A retrospective pilot study of the performance of mammographers in interpreting screening mammograms

S Moran and H Warren-Forward

Discipline of Medical Radiation Sciences, Faculty of Health Sciences, Callaghan Campus, University of Newcastle, New South Wales 2308, Australia.
Correspondence sheila.moran@uon.edu.au

Abstract Purpose: This paper provides pilot data from one Breast Screening Program on whether mammographers have the ability to review mammograms with similar accuracy to screen readers. Methods: The participant group consisted of 11 mammographers (experienced in assessing images for technical adequacy, but no specialised training for detecting abnormalities) and three current screen readers employed at Hunter BreastScreen. Fifty sets of mammograms performed during 2003 were used in the retrospective study. The mammograms were chosen to represent a range of review outcomes. Each participant reviewed each set of mammograms using the BreastScreen Australia assessment scale; these results were then categorised into two groups – Rescreen or Recall. The consensus review outcome of the original screen readers was used as the “Gold Standard”. Patient outcome was assessed by following up on the results of any histology or pathology tests in 2003 or the 2006 screening results. Results: Compared with the Gold Standard outcomes, the three current screen reader participants had sensitivity values of 67%, 83% and 94%; mammographer sensitivity values ranged from 61% to 89%. Specificity for the screen reader participants was 81%, 84% and 87% and mammographers ranged from 45% to 97%. Accuracy for the three screen readers was 76%, 84% and 90% while the mammographers attained between 55% and 86%. Conclusion: Without any training, the sensitivity obtained by three mammographers and the specificity obtained by six mammographers were similar to those of the current screen reader participants. Accuracy rates of the mammographers indicate that screen reading by selected and appropriately trained mammographers may be feasible.

Introduction

More than 1.3 million women worldwide are diagnosed with breast cancer every year. This accounts for 28% of all female cancer cases and 13.5% of all new cancers.1 In Australia, breast cancer is the second most common invasive cancer diagnosed in females;2 it is also a leading cause of cancer death in females.3,4 Breast cancer deaths declined significantly in the 1990s, largely due to mammographic screening (earlier detection) and more effective treatments.5

Mammography is the most common modality for detection of early breast cancer and is the Gold Standard to which other imaging modalities are compared. Screening mammography has been proven in clinical trials to reduce mortality from breast cancer by at least 30% in the over 50 age group.6,7,8

There have been many articles written over the last few years about the need to expand the roles of radiographers in Australia.9,10 There are a number of reasons for this support for changes to the career pathways open to radiographers. First there is a known shortage of radiologists whose workloads are increasing with new technological advances11,12,13 and the supply and demand of radiologists in the future will be a complex issue to address.14,15 Making use of some of the experienced radiographers in certain specialties certainly makes sense. Second, there are very few opportunities for experienced radiographers to expand their skills and put them into practice, thereby maintaining an enthusiasm for their work. As Smith and Lewis have stated,16 “the majority of radiographers have few prospects of increasing their responsibility” in the clinical field. There are many radiographers who do not want to go down the path of management roles, and who are frustrated at having clinical skills that they are not allowed to use.

The shortage of radiologists is very evident in the field of breast imaging, mainly due to the perception of mammography as having a high risk of litigation, high stress levels, low re-imbursement and to the difficulties associated with reading mammogram images.16,17,18,19 Mammography is a highly specialised area: the detection of abnormalities on breast images is difficult because of the subtlety of changes in breast tissue and every radiologist is likely to periodically overlook a cancer or to misinterpret a finding, particularly if the image quality is not optimal.20 Screening programmes have qualified mammographers who produce mammograms every day. Some have many years of experience of critiquing images, and may be proficient at detecting abnormalities.

Radiographer reporting has been undertaken in other countries for two decades. Van den Biggelaar, et al. undertook a review of six studies focusing on the performance of mammographers interpreting mammograms between 1987 and 1996.21 Three studies were from the UK and the other three from the USA. All the studies were conducted in a screening setting and three of them measured the effect of training on the mammographers. The results showed a sensitivity range of 73% to 86% for the radiologists and 73% to 90% for the mammographers. Specificity for the radiologists was 81% to 95% and the mammographers achieved 64% to 91%. It was indicated that training programmes could improve the mammographers’ performance not only to increase cancer detection, but to identify benign lesions, which would increase specificity and keep recall rates low.22
Table 1: BreastScreen Australia results system.

<table>
<thead>
<tr>
<th>BreastScreen Australia</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Normal/No significant finding</td>
</tr>
<tr>
<td>2</td>
<td>Benign</td>
</tr>
<tr>
<td>3</td>
<td>Probably benign/Indeterminate</td>
</tr>
<tr>
<td>4</td>
<td>Suspicious for malignancy</td>
</tr>
<tr>
<td>5</td>
<td>Malignant</td>
</tr>
<tr>
<td>6</td>
<td>Technical fault</td>
</tr>
<tr>
<td>7</td>
<td>Clinical recall (symptom)</td>
</tr>
</tbody>
</table>

In the UK Wivell, et al. documented a trial in 2002 by two formally trained mammographers (post-graduate certificate in Image Interpretation and Analysis). The retrospective trial showed the mammographers recalled all 59 cancers and also recalled 32 of 90 interval cancers. The mammographers’ recall rate was 3.9% higher than the radiologists. The mammographers then became prospective second readers for 54000 screening mammograms and there was no significant difference between the radiologist readers and these mammographers in the recall rate (4%) or cancer detection rate (84%).

There has been cautious support for changes to Australian radiographers’ duties and responsibilities from both the Australian Institute of Radiographers (AIR) and the Royal Australian and New Zealand College of Radiologists (RANZCR). Australian evidence is needed to highlight the abilities of radiographers and their willingness to develop their roles, not only for career advancement, but in order to provide more efficient services.

This research aims to assess whether mammographers are capable of reading screening mammograms as accurately as screen readers.

Materials and methods

Ethics and recruitment

Ethics approval was obtained from the University of Newcastle and Hunter New England ethics committees. Participants were issued with a study identification number, which was unknown to the student researcher; however, information has been provided to identify the group of screen readers (R3, R6, R10), to allow for meaningful results in this paper.

Mammographic images

The mammograms chosen were from women who attended the Hunter BreastScreen programme in 2003 and who had either received a follow-up mammogram in 2005/2006 or who had been diagnosed with breast cancer in 2003. Unlike a normal batch of screening mammograms, these sets were chosen to incorporate different categories of lesions and differing degrees of agreement between the original screen readers.

Of the 50 cases in the study, 26 women were rescreened in the normal interval, 24 women were recalled to assessment and 18 of those recalled women were diagnosed with breast cancer.

All the mammograms to be read were set up on the same viewing carousels that screen readers normally use. The carousels used were RADX Coolbrite High Intensity Illuminators (4500–6000 cd. m–2) with dimmer and masking capabilities to eliminate extraneous light. The viewing carousels are in a darkened room (background lighting 5–20 lux) which is away from the main thoroughfare and therefore quieter and less distracting. Each carousel can hold approximately 100 sets of mammograms.

Normal procedure was followed, so that if any of the women had been identified as having breast symptoms at the time of screening (e.g. unusual pain, a lump, nipple discharge) then the front registration sheet for that episode was put up alongside the current study images. This enabled the readers to be aware of any clinical issues associated with that mammogram.

Any prior recall images were hung with the previous mammogram; participants were not able to access the packets for their own information as these contained the 2003 patient outcomes.

Assessment of screening mammograms

In Australia, mammographic images are reviewed independently by at least two screen readers who provide results according to the system used in BreastScreen Australia (Table 1). Each screen reader allocates a number (#1–7) to the mammogram. The results from the two screen readers are combined into a single recommendation. A technical fault (#6) is usually not a full repeat mammogram but simply one or two views that were not of optimal quality. A clinical recall (#7) may result in recall to assessment or a letter to the woman’s doctor recommending further investigation.

Gold Standard

When both screen readers provide a result of #1 and/or #2 (Table 2), the woman will receive a letter recommending routine rescreen in two years. If both readers allocate a value of #3, #4 or #5, the woman is recalled for further investigation. If one reader indicates a routine rescreen and the second reader indicates a recall, then a third reader is asked to review the mammogram. This third reader is expected to be an experienced senior breast radiologist. If either reader indicates a #4 or #5, the woman is automatically recalled without a third read. The term “Gold Standard” in this context is defined as “the single recommendation outcome” after combining the results of all readers (Table 2).

Table 2: Example of Gold Standard calls.

<table>
<thead>
<tr>
<th>Radiologist 1</th>
<th>Radiologist 2</th>
<th>Radiologist 3</th>
<th>Gold Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>Rescreen</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td></td>
<td>RECALL</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>Rescreen</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>RECALL</td>
</tr>
</tbody>
</table>

Methods

This retrospective study was undertaken to allow the researchers to pilot the research methodology and associated data collection protocol.

The normal result sheet was modified in order to allow readers more space to write comments or ask questions as they read each set of mammograms. This was important as it allowed each reader to “think out loud” and would remind them of their thought process when they later attended a review meeting. There was also space to record the length of time it took to read the mammograms (Table 3).

Each reader was required to identify the corresponding details on the data sheet with the correct mammogram images on the carousel. For each mammogram, the reader gave a result code (#1–7) and for results that were 3 or higher, identified the side (L or R), the site (#1–5) and gave a reason for the recall (#1–7). In the case of a result of #6 (Technical recall) the readers were also required to give another result as these mammograms had already been reported on with the available images. A sealed box was available for the readers to insert their completed result sheets.
The first step in a diagnosis of breast cancer is the perception of an abnormality on the mammogram images, which results in a recall to assessment. A breast cancer cannot be diagnosed from a mammographic image; lesions can appear to be malignant on imaging, but only cytology or histopathology can provide proof of diagnosis.

The researcher accessed follow-up records of the women to ascertain whether the recall had resulted in a definitive diagnosis of breast cancer via a pathologic report. In the cases where the women had been returned to routine rescreen in 2003, the researcher accessed the records from each woman’s return screening episode (2005 or 2006) to ascertain the outcome of those calls.

Data
The information was entered into an Excel spreadsheet. The sensitivity, specificity and accuracy were calculated and graphed.

Sensitivity
Sensitivity is a measure of the probability of detecting an abnormality when it is present in the woman. It is also known as the true-positive fraction.²⁶

Specificity
Specificity is a measure of the probability of a normal interpretation of an image when no abnormality exists. This is also known as the true-negative fraction.²⁶

Accuracy
Diagnostic accuracy is the number of women correctly diagnosed (true positives and true negatives) compared with the total number of women examined. The diagnostic accuracy of the imaging system depends upon a number of factors including the density of the breast tissue, the quality of the mammogram, the viewing conditions and the training and experience of the reader to correctly interpret any abnormality on the image.²⁶ Accuracy is a combination of sensitivity and specificity and provides a truer rating of a reader’s abilities. The sensitivity of a reader would be 100% if they detected every cancer present, and their specificity would be 100% if they recalled to assessment only those women.

<table>
<thead>
<tr>
<th>Result Codes</th>
<th>Reason for Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Normal</td>
<td>1 Circumscribed lesion</td>
</tr>
<tr>
<td>2 Benign</td>
<td>2 Stellate lesion</td>
</tr>
<tr>
<td>3 Probably benign</td>
<td>3 Lesion/mass NOS (not otherwise specified)</td>
</tr>
<tr>
<td>4 Probably malignant</td>
<td>4 Asymmetry</td>
</tr>
<tr>
<td>5 Malignant</td>
<td>5 Calcification – granular/casting</td>
</tr>
<tr>
<td>6 Technical recall</td>
<td>6 Calcification – lobular</td>
</tr>
<tr>
<td>7 Clinical recall (symptoms)</td>
<td>7 Calcification – other/NOS</td>
</tr>
</tbody>
</table>

Table 3: Modified result sheet.

<table>
<thead>
<tr>
<th>Bay</th>
<th>NAME</th>
<th>P.I.D</th>
<th>Unit</th>
<th>Result</th>
<th>Side</th>
<th>Site</th>
<th>Reason</th>
<th>Comments</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones</td>
<td>123</td>
<td>2</td>
<td>3</td>
<td>L</td>
<td>1</td>
<td>1</td>
<td>Probable cyst</td>
<td>VC/FC/NC</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Reading times for 50 mammograms.
with cancer present. A good diagnostic imaging system is a compromise between not missing cancers (sensitivity) and not recalling too many women unnecessarily (specificity).

**Review**

The mammograms with lower than 50% agreement between participants and the Gold Standard, and all 18 diagnosed cancers, were critically reviewed at a meeting with a radiologist and breast physician in attendance.

**Results**

**Time of reading**

Varying degrees of experience can be assumed to result in a wide variation in the time taken to read the mammograms. The times ranged from 35 minutes to 180 minutes, with a mean time of 90 minutes (Figure 1).

The ages of the mammographer participants ranged from 25 to 29 (one only) to 50 to 59 (five participants). The demographic information included age, experience and initial confidence levels – none of which had any relationship to the time taken to assess the images.

**Measure of agreement**

When reporting on a mammogram, the reader is required to provide more detailed information for each woman they wish to recall for further assessment. This information consists of which breast, the site of the lesion and a description of the abnormality noted.

Readers’ perceptions of exactly what and where a lesion/abnormality is, varies greatly. The illustrated mammogram images of a right breast (Figure 2) show a lesion behind the nipple and slightly medial on the CC view. These images show nothing else of any interest, and yet the resulting recall codes given for site and reason covered almost the entire range (Table 4).

Although the level of agreement of classification varied considerably, the area of the breast tissue in question was considered to be the same area for all participants. While all readers should endeavour to standardise their classifications, the overall benefit to the women screened always results in one of two choices – Rescreen or Recall. The final “level of agreement” decided on by the researchers was therefore restricted to agreement with the Gold Standard only for Rescreen or Recall and the correct side.

**Sensitivity**

The Gold Standard used in the pilot study has an inherent sensitivity of 100%; it is assumed that the Gold Standard detected all the cancers. The original screen readers in 2003 had sensitivities of 89% and 94% whereas the range for the participant screen readers in the study was 67%, 83% and 94%, which indicates there were between 1 and 6 cancers missed out of a total of 18 (Figure 3). The mammographers varied from 61% to 89%, with a median of 72%; with between 2 and 7 missed cancers. Remember, the mammographer participants have had no training in the diagnostic assessment of mammograms.
Figure 4: Specificity in reading 50 mammograms (screen reader participants denoted by arrow).

Figure 5: Accuracy of reading 50 mammograms (screen reader participants denoted by arrow).

Figure 6: Agreement per mammogram.
Specificity
The Gold Standard specificity was 81% (Figure 4). The participant screen readers had specificities of 81%, 84% and 87% and the mammographers a range of 45% to 97% (median of 78%). The mammographer with the lowest specificity recalled 14 of the 26 women who had no cancer.

Accuracy
In this pilot study, a reader would have to have recalled only those 18 women who had a diagnosis of cancer to attain a 100% accuracy rating. The Gold Standard accuracy rate was 88% (Figure 5), which is a correct diagnosis for 44 of the 50 women in the study.

The screen reader participants scored 76%, 84% and 90% accuracy, while the range for the mammographers was between 55% and 86%. Three mammographers attained equal or higher accuracy than the lowest scoring screen reader. Accuracy provides a reading of how efficient the reader is overall and while some mammographers are lagging behind the experienced screen readers, it appears that others have the ability to detect cancers without excessive recalls.

Agreement per film
All participants agreed with the Gold Standard on the outcome of nine sets of mammogram images, five of which were cancers (Figure 6). There were seven sets of mammograms (#2, 13, 22, 29, 43, 46 and 47) where the agreement was less than 50%, of which one was a cancer detected in 2003 (#47).

Mammograms # 2, 13, 29 and 46 were all recalled in 2003; more than 50% of participant readers reported these mammograms as normal or benign. No cancer was detected at recall assessment and all had a normal rescreen in 2006 (Table 5).

Mammogram #47 was also recalled in 2003 and a cancer was detected; this was missed by all the mammographers in the study and by one of the three screen readers (Table 5).

Mammograms #22 and 43 were both considered normal in 2003. The majority of participants disagreed with the Gold Standard; mammogram #22 proved to be a normal rescreen in 2006. Mammogram #43 had three screen readers and five mammographers recommend recall for the right breast; when the woman returned in 2006 a cancer was detected – but in the left breast (Table 5).

2006 cancers
There were four women who had a cancer detected when they returned in 2006. Mammogram #43 has been discussed; the others were mammograms #16, 24 and 44. For two of these women (mammograms #16, 24), almost all of the participants agreed with the Gold Standard in 2003; however for mammogram #16 there were two mammographers who noted on the result sheet, the exact side and site of the cancer detected in 2006; one of these mammographers failed to recognise it as a cancer, and did not recommend a recall. Mammogram #24 had one mammographer who detected an abnormality at the site of the 2006 cancer. These were three different mammographers and these two 2006 cancers have been considered “new cancers” (Table 5).

Mammogram #44 had two screen readers and two mammographers who recommended recall; two more mammographers commented on the abnormality, but failed to recognise it as malignant. These six participants noted the exact side and site of the cancer detected in 2006; this could be classified as a “missed” cancer, which for the purpose of this research, can be defined as a false-negative result with the cancer visible in retrospect (Table 5).

Table 5: Participants in agreement with GS (RC = Recall; RS = Rescreen).

<table>
<thead>
<tr>
<th>Film numbers</th>
<th>#2</th>
<th>#13</th>
<th>#16</th>
<th>#22</th>
<th>#24</th>
<th>#29</th>
<th>#43</th>
<th>#44</th>
<th>#46</th>
<th>#47</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003 GS result</td>
<td>RC</td>
<td>RC</td>
<td>RS</td>
<td>RS</td>
<td>RS</td>
<td>RC</td>
<td>RS</td>
<td>RS</td>
<td>RC</td>
<td>RC</td>
</tr>
<tr>
<td>Mammographers agree with GS (n = 11)</td>
<td>3</td>
<td>6</td>
<td>10</td>
<td>4</td>
<td>10</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Screen readers agree with GS (n = 3)</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2003 result</td>
<td>Cancer</td>
<td>N</td>
<td>N</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>N</td>
<td>–</td>
<td>–</td>
<td>N</td>
</tr>
<tr>
<td>2006 result</td>
<td>Cancer</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 6: Reasons for missed cancers.

<table>
<thead>
<tr>
<th>Mammogram</th>
<th>#6</th>
<th>#12</th>
<th>#17</th>
<th>#20</th>
<th>#21</th>
<th>#23</th>
<th>#26</th>
<th>#27</th>
<th>#35</th>
<th>#41</th>
<th>#47</th>
<th>#48</th>
<th>#49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circumscribed lesion (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stellate (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesion NOS (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymmetry (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micro-calcifications (5,6,7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mammographers who missed (n)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

A retrospective pilot study of the performance of mammographers in interpreting screening mammograms

The Radiographer

17
Discussion

When conducting this retrospective trial, the standard achieved by the participants has been compared to the Gold Standard result in 2003. This implies that the Gold Standard is 100% correct; however, the fact is that no screening programme will detect 100% of the cancers because all readers will, at some stage, fail to detect an abnormality. This psychovisual phenomenon is one of the main reasons for the introduction of double reading in mammography to improve the cancer detection rate and reduce the error rate. The advantage of employing two or more readers has been the subject of many papers and trials and in BreastScreen Australia, double reading is accepted as the benchmark, where “specifically trained non-radiologist readers” can be employed.

Missed cancers

Two major causes of “missed cancers” are interpretation and perception. Interpretation errors may be explained by readers who “see” the abnormality, but fail to recognise it as a malignancy. The perceptual error is when a lesion is evident in the field of view, yet is overlooked by the reader, most often due to “busy” or dense parenchyma.

The majority of cancers missed by the mammographers fell into two categories – circumscribed lesions and micro-calculifications (Table 6), both of which can be difficult to detect when their density is similar or lower than the surrounding tissue. A trial in London in 1996 which involved over 17 000 mammograms showed that seven radiographers who had undergone a course in mammogram interpretation at a National Health Service Breast Screen Training Centre, attained sensitivity equivalent to a single radiologist reader (73%) but 9% lower (86%) specificity, and that excess recall could be limited by radiologist review of radiographer reports of potential findings. This current pilot study shows nearly 10% lower mean sensitivity by the (untrained) mammographers compared to the participant screen readers; however, the loss of specificity was equivalent to that of the study by Pauli, et al.

In 2003 in Pittsburgh, USA, Jules Sumkin and co-workers investigated the possibility of technologists reading screening mammograms. In a prospective study, 33 technologists over six facilities were required to assess over 3000 mammograms in total. They were to provide either a negative result or an indication that further procedures were necessary, and these results were compared with the radiologists’ interpretations (82% agreement). That trial showed that, even without any additional training, they could perform at the radiologists’ interpretations (82% agreement). That trial showed that, even without any additional training, they could perform at

Results of the current study show that seven mammographers scored over 72% sensitivity and six obtained higher than 78% specificity; though only two mammographers attained both a high sensitivity and specificity. The majority (eight) of mammographer participants were somewhat successful in attaining a balance between cancer detection and recognising a “normal” mammogram by achieving over 70% accuracy. All the mammographers would benefit from specialised training to develop their skills to discriminate between significant and insignificant abnormalities. The researcher has developed a training package specifically for mammographers to assess screening mammograms. This package demonstrates examples of benign and malignant lesions, and also provides guidance on how to approach assessment in a methodical manner.

During review of the pilot study, it was apparent that the majority of mammographers had acquired a greater appreciation of the difficulties of the screen reader’s work and the importance of high quality mammograms for accurate assessment; this may encourage a higher standard of image quality in the screening setting. The small number of mammograms in this pilot study may have skewed the results to some extent. The relatively large number of cancers may have encouraged the mammographers to detect something in every mammogram (more False Positives) and for the screen readers it may have had the opposite effect (increased False Negatives) as this is not a usual screening scenario. It is noted that the results of this study only concentrated on one BreastScreen programme and the results cannot be generalised to other programmes.

All participant reader results were analysed on an individual basis. If the participants had been paired with another participant (as is standard practice with screen readers) it may have improved the outcome as compared with the “consensus Gold Standard”.

Conclusion

Due to a shortage of radiologists and possible increased role expansion of mammographers, this research looked at the suitability of mammographers to read screening mammograms. This could help to provide the “flexibility when the need arises” recommended in the NAS which enables other medical personnel to report on mammograms providing they have the necessary training and experience. Experienced mammographers, because they are looking at mammographic images every day, become familiar with “normal” and “abnormal” findings. They are also involved in the recall clinics every week, where they locate lesions by doing extra views. The fact that some experienced mammographers may be capable of assessing breast images should encourage us to consider whether they could enhance the service provided by a breast screening programme.

This study has shown that, without any formal training, some women who are screened, sensitivity is the most important aspect of their visit – if they have a cancer, they expect it to be detected. If true cancers are being missed, the screening programme is not being effective. If any cancers were missed by the original screen readers, it would not be apparent until the women returned for their next screening or if they were diagnosed with an interval cancer.

Breast screening programmes in the UK have reported specificity values of between 82% and 97%, and in Australia, the Australian Institute of Health and Welfare (AIHW) reported similar specificity of 80 to 95%. Three mammographers attained over 80% with three others at 78%. If the specificity is low, too many women are being recalled unnecessarily. This is not only expensive but it also can increase anxiety in the women.

Experienced mammographers may be capable of assessing breast images should encourage us to consider whether they could enhance the service provided by a breast screening programme.
mammographers have shown promise of being able to read mammograms with accuracy equivalent to that of current screen readers. Since completion of the pilot, a full retrospective study has been completed, with measurement of inter and intra consistency of reader accuracy. The training package will be expanded and will possibly be offered to mammographers throughout Australia. This research project will also conduct a prospective stage, which will consist of selected mammographers assessing current mammograms under the same conditions as the radiologists. This study will concurrently examine the differences between analogue and digital images and the difficulties of changing modalities.

There is no point in pursuing this expansion of roles if mammographers are not interested; therefore, a questionnaire has been distributed Australia wide, to ascertain the thoughts and opinions of mammographers currently working. It is important to discover in which direction they have most interest.

Acknowledgements
I would like to extend my heartfelt thanks to my colleagues at Hunter Breast Screen (now known as BreastScreen New South Wales Hunter New England) who enthusiastically participated in this trial and donated their time and effort. My thanks also go to the Australian Institute of Radiography for the 2007 Diagnostic Research Scholarship award. This enabled me to present the findings from this pilot study at the 5th ASMMIRT conference in Melbourne in April 2008.

The authors
Sheila Moran MRS
Associate Professor Helen Warren-Forward (Supervisor)
BSc, PhD

References