.5.4.2.1 Effect of Breast Density on Sensitivity

Clients with dense breast tissue have a 4–6 times increased risk of getting breast cancer (56-58). Although a higher percentage of young women do have dense breast tissue, this phenomenon can occur at any age. There is also no abrupt breast tissue density change as this occurs gradually over time.
Breast density affects mammography performance (23, 36, 53-55, 59, 65, 160) as increased breast density may obscure findings related to malignancy (23, 160). Dense breast tissue is associated with poor mammography sensitivity (54, 161, 162) and higher recall rates than for clients with fatty breast tissue (36, 54, 55).

In a study using the BI-RADS density scale, Carney et al demonstrated a 43% increase (CI 1.33-1.55) of needing additional films (being recalled) for dense versus fatty breast tissue (36). Yankaskas et al also found that with decreasing breast density, the recall rate decreased from nearly 7% (6.8%) recalls in extremely dense to 2.4% recalls in the almost fatty density group (55). Breast density is affected by age, use of HRT, menstrual cycle, BMI and familial or genetic tendency.

2.5.4.2.2 Effect of Age on Sensitivity

Sensitivity usually increases with the age of the client; this is due in part to decreased breast density, but may also be due to older women having slow growing, less aggressive tumours that tend to be detected at screening (163).

2.5.4.2.3 Effect of HRT Use on Sensitivity

The modern use of HRT can increase breast tissue density by arresting and often reversing the process of involution (24, 60, 61). Warren’s summary of papers reviewing HRT and breast density reported this increase can occur in up to 25% of women (164). For women exposed to HRT, one study found no evidence of increased breast cancer (165); while other studies have shown that there is an increased probability of subsequent interval cancers (60, 156, 166, 167). An interval cancer is an invasive breast cancer that is diagnosed after a screening episode that detected no cancer and before the next scheduled screening episode, usually one or two years (6).

There has been conflicting research how the use of HRT influences sensitivity of mammography. While some research has shown no difference in sensitivity (155, 156), other studies have shown differences (9, 53, 60, 167-169). Kavanagh et al conducted a study at Victoria BreastScreen examining sensitivity, specificity and small cancer detection according to HRT use in over 100,000 women starting in 1994 (168). This study reported sensitivity was 12.5% lower in HRT users versus non-users (64.8% [95% CI 58%-72%] versus 77.3% [95% CI 74%-81%]) (168); in other words, an additional 20% more cancers would have been detected in the HRT users if sensitivity had been the same as for the non-users of HRT (168). The study also demonstrated that for women without an interval detected cancer, the risk of a recall was higher among HRT users compared with non-users after controlling for potential confounders.
(adjusted OR 1.12 [CI 1.05-1.19], p=0.0004) (168). Other research studies have also reported higher recall rates for HRT users (13, 36, 162, 170, 171).

A review for BreastScreen Victoria reported that between 1994-2001, there was a slight increase among HRT users of mammograms repeated for technical faults or classed as suboptimal (161). As HRT use is associated with increased breast pain and tenderness, this may affect the quality of the image. A study conducted by Nimmo et al hypothesised that women on HRT would find mammograms more painful, tolerate less compression and therefore, be more likely to be recalled for technical reasons (62). Their results indicated no statistically significant difference between HRT and non-HRT users (p = 0.80); although compressed breast thickness had both a statistical and clinical significant relationship with technical recall (62). As blurring of the image was the greatest single reason for technical recall; this could possibly be independently related to compressed breast thickness due to inadequate compression or long exposure times (62).

### 2.5.4.2.4 Interval Cancer Rates and Sensitivity

Another indicator of sensitivity of a mammogram is to measure the interval cancers (25, 68), an indicator of high quality screening is a low interval cancer rate. Taplin et al reported that incorrect breast positioning was the primary reason for not finding interval detected invasive breast cancers, after adjusting for age, film date (year of examination) and breast density (68). Taplin’s study showed that clients with either improper breast positioning or poor overall image quality had subsequent higher interval cancer rates (68). Sensitivity was highest (84%) among clients with proper breast positioning, but when images failed this measure, sensitivity fell to 66.3% (68). The study concluded that invasive breast cancer detection by mammography might be improved through increased attention to correct positioning (68).

### 2.5.5 Mammography Screening Guidelines

Kavanagh et al suggests it is more important to use sensitivity rather than interval cancer rates to describe accuracy of mammography screening as it is independent of the underlying rate of breast cancer in the screened population (168). Indicators of a high quality screening program include a sensitivity of 85% or more, screen detected cancers that are minimal in size (15 mm or less (6)) with less than 25% node involvement (172) and that the detection rate be within a certain standard for both prevalence (1st) and subsequent (2nd) screens (6). The UK NHSBSP advocates finding 3.6 cancers per 1000 prevalent screens and 4 per 1000 subsequent screens (173).
2.5.5.1 BreastScreen Australia Guidelines

To ensure clients receive a high quality mammogram, BreastScreen Australia adheres to mandatory quality assurance (QA) programs. Although technical, QA programs are based on client care and consumer protection. QA programs are designed to examine the mammography equipment, film and processing quality, the expertise of those obtaining and reading the images, recall standards for clients and the physicists involved with testing the equipment.

As part of the QA program, BSA have performance objectives in National Accreditation Standards (NAS) agreed by the Department of Health and Ageing and BreastScreen Australia state and territory programs for individual screening services. Table 2-5 demonstrates NAS guidelines; while Table 2-6 demonstrates BreastScreen Australia statistics with relation to the NAS guidelines for the years concurrent with this research.

2.6 Film / Screen Mammography

Film screen mammography visualizes the internal structures of the breast by using low energy radiation (x-rays) in the 17 to 24 keV range (21, 63), breast compression and film screen combinations to obtain an image. Although mammography has the capability to visualize cancers as small as 3-5 mm (19, 174); there are occult cancers such as lobular cancers not shown on mammograms (24, 53). It is estimated that 10% of lesions are not shown on mammograms (63). The challenge for mammography is to observe subtle differences due to the similarity of the tissue types while retaining detail and reducing dose (21, 63).

2.6.1 History and Development

X-rays were discovered in 1895. By 1913 Albert Salomon, a surgeon at the University of Berlin, had made observations of 3000 mastectomy specimens with the use of x-rays (71). In the 1920’s there were several groups that thought there was potential in breast x-rays; however, the results were disappointing due to the inferior quality of the radiograph (19, 25, 71). The lack of adequate equipment and adequate technique delayed the development of mammography at this time.

One notable pioneer, Stafford Warren (Rochester, New York), developed a stereoscopic technique for breast radiography in 1930. Also by 1930, Raul Leborgue (Uruguay) had written his classic textbook on categorizing calcifications and stressed producing high quality images. It was not until 1956 that Robert Egan introduced a standard mammographic technique and became the “father of mammography” (71).
When Egan began his work, the technical quality of the radiographs was a secondary consideration. It was apparent, however, that the diagnosis of breast lesions was directly proportional to the quality of the radiographs and unless improvement commenced, the procedure was doomed (71). Egan spent several years of experimentation with various technical factors that included an applied voltage range from 12 to 22 kVp, various intensifying screens, focal spot sizes from 0.3 mm to 2 cm, anode receptor distances of 22 cm to 1.8 metres, exposure times of 1/20 to 20 seconds, stationary grids, use of approximately 75 film emulsions, numerous positions, stereoscopy, compression and floating the breast in liquid (19).

While x-rays for mammorams are produced in the same manner as conventional x-rays, conventional techniques are useless due to the fact that mass density and atomic number of soft tissue components of the breast are so similar. Egan eventually used a conventional diagnostic x-ray unit with tungsten target and filtration limited to the inherent filtration of the tube (1 mm Al). The focal film distance was kept above 18" to prevent geometric unsharpness and the breasts remained uncompressed. The technique employed low kVp, high mAs values and resulted in high skin doses (between 40 and 160 mGy per examination) compared to less than 10 mGy achievable today, yet produced the finest contrast and detail and were considered safe for that time (15, 77). Egan’s technique was easily reproducible, which led to en masse training of radiologists and radiographers. Although the method has been refined and improved, the same principle with the addition of breast compression is still used today. In order to gain popularity, the name changed from “breast roentgenology” to “mammography” (77).

According to Egan, mammographic technique is 90-95% of the procedure and interpretation 5-10% (71), so that quality is paramount and mandatory to the examination. This begins in the specialization of the equipment with the ability to produce a quality image and culminates with optimal positioning to allow maximum breast tissue coverage.

By the mid 1960’s, Charles Gros, MD and the CGR Company had developed the first dedicated mammography unit. The new machine introduced commercially in the US in 1969 used a molybdenum target, had an integral device for compressing the breast and employed low kVp’s. Other equipment companies were constantly refining and improving the mammography unit. Table 2-7 shows some of the major advances of mammography equipment, while Figure 2-11 shows a mammography machine and its major components.
Table 2-5: National Accreditation Standards Guidelines (84, 88)

<table>
<thead>
<tr>
<th>National Accreditation Standards Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensitivity - No NAS guidelines</strong></td>
</tr>
<tr>
<td><strong>Recall to assessment – NAS guidelines</strong></td>
</tr>
<tr>
<td>&lt;10% of women aged 50–69 years who attend for their first screen are recalled for assessment</td>
</tr>
<tr>
<td>&lt;5% of women aged 50–69 years who attend for their second or subsequent screen are recalled for assessment</td>
</tr>
<tr>
<td><strong>Interval Cancers – NAS Guidelines</strong></td>
</tr>
<tr>
<td>≤7.5 interval cancers per 10,000 women aged 50–69 years who attend for screening less than 12 months following a negative screening episode</td>
</tr>
<tr>
<td><strong>Detection of small invasive cancers (≤15mm) – NAS guidelines</strong></td>
</tr>
<tr>
<td>≥25 per 10,000 women aged 50–69 years who attend for screening are diagnosed with small (≤15 mm) invasive breast cancer</td>
</tr>
<tr>
<td><strong>Detection of DCIS – NAS guidelines</strong></td>
</tr>
<tr>
<td>≥12 per 10,000 women aged 50–69 years who attend for their first screen are diagnosed with DCIS.</td>
</tr>
<tr>
<td>≥7 per 10,000 women aged 50–69 years who attend for their second or subsequent screen are diagnosed with DCIS</td>
</tr>
<tr>
<td><strong>Rescreen requirements – NAS guidelines</strong></td>
</tr>
<tr>
<td>≥75% of women aged 50–67 years who attend for their first screening round within the program are rescreened within 27 months.</td>
</tr>
<tr>
<td>≥90% of women aged 50–67 years who attend for their second and subsequent screen are rescreened within 27 months of their previous screening episode</td>
</tr>
</tbody>
</table>

According to AIHW, from 1999-2001, for women 40+, 6.7 invasive cancers per 1000 prevalent and 3.93 invasive cancers per 1000 subsequent were detected
Table 2-6: BreastScreen Australia Statistics (84, 88)

<table>
<thead>
<tr>
<th>Table 2-6: BreastScreen Australia Statistics (84, 88)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BreastScreen Australia Statistics</strong></td>
</tr>
<tr>
<td><strong>Round</strong></td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
</tr>
<tr>
<td>24 months after 1st screen</td>
</tr>
<tr>
<td>24 months after 1st screen</td>
</tr>
<tr>
<td>24 months after subsequent screen</td>
</tr>
<tr>
<td>24 months after subsequent screen</td>
</tr>
<tr>
<td><strong>Recall to Assessment</strong></td>
</tr>
<tr>
<td>First</td>
</tr>
<tr>
<td>Subsequent</td>
</tr>
<tr>
<td>**Interval Cancers *</td>
</tr>
<tr>
<td>0 – 12 months after 1st screen</td>
</tr>
<tr>
<td>0 – 12 months after subsequent screen</td>
</tr>
<tr>
<td>**Invasive Small Diameter (&lt;15 mm) *</td>
</tr>
<tr>
<td>First</td>
</tr>
<tr>
<td>Subsequent</td>
</tr>
<tr>
<td>**DCIS *</td>
</tr>
<tr>
<td>First</td>
</tr>
<tr>
<td>Subsequent</td>
</tr>
<tr>
<td><strong>Rescreen Rate</strong></td>
</tr>
<tr>
<td>First</td>
</tr>
<tr>
<td>Second</td>
</tr>
<tr>
<td>Subsequent</td>
</tr>
</tbody>
</table>

* = number detected per 10,000 women

All statistics relate to the target age group (50-69), excluding the rescreen rate (50-67)

Table 2-7: Advances in Mammography (adapted from information (19, 21, 25, 52, 71, 175))

<table>
<thead>
<tr>
<th>Year</th>
<th>Device</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior 1969</td>
<td>Conventional tungsten x-ray tube with direct exposure industrial film was used (increasing dose)</td>
<td></td>
</tr>
<tr>
<td>1969</td>
<td>Dedicated mammography unit with molybdenum tube, low kVp’s, an integral compression device and C-arm introduced</td>
<td>CGR Senographe</td>
</tr>
<tr>
<td>1971</td>
<td>Xeroradiographic technique developed. Xerography was phased out eventually as doses were higher than for new film screen combinations developed in the early 1980’s</td>
<td>Xerox</td>
</tr>
<tr>
<td>1972</td>
<td>Screen–film system developed (a mammography cassette with a single intensifying screen and single emulsion film)</td>
<td>RE Wayrynen, Dupont</td>
</tr>
<tr>
<td>1976</td>
<td>Rare earth screen-film system and special cassette</td>
<td>Kodak Min-R</td>
</tr>
<tr>
<td>1977</td>
<td>Magnification with micro-focal spot introduced</td>
<td>Radiological Sciences Inc</td>
</tr>
<tr>
<td>1978</td>
<td>A moving grid introduced with special carbon fiber interspace material for low absorption of the soft x-ray beam necessary for screen film mammography</td>
<td>Philips</td>
</tr>
<tr>
<td>1980</td>
<td>First motorized compression device used</td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td>High frequency / constant potential generation of dedicated mammography machines introduced</td>
<td>LoRad Medical Systems</td>
</tr>
<tr>
<td>1992</td>
<td>First multi-anode tube (tungsten and molybdenum) &lt;br&gt; Another dual target tube (molybdenum and rhodium). Rhodium enables better penetration of dense breast tissue with less radiation dose to client</td>
<td>Siemens, General Electric</td>
</tr>
<tr>
<td>1998</td>
<td>FDA approves use of computer-aided diagnostic systems (CAD)</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>FDA approves use of digital mammography</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2-11: (A) Mammography Unit (176), (B) Schematic Drawing (177)
2.6.2 Obtaining the Mammographic Image

To be an asset in screening, mammographers must be interested, reliable and enthusiastic about breast imaging. They must be well trained, with defined responsibilities and need patience, compassion and sensitivity to make the client more comfortable. They need to establish rapport quickly to gain a client’s cooperation to position her correctly (16, 17, 19). They need to be willing to continue their education, as breast imaging is a rapidly changing field.

When a client attends for her screening, both verbal and non-verbal communication must be effective and efficient, as it is the mammographer’s responsibility to obtain:-

- the history of the client
- explain the procedure
- answer questions
- visually inspect the clients breasts
- position the breasts
- obtain quality images and
- review those images.

This all needs to be completed in a 10-15 minute time frame.

The history of the client includes personal and family medical history. Figure 2-14 illustrates one type of questionnaire (used by Hunter BreastScreen). After completion of the questionnaire, the mammographer visually inspects the client’s breasts noting anything that may affect / appear on the image and meticulously marking these on the diagram. These markings may include scars, skin lesion, sebaceous cysts, tattoos and piercings. Records should also include any nipple changes, breast symptoms and note if one breast is larger than the other. Any issues should be documented on the questionnaire.

2.6.2.1 Standard Projections

The two standard projections taken include the cranio-caudal (CC) and the medio-lateral oblique (MLO) (Figure 2-12). Basically the CC involves compressing the breast from the superior through to the inferior portion of the breast and the MLO involves compressing the breast from the medial aspect to the lateral aspect (18). The maximum amount of tissue must be demonstrated in each view (19, 174). If all the breast tissue has not been visualized (i.e infra-mammary angle is not visualized) or more information is required, a true lateral must be taken. The most mobile portions of the breast are the lateral border of the breast and the infra-mammary
crease (25). Compression of the breast is most effective and least uncomfortable when applied in these areas.

The CC provides better visualization of the medial aspect of the breast and better image detail because greater compression of the breast is usually possible (25, 174). The breast should be pulled straight forward and not exaggerated laterally (as this is the only view showing the medial portion of the breast) and the breast tissue is well compressed. The difference between the posterior nipple line (PNL) measurement on the medial lateral oblique and cranio-caudal views should be within 1 cm.

The MLO is the most effective single view as it includes the greatest amount of tissue and is the only whole breast view to include all of the upper quadrant and axillary tail (25). The structure of the pectoralis major is important for imaging. The MLO is positioned so that the plane of compression is parallel to the oblique fibres of the free margin of the muscle as it extends from the humerus. This permits maximum traction on the breast so that it can be fully positioned over the detector and completely compressed (61). The breast will fall or “sag” if not held by the mammographer until adequate compression is applied (16, 17, 52). Findings on the MLO that indicate proper positioning include visualization of the pectoralis muscle to the level of the nipple, a convex appearance of the pectoralis major muscle, complete visualization of posterior breast tissue, an open infra-mammary fold, well compressed, “non-saggy”, pulled up and out orientation (40, 178).

![Cranial-Caudal (CC) and Oblique (Obl)](image)

*Figure 2-12: Standard Mammography Positions (179)*
2.6.2.2 Applying Compression

The primary factors controlled by the mammographer are positioning of the client, including applying compression to the breast. The amount of breast compression applied during mammography influences image quality and the discomfort experienced by the client. Although there is the negative aspect that compression causes a degree of discomfort, compression has many positives in mammography that should be explained and emphasized to the client. Many departments have signs posted for client education (Figure 2-13). Compression decreases motion unsharpness, decreases geometric unsharpness, increases contrast (by reducing scatter), separates superimposed areas of glandular tissue, reduces radiation dose and helps produce a more uniform film density (21, 63).

To obtain maximum compression, the client should be advised of the importance and the benefits of being relaxed. This can be achieved by informing the client before the initiation of compression, applying compression gradually, heeding input from the client regarding tolerance and increasing compression until breast thickness is minimized (180). Traditionally, this has been until the breast is taut, like a firm tomato, but not painful to the client (174). After completion of the mammogram, the images should be reviewed for motion, positioning, technique and film screen contact.

Figure 2-13: Compression Sign at HBS (BreastScreen NSW Hunter Region & Wyong Shire; 2000)
BreastScreen NSW

REGISTRATION FORM
(To be Completed for All Women)

Ms
Miss
Mrs
Maiden Name

ADDRESS

PHONE (Home) (Work)

Date of Birth

Language spoken at home

Aboriginal (If Yes, please tick)

GP IDENTIFICATION
Dr

Address

Phone

Reason for Mammogram
(Please tick one only)

Routine Screening
Diagnostic Referral
X-Ray Request Only

Clinical History
(Please tick if applicable)

Hysterectomy
Ever taken BCP or hormones
Previous mammography

If YES, where

Family History
(Please tick if applicable)

Any cancer (includes breast cancer)
Breast cancer (1st degree relatives only)

Any Breast Symptoms
(If Yes, please tick)

Right Left

Lump
Nipple discharge
Any persistent worrying pain
Other abnormality

Past Breast Procedures
(If Yes, please tick)

Right Left

Breast biopsy/needle cytology
Mastectomy
Breast prosthesis (surgical implant)
Surgical reconstruction (no surgical implant)
Other (specify) (include previous breast cancer)

Radiographer: Total Films:
Receptionist: Inadequate Films:

Fig. 2-14: Client Questionnaire for Hunter BreastScreen

Nov 2007

48
2.6.3 **Breast Anatomy Appearance**

In conjunction with use of the specialized mammography equipment, an understanding of breast anatomy is relevant in order to obtain quality images by the mammographer. Anatomy generally “seen” on the mammogram includes the nipple, adipose tissue, glandular tissue, blood vessels, lymph nodes, pectoral muscle and Coopers ligaments (25, 71) (Figure 2-15).

![Figure 2-15: A 45º Oblique Mammography Projection Demonstrating Anatomical Structures that are Visualized with Mammography (52)](image)

A. Pectoral muscle; B. Nipple; C. Fat; D. Glandular tissue; E. Blood vessel; F. Lymph nodes; G. Cooper’s ligaments

The mammographic image is a reflection of breast anatomy and its occasional alteration by pathological processes (24). Tabar proposes that diverse breast tissue is reflected on a mammogram by a mixture of four structural components or building blocks. These four components are: nodular densities, linear densities, homogenous structureless densities and radiolucent areas. The corresponding breast anatomy includes the terminal ductal lobular units, the lactiferous ducts / vessels / fibrous strands, the supporting fibrous connective tissue and adipose tissue (fat) (24, 25).
Five patterns of density are represented by using Tabar’s building blocks (24). Descriptions and images in Figure 2-16 demonstrate how Tabar’s building blocks relate to BI-RADS and Wolfe’s classifications (Table 2-1).

Pattern I is the most common pre-menopausal breast tissue where all four building blocks are fairly represented on the mammogram. This corresponds to “mixed” tissue. Pattern II and III are represented as the end results of the process of involution. Images are dominated by radiolucent adipose tissue and linear densities, making the detection of abnormalities relatively easy. More than 70% of postmenopausal, asymptomatic women have Pattern II or III. This corresponds to “fatty” breast tissue. Pattern III is similar to II with the exception of retro-areolar prominent ducts.

Pattern IV is dominated by prominent nodular and linear densities, making the perception of pathological lesions difficult. This occurs in 10 – 12% of screening women and is “dense” breast tissue. Pattern V is dominated by extensive fibrosis. The overwhelming dominance of this homogenous, structureless fibrous tissue limits the capabilities of mammography to demonstrate the details of normal anatomy and to reveal small pathological lesions. This corresponds to “very dense” breast tissue and affects 6% of asymptomatic women.

2.6.4 TECHNICAL ASSESSMENT

After completion of the mammogram, generally the mammographer taking the images assesses them for technical quality. During the initial trials in the 1980’s, the National Breast Screening Study of Canada made a decision that the quality of mammography was not to be of major importance (61). Centres were permitted to use any mammography equipment available without training either mammographers or screen readers; this resulted in a trial that failed to show any benefit of breast screening. An external review reported that during the trial more than 50% of the mammograms were considered poor or unacceptable due to poor positioning, poor contrast and image unsharpness (61). This decision reiterated the need for quality screening mammograms in all aspects.
Figure 2-16: Patterns of Breast Density per BI-RADS, Wolfe & Tabar (24)

In 1997, the Physician Insurers Association of American (PIAA) evaluated data regarding 472 claims and lawsuits covering screening and diagnostic mammography, adult barium enema examinations, obstetrical ultrasound and results communications (181). From these claims, diagnostic errors were reported in 62% of screening mammography, 54% of diagnostic mammography and 50% of adult barium enema examinations (181). Ten percent to 24% of diagnostic errors per modality were associated with poor image quality (181). These almost always adversely affected the clinical outcome in screening mammography and obstetrical ultrasound clients (181). By 2002, the PIAA had reported a total of 450 paid cases that involved a delay in the diagnosis of breast cancer (an 9% rise from the 1995 data); although more than 68% of the women represented in this study were under the age of 50 (182). Based on the findings from this latest report, one of the suggestions for improving patient care and minimizing malpractice losses states that if a mammogram results in a film of poor technical quality, the study needs to be repeated (182).
2.6.4.1 Methods

Mammography images must cover all the breast tissue, have the correct film identification (including date of exam, client identification, side markers, positional markers and mammographer identification), correct exposure, good compression, absence of motion, correct processing, absence of artefacts, no skin folds and the images must be symmetrical. These requirements are essential for a quality mammogram and mammographers worldwide assess images by similar standards. Two systems for assessing technical quality in mammography include the method used in the United States under the guidelines of the American College of Radiology (178) and the method used in the United Kingdom and Australia, known as PGMI (40).

2.6.4.1.1 American College of Radiology

One method of assessing mammograms is the criteria devised by the American College of Radiology for quality mammography. These guidelines include:

- The length of the posterior nipple line (PNL) on the CC view must be within 1 cm of its length on the MLO.
- Any film that exhibits a non-essential image (i.e. chin, shoulder, earring, etc.) in or over the breast or adjoining tissue will FAIL the image review process.
- Exposure: There must be adequate penetration of all breast tissue.

Separate requirements for the projections include:

CC:
- The breasts should be symmetrically located on the film.
- Nipple should be midline.

MLO:
- The inferior extent of the pectoralis muscle should be visible to the PNL.
- The breasts should be symmetrically located on the film.
- The breast should be pulled out and up with an open infra-mammary fold and the nipple parallel to the floor.
- The nipple should be in profile.
- The pectoralis muscle should have a convex or straight anterior margin.
All clinical images must be permanently labelled:

- For laterality and projection the marker should be placed near the aspect of the breast closest to the axilla.
- Facility name, city, state, zip code, patient name, patient ID number and/or social security number, patient date of birth, technologists (mammographers) initials, cassettes/screen identification, unit identification (if more than one mammography machine) and date of examination; this label should be placed near the edge of the film (178).

2.6.4.1.2 PGMI

Another method of assessing and grading mammograms is the PGMI system, devised by the United Kingdom Mammography Trainers Group, with the support of the College of Radiographers and The Society of Radiographers, UK (40, 183). This method categorizes mammograms as ‘Perfect’, ‘Good’, ‘Moderate’ or ‘Inadequate’. Criteria for PGMI are shown in Table 2-8. BreastScreen Australia uses this system to assess image quality.

2.6.4.1.2.1 BreastScreen Australia Requirements

The NAS guidelines for accreditation for BreastScreen Australia state that a minimum of 50% of a random audit of 50 films sets per mammographer should be graded P and G categories with 75% being desirable (40).

The NAS guidelines for BS Australia for the overall repeat rate is less than 3% of all screening films for each screening and assessment service (40), as repeat mammograms increase anxiety, discomfort, screening costs and radiation dose for the client.
Table 2-8: PGMI Evaluation of Clinical Image Quality (NAS, 2001) (40)

PGMI EVALUATION OF CLINICAL IMAGE QUALITY

Criteria for image assessment

1. All breast tissue imaged (fat visualized posterior to glandular tissue)
2. Correct film identification clearly shown:
   • date of examination
   • client identification – name and (number and/or date of birth)
   • side markers
   • positional markers
   • radiographer identification
3. Correct exposure according to workplace requirements
4. Good compression
5. Absence of motion
6. Correct processing
7. Absence of artefacts
8. No skin folds
9. Symmetrical images

Specific positioning criteria

[Diagram showing different views: CC, MLO, Rcc, Lcc, PNL, Rmlo, Lmlo]
**P = Perfect Images**

**Cranio-caudal view (CC)**
1. All breast tissue imaged
   - medial border well demonstrated
   - nipple in profile (retro-areolar tissue well separated)
   - nipple in midline of imaged breast
   - posterior nipple line (PNL) within 1 cm of PNL on MLO view

**Medio-lateral oblique view (MLO)**
1. All breast tissue imaged
   - pectoral muscle shadow to nipple level
   - full width of pectoral muscle
   - nipple in profile (retro-areolar tissue well separated)
   - infra-mammary fold well demonstrated
   - PNL within 1 cm of PNL on CC view
   * Both CC and MLO images meet criteria for image assessment 1-9

**G = Good Images**

**Cranio-caudal view (CC)**
1. All breast tissue imaged *
   - all postero-medial tissue visualized (* axillary portion of breast not to be included at expense of medial portion)
   - nipple in profile
   - nipple in midline of imaged breast

**Medio-lateral oblique view (MLO)**
1. All breast tissue imaged
2. Pectoral muscle well demonstrated
3. Nipple in profile
4. Infra-mammary fold (IMF) well demonstrated
2 – 6. Both CC and MLO images meet criteria for image assessment 2-6 inclusive for categorization as G
7 – 9. Both CC and MLO images displaying minor degrees of variation in criteria for imaging assessment 7, 8 and 9 will be accepted for categorization as G

**M = Moderate Images**

(acceptable for diagnostic purposes)

**Cranio-caudal view (CC)**
1. Most breast tissue imaged (however, all breast tissue must be imaged on MLO film)
   - Nipple not in profile but clearly distinguishable from retro-areolar tissue (however, nipple must be in profile on MLO film)
   - Nipple not in midline (significant bias)

**Medio-lateral oblique view (MLO)**
1. Most breast tissue imaged
   - Pectoral muscle not to nipple level but posterior breast tissue adequately shown
   - Nipple not in profile but clearly distinguishable from retro-areolar tissue (however, nipple must be in profile on CC film)
   - IMF not clearly demonstrated but breast tissue adequately shown
2. Correct (ed) film identification
3. Correct exposure
4. Adequate compression
5. Absence of movement
6. Correct processing
7. Artefacts which do not obscure the image
8. Skin folds which do not obscure the breast tissue
9. Asymmetrical images

**I = Inadequate Images**

(applies to both CC and MLO images)
1. Significant part of the breast not imaged
2. Incomplete or incorrect identification
3. Incorrect exposure
4. Inadequate compression which hinders diagnosis
5. Blurred image
6. Incorrect processing
7. Overlying artefacts
8. Skin folds which obscure the image

**Recommended Standard:**
A minimum of 50% of an audit of 50 randomly selected film sets should be graded in the P or G categories (75% desirable).

**Repeat Rate:**
<3% of consecutive images to be classified 'Inadequate'.
2.6.4.2 Image Quality Characteristics

Image quality is the measure of the imaging device to faithfully record each point in the object as a point in the final image. Image quality can be divided into photographic and geometric aspects. Photographic factors include those that contribute to the production (information) of the image - density and contrast; while geometric factors relate to clarity of the image and are influenced by resolution, distortion, noise and artefacts (21, 22, 63, 174, 184). All these factors interact and affect image quality and will be discussed in relation to mammography equipment and operator performance (Figure 2-17).

![Image Quality Parameters Diagram]

Figure 2-17: Image Quality Parameters

While technical quality of mammography is associated with the physics of the x-ray imaging process and with performance of the examination by the mammographer; both of these factors can be influenced by breast density, breast composition, cooperation and mobility of the client.

2.6.4.2.1 Optical Density

Optical density (OD) is defined as the degree of blackening on the mammogram. Adequate optimal density is necessary because, if either insufficient or excessive, inform may be lost. In 2008 for BreastScreen Australia, the optimal optical density used in mammography is between 1.6–2.0. Due to studies conducted in the United Kingdom, the optical density has increased progressively from 1.30 (during 1991/92) to 1.57 (during 1994/95) to 1.4-1.8 (during 1995/96) in the National Health Service.
Breast Screen Programme (185). The recommendation in 1997 stated that the detection rate of small invasive cancers would be improved by 50% by using a film density between 1.4 – 1.8 (Table 2-9) (185).

Table 2-9: Detection of Small Cancer Detection Rate on Optical Density (186)

<table>
<thead>
<tr>
<th>Optical Density</th>
<th>Small Cancer Detection Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80-0.90</td>
<td>0.13±0.02</td>
</tr>
<tr>
<td>1.00-1.19</td>
<td>0.11±0.01</td>
</tr>
<tr>
<td>1.20-1.39</td>
<td>0.16±0.01</td>
</tr>
<tr>
<td>1.40-1.59</td>
<td>0.17±0.01</td>
</tr>
<tr>
<td>1.60-1.79</td>
<td>0.18±0.02</td>
</tr>
<tr>
<td>1.80-1.99</td>
<td>0.20±0.02</td>
</tr>
</tbody>
</table>

Data from the United Kingdom Breast Screening Programme

Optical density is primarily controlled by mAs, but also influenced by kVp, distance, film-screen combinations and processing. While increasing kVp does increase density; this is an unreliable factor, especially in mammography where low kVp’s are used. Also, while source to image receptor distance (SID) can influence density in conventional radiography, SID is fixed in mammography (between 60 to 80 cm). Methods of obtaining consistent density have included use of the anode heel effect to optimize breast positioning, use of an automatic exposure control device (AEC) and dedicated processing.

The radiation intensity on the cathode side of the x-ray field is higher than on the anode side as a result of the line focus principle and this is known as the anode heel effect. This is due to the interaction of the electrons and the absorption of x-rays in the “heel” of the target. During a mammography procedure, the cathode is against the chest wall with the x-ray tube tilted for greater penetration for tissue at the chest wall and the nipple is near the anode side of the tube (21). Figure 2-18 shows a mammography tube and some of the specialized components used in this modality.
Dedicated mammography units have an automatic exposure control (AEC) (Figure 2-19) - a device that controls the amount of exposure to result in the desired optical density. Radiation that is transmitted through an object (the breast) is converted into an electronic signal, which terminates the exposure when the predetermined level of radiation has been reached.
To be effective, the photo-timer cell must be positioned over the most dense part of the breast tissue, otherwise the image may be over or underexposed (174) (Figure 2-20). This device eliminates any guesswork about the correct exposure factors as long as it is correctly positioned. Manufacturers now provide photo-timers that select anode, filter and kVp based on breast thickness or pre-sampled exposure of the compressed breast.

![Figure 2-20: (A) Compression Plate with AEC Chambers (B) Showing Correct Position on Breast Tissue (16)](image)

Optical density is dependent on developer temperature and dwell time. Conventional or 90 second automatic film processors operate at a developer temperature in the range 34°C to 37°C with a dwell time of 20-25 seconds. Under these circumstances, the development of mammography film may not be completed. By adopting an extended cycle processing in mammography, with a temperature of 33°C and the dwell time doubled the film density is increased. This processing method has the advantage of producing images with greater contrast at a lower dose for clients. With single emulsion film in an extended processor, contrast is increased by 15% and dose reduced by 30% (63, 174).

There has been debate in recent years about the advantages / disadvantages of dedicated mammography processors; as even in large mammography screening centres the total area used in the processor is quite small compared to general radiography. With the increasing use of digital mammography, processing will no longer be an issue.
### 2.6.4.2.2 Contrast

Contrast is the degree of density difference between adjacent areas on a radiograph and an important factor affecting the ability to visualise minute detail and distinguish subtle differences on a mammogram. The assumption is that any factor affecting optical density and contrast affects visibility of image detail. Radiographic contrast depends on subject contrast, film contrast, fog and scatter (21).

The structures of the breast influence the contrast of a mammogram. In mammography, images result from the difference between the x-rays absorbed by the photoelectric interaction and those that pass through the breast as image forming radiation. Attenuation is the reduction of the x-ray beam intensity as it penetrates through tissue. Differential absorption and attenuation depend on three factors – the atomic number of atoms in the tissue, the mass density of the atoms in tissue and the x-ray energy (21, 64).

Nearly all image contrast on a mammogram image is the result of the difference in the attenuation coefficients of fat and water. Glandular and connective tissue of the breasts are similar to water density structures. The maximum difference between fat and water attenuation occurs in the low kVp range (21, 64, 71). With increased kVp, attenuation coefficients of fat and water become similar and the natural contrast between these two substances decrease. Adipose tissue has a lower density and higher x-ray transmission than fibro-glandular tissue but the difference in transmission between carcinoma and glandular tissue can be small (Figure 2-21).

![Diagram of Linear Attenuation Comparing Tissue Type and Carcinoma](image)

*Figure 2-21: Linear Attenuation Comparing Tissue Type and Carcinoma (189)*
For breast imaging, the use of low peak kilovoltage and small amounts of beam filtration is essential to enhance tissues and lesions that are similar in atomic composition and physical density (Table 2-10). Mammography equipment has been designed to accommodate these similarities.

Table 2-10: Density and Linear Attenuation Coefficient for Mammary Tissue and Lesion Type (190)

<table>
<thead>
<tr>
<th>Breast Tissue Type</th>
<th>Density</th>
<th>Linear Attenuation Coefficient at 20 keV (cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammary Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibro-glandular (dense)</td>
<td>1.035</td>
<td>0.80</td>
</tr>
<tr>
<td>Adipose (fatty)</td>
<td>0.93</td>
<td>0.45</td>
</tr>
<tr>
<td>Mixed</td>
<td>0.98</td>
<td>0.62</td>
</tr>
<tr>
<td>Lesions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carcinoma</td>
<td>1.045</td>
<td>0.85</td>
</tr>
<tr>
<td>Calcifications</td>
<td>2.2</td>
<td>12.5</td>
</tr>
<tr>
<td>Water</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Fat</td>
<td>0.91</td>
<td></td>
</tr>
</tbody>
</table>

The x-ray energies most useful for enhancing differential absorption in breast tissue (maximizing radiographic contrast) are those in the 17 to 24 keV range (21, 22). These energies are obtained by capturing the characteristic x-rays associated with anode materials of a relatively low atomic number. Initially for mammography, molybdenum, with an atomic number of 42, was used as a target because it produces characteristic energies at 17.5 and 19.5 keV. Figure 2-22 depicts the characteristic x-rays of Mo superimposed on the continuous Bremsstrahlung radiation.
While tungsten targets (used in conventional tubes) supply sufficient x-rays in this energy range, tungsten produces no characteristic energy in the mammographic range. Figure 2-23 compares the spectra of a molybdenum anode and filter combination with that of a conventional tungsten anode and aluminium filter combination. Molybdenum is used as a filter to remove much of the radiation with energy below and above the desired range, thereby producing a nearly mono-energetic beam. Molybdenum K-edge energy is at 20 keV; remembering that at the energy of the K-edge, x-ray attenuation increases and transmission decreases.
To penetrate more dense breast tissue, rhodium is used as both an anode and filter material. Rhodium has characteristic energies of 20.2 and 22.7 with a K-edge at 23.2 keV. There is also equipment that now offers a tungsten anode with rhodium as a filter.

Contrast is influenced by the amount of scatter and influenced by kVp, collimation and the thickness of the compressed breast. Grids improve contrast by allowing primary x-rays to pass through their interspace material while the lead strips absorb the secondary scatter. A moving grid with a grid ratio of 4:1 or 5:1 is typically used in mammography. Although grids reduce the amount of scattered radiation; their use requires a 2-4 times increase in exposure dose. Grids significantly improve image quality and while used routinely for standard positioning, are not used for magnification views, as an air gap technique is employed.

2.6.4.2.3 Sharpness

Sharpness, resolution or detail of a system is the ability to define an edge or to distinguish one object from another (21, 63, 71, 184). Good resolution is imperative as micro-calculifications can be as small as 50-100 µm. Motion is the greatest single problem in the production of good radiographic detail (21, 22, 63) and the primary effect of motion is to reduce contrast and hence small detail visibility. Sharpness is largely influenced by the focal spot size with mammography equipment using both a standard (broad) and small focal spot (0.3 / 0.1 mm).

The small focal spot is needed with magnification to maintain sharpness (i.e. the focal spot size must be reduced as object-film distance is increased to avoid geometric unsharpness). Sharpness is affected by other influences including focal film distance, object film distance, improper or inadequate compression (22, 174), film / screen combinations and poor film-screen contact (174).

Design changes to promote good resolution in mammography equipment have included using a fixed source image distance (SID) and a sloped target to give a small effective focal spot, as any increase in focal spot size increases geometric unsharpness (21, 63, 174). The actual focal spot is the area on the anode that is bombarded by the electrons discharged from the filament; while the effective focal spot is the x-ray beam projected towards the patient and the film (Figure 2-24). A tube current of approximately 100 mA is used with the standard focal spot (0.3 mm) which gives a shorter exposure time, thus voluntary motion can be controlled by keeping the exam short and the client still (17, 22); although compression remains the best solution to minimize motion.
A compression device is usually made of polycarbonate or Perspex with a thickness of a few mm. The device is used to spread the breast tissue so the breast is “taut”, like a firm tomato (16) or until the skin blanches (18) and is uniform thickness (Figure 2-25). Compression of the breast reduces geometric and motion unsharpness, increases contrast, separates superimposed areas of glandular tissue, reduces scatter and reduces radiation dose (21, 63). Compression is controlled by the mammographer and is typically applied at 25-40 pounds (11-18 kg) of force or up to 200 Newtons (21).
Film screen unsharpness might also be the result of poorly designed or damaged cassettes, improper placement of film in cassette, dirt or air trapped between film and screen (174). Consistent QA should minimize unsharpness caused by these instances. Other methods of controlling motion include the use of high speed, high contrast and high resolution single emulsion film screen combinations. This combination reduces radiation dose and gives shorter exposure time which means less object motion. Particle size of the rare earth phosphor in the intensifying screen affects blur (i.e. the larger the size, the greater the speed, the less detail). Mammography comprises a slower system to obtain better images, but at an increased dose.

2.6.4.2.4 Noise, Distortion and Artefacts

Noise, distortion and artefacts cause degradation of the image without any useful information and should be kept to a minimum. Noise tends to give the image a textured / grainy appearance and diminishes the ability to discern small details (such as calcifications) and lowers contrast detail (21, 22, 63). A number of factors contribute to noise; inherent factors include the size, shape and thickness of phosphor in the intensifying screens and film graininess.

Another degradation in mammography is quantum mottle which is due to a statistical fluctuation in the number of x-ray photons absorbed at differing locations in the intensifying screen (63, 174). The lower the number of x-ray photons (or low dose), the higher the amount of quantum mottle.

Distortion is the misrepresentation of an objects size or shape (19, 21), used for advantage in mammography with the magnification device. An artefact is anything that may degrade an image. This includes, but is not limited to, dust, dirt, powder, earrings, scratches, fingerprints, fog, processor marks or grid lines (21). Any degradation for any reason should be kept to a minimum.

Table 2-11 shows a summary of the relation between image quality factors to dedicated film-screen mammography equipment.
Table 2-11: Condensed Version of Equipment Relating to Image Quality and Mammography

<table>
<thead>
<tr>
<th>What</th>
<th>Why</th>
<th>How Achieved</th>
<th>How Affects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anode material</td>
<td>Low energy needed (17-24 keV range)</td>
<td>Molybdenum (Mo atomic number 42)</td>
<td>Characteristic peaks - Mo at 17 &amp; 19 keV &amp; Rh at 20 &amp; 23 keV</td>
</tr>
<tr>
<td></td>
<td>To utilize characteristic x-rays, need element with atomic number between 40-50</td>
<td>Rhodium (Rh atomic number 45) – to penetrate dense breast tissue</td>
<td></td>
</tr>
<tr>
<td>Filtration</td>
<td>To remove low or high energy x-rays (Bremmstrahlung) to change beam spectrum</td>
<td>K-edge filtration – Mo or Rh</td>
<td>Eliminates high energy Bremsstrahlung which would ↓ subject contrast ↓ Contrast by ↓ Scatter ↓ Dose – absorption of low energy x-rays</td>
</tr>
<tr>
<td>Focal spot size</td>
<td>To observe fine detail (such as 50-100 µm microcalcification)</td>
<td>0.3 mm (standard)</td>
<td>↑ Resolution by ↓ Geometric blur</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1 mm (magnification)</td>
<td></td>
</tr>
<tr>
<td>Envelope and exit window</td>
<td>At low kVp it is important that x-ray tube window not attenuate the x-ray beam significantly</td>
<td>Beryllium (z=4) or very thin borosilicate glass window (Inherent filtration in window ~ 0.1 mm of Al)</td>
<td>Allows use of greater portion for low energy x-rays, producing ↑ Contrast at ↑ Dose</td>
</tr>
<tr>
<td>Tube / anode angulations</td>
<td>Central ray more intense chest wall</td>
<td>Tube tilt 6 degrees, 16 degrees anode angle</td>
<td>↑ Density (penetration) at chest wall</td>
</tr>
<tr>
<td>Compression paddle</td>
<td>To reduce thickness of breast tissue</td>
<td>Paddle made of Polycarbonate or Perspex – a few mm thick</td>
<td>If taut and uniform: ↑ Contrast ↑ Resolution ↓ Motion unsharpness ↓ Dose</td>
</tr>
<tr>
<td></td>
<td>Separates superimposed areas of glandular tissue</td>
<td>After trying to relax the client, the mammographer gradually applies compression until the breast is taut. Compression is typically applied at 25-40 pounds of force or up to 200 Newtons Compression should be firm, but not painful</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reduces scatter</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>If taut, decreases motion unsharpness</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reduces radiation dose</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increases images quality if correctly applied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grids</td>
<td>Reduce scatter created in the breast and improve contrast by placing grid between breast and image receptor (Not used with magnification device as air gap technique employed)</td>
<td>Carbon fibre interspace material, moving grid usually either 4:1 or 5:1 grid, ↑ Dose by 2-3 times</td>
<td>↑ Contrast ↑ Dose</td>
</tr>
<tr>
<td>Automatic exposure control (AEC)</td>
<td>Provides consistent film density by measuring instantaneous dose rate (as long as positioned appropriately)</td>
<td>Either 1 or 2 detectors (placed behind film cassette) to instantaneously measure dose rate</td>
<td>Provides uniform exposure and density on imaging device (film)</td>
</tr>
</tbody>
</table>
2.6.5 **SCREEN READER REVIEW**

Using dedicated equipment, after the mammographer has produced the images and assessed them for technical quality, the mammograms are given to the screen reader(s) to report for interpretative purposes. The ability of the screen reader to detect signs of pathological processes depends on a combination of three major factors - image quality, the image viewing conditions and the ability of the observer to interpret what he / she sees (71). For optimal viewing of images there must be adequate view box luminance (3,000 candela per square meter), low ambient room lighting and masking of films and viewers (16, 40, 178, 191).

**2.6.5.1 Classification Systems**

When reviewing mammograms for interpretative purposes, the screen readers must determine whether supplemental imaging is necessary for clarification of any areas on the mammogram. Similar to feature analysis (section 2.6.6), there are a number of categories for management of the findings. Table 2-12 compares BI-RADS and Hunter BreastScreen assessment for management of any findings on the mammograms. If an area(s) of suspicion is detected, the screen readers must include the site(s), degree(s) of suspicion and reason(s) for follow-up.

Table 2-12: **Comparison of BI-RADS and Hunter BreastScreen Categories of Mammographic Breast Assessment / Interpretation (75)**

<table>
<thead>
<tr>
<th>Management of Findings</th>
<th>BI-RADS Categories</th>
<th>Hunter BreastScreen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Needs Additional Imaging</td>
<td>0</td>
<td>3, 4 or 5</td>
</tr>
<tr>
<td>Negative or Normal</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Benign Findings</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Probably Benign Findings</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Suspicious, Probably Malignant</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Highly suspicious, Malignant</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
If a screening mammogram shows a possible abnormality (for both BI-RADS and HBS categories 3-5), the client is asked to return for more tests to assess the area (a recall). At recall, additional x-rays may be taken to confirm or exclude a lesion; other procedures such as ultrasound, fine needle aspiration or core biopsy may also be performed to determine the nature of the area under review. After these additional procedures, approximately 90% of clients return to screening; with the remainder progressing to open surgical biopsy.

2.6.5.2 Review Processes

While some diagnostic and screening mammograms are interpreted by only one screen reader (such as in the United States); BreastScreen Australia images are always reviewed by two screen readers. Clients given a category 4 or 5 (even if only by one reader) are always recalled; however, the National Accreditation Standards of BreastScreen Australia state that where there is discordance between two independent screen readers on whether further assessment for the presence of cancer is required (example either 1/3 or 2/3), a single recommendation must be achieved. This can be through either a third reader or by a consensus read by the two readers (40).

2.6.5.2.1 Definitions of Consistency

Systems of classification are helpful in decision making and outcomes if there is to be a high level of agreement among observers. A well defined measure of agreement describes how well one readers observation of an image agrees with what other readers would have reported (inter-observer reliability) or how well a readers observation of an image at one time agrees with the same readers report at another time (intra-observer reliability) (192, 193). A reader is, however, likely to interpret the same image (radiograph) differently on a second reading 5-20% of the time (22, 194, 195), with 2 readers differing 33% of the time (195).

2.6.5.3 Variations in review

Screening mammography is known to have one of the highest interpretation disagreement rates between screen readers (196). Variability in interpretation can be attributed to the differences in visual observation (detection), the variation in lesion characterization (perception) and subsequent management (levels of confidence) (22, 103, 197-199).
From an interpretive perspective, one inherent disadvantage in mammography occurs due to the variability of breast composition. During review for detection, screen readers may overlook 11-25% of all cancers; not due to negligence but rather due to the fact that human observations have limitations (198). Possible causes for misinterpretation (missed cancers) include dense parenchyma obscuring a lesion and poor positioning or technique (58, 199, 200). There are also cancers that are not seen on mammograms (24, 53, 103), such as lobular cancers which comprise about 5 - 10% of all breast cancers (61).

2.6.5.3.1 BI-RADS

Breast Imaging Reporting and Data System (BI-RADS) was developed by the American College of Radiology (ACR) as a standardized method for evaluating the morphology of breast lesions and categorizing the findings in an unambiguous report (201). Use of the standardized feature analysis of BI-RADS does not preclude disagreement; one study conducted between 1997-2001 consisted of 26 radiologists daily double reading over 300,000 cases. When comparing 2% of all modalities between two readers, screening mammography had the highest disagreement rate (5.8%) compared with ultrasound (4%), diagnostic mammography (3.6%) and general radiography (3%) (196).

Studies have shown only moderate agreement for mammographic interpretation (197, 202) and for multiple screen readers assessing descriptors of breast lesions (203-206). Even with lesions clearly marked, studies have shown inter-observer reliability variation for assessing micro-calcification distribution and number between moderate (kappa 0.47) (204, 206) to substantial (kappa 0.77) (203).

2.6.5.3.2 Breast Density

Studies comparing BI-RADS four visual categories of breast density have shown only moderate agreement with that system. Ciatto et al reported that 12 breast radiologists assessing 100 mammograms for breast density had an overall Kappa 0.54 (range 0.02-0.77) (207); while Berg et al reported that 5 experienced radiologists describing breast density for 103 mammograms had an overall Kappa 0.43 (204). Berg et al demonstrated that more agreement was shown in the use of the term “fatty” (Kappa 0.76), than “minimal scattered fibro-glandular elements” (Kappa 0.43), “extremely dense” (Kappa 0.45) or “heterogeneously dense” (Kappa 0.17) (204).
A study by Gao et al assessing inter and intra observer reliability using density classifications of Wolfe, Boyd and a percentage of total densities in the breast demonstrated high reproducibility with appropriate training (74). For the inter-observer agreement, weighted Kappa was 0.89 (p<0.0001) for Wolfe classification and 0.84 (p<0.0001) for Boyd for the two raters; with an intraclass correlation coefficient 0.94 for percentage densities (74). For the intra-observer reliability, weighted Kappa was 0.87 (p<0.0001) for Wolfe classification and 0.86 (p<0.0001) for Boyd; with an intraclass correlation coefficient 0.96 for percentage densities (74). This study concluded that both visual qualitative and quantitative measurements on mammographic density were highly reproducible if appropriate training was provided (74).

2.6.5.3.3 PGMI

While assessment for image quality differs from assessment for interpretation, both have similarities of using established systems. Although PGMI appears to have clear, well-defined standards and explicit criteria, the classification allows for a measure of subjective review. This is consistent with other studies that show individual interpretations of guidelines can suffer from a lack of consistency (193, 208-210).

Other medical professional classifications / standards have shown the same difficulty in inter-observer agreement. A study of four joint arthroplasty surgeons who independently reviewed radiographs for 60 patients agreed on the presence of lesions in at most 57% of the zones (211). Another study found only fair to moderate agreement (inter-observer overall median Kappa coefficient 0.42 (range 0.20-0.71)) for 13 observers selection of one of 5 treatment modalities in 46 wrist fractures, with an intra-observer overall median Kappa coefficient 0.55 (range 0.24-0.79) (193).

Andersen et al found low inter-observer agreement between four specialists (2 orthopaedic hand surgeons and 2 radiologists) comparing four classification systems for distal radius fractures in the assessment of 55 sets of using plain radiographs (208). Given the low degree of inter-observer agreement for each of the study classifications, Andersen et al expressed that use of the classifications as a sole means of determining the direction of the treatment or for the direct comparison of results among different studies is unreliable (208).

During the literature review, one study located was similar to the present research. Moreira et al conducted a study comparing the validity and reliability of PGMI to another system EAR (Excellent, Adequate, Repeat) for the assessment of mammographic image quality (212). The study rated 30 sets of mammograms with
both systems and compared assessment results between twenty-one tutor mammographers with an expert panel (consisting of two senior mammographers and the NSW BS State radiologist). Inter-observer Kappa values for both classification systems were low (0.05 to 0.15), although agreement was over 50%. PGMI had significantly higher values than EAR (for MLO, 77% versus 74%, p<0.05 and 76.5% versus 75.9% for the CC view). It does seem counter-intuitive that with the reduction of categories, agreement did not increase in this study. As an aside, 54% of the 30 sets of mammograms were rated P and G; while the overall inadequate rate was 6% (212).

2.6.5.3.4 Simplifying classifications

Increasing either the number of observers or the number of categories generally increases the rate of disagreement. In the study by Ciatto et al, for example, once BI-RADS four grade scale for breast density was converted to a two grade scale, inter-observer consistency ranged from moderate (k=0.54, range 0.02-0.77) to substantial (k=0.71, range 0.31-0.88) (207). This was similar to a study conducted by Wainwright et al comparing inter and intra observer consistency for nine specialists reviewing 33 sets of distal humerus fractures (210). One of the systems assessed could be divided into three separate groups containing 27 subgroups, 9 groups or 3 groups. As the number of subgroups in the system decreased, overall Kappa for the nine specialists increased from 0.34 (range 0.27-0.33) (27 subgroup), to Kappa 0.52 (range 0.46-0.61) (9 groups) and final Kappa 0.66 (range 0.61-0.80) for 3 groups (210).

2.6.5.3.5 Intra-observer consistency

The study conducted by Ciccone et al comparing 7 radiologists interpreting the same 45 mammograms 2 years apart reported an overall intra-observer Kappa 0.56 (range 0.35-0.67) (202). The study conducted by Ciatto et al showed substantial intra-observer agreement of the 12 radiologists assessing 100 mammograms with BI-RADS 4 scale breast density classification (k=0.71, range 0.32-0.88) (207); while Baker et al showed that with one radiologist re-reading images for assessment of lesion interpretation, Kappa was substantial 0.65 ± 0.08 (203). The study conducted by Wainwright et al demonstrated an intra-observer overall Kappa of 0.40 (range 0.25-0.53) (27 subgroup), 0.58 (range 0.47-0.75) (9 groups) and final is 0.70 (range 0.49-0.84) for 3 groups (210).
2.6.5.3.6 Effect of Training on Variation

To achieve inter-observer consistency and maintain intra-observer reliability, proper training and agreement of the criteria is essential and may require additional training, as well as experience. Studies have demonstrated that training in classification systems can improve accuracy. Berg et al found that after attending a day lecture on BI-RADS, there was improved agreement with consensus feature analysis for both mass margins and asymmetries (pre-training overall mean Kappa 0.36, range 0.07-0.61 to post-training overall mean Kappa 0.41, range 0.06-0.54) and final interpretative assessment (pre-training overall mean Kappa 0.31 to post-training overall mean Kappa 0.45) for 23 mammography screen readers assessing 54 sets of mammograms (206). Audit results by Linver et al found that sensitivity increased from 80-87% over a 2 year time frame for 11 radiologists interpreting over 38,000 mammograms (172, 213); this increase was attributed to attendance in dedicated mammography courses over that 2 year time period (172, 213). The effect of training in any of these classification systems should be to standardize methods and improve client outcomes.

2.6.6 Reasons for Recall

There are differing methods to categorize why women are recalled (24, 25, 71, 201). There are primary, secondary and indirect or subtle signs of breast abnormalities. Primary signs include mass or calcifications. Secondary signs include skin thickening, skin retraction, nipple and areolar changes and increased vascularity of the breast. Subtle signs include new or evolving densities, architectural distortion and asymmetrical breast tissue (25, 71).

The American College of Radiologists use the Breast Image and Reporting Data System (BI-RADS) which was developed to help standardize feature analysis and final management of mammographic findings. BI-RADS uses the categories of masses, calcifications, architectural distortion, special cases (including tubular density / solitary dilated duct, infra-mammary lymph node, asymmetric breast tissue and focal asymmetric density), associated findings (including nipple or skin retraction, skin or trabecular thickening, skin lesion, axillary adenopathy, architectural distortion and calcifications) and location (75) of the suspicious area.

One study conducted in 1986 comprising 300 non-palpable cancers in mammography included 42% of calcifications, 39% of masses and 20% indirect signs (214).
Hunter BreastScreen recalls for 5 broad categories, excluding technical or clinical reasons. These categories include:

- circumscribed lesions
- micro-calcifications
- asymmetry
- stellate lesions and
- lesions not otherwise specified (NOS), which include non specific densities and distortion.

2.6.6.1 Masses

Any space occupying lesion is considered a mass and is described by its shape, margins and density (25, 71, 75). Shape descriptors include round, oval, lobulated or irregular. Margin descriptors include circumscribed, microlobulated, obscured, ill-defined and spiculated (or stellate like) (Figure 2-26).

![Figure 2-26: Descriptors of Shapes and Margins of Mammographic Lesions](25)
2.6.6.1 Circumscribed Lesions

Masses that are circular / oval shaped are easy to perceive as the contour is convex (24). The majority of these lesions are benign. However, if differentiating is the challenge, ultrasound becomes the final assessment and tool to determine whether the lesion is cystic or solid (Figure 2-27).

![Image](A) ![Image](B)

Figure 2-27: Circumscribed Lesions on CC Mammograms – (A) Benign Cyst and (B) Malignant Circumscribed Lesion (78)

2.6.6.2 Micro-calcifications

Micro-calcifications are the deposit of calcium salts in tissues (25). Micro-calcifications are easy to perceive due to high atomic number, but difficult to interpret (24). Micro-calcifications are one of the most common mammographic features resulting in false-negative assessment (215). There are three general groups of micro-calcifications in the breast: typically benign, indeterminate and those having a higher probability of malignancy. When viewing micro-calcifications, their form, size, density and distribution need to be considered.

Characteristics and an image of typically benign micro-calcifications are shown in Figure 2-28. Typically malignant micro-calcifications present as fine, linear, granular, irregular in size and shape, branching (casting) that are scattered or clustered. An observation by a radiologist suggests that these micro-calcifications would be sharp if one attempted to touch them. Figure 2-29 shows images of both indeterminate and malignant micro-calcifications.
Figure 2-28: Descriptors and Image of Typically Benign Micro-calcifications (25)

Figure 2-29: Descriptors of Indeterminate / Malignant Micro-calcifications and image of Malignant Micro-calcifications (25)
Radiological interpretation of indeterminate micro-calcifications as benign or malignant is unreliable (215). An isolated cluster of micro-calcifications require an image guided core biopsy with representative micro-calcifications obtained on the specimen (215) (Figure 2-30).

![Core Specimen Radiograph with Calcifications](image)

**Figure 2-30: Core Specimen Radiograph with Calcifications (177)**

### 2.6.6.3 Asymmetry

Although there is a wide variation in breast size and parenchymal pattern, the breasts are generally symmetrical structures with similar density and architecture. Asymmetric breast tissue refers to a greater volume or density of breast tissue in one breast than in the corresponding area in the contra-lateral breast (Figure 2-31). Although this is often a normal finding, additional evaluation may be required. Asymmetry may be secondary to removal of tissue or to a lack of development or more prominent parenchyma in one breast (54)

### 2.6.6.4 Architectural Distortion

In architectural distortion, a focal area of breast tissue appears distorted with no definable central mass (25). BI-RADS describe architectural distortion as “the normal architecture is distorted with no definite mass visible. This includes speculations radiating from a point, and focal retraction or distortion of the edge of the parenchyma” (201). Spot compression views should be carefully analysed to ensure areas of architectural distortion have truly resolved (75, 215). Architectural distortion can also be associated with surgical scars, so it is imperative for the mammographer to mark for the screen reader any scars and their site. Alternatively, it can be associated with slow growing carcinomas (Figure 2-32).
2.6.6.5 Stellate Lesions

Stellate (speculated or star-like) lesions are typically characterized by an abnormal pattern of linear structures and a central mass (25, 216) (Figure 2-33). There is an area of focal retraction and tethering of normal parenchyma. The majority (93-95%) of stellates are malignant which makes perception of these lesions a priority (23, 24).
2.6.6.6 Lesions Not Otherwise Specified (NOS)

Lesions not otherwise specified include any lesions unable to fit into the previous categories. A poorly defined mass is at an increased risk of being a false negative interval cancer, which may be related to these features being more difficult to locate in screening mammography (217).

2.6.6.7 New or Evolving Density

To be considered a new density access to previous mammograms are required. An evolving density may have been present in retrospect on a previous mammogram, but may have been smaller and perhaps not recognized as significant (25). A density that is seen in only one standard mammographic view is referred to as a density seen in one projection. True lesions may appear on only one view because they are either obscured by overlapping dense parenchyma or located posteriorly and thus outside the field of view. Perception of any lesion is enhanced by the elimination of extraneous light, focusing attention by decreasing peripheral distraction and magnification of image detail (by using a viewer / panascope) (24). When an area of suspicion is detected in a mammogram, those clients must return for further assessment of the area and final diagnosis using a multi-disciplinary team approach.
2.7 SUMMARY

Early detection of breast cancer is still the best hope for survival and the most effective method of detection is mammography. Screening was introduced to facilitate detection. Approximately 10% of clients screened are recalled for additional procedures leading to increased costs to the programs, anxiety and discomfort for those women.

Many variables influence the ability to detect a cancer on a mammogram. The major factors include the knowledge and skill of the operator, x-ray tube features, image receptors (film screen has more limitations than digital due to noise which degrades the contrast resolution and processing, as far as being unable to enhance the information once it is processed) , the positioning technique employed, the amount of compression, the type, size and position of breast cancer, the radiographic appearance of the tissue and knowledge of the interpreting physician (15-19, 21). This research specifically targets the contribution by the mammographer.

Screening mammograms are assessed for image quality. In Australia, technical image quality is assessed using the established grading system of PGMI. Although PGMI is a useful tool, there has been little research undertaken to determine its reliability. During the literature review, only one study could be found assessing this system. This research, therefore, was undertaken to investigate reliability of PGMI and to assess the consistency of a panel using this established method.

Clients are recalled for further assessment if an area has any suspicious features. Unless clearly benign, clients whose mammograms exhibit characteristics of microcalcifications and / or circumscribed lesions are recalled. This research chose to include only those clients recalled for the more subtle reasons of asymmetry, lesions not otherwise specified and stellate lesions.

As mammography is operator dependent with many variables affecting quality, this research was undertaken to review sets of mammograms for both recalled and non-recalled clients to assess image quality. Image quality affects image interpretation and if the number of “missed cancers” might be reduced by better positioning technique as the study by Taplin et al suggests (68), then maybe the number of false positive recalls might be reduced in the same manner. It is hoped this research will provide discussion and promote quality imaging for mammographers in screening mammography.
3 CHAPTER THREE: METHODS
3.1 INTRODUCTION
During the process of recall at HBS, a question often posed among the physicians and mammographers was “If the initial mammogram had been of better quality, would this client have been recalled?” Therefore, the initial purpose of this study was to determine if image quality is a factor in the recall of women and to investigate other factors attributing to recall. A number of specific research questions were asked during the study; these included:

- What is the percentage of inadequate images?
- What are the reasons for inadequate images?
- What is the consistency (inter and intra) amongst panel members in using PGMI?
- What factors influence the likelihood of clients being recalled?

This research comprised of three separate, but inter-related phases assessing sets of mammograms for both recalled and non-recalled clients in screening mammography. These were Phase A (Pilot Study), Phase B (Main Study) and Phase C (Control) (Figure 3-1). Sets of recalled images (Phases A and B) include those recalled for asymmetry, stellate lesions, lesions not otherwise specified (NOS) or a combination of these factors. Phase C consists of images from clients who were not recalled and acts as the control component.
3.2 ETHICAL ISSUES

Application for ethical approval for this study was sought from the University of Newcastle Human Research Ethics Committee (HREC) and the Hunter Area Research Ethics Committee (HAREC); it was approved in November 2003 (Appendix One). The main ethical issues were informed consent from the clients to use their mammograms and clinical history and the storage and disposal of research data.

Clients of HBS are given the opportunity to have their information placed on a central register and provide permission for their mammograms to be used for statistical purposes and research. Only those clients of HBS who had provided written consent for their information and mammograms to be used were eligible to be included in this study. The Data Manager at HBS provided information of the clients meeting eligibility criteria. The sets of images for clients were selected by the researcher to match the inclusion criteria. All client data entered into an Access database contained only de-identified information and was password protected. Electronic data and consent forms will be stored for a period of 5 years in accordance with the National Statement on Ethical Conduct in Research Involving Humans (1999).

3.3 RESOURCES AND MATERIALS

3.3.1 EXPERT PANEL

Two screen readers and four mammographers were invited to participate in this research. They were chosen for their expertise and experience in screening mammography and availability (proximity) to the researcher. All accepted the invitation to participate and will be referred to as the panel members and are identified randomly as R1, R2, R3, R4, R5 and R6.

All panel members had worked for BreastScreen Australia for over 10 years; both screen readers and one of the mammographers worked part time. During Phase A, the panel received no formal training on PGMI, although the mammographers routinely graded images using PGMI. The screen readers did not routinely use this system, though both were aware of PGMI and how to use it.

3.3.2 SETS OF IMAGES

The mammograms (sets of images) used in this research were from clients at BreastScreen NSW Hunter Region and Wyong Shire (Hunter BreastScreen – HBS). This retrospective research study involved the assessment of mammograms performed during 2002 and 2003 for both non-recalled and recalled women. Until August 2004, HBS had used both public and private radiography sites for screening mammography. The public system involved the use of mobile screening vans as well
as fixed sites; while the private incorporated only fixed sites. Whether initial images were undertaken through public or private sites, all recalls were performed at the same public site.

The recalled images belonged to clients recalled for asymmetry, stellate lesions and lesions NOS. The non-recalled client images acted as a control group. The panel members graded image quality using the PGMI classification system. All statistical analyses were conducted using Stata software (218).

3.3.2.1 Inclusion Criteria

Inclusion criteria for recall images

The appearance of micro-calcifications and circumscribed lesions are generally easy to perceive. The majority of circumscribed lesions are benign and while micro-calcifications are generally easy to perceive, they are challenging to assess. This research therefore chose to assess how image quality was affected by the more subtle signs of asymmetry, stellate lesions and lesions NOS.

The inclusion criteria included sets of images from asymptomatic clients who:

   1. were 40 years and over,
   2. participated in screening mammography at HBS during 2002 or 2003,
   3. provided written consent for their data and mammograms to be used for research,
   4. were recalled for one or a combination of the following:
      a. asymmetry,
      b. stellate lesions and
      c. lesions not otherwise specified (NOS)

Inclusion criteria for non-recall images

The inclusion criteria included sets of images from asymptomatic clients who:

   1. were 40 years and over,
   2. participated in screening mammography at HBS during 2003,
   3. provided written consent for their data and mammograms to be used for research and
   4. returned to screening (given normal or benign results)
3.3.2.2 Exclusion Criteria

The exclusion criteria included sets of images from clients who were recalled for: -

1. micro-calcifications,
2. circumscribed lesions,
3. any combination with micro-calcifications or circumscribed lesions,
4. any clinical issues including lumps, nipple discharge and itchy nipples
5. any client with a previous history of breast cancer
6. technical faults or repeats

3.4 PILOT STUDY (PHASE A)

The primary question of this study was to determine if inadequate image quality was a contributing factor for recalled clients at HBS. The pilot study (Phase A) was undertaken to validate the need of the research, assess panel consistency and to test the research methodology.

3.4.1 SELECTION OF SETS OF IMAGES

The Data Manager of HBS provided the researcher with Round 1 (initial) and Round 2+ (subsequent) screening lists containing identification numbers for recalled women during the time frame January to June 2003. This time-interval was chosen to coincide with the six-monthly period immediately preceding data collection. The lists included the Patient Identification Details (the filing number), which is unique for each client; the unit number where the mammogram had been performed; screen readers result; reason, site and side of recall.

To provide a cross section for both recall reason, round of screening and screen reader recall, the researcher aimed to choose 20 sets of images for clients recalled for asymmetry, 20 for stellate lesions and 20 for lesions NOS. Ideally, each set of 20 contained 10 sets from Round 1 and 10 sets from subsequent rounds (Round 2+). Each set of 10 were further divided by half where both screen readers returned the client for recall and half where only one screen reader (plus third reader) returned the client for recall (Figure 3-2).

The images assessed were those leading to recall, including cranio-caudal, medial lateral obliques and occasionally laterals if taken to complete a mammogram. The selection for the sets of images, at this point, was done systematically from the beginning of the list for matching criteria with no regard for where the image was produced (public or private system). All sites were screening at normal capacity. Phase A contained only images of clients who had returned to HBS for their recalls.
Figure 3-2: Phase A - Selection of Sets of Images

HBS Recalls (60)

- Asymmetry (20)
  - Round 1 (10)
    - Both readers Recalled (5)
    - One reader + 3rd reader Recalled (5)
  - Round 2+ (10)
    - Both readers Recalled (5)
    - One reader + 3rd reader Recalled (5)

- Stellate Lesions (20)
  - Round 1 (10)
    - Both readers Recalled (5)
    - One reader + 3rd reader Recalled (5)
  - Round 2+ (10)
    - Both readers Recalled (5)
    - One reader + 3rd reader Recalled (5)

- Lesions NOS (20)
  - Round 1 (10)
    - Both readers Recalled (5)
    - One reader + 3rd reader Recalled (5)
  - Round 2+ (10)
    - Both readers Recalled (5)
    - One reader + 3rd reader Recalled (5)
3.4.2 RETRIEVAL, DATA COLLECTION AND DISPLAY OF SETS OF IMAGES

The selected packets were retrieved and the images were hung on a multi-viewer under standard screening conditions for the panel members to assess. Information for recalled clients was obtained and recorded on Data Collection Sheets 1A (Table 3-1).

3.4.2.1 Data Collection Sheet 1A

Client Identification: Information recorded for clients include their patient identification number (PID), film number (FID), date of birth, unit where mammogram was performed, radiographer performing examination and the two screen readers.

History of Client: Information was also collected relating to client background. Data collected included: history of breast cancer, 1st degree history of breast cancer, client status in relationship to hormone replacement, any current history of clinical breast issues, previous breast or heart surgery and whether previous mammograms (other than screening) have been performed.

Technical Factors: Information printed on images generated from the mammography equipment included technical factors such as kVp, mAs, target material and the level of compression used. Not all mammography equipment generates this information. This data was unavailable for the majority of the images reviewed; therefore this information was not able to be used for the research.

Film review (Data Collection Sheet 1A): Information included the screen readers assessment including side, site(s), reason(s) for recall and view that showed possible abnormality.

Recall (Data Collection Sheet 1A): Information collected related to images performed at recall, other modalities used and final outcome for client, including additional images, ultrasounds, fine needle aspirations (FNA’s), core biopsies, localizations and surgeries.
### Table 3-1: Data Collection Sheet A

#### Client Identification

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PID</td>
<td></td>
</tr>
<tr>
<td>FID</td>
<td></td>
</tr>
<tr>
<td>DOB</td>
<td></td>
</tr>
<tr>
<td>Unit</td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td></td>
</tr>
<tr>
<td>Drs</td>
<td></td>
</tr>
</tbody>
</table>

#### History of Client – Yes or No

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td></td>
</tr>
<tr>
<td>HRT</td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td></td>
</tr>
<tr>
<td>Lump</td>
<td></td>
</tr>
<tr>
<td>Discharge</td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td></td>
</tr>
<tr>
<td>Prev Mammo</td>
<td></td>
</tr>
</tbody>
</table>

#### Technical Factors (on equipment where recorded)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>kVp</td>
<td></td>
</tr>
<tr>
<td>mAs</td>
<td></td>
</tr>
<tr>
<td>Mo/Rh</td>
<td></td>
</tr>
<tr>
<td>Rh/RH</td>
<td></td>
</tr>
<tr>
<td>Mo/Mo</td>
<td></td>
</tr>
<tr>
<td>Compression</td>
<td></td>
</tr>
</tbody>
</table>

#### Film Review

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast(R/L)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reason</td>
<td></td>
<td></td>
</tr>
<tr>
<td>View</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Recall

<table>
<thead>
<tr>
<th>Recall</th>
<th>Radiologist</th>
<th>Mammographer</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGMI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other films</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spot/Mag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other exams</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.4.3 ASSESSMENT OF IMAGES

Prior to assessment, each panel reader received a packet containing the list of images hung (given by both PID and client name) in standard screen reading conditions, a copy of the PGMI requirements and image quality assessment sheets (Data Collection Sheet B) to be completed if an image was considered inadequate (Table 3-2). Extra Data Collection Sheets, a magnifying glass and panascope were available at the multi-viewer. No formal training on use of the PGMI occurred prior to Phase A.

The panel independently assessed each film according to the PGMI criteria. If a set of images were assessed as a P, G or M (perfect, good or moderate) no reason for the decision was needed. However, if any image in a set was considered I (inadequate), the view(s) and reason(s) for inadequacy were documented and recorded on Sheet B.

The images remained on the multi-viewer until all panel members assessed them. The panel members were able to view the images at their convenience generally over a 6-week time frame. Phase A images were available for assessment from December 2003 to January 2004, taking approximately 20-30 minutes for completion.

Table 3-2: Data Collection Sheet B

<table>
<thead>
<tr>
<th>PGMI</th>
<th>Inadequate</th>
<th>Why Inadequate</th>
</tr>
</thead>
<tbody>
<tr>
<td>All breast tissue visualized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct film identification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct exposure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good compression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absence of motion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct positioning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absence of artefacts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No skin folds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symmetrical images</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.4.4 Consensus Agreement

After all panel members had assessed the images, a meeting was held to review set(s) of images ranked as inadequate. During the discussion, the researcher explained the reason(s) why an image had been classed as inadequate. Panel members R2 and R6 reviewed those images with the researcher after the initial group discussion, as they had been unable to attend the group discussion. The reviews consisted of a discussion of the merits of each set of images and whether the inadequate rating would remain inadequate by consensus. This led to an in-depth discussion and a modification of the PGMI for the remainder of the research (Table 3-3). Any image originally classified as “inadequate” due to a reason not directly related to image quality was re-graded to “moderate” for the final analysis.

During this meeting, a training session between panel members was conducted using the mammography sets to evaluate the PGMI criteria and then make a decision on how to classify those images. This was done with the intention of training to increase member reliability with the evaluation system.

### Table 3-3: Modified PGMI Sheet

<table>
<thead>
<tr>
<th>Assessor Code: ____</th>
<th>PID:_______</th>
<th>Client Name:_____________</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PGMI</strong></td>
<td><strong>Inadequate (Please Tick)</strong></td>
<td><strong>Inadequate View/s</strong></td>
</tr>
<tr>
<td>All breast tissue visualized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compression/ Unsharpness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absence of artefacts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No skin folds</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.4.4.1 Modified PGMI

The modified PGMI removed the categories of film identification and symmetrical images, while combining the criteria’s of compression / motion and exposure / processing. The reasons for these changes to the criteria are as follows:

**Film Identification**
Criteria for PGMI states that correct film identification should contain the date of examination, client identification-name and (number and or date of birth), side markers, positional markers and radiographer identification. An image using PGMI would therefore be classified as inadequate if there were incomplete identification. As this does not directly affect the image quality of the mammogram, the criterion was subsequently removed. After Phase A, no inadequate rating was given from the panel for images with incomplete or partial identification.

**Symmetrical Images**

Symmetry of the projections is good mammography practice as it allows for more consistent visualization of any abnormalities of breast areas. As the PGMI already states that no image will be classified as inadequate for non-symmetrical projections; this category was removed for the Modified PGMI Assessment Sheet.

**Compression and Motion**

The criteria of compression and motion are independent using PGMI. However, it is not always apparent whether a blurred image is due to inadequate compression or movement. These criteria were joined in the modified PGMI as compression/unsharpness.

**Exposure and Processing**

The criteria of correct exposure and correct processing are also independent using PGMI. As these are both involved with image production, they were joined in the modified PGMI as exposure.

### 3.4.5 Analysis of Results

After the panel members completed image assessment, the researcher compiled the results and entered them into both an Excel Spreadsheet and an Access Database.

#### 3.4.5.1 Assessment of Inadequate Images

As previously stated the main reason for this pilot study was to assess if the number of inadequate images were really an issue in recalled images at HBS. This was achieved by calculating the frequency of inadequate images from all the images assessed; the frequency was performed for both individual readers and the collective consensus. Reasons for inadequate image quality was also assessed using the collective consensus.
### 3.4.5.2 Assessment of Panel Consistency Using Kappa

For any study involving a number of assessors it is important to ensure that all assessors are consistent in the use of any research tool. This can be done by calculating the level of agreement and by using the KAPPA statistic. Kappa is a measure of agreement beyond the level of agreement expected by chance alone. It has a maximum of 1 when there is perfect agreement, a value of 0 indicating no agreement better than chance and a value of -1 shows complete disagreement. In this study, care needs to be given in assessing the final KAPPA statistic, as it may be misleading for several reasons.

Firstly, one would assume with inter-observer agreement that observers make their assessment on some basis other than random guessing. Thus, the common statement that kappa is a "chance-corrected measure of agreement" is misleading. As a test statistic, kappa can verify that agreement exceeds chance levels. But as a measure of the level of agreement, kappa is not "chance-corrected". In fact, it is not clear how chance affects the decisions of actual observers and how one might correct for it.

Secondly, kappa may be low even though there are high levels of agreement and even though individual ratings are accurate. Whether a given kappa value implies a good or a bad rating (Table 3-4) depends on what model one assumes about the decision making of the observers.

<table>
<thead>
<tr>
<th>Strength of Agreement</th>
<th>Value Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>&lt; 0</td>
</tr>
<tr>
<td>Slight</td>
<td>0 – 0.20</td>
</tr>
<tr>
<td>Fair</td>
<td>0.21 – 0.40</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.41 – 0.60</td>
</tr>
<tr>
<td>Substantial</td>
<td>0.61 – 0.81</td>
</tr>
<tr>
<td>Almost Perfect</td>
<td>0.81 – 1.00</td>
</tr>
</tbody>
</table>

### 3.5 MAIN COMPONENT (PHASE B)

Phase B is the main study component required to achieve statistical power. This number was calculated to be 550 using the five criteria in the modified PGMI and two effects (recall or non-recall).
3.5.1 SELECTION OF SETS OF IMAGES
Image selection was similar in process to Phase A in that lists were provided by the Data Manager of HBS. The differences included the number of images selected and the time frame. All the images from January – December 2002 (the time frame immediately preceding Phase A), meeting the inclusion criteria were retrieved, including those for clients who had gone elsewhere for their recalls. From a possible 946 clients, 681 were eligible for inclusion. Packets were only available for 640 clients.

3.5.2 RETRIEVAL, DATA COLLECTION AND DISPLAY OF SETS OF IMAGES
The selected packets were retrieved and the images hung on a multi-viewer under standard screening conditions for the panel members to assess. Information for recalled clients was obtained and recorded on Data Collection Sheets A by the researcher (Table 3-1).

3.5.3 ASSESSMENT OF IMAGES
The panel members assessed the images during five separate settings between September 2005 and September 2006. Each set was displayed for approximately eight weeks taking between one to one and a half hours to complete. Assessment was conducted as described in the previous phase (section 3.4.3) using the modified PGMI.

3.5.4 CONSENSUS ASSESSMENT
In this phase, consensus on “inadequate” images was determined where three or more panel members assessed any image as “inadequate”. No consensus meeting was held.

3.5.5 ANALYSIS OF RESULTS
After the panel members completed assessing the images, the researcher compiled the results and entered data into both an Excel Spreadsheet and an Access Database.

3.5.5.1 Assessment of Inadequate Images
As in the previous phase, assessment of inadequate images was achieved by calculating the frequency of inadequate images from all the images assessed; the frequency was performed for individual readers and the collective consensus assessment.
3.5.6 **Intra-Observer Consistency**
A subset of 30 sets from Phase A referred to as (A*) were interspersed during Phase B to calculate intra-observer consistency. The panel members were unaware of this phase of the study.

3.6 **Control Component (Phase C)**
Phase C was conducted to assess if there was a difference in the number of inadequate sets of images between recall and non-recalled sets. The number required for statistical power was calculated to be 200.

3.6.1 **Selection of Sets of Images**
At HBS, screen readers sheets for previously read mammograms are kept in boxes in a locked storeroom. The researcher randomly chose a box that contained reading of images during 2002 (the same time frame as Phase B). This box contained reading of images performed during October to December 2002.

The researcher chose 35 sets of images from each of the three public sites and three private sites for a total of 210 sets of images. These sites were representative of where the majority of images had been produced. Sets of images were systematically selected by reviewing the lists for each site and obtaining the first set each day that had not been recalled. This selection process was done to ensure that the sample was representative of the women screened.

3.6.2 **Retrieval, Data Collection and Display of Sets of Images**
The selected packets were retrieved and the images hung on a multi-viewer under standard screening conditions for the panel members to assess. Information for non-recalled clients was obtained and recorded on Data Collection Sheet C (Table 3-5). Data Collection Sheet C was a modified version of Data Collection Sheet A (Table 3-1) - minus recall information and including information relating to subsequent outcomes for these initially non-recalled clients.

3.6.3 **Assessment of Images**
The panel members assessed the images between September 2006 and January 2007. Assessment was conducted as described in Phase A (section 3.4.3).

3.6.4 **Consensus Assessment**
Consensus on “inadequate” images was similar to Phase B and was determined where three or more panel members assessed any image as “inadequate”. No consensus meeting was held.
3.6.5 ANALYSIS OF RESULTS
After the panel members completed assessing the images, the researcher compiled the results and entered data into both an Excel Spreadsheet and an Access Database.

3.6.6 INTRA-OBSERVER CONSISTENCY
A subset of 62 sets from Phase B referred to as (B*) were interspersed during Phase C to calculate intra-observer consistency. The panel members were unaware of this part of the research.
## Table 3-5: Data Collection Sheet Phase C

<table>
<thead>
<tr>
<th>Client Identification</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PID</td>
<td></td>
</tr>
<tr>
<td>FID</td>
<td></td>
</tr>
<tr>
<td>DOB</td>
<td></td>
</tr>
<tr>
<td>Unit</td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td></td>
</tr>
<tr>
<td>Drs</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>History of Client – Yes or No</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td></td>
</tr>
<tr>
<td>HRT</td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td></td>
</tr>
<tr>
<td>Lump</td>
<td></td>
</tr>
<tr>
<td>Discharge</td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td></td>
</tr>
<tr>
<td>Prev Mammo</td>
<td></td>
</tr>
</tbody>
</table>

| Return Screen Outcome 1    |  |
| Return Screen Outcome 2   |  |
| Return Screen Outcome 3   |  |
| Return Screen Outcome 4   |  |
3.7 ASSESSMENT OF PANEL CONSISTENCY
Assessment of panel consistency was evaluated for both inter and intra-observer reliability. Inter-observer consistency was assessed pre-training (Phase A) and post-training (Phases B and C). Intra-observer was assessed twice using information collected from Phase A and Phase B.

3.8 ASSESSMENT OF CLIENT RECORDS FOR THE DETERMINATION OF CANCER STATUS
The researcher evaluated the outcomes for all clients in the research, from their initial outcome in 2003 to present outcomes in 2006. This included all surgeries performed and the presence or absence of cancer. All images were assessed to determine if inadequate image quality influenced the cancer detection rate.

3.9 CASE CONTROLLED STUDY
One matched case-control study was conducted; the case (recalled images) was from Phase B and the control (non-recalled images) was from Phases C. One control was matched to each case by same age and date of screening (within 4 months). Conditional logistic regression analyses were performed to obtain the estimates of odds ratio and their 95% confident intervals.

The two broad categories of client history and image assessment were used to investigate factors attributing to recall for the case-control study. Client history included HRT status and image assessment data included the site of image production, mammographic density and image quality. Density of the breast tissue was measured using a VIDAR VXR Scanner (Version 4.2, VIDAR System Corp., Herndin, VA) to produce digitized forms of the images. The scanner was automatically calibrated after turning on the machine, so the quality of the scanned images is assumed to remain constant for all films. All assessments of the mammographic density were done by the same expert (Dr. Jinnan Gao – co-supervisor), which ensured that the same criterion of measurement was adopted.
CHAPTER FOUR: RESULTS
4.1 INTRODUCTION TO RESULTS
This chapter reports on the results of the research; and consists of a number of sections. After summarising the background of the research material and collection of data; the results will be given in four sections. The first section assesses the consistency of readers using the established classification PGMI system for grading mammographic radiographs and the affect of training on consistency. The second section investigates the number of views considered inadequate as well as the reasons why images were classed as inadequate. The third section evaluates the relationship between components of recalled / non-recalled clients comparing client demographics (age, HRT status), breast density and site taken (public / private) using a case-controlled study. The final section assesses whether inadequate images influence cancer detection rates. A quick reminder to the separate phases of the research is given below (Figure 4-1).

![Figure 4-1: Phases of Research Methodology](image_url)

4.2 RESOURCE AND MATERIALS

4.2.1.1 Expert Panel Members
Six panel members consisting of 2 screen readers (SR) and 4 mammographers (M), experienced and competent in all aspects of mammography participated in the study. The members were assigned assessor codes R1, R2, R3, R4, R5 and R6. Summary demographics of the panel are given in Table 4-1.
Table 4-1: Panel Members Demographics

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profession</td>
<td>SR #1</td>
<td>SR #2</td>
<td>M #1</td>
<td>M #2</td>
<td>M #3</td>
<td>M #4</td>
</tr>
<tr>
<td>Employment in mammography (yrs)</td>
<td>18</td>
<td>22</td>
<td>18</td>
<td>16</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Employment in BreastScreen (yrs)</td>
<td>15</td>
<td>16</td>
<td>11</td>
<td>16</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>Trained in Australia (initial)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Method used to assess images</td>
<td>PGMI</td>
<td>PGMI</td>
<td>PGMI</td>
<td>PGMI</td>
<td>PGMI</td>
<td>PGMI</td>
</tr>
<tr>
<td>Length of time using PGMI (yrs)</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>10</td>
</tr>
</tbody>
</table>

SC – Screen Reader    M - Mammographer

4.2.1.2 Sets of Mammographic Images

The 904 sets of images used in the research comprised 698 recalled and 206 non-recalled clients of BreastScreen NSW Hunter Region and Wyong Shire (HBS) radiographed between January 2002 and June 2003 (Figure 4-2). During this interval, HBS used both public (six sites including 2 fixed and 4 mobiles) and private facilities (three fixed sites) for screening. The sites will be classified as either public and private denoting where the mammograms had been produced.

Figure 4-2: Dates for Research Images
Selection of images: Phase A

The distribution of the sets of images in Phase A is given in Table 4-2. Two sets out of the 60 sets displayed were subsequently removed for a clinic prior to assessment; thus Phase A, consisted of 58 sets of images. The researcher was unaware that they had been removed until after panel assessment – therefore these sets were not replaced.

The number for all categories matched set criteria (section 3.4.1) with the exception for lesions NOS. Twenty-eight sets (46.7 %) were taken at the public sites and thirty-two sets (53.3 %) were taken at the private sites.

Table 4-2: Breakdown for Image Selection Phase A

<table>
<thead>
<tr>
<th>PHASE A</th>
<th>Two Readers Recalled</th>
<th>One Reader Recalled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asymmetry</td>
<td></td>
</tr>
<tr>
<td>Round 1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Round 2</td>
<td>5</td>
<td>5 (3) *</td>
</tr>
<tr>
<td></td>
<td>Stellate Lesion</td>
<td></td>
</tr>
<tr>
<td>Round 1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Round 2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Lesion NOS</td>
<td></td>
</tr>
<tr>
<td>Round 1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Round 2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>29</strong></td>
<td>**31 (29) ***</td>
</tr>
</tbody>
</table>

* Indicates that 2 sets were removed prior to assessment

Selection of images: Phase B

The distribution of the 640 sets of images in Phase B with reference to the reason for recall is given in Table 4-3. Two-thirds of these clients (66%) had been initially recalled by one screen reader, with an additional third screen reader indicating a recall. Seventy-two percent of these images were taken at public sites.
Table 4-3: Phase B: Reasons for Recall

<table>
<thead>
<tr>
<th>REASONS FOR RECALL 2002</th>
<th>NUMBER</th>
<th>PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stellate</td>
<td>198</td>
<td>30.9%</td>
</tr>
<tr>
<td>Lesion NOS</td>
<td>154</td>
<td>24%</td>
</tr>
<tr>
<td>Asymmetry</td>
<td>152</td>
<td>23.8%</td>
</tr>
<tr>
<td>Asymmetry / Lesion NOS</td>
<td>56</td>
<td>8.8%</td>
</tr>
<tr>
<td>Asymmetry / Stellate</td>
<td>42</td>
<td>6.6%</td>
</tr>
<tr>
<td>Lesion NOS / Stellate</td>
<td>38</td>
<td>5.9%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>640</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

### 4.2.1.2.1 Selection of Images: Phase C

Phase C consisted of 210 sets of control films for non-recalled women taken between 1/10/2002 – 19/12/2002; incorporating 35 sets of images each from 3 public (1 fixed and 2 mobiles) and 3 private sites. These units represented the sites where most mammograms were performed. Four sets displayed were removed for a clinic prior to assessment. The researcher was unaware that they had been removed until after panel assessment – therefore these sets were not replaced.

### 4.3 RESULTS FOR LEVEL OF AGREEMENT

#### 4.3.1 Pilot Study (Phase A)

This section reports the results for Phase A with regards to the level of agreement for both PGMI and Adequate/Inadequate (A/I).

#### 4.3.1.1 PGMI Results

The individual panel member grading for each set of images is given in Table 4-4; five sets were removed and returned for clinics during the assessment period (shaded in dark blue). As none of these images received a rating of Inadequate by other panel members, it was not thought necessary to follow up their assessment.

For the 343 ((58x6) - 5)) sets of images assessed, the majority were ranked as M, with 87% being assessed as either G (42.3%) or M (44.8%) (Figure 4-3).
### Table 4-4: Assessment by Panel Members

<table>
<thead>
<tr>
<th>Film Set</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>G</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>5</td>
<td>G</td>
<td>I</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>G</td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>I</td>
<td>G</td>
<td>I</td>
<td>G</td>
<td>I</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>P</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>G</td>
</tr>
<tr>
<td>8</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>9</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>M</td>
</tr>
<tr>
<td>10</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>13</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>P</td>
</tr>
<tr>
<td>14</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>15</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>16</td>
<td>G</td>
<td>I</td>
<td>M</td>
<td>G</td>
<td>G</td>
<td>P</td>
</tr>
<tr>
<td>17</td>
<td>G</td>
<td>I</td>
<td>I</td>
<td>G</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>18</td>
<td>G</td>
<td>I</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>G</td>
</tr>
<tr>
<td>19</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td>G</td>
<td>I</td>
<td>M</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>21</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>M</td>
</tr>
<tr>
<td>22</td>
<td>M</td>
<td>I</td>
<td>I</td>
<td>M</td>
<td>I</td>
<td>M</td>
</tr>
<tr>
<td>23</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>24</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>G</td>
<td>M</td>
</tr>
<tr>
<td>25</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>I</td>
</tr>
<tr>
<td>26</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>27</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>28</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>29</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>30</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>M</td>
</tr>
<tr>
<td>31</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>32</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>G</td>
<td>M</td>
</tr>
<tr>
<td>33</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>34</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>35</td>
<td>G</td>
<td>P</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>36</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>37</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>I</td>
<td>M</td>
</tr>
<tr>
<td>38</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>I</td>
</tr>
<tr>
<td>39</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>G</td>
<td>M</td>
</tr>
<tr>
<td>40</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>41</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>42</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>P</td>
<td>G</td>
<td>M</td>
</tr>
<tr>
<td>43</td>
<td>G</td>
<td>P</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>G</td>
</tr>
<tr>
<td>44</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>I</td>
</tr>
<tr>
<td>45</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>46</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>47</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>48</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>49</td>
<td>G</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>50</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>I</td>
</tr>
<tr>
<td>51</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>52</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>53</td>
<td>G</td>
<td>I</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>M</td>
</tr>
<tr>
<td>54</td>
<td>G</td>
<td>I</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>I</td>
</tr>
<tr>
<td>55</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>56</td>
<td>G</td>
<td>M</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>57</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>I</td>
</tr>
<tr>
<td>58</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>P</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>G</td>
<td>42</td>
<td>22</td>
<td>12</td>
<td>39</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>M</td>
<td>15</td>
<td>22</td>
<td>40</td>
<td>13</td>
<td>34</td>
<td>30</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>
A histogram for the frequency profile of P, G, M and I for each panel member clearly shows the differences between assessors preference for grading of images (Figure 4-4). Some appear to be similar (R1/R4 and R5/R6), while others are very diverse (R1/R3); though in itself these profiles do not give a measure of the level of agreement between panel members for individual images.

Figure 4-3: Phase A: Distribution of Grades

Figure 4-4: Phase A: Individual Reader Assessment
4.3.1.2 Panel Assessment – PGMI Level of Agreement

The level of agreement for individual sets of images was assessed for the panel and varied between 30% (R3/R4) and 81% (R3/R5) (Figure 4-5). The majority of panel pairs (10/15) obtained less than 50% agreement (red horizontal dotted line).

Reassessing the profiles, some panel members gave similar percentages of P, G, M and I (Figure 4-4); for example R1 and R4 both ranked the majority of images as G (72%, 67%), with fewer M (26%, 22%) and even less P/I. Their overall level of agreement was 45.5% (Figure 4-5); illustrating only a moderate level of agreement.

In contrast, panel members R2 and R3 also obtained a level of agreement of 51%, however their profiles were quite different; with panel member R2 giving equal preference (38%) to G/M and R3 giving the majority (69%) of images a M grading.

In addition to the assessment of the level of agreement, the Kappa statistic was also calculated (Table 4-5). The overall Kappa value for individual pairs ranged from 0.13 to 0.64 with the highest level of agreement between R3/R5. The overall Kappa was 0.11 with the mammographers (R3, R4, R5, R6) Kappa being 0.20.
Table 4-5: Phase A: Kappa Statistics PGMI

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>1</td>
<td>0.05</td>
<td>0.1</td>
<td>-0.13</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>R2</td>
<td>1</td>
<td>0.22</td>
<td>-0.03</td>
<td>0.1</td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>R3</td>
<td>1</td>
<td>0</td>
<td>0.64</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>1</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>R5</td>
<td>1</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R6</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall Kappa for all six readers – 0.11
Overall Kappa for mammographers (R3 – R6) – 0.20

4.3.1.3 Panel Assessment – Adequate / Inadequate Level of Agreement

When the categories were dichotomised into Adequate (P, G and M) and Inadequate (I), the level of agreement (Figure 4-6) significantly improved (45.5% to 84.9% p < 0.0001) when assessed using the Wilcoxon matched-pairs two-tailed test.

![Figure 4-6: Phase A: Level of Agreement A/I](image)

The Kappa statistics also improved for most pairs (9/15), however the difference was not significantly different (p = 0.27) when assessed using the Wilcoxon matched-pairs two-tailed test (Table 4-6). A comparison of the Kappa statistics for PGMI versus A/I is shown in Table 4-7.
Table 4-6: Phase A: Kappa Statistics A/I

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R2</td>
<td></td>
<td>0.28</td>
<td>-0.1</td>
<td>0.24</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td></td>
<td></td>
<td>0.16</td>
<td>1</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>R6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4-7: Phase A: Comparison for Kappa PGMI & Kappa A/I

<table>
<thead>
<tr>
<th>Pairs</th>
<th>Kappa (PGMI)</th>
<th>Kappa (A/I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1/R2*</td>
<td>0.05</td>
<td>0</td>
</tr>
<tr>
<td>R1/R3*</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>R1/R4</td>
<td>-0.13</td>
<td>0</td>
</tr>
<tr>
<td>R1/R5*</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>R1/R6*</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>R2/R3</td>
<td>0.22</td>
<td>0.28</td>
</tr>
<tr>
<td>R2/R4*</td>
<td>-0.03</td>
<td>-0.1</td>
</tr>
<tr>
<td>R2/R5</td>
<td>0.1</td>
<td>0.24</td>
</tr>
<tr>
<td>R2/R6*</td>
<td>0.1</td>
<td>0.03</td>
</tr>
<tr>
<td>R3/R4</td>
<td>0</td>
<td>0.16</td>
</tr>
<tr>
<td>R3/R5</td>
<td>0.64</td>
<td>1</td>
</tr>
<tr>
<td>R3/R6</td>
<td>0.35</td>
<td>0.42</td>
</tr>
<tr>
<td>R4/R5</td>
<td>0.07</td>
<td>0.14</td>
</tr>
<tr>
<td>R4/R6</td>
<td>0.04</td>
<td>0.24</td>
</tr>
<tr>
<td>R5/R6</td>
<td>0.36</td>
<td>0.4</td>
</tr>
</tbody>
</table>

* Signifies a decrease in Kappa from PGMI – A/I

4.3.1.4 Discussion and Consensus Agreement

Initially, 20 (34.5%) of the 58 sets were recorded as inadequate by at least one panel member; no mammogram was classified as inadequate by all 6 members (Table 4-8). The panel reviewed these 20 mammograms and a consensus reached where only 7 (12%) remained inadequate (shaded in orange).
Table 4-8: Phase A: Assessment by Panel Members

<table>
<thead>
<tr>
<th>Film Set</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
<th>Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>M</td>
<td>I</td>
</tr>
<tr>
<td>5</td>
<td>G</td>
<td>I</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>I</td>
<td>I</td>
<td>G</td>
<td>I</td>
<td>G</td>
<td>I</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>A</td>
</tr>
<tr>
<td>16</td>
<td>G</td>
<td>I</td>
<td>M</td>
<td>G</td>
<td>G</td>
<td>P</td>
<td>A</td>
</tr>
<tr>
<td>17</td>
<td>G</td>
<td>I</td>
<td>I</td>
<td>G</td>
<td>I</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>18</td>
<td>G</td>
<td>I</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td>G</td>
<td>I</td>
<td>M</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>22</td>
<td>M</td>
<td>I</td>
<td>I</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>I</td>
</tr>
<tr>
<td>25</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>27</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>A</td>
</tr>
<tr>
<td>37</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>I</td>
</tr>
<tr>
<td>38</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>44</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>49</td>
<td>G</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>50</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>G</td>
<td>M</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>53</td>
<td>G</td>
<td>I</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>M</td>
<td>A</td>
</tr>
<tr>
<td>54</td>
<td>G</td>
<td>I</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>56</td>
<td>G</td>
<td>M</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>57</td>
<td>G</td>
<td>M</td>
<td>M</td>
<td>I</td>
<td>M</td>
<td>I</td>
<td>A</td>
</tr>
</tbody>
</table>

Of the 13 reclassified adequate: 11 although not aesthetically pleasing (including some films of low optical density), were considered diagnostic by the majority of the panel and 2 had been initially considered inadequate on criteria not directly related to image quality (non-symmetry of images and part of patient identification details obscured). One set of images (film set 17) received 4 inadequate ratings for “saggy” MLO’s. After discussion for this particular set, although not considered good practice, the set was reclassified as adequate. Reasons for inadequate images are discussed in detail in section 4.4.

Two film sets (3, 37) remained inadequate although only one panel member had classed them as I. One set was underexposed and had screen artefacts that detracted from the images and the other set had missing breast coverage for the left MLO.
The discussion section during Phase A of PGMI revealed each readers biases. R1 indicated a reluctance to assess a set as inadequate as these images had already been reviewed by two, and sometimes three, other screen readers. R2 routinely assessed CC projections as being inadequate for coverage laterally, even though all the tissue was imaged on the medial aspect and in the MLO projection (according to PGMI standards (Table 2-8: PGMI Evaluation of Clinical Image Quality (NAS, 2001) (40), CC images are ranked as G as long as all postero-medial tissue is visualized and the axillary portion of the breast is not to be included at the expense of the medial portion). R3 and R5 commented when faced with an image that was borderline G or M, they veered toward M; while R4 commented when faced with the same dilemma, they veered towards G; this may have subconsciously been influenced by the NAS guidelines which states 50% of images per mammographer per year need to be classified as P/G, with the aim of 75%. R6 tended to assess any image as inadequate for skin folds, even though the skin folds may not have obscured any information.

In general, the mammographers appeared to be more critically aware of clientele presentation and how body habitus influences PGMI than the screen readers. The mammographers commented that some images awarded M were better images than some awarded G. For example, any MLO that does not have pectoral muscle to the nipple is considered M according to PGMI standards, even if the breast tissue is all imaged, pulled up, out, well compressed and correctly exposed. For some clients, their body habitus would always exclude pectoral muscle being seen to the nipple.

The discussion during Phase A could be classed as an informal “training” session, where the panel members were given the opportunity to standardize their understanding and use of PGMI. This led to a more systematic plan and a revised PGMI (Table 3-3: Modified PGMI Sheet, section 3.4.4.1, p. 89) for what would be considered adequate and diagnostic (i.e. agreement over which skin folds obscure, how much motion artefact is acceptable and in what situations, and the recognition that acceptable optical density had increased from images used in this research to current images – from 1.4-1.8 to 1.6-2.0).
4.3.2 MAIN STUDY (PHASE B)

This section reports the results for Phase B, the main study with regards to the level of agreement for both PGMI and A/I.

4.3.2.1 PGMI Results

For the total number of sets assessed by panel members (640 sets read by 5 readers and 639 sets read by 1 reader = 3839 sets), the majority were ranked as G, with 94% being assessed as either G or M (Figure 4-7). The number of images considered inadequate in Phase B (5.3%) was almost half of those considered inadequate in Phase A (10.5%). These values indicate that the panel members assessed this set of images to be of better quality than those assessed in Phase A (Figure 4-8), and this difference was significant at the 95% level (p < 0.001) when assessed using chi-square test for independence.

![Figure 4-7: Phase B: Distribution of Grades](image)

4.3.2.2 Panel Assessment – PGMI Level of Agreement

A histogram for the frequency profile of P, G, M and I for each panel member indicates the similarity for grading of this phase of images (Figure 4-9). This is different from the profile seen in Phase A (Figure 4-4), and could be a result of the informal training discussion after Phase A.
The level of agreement for individual sets of images was assessed for the panel and was seen to vary between 50% (R1/R6) and 65% (R4/R5) (Figure 4-10). In contrast to Phase A, no panel pairs obtained less than 50%. While the range (50% to 65%; SD ± 5.4) in agreement between pairs was lower in Phase B compared to Phase A (30% to 81%; ± SD 13.7), the overall level of agreement significantly increased in Phase B (45.5% to 57.7%, p = 0.0043) when assessed using the Wilcoxon matched-pairs two-tailed test.
The Kappa statistic confirms this with significantly higher values of overall Kappa (0.11 to 0.21, \( p = 0.05 \)) than Phase A (Table 4-9). Individual pairs for the mammographers ranged from 0.27 to 0.4 in Phase B, with the overall Kappa for the mammographers increasing between Phases A and B (0.20 to 0.34, \( p = 0.43 \)), though this was not significant when tested with the paired two-tailed t-test.

![Figure 4-10: Phase B: Level of Agreement PGMI](image)

Table 4-9: Phase B: Kappa Statistics PGMI

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>1</td>
<td>0.07</td>
<td>0.11</td>
<td>0.05</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>1</td>
<td>0.25</td>
<td>0.25</td>
<td>0.31</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>1</td>
<td></td>
<td>0.35</td>
<td>0.35</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>1</td>
<td></td>
<td></td>
<td>0.37</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>R6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Overall Kappa for all six readers – 0.21
Overall Kappa for mammographers (R3-R6) – 0.34
4.3.2.3 Panel Assessment – Adequate / Inadequate Level of Agreement

When the categories were dichotomised into Adequate (P, G and M) and Inadequate (I), the average level of agreement significantly increased (57.7% to 92.4%, p < 0.0001) when assessed using the Wilcoxon matched-pairs two-tailed test, with all 15 pair combinations agreeing over 85% (Figure 4-11). In progressing from Phase A to Phase B, the A/I level of agreement also significantly increased (84.9% to 92.4%, p = 0.001).

![Figure 4-11: Phase B: Level of Agreement A/I](image)

The Kappa statistic for A/I (0.1 to 0.34) (Table 4-10) showed a slight increase in consistency among panel members changing from PGMI (0.05 to 0.37), though this was not significant at the 95% level (p = 0.26) (Table 4-11). There was also no significant increase (p = 0.51) between Phase A (-0.1 to 0.42) and Phase B (0.1 to 0.34) for the dichotomised classification.
### Table 4-10: Phase B: Kappa Statistics A/I

<table>
<thead>
<tr>
<th>Phase B – A/I</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>1</td>
<td>0.16</td>
<td>0.15</td>
<td>0.1</td>
<td>0.15</td>
<td>0.11</td>
</tr>
<tr>
<td>R2</td>
<td>1</td>
<td>0.28</td>
<td>0.22</td>
<td>0.27</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>1</td>
<td>0.34</td>
<td>0.34</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td></td>
<td>0.33</td>
<td></td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td></td>
<td></td>
<td>1</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

### 4.3.2.4 Consensus Assessment

Initially 129 (20.2%) of the 640 sets were considered inadequate by at least one of the panel, with three mammograms classified as inadequate by all six panel members. Using consensus agreement where three or more of the readers assessed a set as inadequate reduced the number of I’s to 19 (3%). There is a possibility that this number of inadequate images is underrepresented, as in Phase A during the panel discussion, 2 sets remained inadequate although only one panel member had assessed them as such. There was no panel discussion after Phase B, though there were continued informal discussions on the use of PGMI throughout the research.

### Table 4-11: Phase B: Comparison for Kappa PGMI & Kappa A/I

<table>
<thead>
<tr>
<th>Pairs</th>
<th>Kappa (PGMI)</th>
<th>Kappa (A/I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1/R2</td>
<td>0.07</td>
<td>0.16</td>
</tr>
<tr>
<td>R1/R3</td>
<td>0.1</td>
<td>0.15</td>
</tr>
<tr>
<td>R1/R4</td>
<td>0.11</td>
<td>0.1</td>
</tr>
<tr>
<td>R1/R5</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>R1/R6</td>
<td>0.1</td>
<td>0.11</td>
</tr>
<tr>
<td>R2/R3</td>
<td>0.25</td>
<td>0.28</td>
</tr>
<tr>
<td>R2/R4*</td>
<td>0.25</td>
<td>0.22</td>
</tr>
<tr>
<td>R2/R5*</td>
<td>0.31</td>
<td>0.27</td>
</tr>
<tr>
<td>R2/R6*</td>
<td>0.22</td>
<td>0.15</td>
</tr>
<tr>
<td>R3/R4*</td>
<td>0.35</td>
<td>0.34</td>
</tr>
<tr>
<td>R3/R5*</td>
<td>0.35</td>
<td>0.34</td>
</tr>
<tr>
<td>R3/R6</td>
<td>0.27</td>
<td>0.22</td>
</tr>
<tr>
<td>R4/R5*</td>
<td>0.37</td>
<td>0.33</td>
</tr>
<tr>
<td>R4/R6*</td>
<td>0.33</td>
<td>0.28</td>
</tr>
<tr>
<td>R5/R6*</td>
<td>0.34</td>
<td>0.31</td>
</tr>
</tbody>
</table>

* Signifies a decrease in Kappa from PGMI – A/I
4.3.3 CONTROL COMPONENT (PHASE C)
This section reports the results of the 206 film sets used in Phase C, the control (non-recalled) component, with regards to the level of agreement for both PGMI and A/I.

4.3.3.1 PGMI Results
For the 1236 sets assessed by panel members (206 x 6), the majority were ranked as M, with 95.1% being assessed as either G or M (Figure 4-12). The number of images considered inadequate in Phase C (4.6%) was similar to that of Phase B (5.3%). The panel members assessed this set of images to be worse overall than those assessed in both Phases A and Phase B (Figure 4-13), and these differences were significant at the 95% level ($p < 0.001$) when assessed using chi-square test for independence.

![Figure 4-12: Phase C: Distribution of Grades](image1)

![Figure 4-13: All Phases: Comparison of Grades](image2)
4.3.3.2 Panel Assessment – PGMI Level of Agreement

A histogram for the frequency profile of P, G, M and I for each panel member is given in Figure 4-14. The panel members’ profiles appear similar with the exception of R1 (who gave a majority of M’s (96%)).

The level of agreement for individual sets of images was assessed for the panel and was seen to vary between 43% (R1/R2) and 67% (R3/R5) (Figure 4-15). Thirteen panel pairs obtained more than a 50% agreement (red horizontal dotted line). While the range (43% to 67%; SD ± 6.3) between pairs was slightly larger than that in Phase B (50% to 65%; ± SD 5.4), the overall level of agreement was not significantly different between the two phases (57.7% to 52.2%, p = 0.72) when assessed using the paired two-tailed t-test.

The overall Kappa significantly reduced between Phases B and C (0.21 to 0.18, p = 0.02) when tested with the paired two-tailed t-test (Table 4-12). While higher than Phase A (0.11, 0.21, p = 0.52) the difference was not significant at the 95% level when assessed with the Wilcoxon matched-pairs two-tailed test.

Individual pairs for the mammographers ranged from 0.14 to 0.38 in Phase C, with the overall Kappa for the mammographers increasing between Phases B and C (0.34 to 0.38, p = 0.17), though this was not significant when tested with the paired two-tailed t-test.
4.3.3.3 Panel Assessment – Adequate / Inadequate Level of Agreement

When the categories were dichotomised into Adequate (P, G and M) and Inadequate (I), the average level of agreement significantly increased (57.2% to 92.5%, \( p < 0.0001 \)) when assessed using the Wilcoxon matched-pairs two-tailed test, with all 15 pair combinations agreeing over 85% (Figure 4-16). In progressing from Phase B to Phase C, the A/I level of agreement remained constant (92.4% to 92.5%, \( p = 0.97 \)).
The Kappa statistic for A/I (-0.03 to 0.52) (Table 4-13) showed a decrease in consistency among panel members changing from PGMI (-0.03 to 0.38), though this was not significant at the 95% level (p = 0.06) (Table 4-15). However, there was a significant decrease (p = 0.017) between Phase B (0.1 to 0.34) and Phase C (-0.03 to 0.52) for the dichotomised classification.

Table 4-13: Phase C: Kappa Statistics A/I

<table>
<thead>
<tr>
<th>Phase C – A/I</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R2</td>
<td>1</td>
<td>-0.03</td>
<td>0.22</td>
<td>-0.03</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>1</td>
<td>0.24</td>
<td>0.52</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>1</td>
<td>0.26</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td>1</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R6</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4-14: Phase C: Comparison for Kappa PGMI & Kappa A/I

<table>
<thead>
<tr>
<th>Pairs</th>
<th>Kappa (PGMI)</th>
<th>Kappa (A/I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1/R2*</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td>R1/R3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R1/R4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R1/R5*</td>
<td>0.03</td>
<td>0</td>
</tr>
<tr>
<td>R1/R6</td>
<td>-0.03</td>
<td>0</td>
</tr>
<tr>
<td>R2/R3*</td>
<td>0.06</td>
<td>-0.03</td>
</tr>
<tr>
<td>R2/R4*</td>
<td>0.35</td>
<td>0.22</td>
</tr>
<tr>
<td>R2/R5*</td>
<td>0.28</td>
<td>-0.03</td>
</tr>
<tr>
<td>R2/R6*</td>
<td>0.23</td>
<td>0.16</td>
</tr>
<tr>
<td>R3/R4*</td>
<td>0.36</td>
<td>0.24</td>
</tr>
<tr>
<td>R3/R5</td>
<td>0.38</td>
<td>0.52</td>
</tr>
<tr>
<td>R3/R6*</td>
<td>0.27</td>
<td>0.11</td>
</tr>
<tr>
<td>R4/R5</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>R4/R6*</td>
<td>0.26</td>
<td>0.19</td>
</tr>
<tr>
<td>R5/R6</td>
<td>0.14</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Signifies a decrease in Kappa from PGMI – A/I

4.3.3.4 Consensus Agreement

Initially 36 (17.5%) of the 206 sets were considered inadequate by at least one of the panel; no mammogram was classified as inadequate by all six panel members. Using a consensus with 3 or more of the readers assessing a set as inadequate, the number of I’s reduced from 36 to 4 out of the 206 sets (1.9%).

4.3.4 Effect of Training on Panel Member Consistency

This research assessed inter and intra-observer consistency over several phases of the study. The results of the inter-observer consistency have already been reported. The panel members were unaware of the intra-observer component of the research.

4.3.4.1 Intra-observer Consistency

Intra-observer consistency was assessed twice; first to assess how training altered the panel members assessment of images (30 sets from Phase A integrated into Phase B – known as A*) and second, to assess post-training consistency (62 sets from Phase B integrated into Phase C – known as B*).
The level of agreement of the 30 sets from Phase A and Phase A\(^*\) for individual sets of images was assessed and was seen to vary between 43.3\% (R2) and 63.3\% (R3) (Figure 4-17). Only one panel member obtained less than a 50\% intra-level agreement. The level of agreement of the 62 sets from Phase B and Phase B\(^*\) was generally higher than Phase A\(^*\), with the exception of R1. The values were seen to vary between 29\% (R1) and 72.6\% (R4). This difference was not considered significant (\(p=0.56\)) when tested with the Wilcoxon matched-pairs two-tailed test. Again, only one panel member obtained less than a 50\% intra-level agreement (red horizontal dotted line) in this phase.

Figure 4-18 and Figure 4-19 show the effect of pre and post training with the PGMI classification for each panel member. The screen readers (R1, R2) both showed a significantly low level of agreement pre-training (A\(^*\)) (R1 \(p=0.02\); R2 \(p=0.0016\)) and screen reader R1 also showed a low level of agreement post-training (B\(^*\)) (R1 \(p=0.0001\)). The mammographers showed a significantly high level of agreement both pre and post training, with the exception of R3 who showed a significantly low level of agreement post-training (pre Kappa 0.41, post Kappa 0.33).

Kappa statistics for these phases is shown in Table 4-15. With the exceptions of R1 and R3, the post training (Phase B\(^*\)) intra-observer consistency (0.00 to 0.56) was either equal or better than pre-training (Phase A\(^*\)) intra-observer consistency (0.22 to 0.41).
Figure 4-18: Phase A / A*: Effect of Training
Figure 4-19: Phase B / B*: Post-Training Effect
Table 4-15: Kappa of Intra-observer Consistency Based on Four Categories

<table>
<thead>
<tr>
<th></th>
<th>Phase A* Kappa</th>
<th>Phase B* Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.23</td>
<td>0.00</td>
</tr>
<tr>
<td>R2</td>
<td>0.22</td>
<td>0.39</td>
</tr>
<tr>
<td>R3</td>
<td>0.41</td>
<td>0.33</td>
</tr>
<tr>
<td>R4</td>
<td>0.37</td>
<td>0.56</td>
</tr>
<tr>
<td>R5</td>
<td>0.33</td>
<td>0.38</td>
</tr>
<tr>
<td>R6</td>
<td>0.31</td>
<td>0.50</td>
</tr>
</tbody>
</table>

4.4 INADEQUATE IMAGES

All inadequate images in this study were due to failure to obtain images classified as adequate, as defined by PGMI. As discussed in the methodology (section 0, p. 89), the PGMI requirements were condensed into 5 broad categories (artefacts, coverage, exposure, motion and skin folds).

4.4.1 PERCENTAGE OF INADEQUATE IMAGES

In Phase A, 12% of the images were considered inadequate by consensus, although in Phases B and C, only 3% and 2% respectively were classed as inadequate by consensus (Figure 4-20). The reduction in number of inadequate images from Phase A to Phases B and C were significant ($p = 0.0005$) using Chi-Square. Due to the small number of overall inadequate images taken in Phases A, B and C, the data was combined to ensure good statistical power.

Of the 904 sets assessed (Phases A, B & C) there were 30 (3.3%) sets containing at least 1 inadequate image. As each of these sets contained four projections, there were a possibility of 120 inadequate images; 80 (75%) were considered inadequate.

4.4.1.1 Distribution by Side and Projection

Of the 80 images considered inadequate, 46% occurred in the right breast and 54% in the left breast; 41% occurred in the CC projection and 59% in the MLO projections (Figure 4-21). The most common projection considered inadequate was the left MLO (30%) (Figure 4-22).
Figure 4-20: Percentage Inadequate for each Phase

Figure 4-21: All Phases: Inadequate Images by Side / Projection

Figure 4-22: All Phases: Inadequate Images by Combined Side and Projection
4.4.1.2 Distribution by Imaging Site

Table 4-16 shows the number of sets taken publicly and privately together with the number of inadequate images. Of the 904 sets of images 593 (66%) were taken at public site, while 311 (34%) were taken at private sites. There were similar numbers of images produced at public and private sites in Phases A and C, however there were over twice the number of images produced at public sites (72%) than private sites (28%) in Phase B (Figure 4-23). Given the higher inadequate rate in private sites, this difference might help explain in part the lower overall inadequate rates seen in Phase B compared to Phase A.

Out of the 593 sets of images taken at the public sites, 582 (98%) were adequate and only 2% were inadequate; out of the 311 sets of images taken at the private sites, 292 (94%) were adequate and 6% were inadequate (Figure 4-24). The number of inadequate images taken at the private sites was significantly higher \( p = 0.0013 \) than those produced at the public sites, when assessed using two-sided Fisher’s Exact.

Nineteen out of the 30 (63%) sets of inadequate images were taken at private sites, with only 37% of the inadequate sets being produced at public sites. Each set of 30 contained 4 images, with 80 images considered inadequate by the panel. From the 80 inadequate images, 53 (66%) were from private sites. The number of inadequate images taken at the private sites was significantly higher \( p = 0.0013 \) than those produced at the public sites, when assessed using two-sided Fisher’s Exact.

Table 4-16: All Phases: Inadequate Images by Public / Private Sites

<table>
<thead>
<tr>
<th>Phase</th>
<th>Number of sets taken (%)</th>
<th>Number of Inadequate Sets (%)</th>
<th>Public</th>
<th>Private</th>
<th>Total of Both</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Public</td>
<td>Private</td>
<td>Total</td>
<td>Public</td>
<td>Private</td>
</tr>
<tr>
<td>A (Pilot)</td>
<td>28 (48.3%)</td>
<td>30 (51.7%)</td>
<td>58</td>
<td>2 (7.1%)</td>
<td>5 (16.7%)</td>
</tr>
<tr>
<td>B (Main)</td>
<td>462 (72.2%)</td>
<td>178 (27.8%)</td>
<td>640</td>
<td>9 (1.9%)</td>
<td>10 (5.6%)</td>
</tr>
<tr>
<td>C (Control)</td>
<td>103 (50%)</td>
<td>103 (50%)</td>
<td>206</td>
<td>0 (0%)</td>
<td>4 (3.9%)</td>
</tr>
<tr>
<td>Phases Combined</td>
<td>593 (65.6%)</td>
<td>311 (34.4%)</td>
<td>904</td>
<td>11 (1.8%)</td>
<td>19 (6.1%)</td>
</tr>
</tbody>
</table>
4.4.2 REASONS FOR INADEQUATE IMAGES

Although there are 5 main reasons why images could be considered inadequate (exposure, motion, coverage, artefacts and skin folds); the varying combinations produced 17 reasons (Table 4-17). The reasons why images were classified as inadequate were mainly due to exposure and motion.
Exposure (alone and in combination) accounted for 56% of inadequate images; motion (alone and in combination) accounted for 36% of inadequate images. Coverage (alone and in combination) accounted for 30% of inadequate images; while folds and artefacts (alone and in combination) accounted for 17.5% and 11% of inadequate images respectively. Thus, underexposure and motion represents over 50% of the reasons why images were classified as inadequate; and 76% of inadequate images were associated with both / or either of these two reasons, along with a combination of coverage, artefacts and folds.

Table 4-17: All Phases: Reasons for Inadequate Images

<table>
<thead>
<tr>
<th>Reasons</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure</td>
<td>31.25</td>
</tr>
<tr>
<td>Motion</td>
<td>12.5</td>
</tr>
<tr>
<td>Exposure / Motion</td>
<td>12.5</td>
</tr>
<tr>
<td>Coverage</td>
<td>7.5</td>
</tr>
<tr>
<td>Artefacts</td>
<td>5</td>
</tr>
<tr>
<td>Coverage / Exposure</td>
<td>5</td>
</tr>
<tr>
<td>Coverage / Folds</td>
<td>5</td>
</tr>
<tr>
<td>Coverage / Motion</td>
<td>3.75</td>
</tr>
<tr>
<td>Folds</td>
<td>3.75</td>
</tr>
<tr>
<td>Exposure / Folds</td>
<td>2.5</td>
</tr>
<tr>
<td>Coverage / Motion / Folds</td>
<td>2.5</td>
</tr>
<tr>
<td>Coverage / Artefacts / Folds</td>
<td>2.5</td>
</tr>
<tr>
<td>Artefacts / Exposure</td>
<td>1.25</td>
</tr>
<tr>
<td>Coverage / Motion / Artefacts</td>
<td>1.25</td>
</tr>
<tr>
<td>Coverage / Exposure / Motion</td>
<td>1.25</td>
</tr>
<tr>
<td>Exposure / Motion / Artefacts</td>
<td>1.25</td>
</tr>
<tr>
<td>Coverage / Exposure / Motion / Folds</td>
<td>1.25</td>
</tr>
</tbody>
</table>
4.5 DO INADEQUATE IMAGES INFLUENCE CANCER DETECTION RATES

Although mammography is instrumental in reducing breast cancer, poor quality mammograms may fail to achieve the mortality reductions and have adverse consequences such as missed cancers, increased false positives, increased costs, anxiety and discomfort and an increased radiation burden for the women who undergo additional procedures.

To assess whether inadequate images influence cancer detection rates, both recalled and non-recalled images were compared in relationship to the assessment of adequate / inadequate and the presence or absence of cancer (Table 4-18). For the 640 recall clients, there were 85 surgeries performed. From 85 surgeries, there were 77 cancers detected (including 2 cancers at 3 month radiology review) therefore there were 8 false positive biopsies performed (Figure 4-25).

The difference for adequate / inadequate images in relation to cancer detection has a relative risk of 1.004 (0.9567 – 1.045). This difference was not considered significant (p = 1.000) when tested with the chi-square.

Table 4-18: Cancer Detection versus Adequate / Inadequate

<table>
<thead>
<tr>
<th>Image Quality</th>
<th>Adequate</th>
<th>Inadequate</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>75</td>
<td>2</td>
<td>77</td>
</tr>
<tr>
<td>No Cancer</td>
<td>546</td>
<td>17</td>
<td>563</td>
</tr>
<tr>
<td>Total</td>
<td>621</td>
<td>19</td>
<td>640</td>
</tr>
</tbody>
</table>
Figure 4-25: Phases B & C: Adequate / Inadequate with Cancer Detection
4.6 CASE CONTROLLED COMPONENT

A case-control study was undertaken to investigate the intrinsic factors contributing to why women are recalled. One control was matched to each case by same age and date of screening (within 4 months). Therefore, the matched case-control study was conducted using images obtained in Phases B and C (recall versus non-recall).

Case-control data included information gathered from the Data Collection Sheets such as client demographics (age), HRT status, breast density, image quality and the site of mammogram production. A total of 110 cases and 110 matched controls out of the 640 recalled and 206 controls were analysed. Conditional logistic regression was performed to obtain the estimates of odds ratio and 95% confidence intervals (Table 4-19).

Table 4-19: Case Control Data

<table>
<thead>
<tr>
<th></th>
<th>Recalled 110 Cases</th>
<th>Non-recalled 110 Controls</th>
<th>OR *</th>
<th>95% CI *</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HRT USE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>52</td>
<td>67</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous</td>
<td>25</td>
<td>24</td>
<td>1.34</td>
<td>0.69 – 2.62</td>
<td>0.40</td>
</tr>
<tr>
<td>Current</td>
<td>33</td>
<td>19</td>
<td>2.24</td>
<td>1.14 – 4.38</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>BREAST DENSITY</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>42</td>
<td>64</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>68</td>
<td>46</td>
<td>2.25</td>
<td>1.31 – 3.87</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>IMAGING SITES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>37</td>
<td>55</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>73</td>
<td>55</td>
<td>1.97</td>
<td>1.14 – 3.40</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>IMAGE QUALITY - READER 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>79</td>
<td>50</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate</td>
<td>25</td>
<td>52</td>
<td>0.30</td>
<td>0.17 – 0.55</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Inadequate</td>
<td>6</td>
<td>8</td>
<td>0.48</td>
<td>0.15 – 1.45</td>
<td>0.25</td>
</tr>
</tbody>
</table>

* OR - Odds ratio   * CI – Confidence Interval
4.6.1 Patient Demographics

4.6.1.1 Age

The youngest client was 40 years and the oldest was 79 years, with the majority (56%) being within the target age group (Figure 4-26).

![Figure 4-26: Case Control: Ages of Clients](chart)

4.6.1.2 Use of HRT

HRT status was self-reported and collected on the Data Sheet as “never used”, “previous user” and “current user”. The distribution between case and control shows similar distributions for “previous use”, with more controls “never” using HRT and more cases “currently” using HRT (Figure 4-27).

![Figure 4-27: Case Control: HRT Status](chart)
The use of HRT can increase breast density by arresting and often reversing the process of involution. The results show that compared to women who never took HRT, the estimated odds ratios assessing the probability of being recalled for those who were former HRT users and current HRT users are 1.34 (0.69 – 2.62) and 2.24 (1.14 – 4.38) respectively. Therefore, for current users of HRT, there is a significant likelihood (p = 0.02) of being recalled than for those clients who have either never or had previously used HRT (Table 4-19).

4.6.2 BREAST DENSITY

As discussed in Chapter 2, studies have shown that clients with high (dense) breast tissue have a 4 to 6 times increased risk of being diagnosed with breast cancer. Dense breast tissue is associated with poor mammographic sensitivity and higher recall rates than for clients with fatty breast tissue.

Two categories of breast density were calculated; low density consisted of breast tissue composed mainly or mostly of fat (Wolfe’s classification N1 and P1), while high density consisted of breast tissue with densities greater than fat (Wolfe’s classification P2 and DY) (76). The distribution of low / high breast density for case and control is shown in Figure 4-28.

The results show that compared to women with low mammographic density, the estimated odds ratio for women with high mammographic density is 2.25 (1.31 – 3.87). Thus, women with denser breasts are significantly (p = 0.004) more likely to be recalled compared to women with low breast density.

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>42</td>
<td>64</td>
</tr>
<tr>
<td>High</td>
<td>68</td>
<td>46</td>
</tr>
</tbody>
</table>

Figure 4-28: Case Control: Breast Density
4.6.3 IMAGING SITES (PUBLIC / PRIVATE)

The site of mammogram production included private and public sectors; the control contained equal amounts from public / private sites (55, 55); the case contained twice the number of public sites (73, 37) (Figure 4-29).

Compared to women who obtained their images at a private site, the estimated odds ratio for those who obtained their images through a public site was 1.97 (1.14 – 3.40; p = 0.02). Thus, women being radiographed at a public site are twice as likely to be recalled compared to women radiographed at a private site; this was significant at the 95% level. One explanation might be that the public sites include the mobile vans; these vans visit rural areas and may be the only facilities that clients use (a client with a breast concern may wait for the van instead of attending her GP).

![Figure 4-29: Case Control: Site of Production](image)

4.6.4 IMAGE QUALITY

As there were very few films classified as P and I for all readers, three categories were created for image quality; good (P+G), adequate (M) and poor (I). As R4 had the most consistent intra-observer reliability post training (Figure 4-19 and Table 4-15), the results from this panel member were used to tabulate the case controlled component of the research (Figure 4-30).

Compared to good quality images, the estimated odds ratios for adequate and poor quality images are 0.30 (0.17 – 0.55) and 0.48 (0.15 – 1.45) respectively. In comparison to good quality images, there is a 70% and 52% lower likelihood of being recalled for adequate or poor images respectively. There was a significant difference (p < 0.0001) comparing good and adequate images; however because of the small number of inadequate images, a Type II error is thought to have occurred when assessing the significance between good and inadequate images.
4.7 SUMMARY

This research assessed both non-recalled images and images recalled for asymmetry, stellate lesions and lesions not otherwise specified. After assessing each set for image quality, inter and intra-consistency was measured for a panel of six readers. Initially, only 5 out of 15 pairs obtained an inter-consistency agreement of 50% or more; this increased to 13 of the 15 pairs agreeing at least 50% of the time. Intra-consistency for post-training was either equal or better than pre-training for the majority of panel members (exceptions being R1 and R3) indicating the positive effect of the training.

During the pilot phase, 12% of images were considered inadequate by consensus; although only 3% and 1.9% were considered inadequate by consensus during Phases B and C, respectively. The major reasons for inadequate images were under-exposure and motion; two-thirds of the images were produced at public sites and accounted for over a third (37%) of the total inadequate images.

A case controlled component of the research was undertaken to evaluate reasons affecting recall. Factors increasing the likelihood of being recalled included HRT use, having dense breast tissue, being screened through the public system and good quality images. Therefore inadequate images did not contribute to an increased likelihood of being recalled. An assessment was also undertaken and found that inadequate image quality did not influence cancer detection rates. The discussion and analysis for these results are given in Chapter 5.
5 CHAPTER FIVE: DISCUSSION
5.1 INTRODUCTION
The research examined 904 sets of mammograms from recalled (698) and non-recalled (206) screening clients x-rayed through Hunter BreastScreen during 2002-2003. A six member panel used the established PGMI classification to assess image quality for these sets of images. This chapter discusses the main findings of the research, relates it to existing research and identifies limitations of the research.

5.2 HOW OBJECTIVE ARE CLASSIFICATION SYSTEMS?
5.2.1 USE OF PGMI
The results gathered from using the PGMI system confirmed that PGMI classification is a subjective, rather than objective, measure of image assessment. Phase A consistency for the 15 pair sets of panel members using this system was low (overall Kappa 0.11, range -0.13-0.64, overall agreement 45.5%); though for the 6 pair sets of mammographers who routinely use the system, the Kappa statistic was slightly higher (overall Kappa 0.20, range 0-0.64, overall agreement 51%).

A previous study conducted by Moreira et al (212) also supports the view that PGMI is a subjective standard. Their study compared the validity and reliability of PGMI to another system (Excellent, Adequate, Repeat – EAR) for the assessment of image quality. Thirty sets of mammograms were assessed by 21 experienced PGMI users (tutor mammographers); Kappa for PGMI was between 0.13-0.19 (overall agreement for CC 76.5% and for MLO 77.2%) and Kappa for EAR was between -0.01-0.04 (overall agreement for CC 75.8% and for MLO 74.1%) (212). While the level of agreement was higher in the Moreira study, Kappa results are similar to the current study.

The Breast Imaging Reporting and Data System (BI-RADS) was developed to standardize methods for evaluating breast density, breast lesions and interpretation; however, studies have shown only moderate agreement for its use with the four categories of breast density (204, 207), the descriptors of breast lesions (203-206) and mammographic interpretation (197, 202).

Non-mammography studies have also demonstrated low agreement in comparing classification systems for distal radius fracture (208) to fair / moderate agreement for treatment modalities in wrist fractures (193). Andersen et al expressed concern that given the low degree of inter-observer agreement for each of the study classifications, their use as a sole means of determining the direction of the treatment or for the direct comparison of results between different studies was not reliable (208).
Decreasing either the number of observers or the number of categories generally increases the rate of agreement. When the categories were converted to an Adequate (PGM) / Inadequate (I) system, all members agreed over 70% of the time for Phase A and over 85% of time for both Phases B and C. With the conversion to A/I during Phase A the overall Kappa improved for most pairs (0.19, range -0.1 to 1), however the difference was not statistically significant.

In contrast, Moreira et al reported that the validity of EAR, shown by mean kappa values, was generally poorer than that of PGMI (212), as were the levels of agreement. This seems to be a counter intuitive result with no explanation given in the paper. When the results were dichotomised as “acceptable” or “needs repeating”, overall levels of agreement for the PGMI were CC 86.4%, MLO 90.4% and for EAR were CC 86.3%, MLO 70.9%. In a study by Ciatto et al, when BI-RADS four grade scale for breast density was converted to a two grade scale, inter-observer consistency ranged from moderate (k=0.54, range 0.02-0.77) to substantial (k=0.71, range 0.31-0.88) (207).

While the level of agreement between readers is higher for a two system classification, this research does not advocate an Adequate / Inadequate system of grading for mammograms. There is the possibility that this may decrease the quality of images, with staff not striving to obtain ‘perfect’ image quality, but rather just accept ‘moderate’ image quality. The National Accreditation Standards for BreastScreen Australia (NAS) guidelines aim to promote excellence by requiring that a minimum of 50% of a random audit of 50 film sets per mammographer are graded P and / or G categories with 75% being desirable (40).

For the current study, approximately thirty mammographers with varying experience produced the images assessed. Phase B was the only phase to contain over 50% P and / or G images (53.58%) (for 640 sets of mammograms assessed). Phase A contained 44.6% (for 58 sets), while Phase C contained 36.65% (for 206 sets) P and / or G images. One screen reader (R1) did modify their pattern of assessing images (from primarily G’s in Phases A and B, to mainly all M’s in Phase C); contributing to the lower overall percentage in Phase C. The similar study by Moreira et al, demonstrated that 54% of the 30 sets of mammograms assessed were considered P and / or G grading (212).
5.2.2 TRAINING IN CLASSIFICATION SYSTEMS

Training in any classification systems is known to increase agreement amongst users. During the discussion after Phase A, all panel members reported becoming aware of their biases. One member gave any image an inadequate rating for skin folds, whether they were obscuring relevant anatomy or not. According to PGMI requirements, an image should only be given an inadequate rating for skin folds which are obscuring breast tissue in the image (40). The goal of a mammographer is to obtain high quality images at the lowest dose possible. This means accepting films that, while not perfect, are diagnostic. If a technical problem is a recurrent one for an individual mammographer (such as “sloppy” folds), then this should be reported back to the mammographer and extra training provided for that individual. This is the standard across BreastScreen, but may not always be practiced.

As expected, the level of agreement increased in the present study after the group discussion. During Phase A, only 5 of the 15 pair combinations agreed more than 50% of the time, while Phase B demonstrated 14 of the 15 pair combinations agreeing more than 50% of the time. Nearly 3 years after training, there was a slight, but not significant, decrease in agreement, where 13 of the 15 pairs agreed more than 50% of the time. This is consistent with other studies which have shown that while the level of agreement increases after training; it decreases again as time progresses.

In the current research, the discussion after Phase A was believed to be responsible for significantly increasing consistency among panel members. The overall Kappa for all panel members post-training (Phases B and C) was 0.21 (range 0.1-0.37) and 0.18 (-0.03-0.38) respectively. These values improved from Phase A – pre-discussion (0.11 range 0.13-0.64). For the mammographers, the overall Kappa increased from 0.20 (Phase A) to 0.34 (Phase B) and 0.28 (Phase C) respectively. Phase A assessment was undertaken from December 2003-January 2004, while Phase B occurred from September 2005-September 2006 and Phase C from September 2006-January 2007. While one mammographer pair (R3/R5) had substantial agreement (0.64) during Phase A, with an increase in the number of sets assessed for Phases B and C, their Kappa decreased to moderate (0.35 and 0.38 respectively).

Other studies have demonstrated that training in classification systems can improve accuracy (206, 213). After attending a day lecture on BI-RADS, 23 mammography screen readers improved agreement for descriptors of breast lesions (for mass
margins and asymmetries) from pre-training Kappa of 0.36 (range 0.07-0.61) to post-training Kappa of 0.41 (range 0.06-0.54) (206). Pre and post-training assessment were conducted 2-3 months apart. Long term improvements in accuracy were not assessed. Another study demonstrated that sensitivity increased from 80-87% over a 2 year time frame for 11 radiologists who had attended dedicated mammography courses (213).

In summary, this study showed that discussion improved consistency among panel members using PGMI. With an increase in time for assessment, consistency decreased slightly; although this decrease was not significant. These results are the same as those shown by other studies.

5.2.3 **Intra-Observer Consistency**

Studies have shown that a reader is likely to interpret the same image differently on a second reading 5-20% of the time (22, 194, 195). Intra-observer consistency was measured twice during the study. The panel assessed sets of mammograms in Phase A prior to consultation about PGMI. During panel consensus for this phase, a training session for PGMI occurred. During Phase B, a subset from Phase A (A*) was assessed without the panel members knowledge. During Phase C, a subset from Phase B (B*) was assessed with the panel members knowledge (Figure 5-1). The expectation was that panel assessment for B* (post discussion) would be more consistent than A* (the effect of the discussion).

![Figure 5-1: Research Methodology for Discussion of PGMI](image-url)
In the current research, B* post-discussion intra-observer consistency (overall Kappa 0.42) was better than A* (overall Kappa 0.35). The mammographers, who consistently use the system, demonstrated an intra-observer consistency overall Kappa 0.44 during A* and an overall Kappa 0.52 during B*.

The level of agreement for intra-observer consistency was seen to vary between 43.3% and 63.3% during A*, to 29% and 72.6% post-discussion. One screen reader (R1) did modify their pattern of assessing images (from primarily G’s in Phases A and B, to primarily M’s in Phase C). As the panel members were unaware of the intra-observer component of the research and the subset of Phase B was incorporated into Phase C; this made a difference as shown for that particular member (R1: Phase A* Kappa 0.23, Phase B* Kappa 0.00). During A*, Kappa values for the other panel members varied between 0.22 (R2) and 0.41 (R3); while post-discussion Kappa values varied between 0.33 (R3) and 0.56 (R4).

The results of this study are similar to a study conducted by Ciccone et al comparing 7 radiologists interpreting the same 45 mammograms 2 years apart; their results showed intra-observer agreement overall Kappa 0.56 (range 0.35-0.67) (202). Other studies have demonstrated higher levels of intra-observer consistency. For example, a study conducted by Ciatto et al showed substantial intra-observer agreement of 12 radiologists assessing 100 mammograms with BI-RADS 4 scale breast density classification overall Kappa 0.71 (range 0.32-0.88) (207); while Gao et al demonstrated a weighted Kappa 0.87 for Wolfe classification and 0.86 for Boyd (74). Baker et al also reported for one radiologist re-reading images for assessment of lesion interpretation, Kappa was substantial 0.65 ± 0.08 (203).

Anecdotal evidence from mammographers in this study indicates that the categories of G and M become blurred with PGMI and are quite subjective. On a first reading, for example, a mammographer may consider an image “good”, yet may consider the same image “moderate” on a second reading. If the categories of G/M had been combined, both inter and intra-observer consistency would have increased for this research.

5.3 ARE IMAGES ADEQUATELY ASSESSED FOR IMAGE QUALITY?

The pilot study was conducted to assess if inadequate images were over-represented at recall. The results demonstrated that with consensus, the panel members considered 12% of the sets of images in Phase A inadequate. The fact that these inadequate images had been assessed for technical review by a mammographer and reported for interpretative purposes by two, and in some cases
three screen readers, demonstrated a need for further research. During Phases B and C, only 3% and 2% respectively of images were classed as inadequate by the panel consensus. The reduction in the number of inadequate images from Phase A to Phases B and C was significant (p=0.0005).

There is a strong possibility that the number of inadequate images was underestimated for Phases B and C. In Phase A, consensus was undertaken with the group reviewing all sets deemed inadequate by any panel member. During this phase, 5 of the 7 sets remaining inadequate were assessed by 3 or more panel members as inadequate. Two additional sets remained inadequate by group consensus even though initially only one panel member had considered them inadequate. For Phases B and C, consensus was determined by 3 or more panel members assessing an image as inadequate from the assessment sheets. While it was known this may have underestimated the overall inadequate rate; time constraints prevented face to face group discussion.

If the same method of assessment had been used for Phase A (consensus from assessment sheets), the number of inadequate images would have been underestimated by nearly 30%; with an inadequate rate of 8.6% instead of 12%. During Phases B and C, assuming the same rate of underestimation, the inadequate rate would become 3.9% and 2.6% respectively, instead of 3% and 2%.

The current study found that of the 904 sets of mammograms assessed, 30 sets (3.3%) (containing at least one inadequate image) were considered inadequate by group consensus. These 30 sets of images constituted 120 individual images. Of the 120 images, 80 images were classified as inadequate. The left MLO had the highest inadequacy rate (30%), although there was not a significant difference between side and projection in terms of inadequacy.

The current study and the study undertaken by Moreira et al both suggest that mammograms need to be more critically assessed. In theory, no inadequate images should be in the packets of screening mammography clients because of the strategies put in place to assess image quality; i.e. the mammographer assesses the images for technical quality, then 2 or 3 screen readers review the sets for diagnostic interpretation and image quality. However, current research demonstrated that 3.3% of images from the 904 sets were inadequate; while Moreira et al reported 6% of inadequate images from 30 sets used in their study (212). More critical appraisal needs to occur initially by the mammographer; if this does not happen, then the task must fall to the screen readers.
Mammographers may need more time to appraise their work or always have a senior mammographer available for consultation of image review. Also, mammographers and screen readers must become more familiar with what is diagnostic. This might include sessions where both mammographers and screen readers of varying experience review mammograms to understand the difficulties of both positioning and interpretation. There is a fine line between technical perfection and diagnostic / dose concerns. These sessions could provide valuable education for all involved.

Screen readers must be a second check for quality. They should not assume that optimal films are always taken as there are always inexperienced mammographers who are unsure and experienced mammographers who simply miss seeing “faults”. Irrespective of who has produced the images, any suboptimal images should be shown to the person who produced them as a learning tool with suggestions on how to achieve better quality.

5.3.1 WHY WERE IMAGES CONSIDERED INADEQUATE?
The reasons for inadequate images included artefacts, coverage, exposure, folds and motion. Each film / or set may have more than one reason for inadequacy. The major single reason for inadequacy was exposure, which accounted for approximately one third of the total (31%). All exposure inadequate images were due to under-exposure of the image.

A main factor that could explain the high percentage of under-exposed images has been the change of recommended optical density over time. Over the last fifteen years, mammography image quality has improved through the recognition of the importance of adequate film optical density in mammograms (185, 186, 220). In 1995/96, the National Health Service Breast Screening Programme (NHSBSP) in the United Kingdom recommended using an increased film density between 1.4-1.8. This increase occurred after a review of the screening services by Young et al which reported that this range could improve the detection of small (less than 1 cm) invasive cancers by up to 50% (185). Increasing optical density increases the visibility in dense breast tissue, thereby improving cancer detection. Due to this research in the UK, standards were also changed in Australia to improve detection by increasing optical density.

During the current study, the panel members assessed images produced in 2002-2003 during late 2003 to early 2007. During 2002, the optimal optical density used in HBS for mammograms was 1.6 (range 1.5-1.7). In 2006, the optimal optical density increased to 1.8 (range 1.6-2.0). Therefore, some images now classed as
inadequate, may have been in the range of acceptable when they were produced. Although all panel members were aware of these facts and were reminded during the study; they seem to have critiqued by today’s standards. For example, inadequate images due to exposure, plus other factors was 6.3% during Phase A, 73% during Phase B and 36% during Phase C. While digital mammography should eliminate most exposure errors, film screen mammography is effective if the automatic exposure control (AEC) is correctly positioned over the densest part of the breast before exposure.

Motion is the greatest single problem in the production of radiographic detail (21, 22, 63). In this study, as a single reason, motion accounted for 13% of the inadequate images, though was reported in 36% of inadequate images. While an unsharp image may be the result of a technical factor (such as damaged cassettes), this is unlikely as quality control on equipment in screening mammography is rigorous. An unsharp image in mammography is usually due to inadequate compression, either causing slight movement from the client or sagging of the breast, as the breast will fall or sag if not held by the mammographer until adequate compression is applied (16, 17, 19).

While compression is uncomfortable, the benefits should be clearly explained to the client in order to obtain quality images and encourage a positive experience for both the client and mammographer. A study by Kornguth et al reported the two variables that consistently predict a painful experience for the small proportion (<15%) of clients who complain of intense pain during mammography are the average pain at the last mammogram and increased breast density (141). A mammographer is unable to alter a client’s shape, size or breast composition; but their attitude and approach can enhance or diminish the mammography experience for the client.

If all breast tissue is not visualized or is obscured in some way, cancers may be missed. As a single reason, coverage (or lack of) accounted for 7.5% of the inadequate images. Artefacts and folds may hinder the diagnosis by obscuring tissue. As a single reason, artefacts accounted for 5% of the inadequate images. As a single reason, folds accounted for the least reason for inadequacy at 3.75%. It is possible that some of these reasons are underrepresented, as some panel members may only have commented on the primary reason for inadequacy.

Inadequate images may adversely affect the clinical outcome for the screening clients (194). A study by Taplin et al concluded that incorrect breast positioning was the primary reason for not finding interval detected invasive breast cancers, after adjusting for age, film date and breast density (68). Image quality affects image
interpretation and if Taplin’s false negative study concluded that the number of “missed cancers” might be reduced by better positioning technique, then maybe the number of false positives recalls could be reduced in the same way.

While only 3.3% of the total images were considered inadequate; this percentage excluded any repeat films and any technical recalls for this time period. The overall repeat rate according to NAS guidelines should be 3% (40).

This research advocates good basic training in mammography with sufficient time to relax the client, gain her confidence and position well the first time. If staff try to produce too many mammograms too quickly, or are not confident in their initial positioning, or are hurrying the clients; all these scenarios increase both client and staff dissatisfaction and may decrease the quality of the mammogram produced.

5.3.2 SITE OF PRODUCTION OF INADEQUATE IMAGES

There is a discrepancy between the numbers of images produced at public and private sites during the phases. Although there were similar numbers of images produced at both sites in Phases A and C, during Phase B over twice the number of images were produced at public sites (72%) compared to private sites (28%). These included all sets that met the research criteria and were available. In total, this research included 65.5% of images produced at public sites and 34.5% at private sites. The amount of inadequate images were disproportional in the private (6.11%) compared to the public (1.85%) sectors and this difference was statistically significant (p = 0.0013).

Several factors may contribute to this inequity. The majority of mammograms for the public sector are taken on mobile vans. Mammographers working on the vans generally inform clients that if they are asked to return to the van for additional imaging, it will be due to a technical fault. After production, these images are sent to a central location to be checked for image quality, usually by a tutor mammographer or senior staff member familiar with the PGMI assessment classification. These images have traditionally been checked on a daily basis, so that if technical repeats need to be obtained, this can happen while the mobile van is located in the area; thereby causing less inconvenience to clients. As the images are assessed by tutor mammographers or senior staff members; standards of quality remain consistently high.

In the private sector, each mammographer is responsible for reviewing the quality of the images produced; regardless of their experience as a mammographer (i.e. there may be no senior staff member to confer with about quality). If a mammographer
does not have the training or experience to understand quality mammographic practice, they will not be able to replicate it. Training and education are the keys to quality mammography.

Although their research did not include public/private sites of production, similar concerns were raised by Moriera et al. Their focus was on the classification system used for mammography and it was suggested that at the ground level inconsistent decision making about quality may result in either films being repeated unnecessarily, or films of suboptimal diagnostic quality being accepted for reporting (212).

5.4 INFLUENCES OF INADEQUATE IMAGES ON CANCER DETECTION
To assess whether inadequate images influence cancer detection rates, both control (non-recalled) and case (recalled) sets of mammograms were compared in relationship to the assessment of adequate or inadequate and the presence or absence of cancer. The difference for adequate/inadequate in relation to cancer detection had a relative risk of 1.004 (0.9567-1.045). This difference was not significant when tested with chi-square. This was a positive finding as this research concluded that no cancers were thought to be missed due to an inadequate image.

5.5 CASE CONTROLLED COMPONENT
A case-control study was undertaken to investigate the intrinsic factors contributing to the recall of clients. The intrinsic factors included HRT use, density of the breast, site of production and quality of the image. The sample size for the case controlled study contained 110 paired sets. The majority of sets (56%) were from the target age group (50-69 years), while 30% of sets were for clients under 50 years of age and 14% of sets were for clients 70+ years of age. Results from R4 were used to tabulate the case controlled component, as this panel member had the most consistent intra-observer reliability post-discussion (level of agreement 72.6%, Kappa 0.56).

5.5.1 HRT Use
HRT use can increase breast density by arresting and often reversing the process of involution. HRT is thought to have an effect on the sensitivity of mammography with HRT users reported to have a greater risk of having additional imaging compared to non-users; this was confirmed in the current study. Compared to women who never took HRT, the estimated odds ratios for those who were former HRT users and current HRT users were 1.69 and 2.16 respectively. That is, in comparison to those women who never took HRT, there were a 1.69 and 2.16 times likelihood of being recalled from screening for those clients who were former and current users of HRT.
These results are higher than those of the study conducted by Kavanagh et al which demonstrated the risk of a recall was higher among HRT users compared with non-users by an adjusted OR 1.12 (1.05-1.19) after controlling for potential confounders (168). The study undertaken by Kavanagh et al reviewed over 100,000 women increasing its statistical power. Other studies have shown HRT users appear to have a greater risk of being recalled compared to non-users (13, 36, 155, 162, 168, 170, 171, 221), although one study reported this was not clinically significant (155). An overview of published literature conducted by Banks demonstrated that the percentage of false positive recalls varied markedly by study, from 2.1% among UK NHSBS for past users to 14.4% for past users in the US (167); and between 3.1% to 17.5% for current users (167).

A review for BreastScreen Victoria by Kavanagh et al reported that between 1994-2001, there was a slight increase among HRT users of images repeated for technical faults (8.4 per 1000 in HRT users versus 7.1 per 1000 in non-users, p<0.001) or classed as suboptimal (3.1 per 1000 in HRT users versus 2.4 per 1000 in non-users, p<0.001) at subsequent round screening (161).

5.5.2 Breast Density

All mammographic images in the case-control component were assessed for breast density and reviewed by one experienced physician (74, 222). After measuring breast density, the current study confirms other research findings that an increased breast density may decrease sensitivity, thus increasing the recall rate (13, 23, 36, 53-55). Higher breast density contributes to the difficulty in either detecting breast cancer or to reliably exclude its presence. The current study found that clients with high breast density were approximately 2.25 (OR 1.31-3.87) times more likely to be recalled than those with low breast density.

These results are higher than those demonstrated by Carney et al which reported a 1.43 (OR 1.33-1.55) likelihood of being recalled for dense versus fatty breast tissue (36). The study conducted by Carney et al reviewed over 37,000 women. Yankaskas et al reported that the recall rate decreased with decreasing breast density from nearly 7% (6.8%) in extremely dense to 2.4% in the almost fatty density group (55). These results are slightly higher than the current study.

While dense breast tissue decreases the sensitivity of film screen mammography, the DMIST trial has proven that digital mammography is more sensitive in women with more dense breast tissue (78% versus 51%), those under 50 years old and pre-menopausal women (104). As BreastScreening programs convert to digital technology, this may reduce the number in recall rates for clients with dense breast tissue.
5.5.3 Site of Production
The current research included 65.5% of images produced at public sites and 34.5% at private sites. For the case control component, the control contained equal amounts from public / private sites (55, 55); the case contained twice the number of public sites (73, 37). This may partially explain why clients were approximately twice (1.97) as likely to be recalled if their images were produced at a public site (OR 1.14-3.40) compared to a private site. Besides the unequal numbers from the public sites, another explanation may be that the public sites include the mobile vans which travel into country areas. While BS does not advocate including clients with any breast related “issues” (pain, lump, discharge, etc); women in these rural areas may not disclose all relevant breast information to be able to obtain their screening mammogram, rather than having to travel any distance to obtain a diagnostic mammogram, thus increasing the number of recalls.

5.6 Limitations
Bias is any form of systematic error that can affect scientific investigations and distort the measurement process. It is impossible to eliminate bias, but care should be taken to reduce bias. Bias for the current study is listed below under selection and information bias.

5.6.1 Selection Bias
A known bias is the panel member selection, as the mammographers were more familiar with the PGMI system than the screen readers. As the screen readers must also be able to distinguish quality images, as well as interpret for abnormalities, this is seen as an acceptable bias. All panel members were aware of the nature of the research and may have scrutinized the images in a different manner for the study as opposed to the level in routine screening.

Also, the panel only assessed screening, not diagnostic, mammograms. Diagnostic mammograms may be more difficult to interpret for abnormalities as many clients are younger with more dense breast tissue or have had previous breast surgery and radiation, causing scarring and distortion of the normal anatomy. For these same reasons, the standards of PGMI image quality may be more challenging to obtain.

Another bias would be that the sets of mammograms assessed were only from one BreastScreening Program in New South Wales. The screening program did, however, incorporate 3 private and 4 public facilities using over 30 mammographers with varying degrees of education and experience in mammography. This group would be representative of most radiography / screening practices.
It is unlikely that selection bias of the sets of images occurred. Phase A was a simple random sample where images that fitted the criteria were independently selected one at a time until the desired sample size was achieved. Phase B included all available images that fit the criteria. Phase C was a systematic sampling whereby the first non-recall of each list was chosen. The selection should therefore make the sample representative of the population in screening mammography. This research excluded clinical and technical recalls.

A possible bias might have been the exclusion of recalls containing microcalcifications and circumscribed lesions. Generally, unless obviously benign, clients must be recalled for these “abnormalities”. This research specifically wanted to assess image quality using the more subtle feature criteria of asymmetry, stellate lesions and lesions not otherwise specified.

5.6.2 INFORMATION BIAS

Data collection strictly complied with study protocol and several quality control measures were implemented to ensure data accuracy. It is possible that there was some recall bias with information obtained for the client sheets, as information relied on client memory. For example, a client might have forgotten her prior use of HRT. Neither physical or mental limitations of the clients were taken into account; although experience has proven that there would only be a small subset of inadequate images due to physical / mental disabilities in screening well women.

All panel members were aware that all of the images assessed had been previously recalled or assessed as normal by two screen readers; this may have influenced their grading an image as inadequate under the PGMI classification. The NAS guidelines may have also influenced the reluctance of panel members to assess an image as inadequate; as an increase may alter previous statistics.

It is highly unlikely that systematic errors in the assessment of mammographic density occurred, as all mammographic images assessed for density were reviewed by one physician experienced in breast density (Dr. Jinnan Gao) (223). The study by Gao et al assessing inter and intra observer reliability using density classifications of Wolfe and Boyd demonstrated an intra-observer reliability of weighted Kappa 0.87 (p<0.0001) for Wolfe classification and 0.86 (p<0.0001) for Boyd classification (74).
CHAPTER SIX: CONCLUSION
6.1 CONCLUSION
This chapter summarizes research which was conducted to assess image quality in screening mammography. The sets of mammograms belonged to 206 non-recalled and 698 clients recalled for asymmetry, stellate lesions and lesions not otherwise specified x-rayed through NSW Hunter Region and Wyong Shire during 2002 and 2003. The main research aims were to:

- assess if inadequate image quality contributed to unnecessary recalls
- identify factors attributing to inadequate images
- investigate the intrinsic factors contributing to why clients were recalled

This was achieved by answering the following research questions:

6.1.1 WHAT IS THE LEVEL OF CONSISTENCY BETWEEN PANEL MEMBERS IN REVIEWING IMAGE QUALITY?
Numerous studies have been conducted to assess the reliability and validity of classification systems. Few, however, have targeted the mammographic grading system of PGMI. This research assessed the consistency of a six panel member using the PGMI classification in reviewing for image quality. This study demonstrated that even with experienced users, agreement was only fair using the Kappa statistic, making PGMI a subjective classification.

6.1.1.1 Recommendations and Future Research
One recommendation is to construct a sharper, clearer description for criteria of assessment reviewing mammographic image quality; this recommendation is to increase the level of objectivity. Regardless of what system is put into practice, it is imperative that mammographers are properly trained and made aware of how differing body habitus influences both positioning and assessment of the images produced. Any new system must be trialled for consistency and accuracy.

6.1.2 CAN THE LEVEL OF AGREEMENT BE IMPROVED WITH TRAINING?
The level of agreement improved significantly among panel members after training. Even after 3 years, there was only a slight, but not significant, decrease in consistency from the after training level.

During the current research, both screen readers and mammographers benefited from the formal and informal discussions involving image quality. These discussions included information on the assessment of positioning of clients, effect of body habitus and limitations and how these factors affect image quality and interpretation.
6.1.2.1 Recommendations and Future Research

Since PGMI is a subjective classification and training improves consistency, it is recommended that regular review sessions are conducted for use of PGMI. It would also be advantageous if more than one staff member assessed each set of images at the same time with this tool for training and educational purposes. This is believed to reduce the degree of subjectivity by having another view. Unfortunately, once film images become redundant, then this will not be feasible.

It is recommended that discussion sessions become a common practice between screen readers and mammographers; as it was so apparent during this research that this increased the level of understanding for both groups.

6.1.3 WHAT PERCENTAGE OF IMAGES IS DETERMINED TO BE INADEQUATE?

There was an overall inadequate rate of 3.3% of the total images assessed (904 sets). Although not a large percentage, these images had been previously assessed by a mammographer, two and sometimes three screen readers; plus 698 sets had been through a recall process and evaluation.

6.1.3.1 Recommendations and Future Research

These findings lead to the recommendation that mammographers and screen readers adapt a more critical approach to image assessment. Future research could include questioning mammographers to determine how they technically evaluate images. This could be conducted both a qualitative and quantitative research approach using focus groups for the development of a questionnaire. A more important goal would be to encourage continuing education at all levels for mammographers on positioning, technology, assessment and communication with clients, other mammographers and the clinical team. One such recommendation would be to ensure all mammographers (working in both public and private sectors) obtain the Certificate of Clinical Proficiency in Mammography (CCPM).

6.1.4 PUBLIC VERSUS PRIVATE SITES OF PRODUCTION FOR INADEQUATE IMAGES

This research demonstrated a significant difference in the inadequate rate between mammograms produced at public and private facilities (1.8% versus 6.1%).

6.1.4.1 Recommendations and Future Research

A recommendation would be to assess reasons for the difference and implement strategies to ensure a reduction in the number of inadequate rates seen at the private sites. Further research would therefore need to assess the differences between the public and private sector in terms of education, training, supervision and levels of
motivation and commitment in the mammographers at the different sites. Again this information could be gained through the use of focus groups and questionnaires.

6.1.5 WHAT ARE THE COMMON REASONS FOR INADEQUATE IMAGES?

This research used a modified version of the PGMI which contained 5 broad categories why an image could be considered inadequate. The most common causes for inadequacy, listed in order were: exposure, motion, coverage, folds and artefacts.

All exposure faults were due to under-exposure. Many of these inadequate images may have been considered acceptable at the time of their production; however, due to an increase in optical density in viewing mammograms, these images would now be considered under-exposed. While the panel members were aware of this fact, they appeared to have critiqued by today’s standards. With the emerging technology of digital mammography, under-exposure should not be an issue.

6.1.5.1 Recommendations and Future Research

The practice of mammography is operator dependent. Motion, coverage and folds are faults that one should learn how to correct during initial training in mammography. While no mammographer ever obtains all perfect mammograms; this should be the standard that one strives to achieve. However in order to obtain a perfect image it is recognised that the client also needs to be perfect in terms of cooperation, body habitus and having no physical limitations. As this is an unrealistic goal; good basic training in mammography is a must.

With changes from film screen to digital technology, a reduction in the number of images being classed as inadequate from exposure should occur. It is expected that the use of digital imaging will bring other challenges to both mammographers and screen readers. It is proposed that further research be conducted to assess reasons why images will be classed as inadequate with this new modality.

6.1.6 WHAT FACTORS INFLUENCE THE RECALL VERSUS NON-RECALL OF CLIENTS IN SCREENING MAMMOGRAPHY?

The research indicated that clients to be recalled are more likely to be former and current users of HRT, women with dense breast tissue, those who had been x-rayed through a public site and those with good quality imaging. Other studies have shown similar results for both HRT use and dense breast tissue. On analysis, the increase in the number of clients recalled for good image quality compared to those who were recalled for adequate image quality is thought to be due to subtle differences in the breast (asymmetry, stellate lesions and lesions not otherwise specified) may be
more visible with good quality imaging therefore increasing the likelihood of a client being recalled. While women who had inadequate images were less likely to be recalled, this difference was not significantly different.

6.1.6.1 Recommendations and Future Research

Given the nature of the findings from this study, it is important that quality images are presented to the screen readers to ensure that all subtle lesions are visible. This may increase the accuracy in interpretation of the images, thereby possibly reducing the number of missed cancers. It is intended that this analysis will be conducted at a later stage with an increased number of sets of mammograms.

6.2 FURTHER RECOMMENDATIONS

Mammography is a regulated field with stringent guidelines for training and education. Mammographers must attain certain qualifications and continue further education; however, should there be a more structured refresher course for “experienced” mammographers? The reality of a working department is threefold:

- there is always more to learn
- clinical team members are interdependent on each other and their knowledge
- the department is only as good as the least experienced person in that department

Ultimately, the responsibility for achieving quality mammography lies with the individual mammographer. Initial training and education must produce efficient, caring and committed mammographers who have a genuine interest in a difficult modality. Continuing training and education is vital in this changing technology.

While the study demonstrated that inadequate images were not overrepresented at recall: there was concern that inadequate images were in client packets. More research should be conducted into:

- the expansion of this type of research in other centres both within and outside New South Wales
- how mammographers actually gain feedback from their department and whether this feedback is adequate.

This research concludes with the quote at the beginning of the research:

"The point at which the learner becomes satisfied with or indifferent to his attained level of competence defines the ultimate level of ability for that person. An attitude towards work that includes continual self-criticism, progressive problem solving and continual re-investment in improvement is the description of a profession, where expertise, accountability, autonomy and authority are interrelated." (1)


45. Pinckney RG, Geller BM, Burman M, Littenberg B, Watanabe JM, Elmore JG. Women with false-positive mammograms are more likely to attend future screenings. Evidence-Based Healthcare. 2003;7(4):190-1.


157. Cardenosa G. Breast imaging Philadelphia: Lippincott Williams & Wilkins; c2004


175. Haus AG. Historical Technical Developments in Mammography. Technology in Cancer Research & Treatment. 2002;Volume 1(Number 2, April).

176. NHS Cancer Screening Programmes. Evaluation and clinical assessment of the Siemens Novation full field digital mammography system. NHSBSP December 2007 [updated December 2007; cited 19 February 2008]; Available from:


APPENDIX ONE: ETHICAL APPROVAL
10 November 2003

Dr H Warren-Forward
Discipline of Medical Radiation Science
University of Newcastle

Dear Dr Warren-Forward,

Re: A Feasibility study into the Reduction of Recall Rates (Further Assessment Investigation In Screening Mammography (03/07/23/3.14))

The above protocol was reviewed by the Hunter Area Research Ethics Committee and the University of Newcastle’s Human Research Ethics Committee at their meetings held on 23 July 2003 and 15 October 2003. Following receipt of the requested clarifications and the revised information sheets and consent forms by the Professional Officer, the Hunter Area Research Ethics Committee has resolved that the protocol A Feasibility study into the Reduction of Recall Rates (Further Assessment Investigation In Screening Mammography, the screening Mammography Consent Sheet, Data Collection sheet A (version 1 dated 20 April 2003), Data Collection Sheet B (version 1 dated 20 April 2003), the Participant Information Sheet and Consent Form (version 3 dated 6 November 2003), the Participant Questionnaire (version 2 dated 13 August 2003), the Authority to Proceed Form (version 2 dated 13 August 2003), the Mammographer Information Sheet (version 2 dated 13 August 2003) and Consent Form (version 1 dated 20 April 2003), the Mammographer Questionnaire (version 1 dated 20 April 2003) and the Pre-Invitation are now approved. You may commence your research.

The University of Newcastle Human Research Ethics Committee will issue a separate certification of approval for the above protocol.

Approval from the Hunter Area Research Ethics Committee for the above protocol is given for a maximum of 3 years from the date of this letter, after which a renewal application will be required if the protocol has not been completed.

The National Statement on Ethical Conduct in Research Involving Humans, (1999), which the Committee is obliged to adhere to, include the requirement that the committee monitors the research protocols it has approved. In order for the Committee to fulfil this function, it requires:

- A report of the progress of the above protocol be submitted at 12 monthly intervals. Your review date is November 2004. A proforma for the annual report will be sent two weeks prior to the due date.

- A final report be submitted at the completion of the above protocol, that is after data analysis has been completed and a final report compiled. A proforma for the final report will be sent two weeks prior to the due date.
• All variations or amendments to this protocol, including amendments to the Information Sheet and Consent Form, must be forwarded to the Hunter Area Research Ethics Committee for approval. The approval letter must be received before the amendment is implemented.

• Adverse events, however minor, must be recorded as observed by the investigator or as volunteered by a participant in this protocol. Full details will be documented, whether or not the Investigator or his deputies considers the event to be related to the trial substance or procedure.

• Serious adverse events that occur during the study or within six months of completion of the trial at your site should be reported to the Professional Officer of the Hunter Area Research Ethics Committee as soon as possible and at the latest within 72 hours.

• Copies of serious adverse event reports from other sites should be sent to the Hunter Area Research Ethics Committee for review as soon as possible after being received. Serious adverse events are defined as:

  Causing death, life threatening or serious disability.
  Cause or prolong hospitalisation.
  Overdoses, cancers, congenital abnormalities whether judged to be caused by the investigational agent or new procedure or not.

If for some reason the above protocol does not commence (for example it does not receive funding), is suspended or discontinued, please inform Dr Nicole Gerrand, the Professional Officer of the Hunter Area Research Ethics Committee as soon as possible.

If you have any enquires please contact Dr Gerrand, as per her contact details at the top of the previous page. I wish you well with your research.

Yours sincerely,

Ms E. Kearns  
Chairperson  
Hunter Area Research Ethics Committee  

Cc: Ms S O'Connor  
University of Newcastle's  
Human Research Ethics Committee
Jennifer Thompson
Hunter NSW BreastScreen

RE: Research Study: “A Feasibility Study into the Reduction of Recall Rates (Further Assessment Investigation) in Screening Mammography”

Dear Jennifer,

Thank you for your letter regarding the above research project. I tabled this along with your HREC/HAREC application at the Radiation Safety Committee meeting of 28th July 2003. The committee has no objection to this proposal and note that the study will not involve any radiation that would not otherwise be received by the study patients.

We appreciate your advice and wish you success with this work.

Regards,

Steve

cc. Radiation Safety Committee
APPENDIX TWO: STATISTICAL ANALYSIS TESTS
A2.1 PAIRED T TEST

The paired t test compares two paired groups. It calculates the difference between each set of pairs, and analyzes that list of differences based on the assumption that the differences in the entire population follow a Gaussian distribution.

The difference between each set of pairs is calculated, keeping track of sign. If the value in column B is larger, then the difference is positive. If the value in column A is larger, then the difference is negative. The t ratio for a paired t test is the mean of these differences divided by the standard error of the differences. If the t ratio is large (or is a large negative number), the P value will be small.

The paired t test compares two paired groups so you can make inferences about the size of the average treatment effect (average difference between the paired measurements). The most important results are the P value and the confidence interval.

The P value answers this question: If the treatment really had no effect, what is the chance that random sampling would result in an average effect as far from zero (or more so) as observed in this experiment?

A2.2 WILCOXON MATCHED PAIRS TEST

The Wilcoxon matched test is a non-parametric statistical hypothesis test for the case of two related samples or repeated measurements on a single sample. It can be used as an alternative to the paired Student's t-test when the population can't be assumed to be normally distributed. Like the t-test, the Wilcoxon test involves comparisons of differences between measurements, so it requires that the data are measured at an interval level of measurement. However it does not require assumptions about the form of the distribution of the measurements. It should therefore be used whenever the distributional assumptions that underlie the t-test cannot be satisfied.

The Wilcoxon matched pairs test calculates the difference between each set of pairs, and analyzes that list of differences. The P value answers this question: If the median difference in the entire population is zero (the treatment is ineffective), what is the chance that random sampling would result in a median as far from zero (or further) as observed in this experiment?

In calculating the Wilcoxon test, the differences between each set of pairs are computed, and the absolute values of the differences are ranked from low to high.
Then the sums the ranks of the differences where column A was higher (positive ranks), sums the ranks where column B was higher (it calls these negative ranks), and reports the two sums. If the two sums of ranks are very different, the P value will be small. The P value answers this question: If the treatment really had no effect overall, what is the chance that random sampling would lead to a sum of ranks as far apart (or more so) as observed here?

A2.3 CONTINGENCY TABLES (CHI-SQUARE, RELATIVE RISK AND ODDS RATIO)

Contingency tables summarize results where the outcome is a categorical variable such as disease vs. no disease, pass vs. fail, adequate vs. inadequate. The data from a contingency table is used to compute the relative risk and odds ratio along with 95% confidence intervals. The values entered in a contingency table represent the number of subjects actually observed in this research. Note that the columns define mutually exclusive categories, as do the rows. A subject can be in one or the other, but not both. Most contingency tables have two rows (two groups) and two columns (two possible outcomes).

A2.3.1 CHI-SQUARE ($\chi^2$) AND FISHER EXACT TEST

The Chi-Square is used to test for association between categorical variables, data distribution is based on a sum of squares, therefore the value of $\chi^2$ will always be larger than (or equal to) zero. The null hypothesis (an expectation of some sort) for the (Pearson) chi-squared test is that there is no association between the variables; consequently a significant p-value implies an association. The Pearson chi-squared test can be used with nominal data.

Some chi-squared tests are referred to by names after the originator, for example Fisher exact used for analysing small data sets, and the McNemars test, a chi-square test used for the analysis of matched paired data. This non-parametric test uses matched-pairs of labels ($A$, $B$). It determines whether the proportion of $A$- and $B$-labels is equal for both members.

A2.3.2 RELATIVE RISK

Relative risk is a more direct measure comparing the probabilities in two groups is the relative risk, which is also known as the risk ratio. The relative risk is simply the ratio of the two conditional probabilities. Like the odds ratio, a relative risk equal to 1 implies that the event is equally probable in both groups. A relative risk greater than 1 implies that the event is more likely in the first group. A relative risk less than 1
implies that the event is less likely in the first group. Using the table below (taken from section 4.6) as an example, it can be seen that the risk ratio is more than 1, implying that there were more inadequate images amongst the women not diagnosed with cancer, or put another way, it is shown that inadequate images does not hinder the appearance of cancers.

<table>
<thead>
<tr>
<th></th>
<th>Outcome 1 (Adequate Images)</th>
<th>Outcome 2 (Inadequate Images)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 (Cancer)</td>
<td>A (75)</td>
<td>B (2)</td>
</tr>
<tr>
<td>Group 2 (No Cancer)</td>
<td>C (546)</td>
<td>D (17)</td>
</tr>
</tbody>
</table>

\[
RR = \frac{A}{C} \div \frac{B}{D} = \frac{75}{546} \div \frac{2}{17} = \frac{0.974}{0.9698} = 1.004
\]

**A2.3.3 Odds Ratio**

The odds ratio is a way of comparing whether the probability of a certain event is the same for two groups. An odds ratio of 1 implies that the event is equally likely in both groups. An odds ratio greater than one implies that the event is more likely in the first group. An odds ratio less than one implies that the event is less likely in the first group. Using the example discussed in section 4.6.2 which investigates the probability of being recalled based on breast density, the odds ratio is calculated by:

<table>
<thead>
<tr>
<th></th>
<th>Number of Women Recalled</th>
<th>Number of Women Not-Recalled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Breast Density</td>
<td>A (42)</td>
<td>B (64)</td>
</tr>
<tr>
<td>High Breast Density</td>
<td>C (68)</td>
<td>D (46)</td>
</tr>
</tbody>
</table>

\[
OR = \frac{A}{B} \div \frac{C}{D} = \frac{AD}{BC} = \frac{42 \times 46}{64 \times 68} = \frac{1932}{4352} = 0.44
\]

In this example, the OR of 0.44 indicates that the probability of being recalled in 0.44 times less if you have low breast density, or put another way, you are more 2.25 times more likely to be recalled if you have high breast density.