

# **COMPUTER AIDED DETECTION IN MAMMOGRAPHY**

**Working Party of the Radiologists  
Quality Assurance Coordinating Group**

**NHSBSP Publication No 48  
January 2001**

First published by:

NHS Cancer Screening Programmes  
The Manor House  
260 Ecclesall Road South  
Sheffield S11 9PS

Tel: 0114 271 1060

Fax: 0114 271 1089

Email: [nhs.screening@sheffield-ha.nhs.uk](mailto:nhs.screening@sheffield-ha.nhs.uk)

Web site: [www.cancerscreening.nhs.uk](http://www.cancerscreening.nhs.uk)

© NHS Cancer Screening Programmes 2001

The contents of this document may be copied for use by staff working in the public sector but may not be copied for any other purpose without prior permission from the NHS Cancer Screening Programmes.

ISBN 1 871997 34 8

## **CONTENTS**

	<b>Page No</b>
<b>1. INTRODUCTION</b>	1
1.1 Case of need	1
1.2 Remit of the working party	2
1.3 Membership of the working party	2
1.4 Methodology	2
<b>2. PROMPTING IN MAMMOGRAPHY</b>	4
2.1 What do we mean by prompting?	4
2.2 What is a prompting system?	4
2.3 Why not use a computer to read mammograms automatically?	4
2.4 Why does prompting work?	5
2.5 Clinical prompting studies	5
2.6 Automatic detection of mammographic abnormalities	7
2.7 Methodology for evaluating algorithm performance	7
2.8 Discussion	9
<b>3. COMPANY REPORTS</b>	11
3.1 R2 Technology Inc	11
3.2 CADx Medical Systems	14
3.3 DBA Systems Inc/Titan Systems Corporation	16
3.4 MedDetect	17
3.5 Cambridge Parallel Processing/OXIVA	18
3.6 SCANIS Inc	18
3.7 International Medical Multimedia	19
3.8 Summary	19
<b>4. FULL FIELD DIGITAL MAMMOGRAPHY AND CAD</b>	24
4.1 Full field digital mammography and CAD	24
4.2 Guidance on equipment for CAD in mammography	25
4.3 Research issues	26
4.4 Ergonomic and psychological issues	28
<b>APPENDICES</b>	32
1 Questionnaire on prompting in mammography: systems performance	32
2 Company contact details and web sites	35
<b>REFERENCES AND BIBLIOGRAPHY</b>	37



## **MEMBERS OF THE CAD WORKING PARTY**

Professor Fiona J Gilbert, Department of Radiology, University of Aberdeen (Chairman)

Dr Alistair Kirkpatrick, Edinburgh Breast Screening Programme

Dr Caroline Boggis, Manchester Breast Screening Programme

Dr Sue Astley, Medical Physics Department, Manchester University

Professor Stuart Field, Chairman, Royal College of Radiologists Breast Group

Professor Alastair Gale, Applied Vision Research Unit, University of Derby

Mr Colin Hancock, Consultant Engineer, NHS Cancer Screening Programmes

Dr Ken Young, National Coordination Centre for the Physics of Mammography

Dr Julie Cooke, Jarvis Breast Screening Training & Diagnostic Centre, Guildford

Dr Sue Moss, Cancer Screening Evaluation Unit

Dr Roger Blanks, Cancer Screening Evaluation Unit

Dr Linda Garvican, Public Health Specialist

## **ACKNOWLEDGEMENTS**

In addition to the Working Party, we acknowledge the assistance of the following:

Julietta Patnick, National Coordinator, NHS Cancer Screening Programmes

Richard Winder, Deputy National Coordinator, NHS Cancer Screening Programmes

Anne Richards, Eastman Kodak

Jeff Byng, Eastman Kodak

Maureen Gillan, University of Aberdeen



# 1. INTRODUCTION

## 1.1 Case of need

The UK NHS Breast Screening Programme (NHSBSP) was started in 1988 following publication of the Forrest report.<sup>1</sup> Over a five year period, 95 breast screening units were set up throughout Scotland, England, Northern Ireland and Wales. This programme resulted in a large number of radiologists undergoing extensive training in mammography in order to become expert in reading screening mammograms. Over the 10 years since the start of the programme, it has become increasingly difficult to attract radiologists into posts with a large component of breast work.

Another major concern is that of the number of false negative errors that can arise in the screening programme, ie cases in which a mammogram containing a significant abnormality is classified as normal. It is thought that up to 25% of cancer cases occurring in the interval between screens are missed by the screening radiologist. It is proposed that prompting with computer aided detection (CAD) would reduce the number of false negative cases. These cases are sometimes subject to litigation, and any reduction would represent a considerable saving to the NHS.

Demand on radiologists' time has increased over the past 10 years for several reasons: an increasing proportion of the population are attending for screening, more centres are moving towards double reading, procedural work at assessment clinics has increased and a greater emphasis on quality assurance measures requires more extensive audit of the performance of both units and individuals. The success of the screening programme has resulted in an increased awareness of breast disease in the general population, resulting in additional workload in symptomatic clinics. The high standards developed by the screening programme in radiological procedural work have been transferred to the symptomatic setting, with a consequent increase in the requirement for radiological resources.

A study of radiology staffing levels in breast screening units found that 41% of breast screening units have unfilled radiology sessions and that staff morale is low because of increased pressure of work or the threat of, or ongoing, litigation.<sup>2</sup> Two-view mammography is being introduced for all screening episodes, and this will result in a 33% increase in the number of films to be read by radiologists. In addition, the age for routine invitation has recently been increased to 70, adding two more screening rounds to the existing five and thus increasing workload by a further 40%.

The workforce problem in radiology is not confined to posts in breast imaging. The Department of Health has already accepted that there is a national radiology workforce shortage and recommended a considerable increase in training numbers in England and Wales. Lack of funding has meant that only a small proportion of these training posts have been filled, but it now appears likely that the target of 70 additional radiology training posts for September 2000 will be achieved.<sup>3</sup> A number of

possible solutions have been put forward to ease the problem of film reading, including the introduction of non-radiologist film readers. However, it will take at least five years to train the extra radiologists and a considerable time to train and mentor radiographers to read mammograms in sufficient numbers to ease the current workforce crisis.

CAD in mammography has shown promising results when used as a 'first reader' to prompt radiologists to likely abnormalities. Considerable research has been carried out in this area, firstly developing algorithms to detect various abnormalities found on mammograms and, secondly, looking at the effects of prompting on radiologists' performance on reading mammograms. It has been suggested that CAD may be particularly effective for less experienced film readers, but as yet there is no evidence to support this hypothesis.

### **1.2 Remit of the working party**

The Radiology Quality Assurance Group of the NHSBSP asked a working party to examine the current status of CAD, to advise whether CAD should be considered within the context of the UK breast screening programme and to identify areas of research that would be needed before CAD can be introduced into the screening programme.

### **1.3 Membership of the working party**

The working party was multidisciplinary and comprised radiologists from the breast screening programme who have had some experience with CAD, medical physicists experienced in screening or CAD, an applied vision psychologist with extensive research experience of radiologist and radiographer film reading performance in the NHSBSP, an engineer with extensive experience in mammography and statisticians from the Cancer Screening Evaluation Unit. The working party was supported by the national coordination team of the NHS Cancer Screening Programmes.

### **1.4 Methodology**

The working party met on three occasions. The first meeting was used to define the aim of the group and the content of the report and to agree how companies involved in the production of CAD equipment would be identified and evaluated. It was agreed that companies would be invited to give presentations and be interviewed using a structured questionnaire (Appendix 1). Companies were identified through personal contact, a review previously carried out by Eastman Kodak and through international radiology meetings (European Congress of Radiology, Symposium Mammographicum, the International Workshops on Digital Mammography and the Radiological Society of North America). A list of companies approached is given in Appendix 2.

The structured questionnaire, devised by Sue Astley, was circulated and agreed by members of the group, and was then distributed to companies. Companies were invited to attend the second meeting of the group in London to give a presentation and undergo a structured interview, at which information was gathered. Companies that were unable to send a representative to London for evaluation were sent the questionnaire and asked to provide as much information as possible on their system. The outcomes of the questionnaires and presentations are given in Chapter 3.

As the technology was still evolving, it was felt it would be useful for the working party to agree a specification for manufacturers, as this has previously been influential in the development of mammography equipment. Research groups currently working on CAD were identified by a literature review and from international meetings, in particular the 4th International Workshop on Digital Mammography (Nijmegen 1998). As there appeared to be a number of unresolved issues with regard to prompting technology and behaviour in response to prompting, it was agreed to recommend further research. The equipment specification and recommendations for further research are discussed in Chapter 4.

## 2. PROMPTING IN MAMMOGRAPHY

### 2.1 What do we mean by prompting?

One method of helping radiologists screen mammograms for signs of cancer is prompting, in which computer based algorithms are used to detect potential abnormalities in digital images and to draw attention to the corresponding regions in the original films. Prompting has the potential to improve an individual's detection performance, provided the prompts are sufficiently accurate.

### 2.2 What is a prompting system?

A prompting system comprises a number of different components, one of which is the acquisition of digital data. At present, most mammograms are produced using film–screen systems. Before films can be processed by a prompting system they must be digitised, ie the film densities must be transformed into a matrix of numerical values (grey levels). To cope with the high throughput of breast screening, films are generally fed into the digitiser via a hopper. Barcoding may be used to facilitate the association of the digital data with the original film. Clearly, the advent of full field digital mammography will facilitate the process and may well increase the acceptability of prompting in a screening environment.

Another essential component is a system that enables automatic identification of locations in the digital image that are likely to correspond to significant abnormalities. This involves manipulating the digital data by applying a series of algorithms designed to identify regions in which abnormal patterns of grey levels are present. Prompts will be generated at the most suspicious locations.

The prompting system must display any prompts in such a way that the radiologist's attention is drawn to the corresponding regions of the original mammogram. For example, a system may present the prompts on a computer screen or printed out on paper, superimposed on a low-resolution version of the original image.

### 2.3 Why not use a computer to read mammograms automatically?

One approach to overcoming the shortage of expert mammographic film readers would be to use a computer to prescreen mammograms to identify those that are unequivocally normal. This would enable radiologists to concentrate their efforts on examining only the small percentage of screening films that are potentially abnormal. Unfortunately, this approach is currently not feasible, as it would require the development of highly sensitive algorithms to detect **all** possible types of abnormality. Such levels of sensitivity are not currently available.

Prompting, however, allows radiologists to exploit some of the benefits of computer based analysis, such as reproducibility and objectivity, without making unrealistic demands on algorithm performance, as it requires neither a complete suite of algorithms nor perfectly sensitive prompts.

### 2.4 Why does prompting work?

Prompting aims to improve radiologists' performance by reducing the number of false negative errors, ie cases in which a mammogram containing a significant abnormality is classified as normal. Research has shown that nearly a quarter of interval cancers correspond to false negative errors, and in approximately a quarter of these clear signs of malignancy were missed by the screening radiologist.<sup>4</sup> These errors represent a failure of the radiologist's search strategy.<sup>5</sup> In the remaining false negative cases, the signs of abnormality are more subtle and the error could either have been one of search or one of interpretation, ie the abnormality could have been detected but dismissed as insignificant.

Psychophysical research has shown that the ability to discern areas of an image that are in many respects similar to the background (such as mammographic abnormalities on a radiographic film) requires a systematic search of the image.<sup>6</sup> However, eye movement studies have shown that radiologists' search patterns in mammograms are far from systematic.<sup>7</sup> The requirement for systematic search can be bypassed by drawing attention to the appropriate locations.<sup>6, 8</sup>

Much fundamental research in this area has involved highly constrained experiments and is thus difficult to relate specifically to mammographic screening, in which both abnormalities and backgrounds can vary considerably. Although some of the outstanding questions – such as how false prompts affect performance – can be answered using realistic synthetic images, it is still necessary to perform experiments using real mammograms for which the ground truth has been determined by informed experts.<sup>9</sup>

The first study of the effects of prompting in mammography focused only on the detection of microcalcification clusters, using prompts overlaid on printed digitised mammograms.<sup>10</sup> Despite a number of problems with the experimental design, this research demonstrated the potential of prompting; a significant improvement in detection performance was achieved using an algorithm giving 87% sensitivity at four false prompts per image.<sup>11</sup>

A subsequent, more realistic, study using screening mammograms and paper prompts showed that any benefit from prompting is reduced when false prompt rates are high.<sup>11</sup> In this study, with an algorithm correctly prompting 90% of abnormalities, prompting significantly improved performance only when the false prompt rate was less than about 0.3 per image. These results are consistent with those of another experiment that examined the detection of spiculated masses between 1.2 and 2.7 cm in diameter. In this case, the algorithm operated at 97% sensitivity with an average of 0.28 false prompts per image, and a highly significant improvement in radiologists' detection performance was achieved.<sup>12</sup>

### 2.5 Clinical prompting studies

The R2 ImageChecker system has undergone extensive trials in the USA and more recently in Europe.<sup>13</sup> Briefly, the trials demonstrated high algorithm sensitivity for microcalcifications, prompting over 98% correctly. They also prompted over 72% of masses, with approximately one false prompt per image (ie four per four-view case). Since these

results were published, R2 claims to have reduced the false prompt rate to 0.6. Although it was shown that using the system did not increase the participating radiologists' recall rate, no improvement in detection performance could be demonstrated. However, a retrospective review of prior screening films of patients with cancer showed that the system is capable of detecting a large proportion of subtle lesions.<sup>14</sup> Although these results appear impressive, it does not automatically follow that the system would improve the detection of such early cancers in a screening context, as radiologist performance with the system depends also on the false prompt rates.

The PROMAM system was evaluated in Scotland.<sup>15</sup> This system correctly prompted nearly 94% of microcalcification clusters, 73% of masses and 82% of cases in which both a mass and microcalcifications were present. False prompts were expressed in terms of the percentage of women prompted; overall, this was 83.6%. Once again, the recall rate was not increased and there was no improvement in radiologists' performance.

The Second Look system is currently under evaluation at the Nightingale Breast Centre in Manchester. Preliminary results show an average of 1.4 false prompts per normal screening film, compared with the manufacturer's quoted 1.3. This difference may be due to the data set selection criteria. In the Manchester tests, all screening films for patients without cancer were used, and the data set contained a significant proportion of benign features (calcifications, cysts, etc). For cancer cases, the average number of false prompts fell to 1.25. The sensitivity of the system was 75.4% (on a per case basis). This is considerably lower than the 85% claimed by the manufacturers. The main reason for the difference is probably because the manufacturer's studies always considered four-film cases, whereas in Manchester only one radiographic view of each breast is taken in approximately two-thirds of women. Current studies are aimed at determining the effect of prompting on radiologists' and radiographers' performance.

The question remains: why has neither the R2 trial nor the PROMAM trial demonstrated any improvement in radiologists' detection performance? It seems likely that the answer lies in the relatively high false prompt rates, although a number of other factors could also have affected the results.

As the sensitivity of an algorithm is increased, the number of false positive prompts also increases, but there is as yet no systematic way of determining the optimum balance between the two. In addition, there is little knowledge about human responses to prompts when multiple types of abnormality are prompted using algorithms with different error rates. There are also a number of different prompting strategies which merit investigation to determine the most effective way of improving performance.

### 2.6 Automatic detection of mammographic abnormalities

Researchers have been developing algorithms to detect mammographic abnormalities for the past 25 years. Some of these algorithms are highly specific to a particular mammographic sign of abnormality, but others are generic methods for detecting patterns of brightness that could appear in several different types of image. For example, methods to detect bright oval blobs in images can be applied at different scales to detect both microcalcifications and larger tumours.

It is not always easy to determine the key characteristics that differentiate normal and abnormal mammographic appearances, as both show a high degree of variability. One method of dealing with this problem involves extracting information from a large training set of images chosen to represent the expected range of appearances. This information is used to produce statistical models that encapsulate the important features of the training data, and can be used to detect these features in previously unseen images. Neural networks have also proved valuable, particularly for integrating evidence and classifying data extracted from images.

The majority of research effort in this field has been directed towards the detection of microcalcifications, followed by masses (particularly spiculated), asymmetry and distortion. This is partly because of the difficulty in understanding what discriminates normal appearance from abnormal for some of the more subtle signs. As a consequence, the success rates quoted for microcalcification detection are much higher than for any other type of abnormality. There is, however, wide variation in the methodology used to evaluate algorithms. This makes direct comparison of algorithms difficult, a problem discussed in more detail below.

### 2.7 Methodology for evaluating algorithm performance

#### 2.7.1 *Selection of data*

There are two main approaches to selecting data with which to train and test algorithms. The first method involves taking sequences of screening films. For example, a consecutive sequence of normal films could be used to establish the false positive rate on normal films, and a consecutive sequence of abnormal films could be used to determine both the sensitivity and the false positive rate on abnormal films. From these it would be possible to estimate the sensitivity and false positive rate that would be achieved in a screening situation. The main advantage of this approach is its integrity; the principal difficulty is ensuring that the number of examples is sufficient to represent adequately the variety of pathologies and normal appearances that could be encountered. In practice, it is often difficult to collate such a test set, as the normal films are likely to be at least three years old (in order to ensure normality at a subsequent screening session) and the abnormal films are frequently removed from archives for clinical purposes. Random selections of films within a given time period are sometimes used instead of a consecutive sequence.

An alternative approach involves the selection of films that are deemed by an expert to be representative of those which might be encountered in routine screening; films are also selected to include examples of a wide range of pathology, including those only infrequently encountered. The main advantage of this approach is that the range of appearances is likely to be broad; however, the method is subjective and results cannot easily be related to routine screening.

Some researchers have also evaluated their results on prior films from interval cancer databases. These are films taken before the detection of cancer on which the abnormalities have been identified with hindsight. Results showing the sensitivity of an algorithm on such a database indicate the potential of the method for detecting very early cancer, but should be approached with caution. Although they may facilitate direct comparison of different algorithms, the relationship of such results to the ability of the algorithm to produce useful prompts for this type of subtle abnormality is complex, and also depends on the false positive rates for different categories of images.

### 2.7.2 *Ground truth*

Another important issue is the definition of the ground truth against which the algorithms are tested. Ideally, pathological evidence should be used, but this is not generally feasible. For practical reasons the ground truth is often obtained by having at least one radiologist annotate the original image, either on a registered acetate overlay, which is then digitised and aligned, or on-screen. This process is tedious, time-consuming, subjective and error-prone. It would be better to obtain multiple independent annotations, so that a truth 'probability' image could be constructed.

### 2.7.3 *Receiver operating characteristic analysis*

Most detection methods take as their input digitised mammographic image and produce as output another digital image in which the grey level values correspond in some way to the probability of there being a target abnormality at the corresponding pixel location in the original image. One of the simplest ways of evaluating the performance of a method is to apply a threshold to the output image and to label as abnormal every pixel with a grey level above that threshold value. The remaining pixels are labelled normal. Given the ground truth (ie those pixel locations that really correspond to abnormalities), it is possible to quantify the success of the algorithm. By repeating the process at a range of thresholds, a receiver operating characteristic (ROC) curve can be constructed, in which the true positive fraction is plotted against the false positive fraction.<sup>16</sup> ROC curves give a useful picture of how an algorithm performs, and enable the selection of an appropriate operating point (in this case corresponding to a threshold value).

There are, however, some practical issues that need to be addressed. As the optimum threshold is likely to vary from image to image, the method must be tested on a set of images and a threshold selected on the basis of overall performance. A number of methods attempt to do this in an unbiased way by splitting the dataset into training and test components. Although better methods are available, many researchers have opted to use a 'leave-one-out' methodology, which slightly

underestimates performance, although some quote results for ‘leave-all-in’, which slightly overestimates performance.<sup>17</sup> Clearly, both the size and the composition of the data set are important; until recently, no large publicly available mammogram databases were available, and many researchers have used only very small, and often unrepresentative, sets of images.

### 2.7.4 *Measuring whether an algorithm produces useful prompts*

The usefulness of pixel classification ROCs is limited as they do not carry any information about the **distribution** of true and false positive pixels.

If the ultimate aim of an algorithm is to enable the system to place a prompt with sufficient accuracy to aid a human observer, then its ability to generate useful prompts must be measured. For this reason, free response operating characteristic (FROC) curves are usually generated.<sup>18</sup> FROC curves plot the true positive rate against the number of false positive detections per image; any contiguous group of pixels with the same label are counted as a single detection.

To generate a FROC curve, instead of simply counting the number of labelled pixels corresponding to positions in the abnormality, it is necessary to define whether or not a contiguous group of pixels labelled by the algorithm as abnormal corresponds to a ‘hit’ on the abnormality or a ‘miss’. A wide range of methods has been proposed, including looking for a 50% overlap between the detected and truth regions and determining the distance between these regions’ centres of mass, but none has been universally adopted by the vision community. Some methods in current use fail to take into account the relative sizes of the lesion and the detected region – a small detected area within a larger lesion might enable the generation of a useful prompt, but a large detected area containing a small lesion may not.

## 2.8 Discussion

The wide variety of methods used to evaluate detection algorithms makes meaningful comparison of detection methods and prompt generators very difficult. The research community has, in some cases, adopted common data sets (most notably the very limited Mammographic Image Analysis Society (MIAS) database, produced in the UK, and the extensive collection of images available from the University of South Florida). However, the use of common data is not widespread. In this country, many researchers prefer to gather data from their local screening centres, in order to specify fully the type of annotations required to train and test their algorithms, and to be able to relate their results directly to the mix of mammographic appearances found in the NHSBSP. One of the potential advantages of initial experience with CAD systems might be the creation of a substantial research database of digitised, annotated screening mammograms.

For microcalcification detection, Karssemeijer’s method, adapted by R2, is widely regarded as one of the most sensitive and specific.<sup>19</sup> The current challenge is to maintain the high sensitivity of this method while reducing the false positive rate by eliminating prompts of clearly benign calcifications. Strategies include modelling particular types of

clusters and exploring the relationship of detected groups of particles with underlying background structures. For mass detection, the situation is less clear cut, partly because the task is more difficult and partly because there is greater subjectivity in selecting data sets and defining ground truth. The wide variety of size and appearance of masses, and of the relationships between the target abnormalities and their backgrounds, means that few academic research groups have tested their algorithms on data sets of sufficient size to enable adequate prediction of clinical performance in the context of the UK breast screening programme. Once again, high sensitivities can be achieved at the expense of specificity.

### 3. COMPANY REPORTS

#### 3.1 R2 Technology Inc. *ImageChecker M1000*

R2 Technology Inc. is an American company founded in 1993 to market the University of Chicago's emerging CAD technology. This was developed at the Kurt Rossman laboratories for Radiologic Image Research at the university's radiology department. The company, based in Los Altos, California, designed and developed the proprietary signal processing neural network technology ImageChecker. In April 1998, R2 Technology Inc. signed an agreement with GE Medical System giving GE rights to develop the use of ImageChecker in conjunction with its digital x-ray equipment.

The ImageChecker uses algorithms originally developed by Dr Nico Karssemeijer and colleagues in Nijmegen in 1992. The algorithms in the proprietary signal processing software identify regions of interest (ROIs) based on physical characteristics with visually perceptible structures containing either clusters of bright spots, indicative of microcalcifications, or dense regions with radiating lines, suggestive of masses or architectural distortions. Thus, the marked areas may have some of the characteristics of masses or microcalcifications but may not actually represent abnormalities.

The prompted images are displayed on low resolution monitors. A barcode reader enables the films to be identified and allows random access to the digitised images, allowing them to be matched up with the original mammographic films displayed on the viewer. The system uses two different markers: a triangle for a cluster of bright spots and an asterisk for a density with radiating lines.

The radiologist reviews the mammographic films in the conventional manner, and then activates the system to display the low resolution marked images. The radiologist may then go back to the original films to reassess any of the marked areas and revise his or her opinion if appropriate. The ImageChecker makes no diagnostic interpretation of the ROI. The average number of false prompts per film is 0.6.

The ImageChecker can be used only for standard screening films, not for magnification or spot compression views. It can analyse craniocaudal, mediolateral oblique and lateral views, on either 18 cm 24 cm or 24 cm 30 cm films. A maximum of four films (usually two views of both sides) can be analysed at one time, so the system is not suitable for a 'mosaic' (ie as in the case of a breast that was too large to be imaged on a single film). The films must not be bent or damaged and should have no labels or tape within 1 mm of the edge.

#### *Microcalcifications*

The system places a triangle over the centre of the ROI on the low resolution image. Normal breast structures such as benign calcified tissues or crossing linear tissues may be shown as false prompts. According to the manufacturer, the system will not mark either very small clusters with fewer than three elements or bright spots separated by more than 2.5 mm.

### *Masses*

The ImageChecker uses an ‘annulus’, consisting of inner and outer concentric circles of diameter 6 and 32 mm, which is moved over the image. When the software encounters line segments radiating from a common origin (the inner annulus) it marks the region of maximum convergence with an asterisk on the low resolution image. It is also sensitive to the presence or absence of a central density. False prompts include ducts and tissue radiating from the nipple and inadvertent scanning of parenchymal tissue. The algorithm has been optimised for masses of 10–20 mm diameter but will detect masses of diameter between 3 and 30 mm. It is more likely to detect a spiculated mass than either an ill-defined mass with irregular border or a lobulated mass with well-defined border.

Formal approval to market the product was granted by the US Food and Drug Administration (FDA) in June 1998. The FDA is concerned that use of a CAD system does not result in a significant increase in the numbers of women being called back for assessment, especially for additional biopsies. It also has to be assured that the system is capable of correctly identifying areas associated with a diagnosis of cancer. The FDA is not concerned with the impact of a device on workflow or any value-for-money issues, which might be considered important to the NHS. FDA approval is a certificate of safety rather than of clinical effectiveness or efficiency in the sense normally understood in the UK.

A summary of clinical studies of ImageChecker is given in Table 3.1. When considering these trials it is important to recognise the inherent differences in mammography in the USA:

- the age range of the screening population is wider and younger women are included; as a result, the cancer detection rate is about 4 per 1000 women screened overall
- the screening population is self-referred, with a recommended screening interval of only one year
- two views of each breast are standard at all screens
- the system of record keeping is different, with no necessary continuity or access to previous screening history
- the level of mammographic expertise varies widely – some radiographers screen only 100 mammograms per month
- more time for reading each film.

There have been no published studies on the reaction of radiologists to the prompts generated by the system, measuring either response to false positive prompts or the possibility of ignoring true positive markers. R2 states that there is no systematic pattern of features likely to result in false negatives other than positioning, poor contrast, motion or experience of the neural network. False positive prompts relate to complex glandular structure, small benign calcifications or crossing vascular structures.

Table 3.1 Clinical trials of the R2 ImageChecker M1000 system<sup>14</sup>

Study aim	Study design	Number of centres/ radiologists/cases involved	Results	Conclusion (according to manufacturers)
To support the claim that ImageChecker does not increase the number of 'patient work-ups'	Multicentre, prospective study	Five institutions Fourteen radiologists	No significant difference in work-up rates, but radiologists' ranged from 2.2% to 15.2% before installation and 3.3% to 11.3% (mean 7.4%) with the R2 system	Demonstrates no significant increase in work-up rate following installation, either by individual radiologists or group as a whole
To determine the ability of the system to detect actionable abnormalities in prior screening films of cancer patients	Multicentre retrospective trial 'Current' films where cancer was detected were compared with 'priors' previously reported as normal	Films from 12 institutions read by a blinded panel of five expert radiologists	Twenty-three per cent of all 'priors' judged actionable by majority of panel. ImageChecker correctly marked 82% of these cases	Demonstrates effectiveness at minimising false negatives due to observational lapses
Determination of overall sensitivity in identification of lesions in proven cancer cases	Multicentre retrospective trial	Thirteen institutions 1010 consecutive cases from over 200 000 screening mammograms	ImageChecker correctly marked 98% of microcalcifications, 74% of masses and 83% of cancers overall	Direct measure of sensitivity to mark microcalcifications and masses from a large representative sample of screening mammograms
Determination of intra- and intersystem reliability to mark the characteristics associated with cancer	Multicentre retrospective trial	Twenty-five current (two view) cases run 10 times each of three systems	Of the 1500 evaluations, ImageChecker correctly marked the feature and location with a precision of 99.93%	Highly reproducible within and across systems

A recent North American study retrospectively reviewed screening mammograms taken over a two-year period prior to the diagnosis of cancer in 430 patients.<sup>14</sup> With the benefit of hindsight, 67% had some abnormality at the lesion site visible on the previous films. On blinded review, at least three of a panel of five radiologists judged 40% of these to be actionable, but ImageChecker marked 60% correctly. In a separate prospective study the recall rate was not increased significantly by the use of ImageChecker, despite the fact that the system placed an average of one mark on each normal screening film.

R2 claims that, after initial familiarisation with the system, experienced radiologists can work faster because they do not spend time looking for microcalcifications. All trial participants were given one month to become familiar with the system before entering into the studies. In the retrospective trials, radiologists were not given prior knowledge of the abnormality rate as this could have adversely affected their performance.

Systems are now installed in about 20 sites in the USA. In Sweden, one system has been in use for about two years in a screening programme situation rather more analogous to that in the NHSBSP. Dr Nils Bjurstam has reported that he finds the system helpful for an expert breast screening radiologist but not for an inexperienced film reader.

In April 2000, the FDA granted R2 clearance to market an upgrade of its Version 2.2 software, which increases the overall sensitivity of the ImageChecker M1000 system to 90%. Specificity remains relatively low. The system is now available for purchase in the UK at a list price of £108 000. A service contract costs about £10 000 per annum, which includes software upgrades.

The ImageChecker is currently one of two commercially available CAD systems for mammography. To date over £20 million has been invested in its development, and as investors are now keen to see a return on their capital the company is marketing the system aggressively in the USA and Europe. Although the manufacturer may be willing to lend systems to the NHSBSP for evaluation, it is unable to contribute to the funding of a research project. The level of investment required to bring such a product to market indicates that it may be difficult for other companies to raise the necessary resources to complete the development and clinical trial phases on alternative products.

### **3.2 CADx Medical Systems** *Second Look*

Qualia Computing Inc. was founded in 1997 by Steven Rogers PhD, a retired smart weapons specialist at Wright Paterson airforce base in Ohio, USA. Qualia's product, Second Look, uses similar artificial intelligence and neural network algorithms to those employed by the US Air Force fighter jets. Qualia Computing is associated with Biochem Pharma, a major pharmaceutical company. The two companies are in a partnership with CADx Medical Systems, a Canadian based company that develops and markets CAD systems for mammography and other health related applications. Second Look is now available for commercial use throughout Europe, Latin America, Canada and

Asia/Pacific and will be available in the USA pending FDA approval. The Karolinska Hospital, Stockholm, Friedrich-Schiller University, Jena, and Groupe Hospitalier Necker, Paris, have been involved in clinical trials.

The Second Look system digitises mammograms and then analyses the digital images for suspicious regions, which are highlighted on a reduced resolution paper printout known as a mammograph. The algorithms target two types of abnormality, microcalcification clusters and focal masses. Each of these is prompted with a different symbol (an ellipse for masses and a rectangle for clusters of microcalcification), sized according to the extent of the detected abnormality. Any detected masses and clusters of more than three particles of microcalcification are prompted. A maximum of five prompts is allowed on a two-film case and nine on a four-film case.

It is intended that the radiologist, after initial inspection of the film mammogram, consults the mammograph printout and then takes a second look at the corresponding original film. CADx has evaluated Second Look on an in-house database of 1300 normal cases and 1200 biopsy proven cancers. A sensitivity of 85% was achieved, with an average of 1.3 marks per non-cancer case and 1.8 per cancer case. Thirty-six per cent of normal images had no marks. Reproducibility was 95% for well-characterised cases, and the average time to analyse a four-film case was 8 minutes.

The Second Look system incorporates a hopper that allows several cases to be loaded at once. The system copes with both 24 × 30 cm and 18 × 24 cm film formats, and the use of case separators allows mixing of two- and four-film cases. Patient identifiers and demographic data can be typed in by an operator. Alternatively, a barcode reader is provided. The system requires regular cleaning of the digitiser rollers, camera window and illuminator at intervals of approximately 250 films; this is a simple task that could be carried out by the person responsible for feeding films and entering data. Should films be entered the wrong way up, or in the wrong order, they can be realigned using a touch screen display.

Two trials of the system are currently under way. A retrospective trial will assess the efficacy of Second Look in the detection of missed breast cancers with the aim of increasing the sensitivity of screening mammography. This trial has the following objectives:

- to assess the claim that the use of Second Look results in a clinically significant increase in the sensitivity of a radiologist compared with the current practice of a single read
- to determine, from biopsy proven cases, the sensitivity of Second Look for mammographically screen-detected cancer
- to verify reproducibility by identifying cases from screen-detected cancers and processing each 10 times on multiple Second Look systems, following which a radiologist will compare these outputs with the pathology-confirmed mammogram findings to assess intra-system and inter-system repeatability.

A prospective trial will assess whether the use of Second Look in a clinical environment significantly alters the work-up rate or the positive predictive value of screening mammography.

The Second Look system is currently under evaluation at the Nightingale Breast Centre. The system's performance has been established on screening mammograms from the NHSBSP, and its effect on reader performance is now being evaluated. In addition, issues of throughput will be addressed using current screening films.

### 3.3 DBA Systems Inc./Titan Systems Corporation *PROMAM*

As a result of a government initiative to encourage technology transfer from pure science to alternative applications, the Royal Observatory, Edinburgh (ROE) identified a CAD system to aid in breast cancer screening as a suitable technology transfer area for consideration. The scientific and technological aspects of the problems facing the breast cancer screening programme were similar to those addressed by the Super Cosmos measuring machine designed and built by the ROE to digitise and analyse photographs of the southern hemisphere night sky. It was recognised at the outset that this was an ambitious and high risk project, but nevertheless a suitable one for the Particle Physics and Astronomy Research Council (PPARC) to address in collaboration with a commercial partner.

The plan was to develop a commercial software package that would use a mammographic film scanner to convert analogue images to digital form and, at some future date, to analyse images directly from a digital imaging system. The application of image processing techniques to digital images would provide prompts on low resolution photographs of the mammograms to draw the attention of the radiologist to suspicious features in the corresponding film images. The primary objective was embodied in the acronym PROMAM: PROMpting for MAMmography.

The PROMAM system was brought to the stage of a preclinical trial with two arms. First, the system was run on a set of 84 interval cancers from the Edinburgh Breast Screening Programme. Secondly, a set of 2002 cases, which included 102 pathology proven cases, again from the Edinburgh Breast Screening Programme, were analysed by the system. The system correctly identified 37 of the 84 interval cancer cases, suggesting a 5% improvement in detection. The prompting system improved the sensitivity for four out of the five readers, with the fifth reader suffering a reduction in sensitivity when prompted.

After the completion of the preclinical trial in October 1997, funding from the PPARC ceased. This and difficulties encountered with the commercial partner, DBA Systems Inc. of Melbourne, Florida, resulted in progress on the system being halted until DBA Systems Inc. was taken over by the Titan Corporation. The company renewed its interest in this field and started to develop a Beta System based on the PROMAM system. Only two algorithms were employed in the PROMAM clinical trial: a calcification detection algorithm developed in conjunction with Nico Karssemeijer and a mass detection algorithm developed by

Lance Miller of the ROE. In a presentation given in November 1998, DBA claimed that the system was reliable and described a number of enhancements it intended to implement. However, the relationship between ROE and the Titan Corporation subsequently broke down and, to our knowledge, the technology developed by ROE has not been exploited elsewhere. We understand that the Titan Corporation has withdrawn financial support from its UK subsidiary and the PROMAM project is no longer viable in this country.

### 3.4 MedDetect

MedDetect is an early stage development company formed by Lockheed Martin Astronautics (LMA) and Rose Biomedical Development Corporation (RBDC) to perform medical image analysis. The company's optical processing technology was originally developed for automated target recognition. MedDetect's processor, a hybrid digital and optical computer, provides the imaging processing power to perform near real-time feature extraction in the multiresolution and multiorientation domain that preserves image details for improved CAD performance. This proprietary technology was licensed to MedDetect for 25 years on a royalty free basis that covers all medical applications and includes all future technology improvements.

MedDetect has a two-stage processing architecture for mammograms. The first stage uses optical processing (OP), bandpass and wavelet based methods with no algorithmic approximations to detect and segment ROI. The OP is used to measure feature metrics for the ROI. The second stage of processing uses digital computing to run a neural network to combine the segmented raw pixel data and the measured features from several domains to determine the context of these data, and to establish the likelihood of the detected lesion being cancerous.

MedDetect's algorithms were developed entirely for spiculated lesions, and at the University of South Florida (USF) Moffitt Cancer and Research Centre for detection of calcifications. USF has also developed algorithms to detect masses that have not yet been incorporated. The USF algorithms use multiresolution wavelet methods. The microcalcification and mass detection algorithms were developed by Dr Laurence Clarke at USF, who has been working in this area since 1993. A modular approach was developed to address noise suppression, feature extraction, feature selection using generic algorithms and classification by multistage neural networks. Each image preprocessing module can be individually optimised to improve sensitivity and reduce the false positive detection range.

The company has pioneered the development of M channel multi-resolution wavelet transforms for microcalcification with image detail preservation for improved detection and classification. Multiorientation N direction wavelet transforms allow for directional feature extraction for both segmentation of suspicious masses and direct measurement of spiculation analysis. USF investigators are completing their CAD methods for classification. It is thought that the multiresolution and multiorientation properties are superior to the single scale approach used by other companies.

MedDetect is currently in the development phase and is about to enter clinical trials. Insufficient preliminary work has been performed to assess the efficacy of the algorithms. There appears to be considerable financial investment in this product. No clinical trials have been performed yet although a multicentre clinical trial based in Denver is planned.

### 3.5 Cambridge Parallel Processing/OXIVA *Xmammo*

Over the past nine years, the Oxford laboratory headed by Professor Michael Brady and Dr Ralph Highnam has developed algorithms for mammographic image analysis. In particular, the laboratory has developed a software package called Xmammo, which provides a physics-based model of the image formation process. The system was developed with radiologists in the breast care unit of the Churchill Hospital, Oxford. The Oxford group linked up with Cambridge Parallel Processing (CPP), a company specialising in real-time algorithm implementation, with the aim of developing a CAD system, but this has not materialised. Brady and Highnam have decided to try to exploit their research commercially by founding OXIVA (Oxford Intelligent Visualisation and Analysis). This company is developing:

- normalisation and quantification of breast images
- understanding and modelling of breast compression
- disease simulation
- construction of three-dimensional images from two-dimensional imagery.

OXIVA is now bringing products and advanced prototypes to the market ready for licensing or sale. However, as yet a complete CAD system is not available.

### 3.6 SCANIS Inc. *MAMMEX TR*

The SCANIS patented technology, derived from photogrammetric image interpretation techniques developed from imaging software experience with satellite and aerial image analysis, was developed and tested with assistance from the University of California at San Francisco and Stanford Medical Centre. The company plans to sell in Europe and the USA, initially for research and education (non-clinical use) until FDA approval is obtained.

A rule based algorithm tests up to 100 breast cancer image features. MAMMEX TR is predominantly concerned with the detection of abnormalities rather than diagnosis. MAMMEX TR has demonstrated better than 95% accuracy in detecting masses and calcifications, with an average false positive rate of less than two per image. Work is ongoing to reduce the false positive rate to one per film without affecting the sensitivity.

The company signed a joint corporation agreement with the Thompson Ramo Woolridge (TRW) Healthcare Information Systems Unit (San Bernadino, California). Under the three-year, \$12 million contract, TRW will provide engineering services such as new product development, testing, and installation for the MAMMEX TR system sold by SCANIS Inc. to hospitals and imaging clinics worldwide. TRW also provides

the operational and clinical training for clinicians using the system and assists SCANIS Inc. in conducting formal clinical trials and compiling data leading to regulatory approval of the system in the US and international markets. The MAMMEX TR has been approved for sale in the EU and Canada and an application to the FDA is being prepared.

Clinical studies have been conducted by Dr Yuri Parisky at the Norris Comprehensive Cancer Centre at the University of Southern California and the results were recently presented at the Computer Assisted Radiology and Surgery (CARS) 2000 conference.

### 3.7 International Medical Multimedia

**This is computer aided diagnosis and is not used for detection.** International Medical Multimedia AS (IMM) was founded in 1993. It is owned by Skaufoss AS, a substantial Norwegian investment holding company, and has operations in two locations: Abingdon (near Oxford), UK, and Jerusalem, Israel. The company manufactures two groups of products: advanced image management systems (AiMS) and computer assisted diagnosis for mammography (MammoViz). The AiMS software, which is already on the market, is a radiology reference library with teaching and assessment tools. Through an acquisition of SophisView Technologies, an Israeli based company, IMM acquired the MammoViz product range. MammoViz Pro, an image manipulation program, is now ready for commercialisation, and MammoViz Class, already developed but requiring FDA clearance, is a software package that uses algorithms to quantify features indicative of breast cancer on the mammogram. The system is a computer-aided diagnosis system: the user identifies a suspicious lesion (microcalcification cluster or mass) and the system gives a diagnosis (benign or malignant). The results quoted are that 79% of microcalcification clusters and 89% of masses are correctly interpreted, with a false negative rate of less than 2% and false positive rates of 48% and 15% respectively. At the moment, the system does not detect lesions on films.

### 3.8 Summary

The two CAD systems currently commercially available are Image-Checker and Second Look. A summary of the trials of these systems and of the PROMAM system is given in Table 3.2. Table 3.3 summarises the results of the working group's evaluation of these systems and Table 3.4 summarises the practical issues in implementing the systems.

Table 3.2 Summary of CAD systems

Question	ImageChecker Precision study 1	Second Look	PROMAM	
			a Interval cancer study	b Preclinical trial
<b>Target abnormality</b>	Microcalcifications Masses	Microcalcification clusters Masses	Microcalcification clusters Masses	
<b>Films in test set</b>	1083 cancers 300 normals	2500 cases (four films per case)	a 84 false negative films b 2000 sets of films	
<b>Selection of normals</b>	Negative screening mammograms with at least one negative follow-up	1300 cases (confirmed negative)	a N/A b 1900 sets without malignancies	
<b>Selection of abnormalities</b>	1083 consecutive biopsy proven malignancies from screening	1200 cases (biopsy proven)	a From 249 interval cancers the subset of 84 were classed as false negative b 100 sets with proven malignancies	
<b>Other film selection</b>	None			
<b>Origin of films</b>	Thirteen different institutions in USA	Three US screening facilities	a SE Scotland BSC b SE Scotland BSC	
<b>Original films or copies?</b>	Originals	Originals	a b Originals	
<b>Were films excluded?</b>	Implants without ID views, mosaics	Cancers had to be biopsy proven with pathology report, otherwise excluded		
<b>Were films from normal NHS screening?</b>	No (asymptomatic US patients)	No (development films from US sites)	Yes	

Table 3.3 Evaluation and results

	<b>ImageChecker</b>	<b>Second Look</b>	<b>PROMAM</b>
<b>Question</b>	Precision study 1		a Interval cancer study b Preclinical trial
<b>How was gold standard defined?</b>	Biopsy proven	Biopsy proven. Pathology reports for all cases	
<b>How are hits and misses defined?</b>	Correct marking of location of cancer on at least one of the films	Correct marking of location in one or more views	
<b>Is a hit on another abnormality rated as a hit or a false positive?</b>	False positive, unless other abnormality is biopsy proven	False positive	
<b>What proportion of abnormal films are correctly prompted?</b>	98.3% (399/406) of calcifications and 85.7% (580/677) of masses were prompted Note that all cancers in this study were classified as calcifications or masses (asymmetries were included as masses)	88% on biopsy proven cancers	a Interval cancers – 44% prompted (30% of masses and 60% of calcifications) b Preclinical trial – 94% of calcifications prompted and 73% of masses
<b>What is the average number of false prompts per image?</b>	Cancers 1.2 prompts per film and normals 0.5 prompts per film. In practice, as only five per 1000 films in real life are cancers, the false prompt rate is 0.54 per film (or 2.19 per four films of one subject)	1.3	Microcalcification algorithm gives 0.56 prompts per film and masses algorithm gives 0.52 prompts per film = 1.08 prompts per film
<b>False negative related to?</b>		Sensitivity is expected to improve over time Current settings balance sensitivity with an acceptable false positive rate	
<b>False positive related to?</b>		Representations in the films that were similar to cancers in the training database	

Table 3.4 Comparison of practical issues in implementing different CAD systems in breast screening

Criteria	ImageChecker	Second Look	PROMAM
<b>Is the system currently a commercially available product?</b>	Yes. In USA, EU and Japan	Yes	No
<b>Price</b>	Depends on the configuration	Established by distributors for each region	N/A
<b>Have preclinical trials been completed?</b>	Yes. Very extensive trials in the US have been reported in the scientific literature <sup>14, 22, 23</sup>	In the UK – Nightingale Screening Centre in Manchester. European data are available but not published	There has been one preclinical trial in Edinburgh which has been presented at a scientific meeting <sup>15</sup>
<b>Has FDA accreditation been received?</b>	Yes	In progress	No
<b>Reliability?</b>	Mean time between failures is 1000 h	Expected four calls per year at launch	
<b>Method of conversion to digital format</b>	Film digitiser. Two configurations – single loader (M1000-SL) and continuous loader (M1000-CL)	Modified Howtek digitiser	DBA CCD film digitiser
<b>Pixel size</b>	50µm	43.5µm	42µm
<b>Digitisation throughput</b>	60 s per film	Approximately 60 s per film	<40 s per film
<b>Batch loading</b>	100 in hopper (M1000-CL)	Fifty in hopper. Continuous loading	250 in hopper
<b>Is film processing fully automatic on a batch basis?</b>	Yes. No supervision required	Yes. Supervision currently required	Yes. No supervision required
<b>Processing throughput</b>	60 s per film	Average 1 min 20 s per film	Currently up to 200 s per film but hardware upgrades should significantly reduce this
<b>Method of relating digital images to patient.</b>	Films are related to patient by means of a barcode label	Barcode or touch screen entry	Labelled with ID number extracted from film (barcoding being considered)
<b>Procedure for making primary diagnosis</b>	Films are mounted on a modified roller viewer, as is the normal practice at screening sites. The viewer is modified to incorporate a CRT display, barcode reader and storage computer	Mammograph printout	Original films on any view box

<b>Form of prompts</b>	Triangle over centre of cluster of micro-calcifications and star over centre of mass or architectural distortion. New optional feature displays a circle around the star when particularly strong features are indicated	Ellipse and rectangles that are size differentiating	Ellipse around masses and an irregular outline drawn around microcalcification clusters
<b>Prompting media</b>	CRT display mounted below film belt	CRT display and hard-copy printout	Printed paper sheet with four views
<b>Hard copy record of prompt</b>	Not standard but optional printer can be configured with the system	Yes	Standard method of prompting
<b>Computer archiving of prompt</b>	No	Next release	Not regarded as a current requirement
<b>Are the prompts weighted to indicate degree of suspicion?</b>	Not in standard configuration. New optional feature will display a circle around masses when particularly strong features are indicated	No	No
<b>Which views are prompted?</b>	MLO and CC	MLO and CC	MLO and CC
<b>Can the system process large format (24 30 cm) films?</b>	Yes	Yes	No
<b>Can the sensitivity and specificity be adjusted?</b>	Yes, in principle (in consultation with R2 and only outside the USA)	No	
<b>When are the prompts displayed?</b>	It is intended that the radiologist should look at the prompt after first making a determination based on the films	After first determination	Before or after is possible
<b>Prompting strategy</b>	Prompts are based on ROC operating points	Limit to number of prompts (five for two-film, nine for four-film cases)	Microcalcifications greater than 100 µm are identified. A cluster is defined as five or more particles within 3.1 mm. Masses are also identified
<b>How are extensive, diffuse or multifocal abnormalities prompted?</b>	Masses may be marked more towards the side. Extensive calcifications are broken	No limit to size of mass that will be prompted	

## 4. ISSUES FOR THE NHSBSP

### 4.1 Full field digital mammography and CAD

The introduction of digital mammography into breast screening will dramatically affect the potential for implementing CAD as part of the breast screening process. There are several reasons for this.

1. Almost all of the current research and practical implementation of CAD in mammography requires mammograms to be digitised before they can be analysed by computer. Only the most expensive digitisers are suitable for the task to avoid degrading the image information available on the original films. This stage, which is time-consuming and cumbersome, would be avoided if the x-ray mammography system produced mammograms already in digital form.
2. Because in a digital management system the images and patient information are already stored on computer, the task of matching prompts to clinical images can be fully automated. Archiving of prompting along with clinical images would also be possible
3. Computer screens used to display the clinical images could also be used to overlay prompting information. Current screening practice requires a printer or small video monitor viewed in parallel with clinical films.
4. Computers used to store and display mammograms could also be used to run the CAD algorithms.

All of the above mean that the additional cost of adding CAD to digital mammography systems would be essentially limited to the software necessary to run the algorithms. This could make CAD much more economically viable.

There are a number of problems to be resolved before CAD can be combined with digital mammography:

1. Digital mammography is not routinely used in any screening centres in the NHSBSP.
2. Although digital mammography is commercially available, most systems are three or four times more expensive than current analogue options.
3. The high capital outlay for digital mammography could be justified if screening centres could reduce current film costs. However, the first digital mammography systems have relied on laser film for the diagnostic reading of the images. With this approach there is no cost saving for film.
4. It would probably be technically possible to produce computer monitors with adequate performance for reading screening mammograms (soft copy reporting), but the current technology is very expensive and this approach remains unproven.
5. The ability to overlay prompting information on clinical images with digital mammography is less useful if soft copy reporting is not available.

6. The latest generation of digital mammography systems are capable of producing excellent mammograms on laser film in a low throughput environment. However, the logistics and computing resources necessary to implement digital mammography in a mass screening environment have still to be developed.
7. The algorithms developed for use with film digitisers would need to be modified to work with images produced by digital systems. For example the pixel size required by the main CAD systems are of the order of 50 $\mu$ m, whereas many digital systems use a pixel size of 100 $\mu$ m. The pixel size may be an important factor in the performance of the calcification algorithms.

It is likely that, in the future, the cost of digital mammography will come down and that soft copy reporting will become possible. At this point the implementation of CAD would be much more straightforward and cost-effective than at present.

### **4.2 Guidance on equipment for CAD in mammography**

#### *4.2.1 Background*

When evaluating the use of new technology, it is essential that prospective users are clear about their expectations for its performance. These expectations can be expressed as guidance notes, which have three principal functions:

- to provide guidance for drawing up the evaluation protocol to be used for equipment assessment
- to provide guidance to the equipment designers about user requirements
- to provide the basis of any subsequent purchasing specification.

#### *4.2.2 Performance level*

The present 'gold standard' for cancer detection using film–screen mammography is achieved by two experienced specialised radiologists, with a third specialist arbitrating in cases of disagreement on whether to recall. This level standard has to be replicated by one experienced radiologist prompted by the CAD system.<sup>20, 21</sup> The improved 'gold standard' for cancer detection (yet to be established) using digital mammographic systems under the same reading standards needs to be replicated using the CAD system.

#### *4.2.3 Range of application*

The CAD system detection algorithm must be capable of adjustment to operate with any of the film–screen combinations in current use, exposed in compliance with the standard optical density levels of 1.4–1.8. Later use with digital mammographic systems is to be anticipated.

#### *4.2.4 Throughput*

The digitiser and detection system must be capable of processing and preparing for display (on paper or VDU) no fewer than 100 mammographic films per hour. The data storage capacity for subsequent VDU display must be not fewer than 700 films, with a retrieval time of not more than 5 seconds per film.

### 4.2.5 *Evaluation of performance level*

Since ‘gold standard’ performance levels cannot be stated in readily measurable terms for comparison against the performance data currently available from CAD system suppliers, a protocol will require to be drawn up which will enable evaluation of the available systems against the ‘gold standard’ to be made in order that their suitability for use in the NHS may be determined.

## 4.3 **Research issues**

### 4.3.1 *Target questions*

The initial target for research questions for CAD systems must include:

- effect of the addition of CAD on the sensitivity of the mammography screening test and the false negative rate
- effect, if any, on the small cancer detection rate and the ability of CAD to prompt small invasive cancers (<10 mm) (those most likely to be overlooked by radiologists)
- effect on the specificity/recall rate
- effect of CAD on readers of different abilities and experience (including radiographers and breast physicians).

There are, in addition, a whole raft of more subtle but nevertheless important research questions to be explored in terms of the relationship between the prompting system and the radiologist using it. It will also be necessary to evaluate whether there is any element of ‘de-skilling’ of radiologists arising from working with prompting systems over a long period of time. There is little doubt that, as experience is accrued in working with such systems, further questions requiring evaluation will become evident.

### 4.3.2 *Current research*

The first ImageChecker system in the UK was installed at the Kent & Canterbury Hospital in November 1998, on extended loan, for evaluation to determine its suitability for use in the NHSBSP. A single retrospective study led by Professor Stuart Field and Linda Garvican looked at 100 interval cancers both blindly, using ImageChecker, and then with the information on the clinical films at the time of diagnosis. Three readers – an experienced radiologist screener, a breast clinician with 10 years’ experience of clinical work and about two years of intermittent second reading, and a recently appointed, trained radiologist with limited screening experience – viewed the mammograms. The data are currently being analysed.

Dr Sue Astley and Dr Caroline Boggis are evaluating the Second Look system in Manchester, supported by the South Manchester University Hospital Trust Cancer Research Endowment Fund. To date, they have focused on the detection performance of the system, by characterising and processing both normal (non-cancer) and abnormal (cancer) screening mammograms, and comparing annotations of the abnormalities with regions detected as suspicious by the system. The emphasis of their research is to assess performance in the context of the UK screening

programme, so that realistic prompting rates can be established. An early finding is that the system's sensitivity is considerably improved if two views of each breast are available. Current work is concerned with studying the effect of prompting on radiologists and radiographers, but because of limited resources this can only be considered to be a pilot study. The researchers are also looking at the practical problems of using such a system in a busy screening unit. The pilot results demonstrate a false prompt rate of 1.45 per film. Initial studies have also identified reproducibility of the prompt as an area for further evaluation. At present the prompt rate is high and its effect on the radiologist's detection rate in a population screening programme is unknown. Thus, it is not yet possible to determine whether Second Look can act as a second reader for screening radiologists. The overall sensitivity for detection of breast cancer is 74% per cancer case; this is a combination of 65.8% for single projection and 92.1% in two-projection cases if the cancer was visible in both projections. A formal evaluation of implementation and practicalities in the NHSBSP is needed, as well as an evaluation of the system's effect on cancer detection rates.

Future work will also include:

- comparison with the ImageChecker system of prompting of mammograms of immediate prior screen of interval cancers
- comparison with radiologists' detection rate for routine NHSBSP screening
- comparison with radiographers' detection rate for routine NHSBSP screening.

There is ongoing research in the USA and Europe. We await evidence to demonstrate that CAD is able to give sustained benefit in the context of a large volume screening centre such as the majority of those in the UK.

### *4.3.3 Future research*

Researchers in a number of academic institutions in the UK have been working on the development of algorithms to detect mammographic abnormalities for the past 12 years, and have made some valuable contributions in this area. This activity is set to continue, as two of the largest computer vision groups with expertise in mammographic analysis (Oxford and Manchester Universities) join forces to tackle some of the more challenging problems. This is part of a larger six-year project aimed at developing methods for extracting clinical information from medical images. This EPSRC funded project, which started in January 2001, will be led by Professor Chris Taylor and the team includes Dr Sue Astley in Manchester and Dr Mike Brady in Oxford.

### *4.3.4 HTA programme*

In February 1999, the Health Technology Assessment (HTA) panel of the NHS Research and Development Programme sought proposals to answer the question 'What is the cost-effectiveness and appropriateness of using computer aided diagnosis in the NHS National Breast Screening Programme?' This working party submitted a two-part trial. The first part assessed CAD systems that were available commercially or as

prototypes, together with ergonomic issues and the effect of CAD on differing reader abilities. The second part was a prospective multicentre trial to test CAD in the screening situation with sensitivity and specificity as the primary outcome measures together with a health economics assessment. This was not funded. However, Dr Paul Taylor's group in London was supported by the HTA programme. Their project will attempt to identify the most cost-effective combination of human reader and CAD system consistent with the need to maintain the existing quality of screening and the need to develop new ways of working in response to the growing manpower crisis. Comparisons will be made of double reading by two radiologists and double reading by a radiologist and a computer-supported radiographer, as well as of double reading by a radiologist and a radiographer and single reading by a computer-supported radiologist. The impact of CAD on each of three professional groups and its implications for the future of the screening programme will also be assessed. Funding has only recently been released owing to budget restrictions on the HTA programme.

#### **4.4 Ergonomic and psychological issues**

##### *4.4.1 Current film reading workstations*

Standard practice in radiology is for cases to be routinely read in a suitably darkened room with extraneous light masked off to minimise potential sources of glare and distractions. Generally, mammographic films (one or two views per case) are examined on a mammographic multiviewer, which facilitates a large number of cases being examined within a short time frame, although a static single or double viewing box is also occasionally used to view particular cases. In the UK, the two views of a case are typically examined together, with both of the mediolateral oblique (MLO) views positioned above the craniocaudal (CC) views (although a small number of centres use the reverse arrangement). This case layout is in contrast to other regimes, eg in the USA, where films are often examined in a row such as right CC, right MLO, left MLO then left CC. Additionally, single views of a case are used, and in follow-up studies both current and previous views are examined together. In the case of some women, more than one film/view of each breast may be required.

With modern multiviewers the case is either delivered directly in front of the seated observer or the seated observer has to physically traverse the multiviewer to examine a particular film. The illuminated multiviewer panel is either vertical or split in two horizontally, with the lower part angled towards the observer so as to increase the ease of reading by a seated individual. Additionally, some multiviewers are designed to allow the observer to stand and view films easily. When several observers are required to examine or discuss cases, this may be done either sitting or standing.

Typically a magnifying glass is used for a part (or for all) of the case inspection time, although there is considerable variation in this behaviour, which reflects the individual's preference. Sometimes, a

hand-held 'restricted field of vision' viewer is used to limit the film areas being studied. A case is often examined in some 6–15 s, with further time then being spent either recording decisions on paper or entering the data directly into a computer system – there are currently a variety of computer record systems available across the UK. With a multiviewer, individuals attend visually to the next case, even as they are examining the current one.

As long as mammographic films are used in the UK screening programme then any CAD system has to fit into this generic film-reading scenario. The costs and benefits of CAD use, in terms of potential changes in observer behaviour, need to be examined. Previous studies of CAD use involved relatively small case series and small numbers of observers, and their results may not be applicable to routine UK screening. Additionally, American studies of CAD in screening differ markedly from the UK screening situation in terms of number of cases typically examined per hour, the motivation for CAD implementation and other factors.

#### 4.4.2 *Impact of CAD on case throughput*

The current screening case throughput is equal to, or in excess of, 60 cases an hour. To maintain such case throughput, the use of CAD must not increase the individual case examination time unduly (a limit should be defined through research) unless there are other benefits in terms of reduced time elsewhere in the case screening cycle. CAD use will entail offering the observer prompts of candidate areas of interest on the mammographic films. Examining such prompted areas in detail and identifying them appropriately (or disregarding them) will take time. Current CAD systems produce many false positive prompts of potential areas of interest – in a screening situation the observer may come to mistrust many or all prompts, negating the potential benefits.

#### 4.4.3 *Impact of CAD on observer posture*

Multiviewers should be well designed ergonomically so as to sustain the long term reporting of cases by any individual without causing musculoskeletal discomfort. There are no ergonomic standards applicable solely to the design and use of multiviewers, although there are data on ergonomic workstation design for general seated inspection work.

The 'bolt on' of CAD prompting systems to current multiviewers is not necessarily the best ergonomic solution and may serve to increase both observer physical discomfort and mental workload. Any such discomfort in the screening workforce could have long term deleterious effects on staffing. Where a CAD display is employed the existing VDU regulations are relevant.

#### 4.4.4 *Effect on observer decisions*

The prime purpose of CAD is to improve correct recall/not recall decisions by the observer by ensuring that suspicious features are not overlooked. It is important to determine whether the overall detectability of cancer/abnormalities is actually improved by CAD and/or whether the decision criteria of the observers are altered. The effect of the false positive prompt rate on these factors must be established.

### 4.4.5 *CAD usability*

CAD systems must be very easy to use, and this can be assessed using usability analyses. The placement of the CAD mimic image (either paper or a computer monitor) needs to be ergonomically appropriate, and ideally individual users should be able to move the display to their preferred location. The CAD images should be presented at the same viewing distance to the observer as the mammograms themselves. The optimum size of the mimic images needs to be determined. What level of customisation of display parameters (including positioning) for the individual observer should be permitted (if any)? The system should allow for different observers, for instance in terms of physical characteristics and experience.

### 4.4.6 *CAD prompt characteristics*

Current CAD systems display prompts on a mimic display. CAD prompts can be displayed in a variety of ways (eg paper, cathode ray tube/flat panel screen). Whether the mimic display should be an outline of the breast under examination, a generic breast outline or some simplified (and how best to simplify?) grey level representation of the breast also needs to be determined. Also, prompts may be black and white or coloured and, if coloured, the optimum colours for the screening situation and whether different colours should be used to prompt different features needs to be determined. Prompts are currently arrows, circles etc, and the optimum iconic display and size of particular prompts has to be decided. Prompts must be displayed in such a way that observers cannot miss them, while at the same time not overburdening the observer with a complex display. Prompts may need to indicate accurately the suspicious feature, its location, size and shape or, more simply, just the broad location of the suspicious feature. Such potential display choices may depend upon the experience and skill of the observer.

There should be standardisation of prompt characteristics across different manufacturers' systems otherwise problems will occur when observers use different systems within the same screening centre or when observers transfer between screening centres. There must be clear and unambiguous correspondence between the mimic display and the actual mammograms so that errors of correspondence are not introduced. The mimic display must be easy to interpret and not, of itself, increase examination time per case/film. Additional information could be displayed along with the prompts, such as the probability of the prompted feature being abnormal.

### 4.4.7 *Accuracy issues*

Different manufacturers claim different accuracies for their systems in prompting for particular mammographic features. The observer should be aware of these. The number of false positives, in particular, will affect the observer's performance. Studies performed in this area need to be expanded greatly before ready acceptance of their initial conclusions. Most CAD approaches only prompt the 'most likely' candidate areas – how many areas should be prompted and should the prompting level vary for different observers or be variable by the observers themselves? Additionally, should the system be intelligent and learn how particular individuals prefer to be prompted? Should the prompted images and associated data be archived for future examination?

### 4.4.8 *Protocols for use*

To overcome any bias or overreliance on the CAD prompts, it is suggested that the observer should examine each case first without any prompts available and then with the prompts. This implies some facility for turning on the prompts or an in-built delay before prompts appear on the display. Given that each film may contain multiple prompts, indicating different mammographic features, the question arises of whether the observer should be forced to cancel all the prompts in some manner to ensure that he or she has addressed them all. While this may be admirable in a training situation, or where the observer is relatively new to screening, it may be impractical for the experienced screening observer. The CAD system should be capable of being adequately integrated into a double reading protocol within a breast screening centre.

### 4.4.9 *User issues*

Potentially the CAD system (prompts and display) will be used by different types of observer (eg radiologist, radiographer, breast physician, 'mammographic reader') and by observers of different experience and skill – skill both in using the CAD system and in interpreting breast screening images. The CAD system must be usable by all such observers and, if necessary, be variable to allow for different classes of users. The CAD system may be used at different stages of the screening process and should be flexible to allow for this (eg radiographers may potentially use it to determine quickly whether extra views are required). Appropriate training needs to be addressed for the different types of uses and system users.

### 4.4.10 *System integration*

The CAD system needs to be capable of appropriate integration, in user terms, into both existing and future hospital radiology information systems and picture archiving communication systems. There is the potential future use of CAD systems in teleradiology.

### 4.4.11 *Future developments*

Direct digital capture and image processing will become available in time, and current CAD systems would need to be able to be employed without enormous financial reinvestment.

Assuming digital capture and display, current multiviewers will be replaced by workstations containing one or several monitors. The number of monitors and their size, resolution and brightness are issues to be addressed in order to produce optimal observer performance. Monitors should be perceptually maximised for the particular viewing conditions of the room.

The ideal way to view an image is for it be presented directly in front of the observer, in a zone that is below the horizontal line of sight and above a line of sight some 40° visual angle below this. This largely prevents undue neck flexion. In the case of a single VDU being examined by a seated observer, an adjustable height chair coupled with a desk of adjustable height can cater for most individuals. This is more complicated for multiviewers and multiple monitor systems. Clearly, if both standing and seated observers of widely differing heights have to be catered for adequately then the ergonomic design is even more complex.

## APPENDIX 1: QUESTIONNAIRE ON PROMPTING IN MAMMOGRAPHY: SYSTEM PERFORMANCE

### 1. Introduction

The primary objective of this questionnaire is to elicit sufficient information about the functioning and evaluation of computer aided mammography systems to enable us to determine whether any such system is appropriate or is likely to be appropriate in the near future for clinical trial in the UK NHSBSP.

There are three main areas to be addressed: practicalities such as throughput and potential for adaptation to UK needs; performance of the detection algorithms; and how the use of a system affects the performance of radiologists.

### 2. Practical issues

- scanner (throughput of films, hopper, capacity of hopper, need for supervision)
- method of relating digitised films to patient (reliability, ease of use)
- cost (equipment, staff; maintenance, upgrades)
- reliability (all components)
- processing (throughput of images, need for supervision)
- storage, archiving, security of data
- how are prompts displayed? (paper, on-screen, projected)
- form of prompts (appearance, strength of evidence, different prompts for different abnormality types)
- film display (specialised viewer, two views, past films, are prompts available for second view or previous films?)
- which radiographic projections and film formats can the system process?
- when are the prompts displayed? (before or after viewing films)
- can the system be adapted to suit a given radiologist or screening centre's needs?
- prompting strategy (prompt only strongest target or several strongest targets, ROC operating point)
- how are extensive, diffuse or multifocal abnormalities prompted?

### 3. Prompt generator evaluation

- what types of abnormality is the system designed to detect?
- what range of sizes of each abnormality is the system designed to detect?
- how well does each detection algorithm perform? (eg ROC results)
- how was each algorithm evaluated?

The NHSBSP needs to know in some detail how each algorithm was evaluated.

### 3.1 Data

- what exactly is the target abnormality? (eg pathology, mammographic appearance, size)
- how many films were there in the test set?
  - unequivocally normal
  - showing target abnormality
  - other (ie showing other type of abnormality or benign changes)
- how were the normal films selected? (sequential/screening/normal on subsequent mammogram after  $x$  months/years)
- how were the abnormal films selected? (sequential for abnormality type/screening/incident/prevalent/biopsy proven/consensus radiological opinion/interval cancers prior films from intervals or screening)
- how were any other films selected?
- did the films originate in a single centre, or in multiple centres? what was the country of origin?
- were they original or copy films?
- were any films excluded, and on what grounds? (eg quality, size, equivocal diagnosis?)
- if the films were not selected sequentially from a screening population of similar composition to the UK breast screening population, give details of the screening strategy, eg invited or self-referred? (any symptomatic?) frequency of screening? distribution of ages?

And also ages of the data set, eg distribution of glandular pattern types? distribution of lesion sizes? information about degree of subtlety of abnormalities

### 3.2 Evaluation and results

In addition to any available results (eg sensitivity, specificity, FROC) for each algorithm, we need to know the following information about the method of evaluation, and about the performance of all the algorithms in combination:

- how was the ‘gold standard’ defined?
- how are ‘hits’ and ‘misses’ defined?
- is a ‘hit’ on another type of abnormality rated as a ‘hit’ or false positive?
- was reproducibility measured?
- what proportion of abnormalities are correctly prompted?
- for each type of image (ie normal, abnormal and other)
- what proportion of images have one or more false prompts?
- what is the average number of false prompts per image?
- are the false negatives related to:
  - type of abnormality
  - size of abnormality

- location of abnormality in breast
    - glandular pattern
    - image quality
    - nature of abnormality
    - multiplicity of foci?
  - are the false positives related to:
    - location within the breast
    - glandular pattern
    - benign changes
    - image quality
    - presence of an abnormality in the same or opposite breast?
- 4. Effect on radiologist's performance**
- what is the percentage increase in time taken by an experienced radiologist to read, say, 100 patients' mammograms?
  - does use of the system significantly alter
    - number of cancers detected
    - rate of referral for further imaging or investigation?
  - if referral rates are increased, is there also a corresponding increase in cancer detection?
  - if referral rates are decreased, are at least the same number of cancers detected?
  - are results available for different classes of film reader, eg experienced vs inexperienced? If so, how were these classes defined?
  - were the radiologists using the system as intended?
  - did they find the system acceptable?
  - what training and practice was given before the start of any trials?
  - was there any evidence that performance improved with experience?
  - were the radiologists given prior knowledge of the abnormality rate in the test set?
- 5. Comment**
- There are some additional points which are of interest, but not essential at this stage.
- for cancers missed by radiologists using the system:
    - did the radiologist dismiss an accurate prompt?
    - were there any false prompts in the image?
  - for radiologists' false positive judgements
    - did these correspond to false prompts?

## APPENDIX 2: COMPANY CONTACT DETAILS AND WEB SITES

R2 Technology Inc. ImageChecker  
325 Distel Circle  
Los Altos  
CA 94022  
USA

Tel: 001 650 254 7200  
<http://www.r2tech.com>

CADx Second Look  
Qualia Computing, Inc.  
2372 Lakeview Drive, Suite G  
Beavercreek  
OH 45431  
USA

Tel: 001 937 431 1464  
Fax: 001 937 431 1465  
<http://www.qualia-computing.com>  
See also <http://www.cadxmed.com>

DBA Systems Division/Titan Systems Corporation PROMAM  
1200 S. Woody Burke Road  
PO Box 550  
Melbourne  
FL 32902-0550  
USA

Tel: 001 321 727 0660  
Fax: 001 321 952 1689  
Email: [sales@dba.titan.com](mailto:sales@dba.titan.com)  
<http://www.titan.com/dbasystems>

MedDetect  
4545E 9th Avenue, Suite 110  
Denver  
CO 80220  
USA

Tel: 001 303 320 2594  
Fax: 001 303 333 7511

OXIVA  
Oxford Centre for Innovation  
Mill Street  
Oxford OX2 0JK  
UK

Tel: 01865 463195  
Fax: 01865 793165  
Email: [rph@oxiva.com](mailto:rph@oxiva.com)  
<http://www.oxiva.com>

Xmammo

SCANIS Inc.  
1111 Triton Drive, Suite 201  
Foster City  
CA 94404  
USA

Tel: 001 650 378 4100  
Fax: 001 650 378 4109  
Email: General Information: [scanis @ best.com](mailto:scanis@best.com)  
Sales: [kilinski@scanis.com](mailto:kilinski@scanis.com)  
<http://www.scanis.com>

Mammex TR

International Medical Multimedia Ltd  
14 Foxcombe Court  
Abingdon Business Park  
Abingdon  
Oxon  
UK

Tel: 01235 552500  
Fax: 01235 552525

### REFERENCES AND BIBLIOGRAPHY

#### References

1. Forrest P. *Breast Screening: Report to the Health Ministers of England, Wales, Scotland and Northern Ireland*. London, HMSO, 1988.
2. Field S. Breast screening issues – a report into workforce, funding, litigation and morale. *The Newsletter of the Royal College of Radiologists* 1998,54: 12–14.
3. Armstrong P. President's Column. *The Newsletter of the Royal College of Radiologists* 2000, 62: 2.
4. Savage CJ, Gale AG, Pawley EF, Wilson ARM. To err is human, to compute divine? In: Gale AG, Astley SM et al (eds) *Digital Mammography*. Amsterdam, Elsevier, 1994: 405–414.
5. Kundel HL, Nodine CF. Studies of eye movements and visual search in radiology. In: Seders JAW et al (eds) *Eye Movements and the Higher Psychological Functions*. Hillsdale, NJ, Erlbaum, 1978: 317–328.
6. Treisman A. Pre-attentive processing in vision. *Computer Vision, Graphics and Image Processing* 1985, 31: 156–177.
7. Krupinski E. Visual scanning patterns of radiologists searching mammographic images. *Academic Radiology* 1996, 3: 1000–1006.
8. Treisman A. Features and objects: the fourteenth annual Bartlett Memorial lecture. *Quarterly Journal of Experimental Psychology* 1988, 40A: 201–237.
9. Astley SM, Zwiggelaar R, Wolstenholme C et al. Prompting in mammography: how good must prompt generators be? In: Karssemeijer N, Thijssen M et al (eds) *Digital Mammography*. Dordrecht, Kluwer Academic Publishers, 1998: 347–354.
10. Chan H-P, Doi K, Vyborny CJ et al. Improvement in radiologists' detection of clustered microcalcifications on mammograms. *Investigative Radiology* 1990, 25: 1102–1110.
11. Hutt I. *The Computer-aided Detection of Abnormalities in screening Mammograms*. PhD Thesis. Manchester, University of Manchester, 1996.
12. Kegelmeyer WP Jr, Pruneda JM, Bourland PD et al. Computer-aided mammographic screening for spiculated lesions. *Radiology* 1994, 191: 331–337.
13. Roehrig J, Doi T, Hasegawa A, et al. Clinical results with R2 ImageChecker system. In: Karssemeijer N, Thijssen M et al (eds) *Digital Mammography*. Dordrecht, Kluwer Academic Publishers, 1998: 395–400.
14. Warren Burhenne LJ, Wood SA, D'Orsi CJ et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology* 2000,215: 554–562.
15. Williams LJ, Prescott RJ, Hartswood M. Computer-aided cancer detection in the UK breast screening programme. In: Karssemeijer N, Thijssen M et al (eds) *Digital Mammography*. Dordrecht, Kluwer Academic Publishers, 1998: 359–362.
16. Macmillan N, Creelman CD. *Detection Theory: A User's Guide*. Cambridge, Cambridge University Press 1991.
17. Efron B, Gong G. A leisurely look at the bootstrap, the jackknife and cross-validation. *The American Statistician*, 1983, 37: 36–48.
18. Metz C E. Evaluation of digital mammography by ROC analysis. *Excerpta Medica* 1996, 1119: 61–68
19. Karssemeijer N, te Brake G. Combining single view features and asymmetry for detection of mass lesions. In: Karssemeijer N, Thijssen M et al (eds) *Digital Mammography*. Dordrecht, Kluwer Academic Publishers, 1998: 95–102.
20. Blanks RG, Moss SM, Wallis MG. A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers for one and two view mammography: results from the UK National Health Service breast screening programme. *Journal of Medical Screening*, 1998, 5:141–145.

21. Blanks RG, Wallis MG, Given-Wilson RM. Observer variability in cancer detection during routine repeat (incident) mammographic screening in a study of two versus one view mammography. *Journal of Medical Screening*, 1999, 6: 152–158.
22. Vyborny CJ, Doi T, O'Shaughnessy KF et al. Breast cancer: importance of spiculation in computer-aided detection. *Radiology* 2000, 215: 703–707
23. Birdwell RL, Ikeda DM, O'Shaughnessy, Sickles EA. Mammographic characteristics of 115 missed cancers later detected by screening mammography and the potential utility of computer aided detection (CAD). *Radiology* (in press).

### Bibliography

Highnam R, Brady M. *Mammographic Image Analysis*. Dordrecht, Kluwer Academic Publishers, 1999.

Karssemeijer N, Thijssen M et al. (eds) *Digital Mammography*. Dordrecht, Kluwer Academic Publishers, 1998.

### Web sites

5th International Workshop on Digital Mammography  
<http://www.sunnybrook.on.ca/~iwdm2000/>

Mammographic Image Analysis Society  
<http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html>

Computer Assisted Radiology and Surgery (CARS 2000)  
<http://www.cars-int.de>