

Maintaining quality in the UK breast screening program

Alastair Gale*

Applied Vision Research Centre , Loughborough University , Loughborough, LE11 3TU, UK

ABSTRACT

Breast screening in the UK has been implemented for over 20 years and annually nearly two million women are now screened with an estimated 1,400 lives saved. Nationally, some 700 individuals interpret screening mammograms in almost 110 screening centres. Currently, women aged 50 to 70 are invited for screening every three years and by 2012 this age range will increase to 47 - 73 years. There is a rapid ongoing transition from using film mammograms to full field digital mammography such that in 2010 every screening centre will be partly digital. An early, and long running, concern has been how to ensure the highest quality of imaging interpretation across the UK, an issue enhanced by the use of a three year screening interval. To partly address this question a self assessment scheme was developed in 1988 and subsequently implemented nationally in the UK as a virtually mandatory activity. The scheme is detailed from its beginnings, through its various developments to current incarnation and future plans. This encompasses both radiological (single view screening, two view screening, mammographic film and full field digital mammography) as well as design changes (cases reported by means of: form filling; PDA; tablet PC; iPhone, and the internet). The scheme provides a rich data source which is regularly studied to examine different aspects of radiological performance. Overall it aids screening radiologists by giving them regular access to a range of difficult exemplar cases together with feedback on their performance as compared to their peers.

Keywords: Breast screening, performance, self assessment, training, visual search.

1. INTRODUCTION

1.1 Background

Breast cancer is now the most common cancer in the UK. Free breast screening was introduced across the UK in 1988 by the NHSBSP (National Health Service Breast Screening Programme) following the 1986 Forrest Report¹ which reviewed the available scientific evidence on screening at that time and recommended screening all women aged 50-64 every three years using a single (MLO - medio-lateral oblique) mammographic view. Some 4.9 breast cancers were predicted over a three year period to occur per 100 women, with screening detecting circa 75% of these. It was estimated that a radiologist could read 55 single view cases an hour and non-radiologists 40 such cases an hour. Subsequently in 1988 the Pritchard Report² further detailed quality assurance aspects in breast screening and argued the need to monitor performance appropriately. The original single mammographic screening view was eventually expanded to two view screening which included the cranio-caudal (CC) view. Double reading of cases became recommended with arbitration where there was disagreement and where practically possible within a screening centre.

Nationwide breast screening was very rapidly established, with data from 1992/93³ showing that the programme was in its second round of screening and already was inviting 1/3rd of the women in the UK aged 50-64 for screening. At this time approximately 1.2 million women were screened with some 6,597 cancers detected. Eventually the screening age range was expanded and now⁴ 50-70 year old women are screened with plans to further increase this range to encompass 47-73 year old women by 2012⁵. Women over the screening age range do not receive screening invitations, instead being encouraged to make their own appointments.

*a.g.gale@lboro.ac.uk

Since screening began in the UK well over 18 million mammograms have been taken and over 100,000 cancers detected⁶, i.e. since 1988 over 100 cancers have been detected every week. The latest UK figures⁴ indicate that in 2007-8 over 2.5 million women were invited for screening with just under 2 million women actually screened and 83,222 being recalled for assessment resulting in the detection of almost 16,500 cancers. Of these 3,257 were in situ and 6,878 were small invasive cancers. Only 1,716 benign biopsies were performed which typically indicates a difficult case. It is estimated that the NHSBSP saves 1,400 lives every year in England alone and the screening programme has been described as a model to emulate⁷.

Screening is performed both in static screening centres (circa 85 centres in England alone) as well as in mobile units to ensure ease of availability for all women. A key feature of screening in the UK is that screening centres have regular multidisciplinary meetings where cases are discussed by all personnel involved (i.e. radiographers, radiologists, pathologists, surgeons etc.) thereby providing performance feedback as well as ongoing training. Screening is organised nationally and different aspects are overseen by various committees with representatives from each English health region, Scotland, Wales and Northern Ireland. Quality assurance is ensured nationally and then regionally with regular performance inspections of each screening centre within each region and by regional quality assurance meetings.

In the UK traditionally radiographers (c.f. technologists) actually screen a woman within six minutes (with it taking an overall 30 minutes for a woman to attend and be screened), develop the mammographic images and check these for any technical issues then consultant radiologists generally examine the resultant mammograms for signs of abnormality. As screening expanded then so too did the workforce. In recent years there has been something of a shortfall in screening radiologists, however at the same time the role of radiographers in the UK has expanded and now some radiographers (known as advanced practitioners) are specifically trained to interpret screening mammography cases.

Although approximately one in nine women in the UK will be affected by breast cancer at some stage in their lives this cancer remains a rare disease to be detected by screening asymptomatic women and screening itself will never be totally accurate. Individuals undertaking breast screening interpretation in the UK have to read at least 5,000 cases a year with many reading far more than this and some reading well over 20,000⁸. This prerequisite minimum number means that annually they will experience a large number of normal variants and therefore build up mental schemata of what constitutes normal appearances. A similar number of cases annually is recommended in Europe⁹ which differs from the approach in the USA¹⁰. A review of the UK's three year screening interval determined it to be appropriate¹¹, however this differs from some countries. For instance, two-yearly mammography was proposed in Europe¹² and in the USA annual screening for women over the age of 40 years has been recommended¹³.

An important issue with screening for a rare disease is how to maintain high visual inspection quality standards. A key factor being that if an indicator of cancer is missed then it may not be picked up until the next screening round unless the woman presents symptomatically. Also, whilst maintaining high sensitivity it is also important to preserve equally high specificity and not over-read cases resulting in potentially carrying out needless biopsies – in the UK only 0.1% of incident women screened have a benign biopsy⁴.

1.2 Mammographic features

Shortly before screening began in the UK I was involved in various radiological inspection/performance research projects primarily revolving around the chest radiograph. Initial studies in mammography began largely accidentally by knocking over a pile of filing cards which contained a mass of data collected by an expert breast radiologist on a series of mammographic cases. The issue at that time, in the early 1980s, was that almost half the women who were recommended for biopsy turned out to be normal or benign, thus we were keen to improve diagnostic specificity.

To do this some 43 mammographic features were first identified by the expert radiologist and split into three categories relating to lesion, reaction and calcification patterns. A series of mammograms of 500 biopsied women were then re-examined by the expert in terms of these features and the data related to subsequent clinical outcome. This yielded an approach that fairly well predicted the outcome for any particular case based simply on identifying whether or not certain mammographic features were present. Discriminant function analysis elucidated 12 key mammographic features as important in predicting outcome. The system, very optimistically termed MAMCAD¹⁴, and based on the one expert breast radiologist's opinions on whether key features were present or not in a case was shown to be promising but not as good as the experienced radiologist himself. Whilst increasing TN (True Negative) decisions the TP (True Positive)

values were lower than the expert. For these cases the original radiologist’s diagnosis was of 49.1% TN with 5.3% False Negatives (FN). The MAMCAD approach achieved 87.6% TN and 21% FN with 79% TP and 12.4% FP (False Positive). This compared with the original radiologist’s 87.2% TP and 33.5% FP¹⁵ (the radiologist also determined some cases to be indeterminate). Subsequent efforts to implement the system for other radiologists to use met with somewhat variable success and served to demonstrate the well known variability both in identifying various mammographic features and also in radiological reporting itself; neither of which melded well with the strict MAMCAD algorithm used at that time.

However, the approach showed promise in having individuals decide whether a case should be recalled for biopsy or not and also have them identify certain case mammographic features. Comparing which features different people identified when they were examining the same mammographic cases would then give insight into aspects of individuals’ performance. This led to the proposal to the NHSBSP to develop and implement as an educational exercise a national self assessment scheme wherein all UK screening radiologists would examine the same set of screening images annually, making various decisions about each case and then receive feedback on how well they have; identified cancer, agreed with a gold standard radiological opinion, as well as agreed with their peers on how every case should be radiologically managed – either recalled for further assessment or returned to routine screening. Thus the approach would offer both a self assessment and training purposes. The scheme, called PERFORMS (PERsonal perforMance in Mammographic Screening) was thus begun^{16,17,18}.

2. PERFORMS

An initial trial of the approach was carried out in three health regions in the UK. The first data set was generated by 29 screening radiologists and figure 1 shows data for the resultant basic screening decision of whether each case should be recalled for further assessment or not for 120 single view mammographic cases. These are split by a gold standard of ‘achievable radiological opinion’ (based on the decisions of four radiologists experienced in mammography) as being ‘normal’, ‘benign’, ‘probably benign’, ‘indeterminate’, ‘probably malignant’ or ‘definitely malignant’. What such a figure should show is an almost orderly rise from the bottom left (of normal, then benign cases – not requiring recall) to the top right (of definitely malignant cases requiring recall). This almost is the case, however it is clearly overlaid with a wide fluctuation in agreement between participants in whether particular cases should be recalled or not.

This then provided the driver for the subsequent development of the scheme as it well illustrates that even for the very simple ‘recall/not recall’ binary screening decision there is often wide disagreement between individuals on a case and also differences between the opinions of experts and those of the less expert participants. These issues may well represent expertise variables between participants and/or individual case presentation factors, all of which can be examined in much further detail.

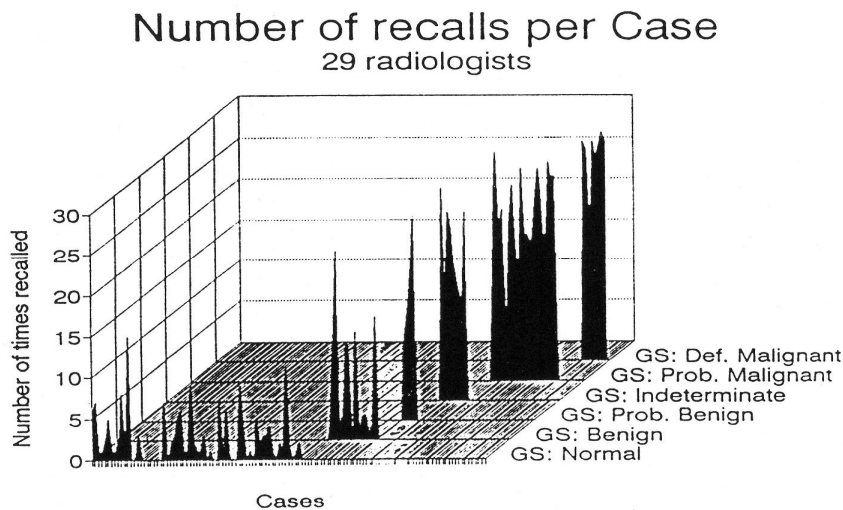


Figure 1: Data from the first 29 radiologists

A woman presenting for screening anywhere in the UK merits the same high quality of screening and so PERFORMS should aid screeners in maintaining their inspection skill levels. The importance of the scheme is for it to have considerable merit to the participating individuals and not just to represent a research tool of interest primarily to academics. Participation for radiologists is very strongly encouraged, virtually mandated, by the Royal College of Radiologists (RCR) and individuals receive CME (Continuing Medical Education) accreditation points for taking part. To help maintain high real life screening skills all screening individuals regularly examine sets of challenging screening images with appropriate feedback on their performance. The scheme is free and confidential to participants. Any differences highlighted between individuals should then essentially be due to differences in their skills of identifying early signs of abnormal appearances. Whilst, admittedly, this is not wholly true as there will inevitably be some uncontrolled factors which will affect performance, such as variations in; the ambient lighting of different reporting rooms, the time of day when individuals have reported on the case set, an individual's workload both before and after reporting the case set which affects how well they attend to the images etc..

2.1 Case selection

A selection of challenging difficult cases are selected each year which represent normal, benign or malignant appearances so as to stretch participants' skills. To do this annually every UK breast screening centre is encouraged to submit interesting and challenging cases for consideration for inclusion in the scheme. These cases are then examined by a panel of experienced radiologists and decisions made about whether each case could potentially be included in the case set as well as agreement reached on an expert decision of each case radiological classification and of which key mammographic features are present. Malignant and benign case pathology is known and normal cases have had a normal subsequent screening result in the next screening round. These selected cases are anonymised by removing all patient and hospital identification marks and replaced with a unique PERFORMS scheme identifier. Typically some 300 difficult cases are initially chosen from which 120 are used.

2.2 Case sets

One of the original ideas was to simply use the same case set for the scheme year after year. The argument here was that as individuals received no feedback per case and in between taking part each year they would have reported at least over 5,000 other real life cases and so they would have little likelihood of remembering individual test cases and their previous responses to them. Thus, ostensibly, over time the performance of an individual could be tracked. Although this was done for the first two annual rounds of the PERFORMS scheme it was subsequently dropped and each year new cases are sourced and then circulated instead, this yielding more interesting data for the participants.

Over 30 sets of identical cases are currently circulated around the country. Originally cases were anonymised and copied photographically which inevitably could produce some fluctuations in appearance. Subsequently cases have been digitised at 50 μ m, initially using a Lumisys digitiser which was subsequently replaced with an Array digitiser. Original mammographic images sent to us can vary in appearance and so these are processed to normalize them. Copies have then been printed, firstly using an Agfa mammographic film processor and subsequently an Agfa mammographic Drystar laser printer. This permits printing of identical case sets which are each randomised before being sent out to a particular screening centre. The randomisation is done with constraints to ensure a spread of benign and malignant cases amongst the normal cases.

2.3 Reporting decisions

The first version of the scheme, from the late 1980s, used 120 single MLO mammographic views of both breasts which had been selected by a group of experienced mammography radiologists. These cases were sent to screening centres together with reporting books of pre-printed forms for each participant (Figure 2). Individuals examined the cases and for each they made various decisions and also identified on a diagram, and a reporting form, which features they thought were present. After all participants at a screening centre had taken part then the various reporting books and the film set were returned to us. The film set and new booklets were then sent out to the next screening centre in the scheme.

Data from each individual were manually entered into a spreadsheet for analysis. When two view screening was introduced then that was easily incorporated into this approach. Whilst this worked well, it was tedious and always

subject to potential transcription error. The option of using an optical reader to input these forms automatically was investigated in detail but not implemented. After data had been analysed then individuals received feedback on their performance. Unfortunately by the time an individual received such feedback then they no longer had the PERFORMS cases and so such feedback was not as useful as it could have been. Once all participants had taken part then they all received anonymous detailed feedback.



Figure 2. The original reporting form used per case and the Psion 3MX with bar code reader and bar coded forms

The second major change to the scheme was to improve the data recording and feedback process in 2000 by using a small hand held computer (PDA) to record decisions. This was linked to a barcode reader and participants simply used a large sheet of card, positioned on their multiviewer desk, containing various pre-printed barcodes to quickly and easily enter their data. On the reverse of the card further barcodes allowed users to enter in questionnaire responses about the scheme and their normal screening practice. At the time this was a very innovational approach.

In the first development of this approach, after an individual had read all the cases, they emailed their data to the research centre which we processed and provided them with feedback (by fax as the feedback included graphical information) when they still had the case set mounted on the multiviewers. Different bar code readers were implemented until a robust system was derived which was not prone to misread errors. However, the email approach had inevitable difficulties.

Initially a Psion Walkabout was used, subsequently replaced by a Psion 3C and eventually a Psion 3MX was settled on¹⁹ (Figure 2). With this PDA, after an individual had read all the cases then, using bespoke software, their data were immediately analysed and summary performance results produced as well as other feedback on each case. The software was written to only allow the PDA to run the PERFORMS scheme. Despite putting considerable effort into the ergonomic design of the reporting system it was found that the Psions were sometimes damaged. The small, lightweight PDA could potentially be accidentally dragged across the multiviewer shelf by the barcode reader cable. This led to the development of a robust ‘box’ containing weighty backup batteries and incorporating an already open Psion 3MX (figure 2). This system performed extremely well and was used successfully for some five years. The downside was the limiting factor of what information the PDA could convey easily back to participants due to the very small size of screen and limited graphics facility. Once all participants had taken part then, as before, detailed anonymous data were made available.

2.4 Current national implementation

Subsequently, in 2005, new software was developed which runs on a ruggedized tablet PC (Xplore iX104) to withstand rough handling. The participant first identifies a particular case to the PC using a pen based GUI which pulls up small images of the two view mammograms on the PC. The individual can then make various decisions about the case and indicate locations of any abnormalities directly on these images (Figure 3). The system is very intuitive and has been

found to require little or no training for participants. Once all the cases have been read then immediate feedback is provided both on the individual's decisions as compared to pathology and on radiological decisions (e.g. recall, features) as judged against the radiological panel. All, or selected, cases can then be reviewed at leisure where the individual's decisions and identified abnormality locations are compared to those of an expert panel. As before, once all participants have completed the scheme then they all receive detailed and confidential reports.



Figure 3. Tablet PC with some of the various reporting and feedback screens.

The growth in UK breast screening can be seen by the increase in number of participants in the PERFORMS scheme. In 1991 approximately 150 radiologists were reported to be involved in screening¹⁶; in 1996 some 250 radiologists took part²⁰. Currently there are over 700 participants. These comprise 417 consultant radiologists, 266 advanced practitioners, 32 breast physicians and breast clinicians, as well as other health professionals. Additionally radiologists and advanced practitioners who only deal with symptomatic women also take part but are not funded directly by the NHSBSP.

3. DIGITAL MAMMOGRAPHY

In the UK successful trials of FFDM mammography have been carried out and digital mammography is now being implemented. However, any change to screening has to be undertaken nationally and so represents considerable financial, logistical and performance issues. The UK's Cancer Reform Strategy⁵ determined that all breast screening centres must have digital mammography by 2010 and be fully digital by 2012. The changeover to digital breast screening introduces some challenges concerning the comparison of current digital images viewed on a workstation and whether images of women from previous screening rounds should be viewed at the same time for comparison purposes on film on radiographic multiviewers or else digitised and then viewed on the same workstation. Research indicates that performance is improved if current digital images are always compared to prior cases if these are available and we have recommended that prior cases should be digitised to facilitate easy comparison to current digital images²¹.

3.1 Digital PERFORMS

As UK breast screening centres increasingly use digital screening and dispense with radiographic film then it is vital that individuals can still participate in the PERFORMS scheme. Consequently, digital versions of the scheme have been under development over the past year for release in early 2010. The ideal implementation may be to display the case sets on a screening centre's reporting workstations and overlay these with the reporting/feedback software. In the UK the latter poses various logistical difficulties and so the image display and reporting are being kept separate. The initial implementation is simply to continue to use the existing tablet PC for participants to record their various decisions but for them to view full DICOM images on their workstations (Figure 4).



Figure 4. Tablet PC with GE workstation

However, the main PERFORMS scheme that is under development will provide a web based self assessment and training system which is platform independent. Anyone wishing to participate will simply log onto a secure part of our website and download both the reporting software and test mammographic images. We are currently investigating cloud-hosted image databases both for image acquisition and delivery to participants. Training systems are also under development in a similar manner so that anyone can log in and download a range of images and training software. In particular we are very interested in making interactive training available to individuals whenever they want and wherever they are, using a range of platforms such as laptops, tablets or PDAs such as the iPhone²².

4. PERFORMANCE OUTCOME MEASURES

There are two main measures from the PERFORMS scheme. As some abnormal cases can appear normal and some normal cases can appear abnormal, the scheme measures how well the radiological decisions made by each individual agree with the decisions of other participants; i.e. if a particular case presented anywhere in the UK would it be dealt with in the same manner? When reporting the cases an individual's data are compared to the opinions of a panel of expert radiologists. However, when constructing the scheme results for a full annual round then the opinions of the panel are dropped and each participant's data are compared to the data of all UK participants. The second measure is how well on the case set an individual detects breast cancer as measured against the known case pathology.

Both of these measures are derived from the classification rating decisions which are based on the UK screening approach²³. This pre-dates BIRADS although it is broadly similar. In scoring performance, abnormality location information can also be used. When the national data are calculated then any individuals who are statistically mild or

severe outliers can be identified, both for cancer detection ability as well as how well agree with their peers. There is a formal procedure agreed with the RCR for dealing with outliers, this is a process which ramps up the consequences if the individual remains an outlier on subsequent self assessment occasions. In practice we find that if someone is an outlier on one occasion then on subsequent case sets they typically perform well²⁴.

When calculating the results for a full round of the scheme then each individual receives very detailed feedback which comprises tables and charts comparing their performance to that of other participants anonymously. The feedback is a result itself of comments from participants about what they would like to know. A breakdown of the overall number of participants who sat the scheme together with a breakdown of the participants per region including their own is given, where each region is anonymous except their own. Their own data are then presented alongside the anonymous results for other participants from their health region, together with the regional and national mean performances. This is done for a range of measures, including; correct recall, correct return to screen, positive predictive and negative predictive values, derived ROC measures and ROC or FROC curves, as appropriate, are also presented. National data are then given as bar charts showing where the individual lay amongst the distribution of all of their peers for various measures.

The individual's cancer detection ability is described in a similar manner where their performance is anonymously compared to their peers in the health region and then nationally. This is followed by feedback on false negative cases and whether their responses arose through particular features being undetected or detected but then misinterpreted. From these data they can see how well they did on detecting and interpreting key mammographic features on this case set. Where individuals fare poorly with certain features then they can be offered dedicated training case sets to help improve their skills²⁰.

We always urge some caution in an individual participant extrapolating their performance too much from the self assessment case set to routine screening as the case set is weighted both with difficult cases as well as containing more abnormal cases than would occur in the same number of cases in routine screening. In terms of feature identification it may well be that across the case set there are small numbers of particular features and so again care is needed in extrapolating performance in identifying case set features to identifying similar features in real life screening. However, despite such prudence these are all routine screening cases which have presented in the UK and have been successfully reported in clinical practice. Consequently, it is perfectly reasonable to make such an extrapolation. This is supported by studies that show that, whilst the scheme is not real life screening, participants' scheme performance does relate to their real life screening performance^{25,26,27}. Additionally, the performance of radiologists and advanced practitioners has been shown to be similar when undertaking the scheme²⁸ which further supports the role development of technologists to interpret screening mammograms.

Data from the scheme can be analysed to examine a range of factors. For instance the number of cases interpreted each year by an individual and their screening experience relate to high performance on the scheme²⁸. Such expertise development through reading high case volumes supports the UK approach of a minimum annual screening case volume. Comparing matched groups of symptomatic radiologists, who do not have such an annual minimum volume requirement, and screening radiologists over several years of the scheme elicited subtle differences in their underlying skills which have been related to volume differences²⁹. In a somewhat related fashion a trial of the scheme in California³⁰, as compared to UK screeners, found that the volume of cases read per month was important in determining sensitivity and specificity. High sensitivity coupled with high specificity was achievable in high volume centres in California but centres with lower volumes fared less well. Trials of the scheme has also been run in other countries including Germany³¹ and Australia.

Interestingly, some disagreements have been found between cases which the radiological panel consider to be difficult and those cases which participants actually find difficult³² (somewhat reminiscent of the scheme's start; c.f. figure 1). The perceived importance of various mammographic features have been investigated³³ and, when particular features are missed, then in general whether these are undetected or misinterpreted can be elucidated^{34,35}. Other related studies have examined the role of visual search in examining such mammographic images^{36,37,38}. Further factors which may underlie performance have been variously investigated. Somewhat surprisingly factors such as the time of day when case sets are read or how the case sets are read (e.g. with or without breaks from reading) are poorly related to performance³⁹. Identifying which particular features individuals find difficulty with has been used as a starting point to develop different

training approaches. These have included offering sets of cases weighted with particular features²⁰, expert system derivatives⁴⁰, computer assisted educational approaches, the use of different display modalities, and grid based systems⁴¹.

5. CONCLUSIONS

The scheme has now been running for over 20 years continuously in the UK. The overall feedback from participants is very positive, acknowledging that the scheme helps them with maintaining and improving screening skills.

Screening exists to identify breast cancer at an early stage offering women the opportunity for successful follow up. The PERFORMS scheme aids in this process and is constantly being changed both in response to other ongoing research findings from colleagues, feedback from participants and as a result of changes within screening in the UK itself. The current development of a platform independent system will lend itself to being offered internationally in 2010 where health professionals can examine sets of cases and have their performance compared to many other experienced screeners. Further details of the scheme are available at www.performs.org.uk.

Acknowledgements

I acknowledge the continual support of the NHSBSP and in particular the help and backing of the Director of the UK NHS Cancer Screening Programmes, Professor Julietta Patnick CBE. Dr Eric Roebuck is the expert radiologist who started it all. I would like to thank all the current and past members of the PERFORMS team who do all the real work. I am grateful to the many healthcare professionals with whom I have the privilege to work and collaborate.

REFERENCES

1. Forrest P., "Breast cancer screening: report to the health ministers of England, Wales, Scotland and Northern Ireland", HMSO, London, (1986).
2. Pritchard J., "Expert sub-committee of the CMO's radiological advisory committee report: guidelines on the establishment of a quality assurance system for the radiological aspects of mammography used for screening", HMSO, London, (1988).
3. Patnick J., (ed.) "NHS Breast Screening Programme Review 1994", NHSBSP, Sheffield, (1994)
4. Patnick J. (ed.) "NHS Breast Screening Programme Annual Review 2009, NHSBSP, Sheffield (2009)
5. Department of Health, "Cancer Reform Strategy", Department of Health, London (2007)
6. Patnick J. (ed.) "Celebrating 20 years of screening", NHS Cancer Screening Programmes, Sheffield (2008)
7. Colditz G.A., "NHS breast screening: a model national programme", *J. Med. Screen.* 13:58 (2006)
8. Scott H.J. & Gale A.G., How much is enough: factors affecting the optimal interpretation of breast screening mammograms. In *Image Perception, Observer Performance, and Technology Assessment*. Y Jiang and B Sahiner (Eds.), Proceedings of SPIE 6515, (2007)
9. European Reference Organisation for Quality Assured Breast Screening and Diagnostic Services (EUREF) certification protocol <http://www.euref.org/>
10. FDA Mammography Quality Standards Act Regulations, Requirements for certification. <http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/Regulations/ucm110906.htm#s90011>
11. The Breast Screening Frequency Trial Group, "The frequency of breast cancer screening: results from the UKCCCR Randomised Trial", *Eur. J. of Cancer*, 38, 1458–1464, (2002).
12. IARC Working Group on the Evaluation of Cancer Preventive Strategies, "Breast cancer screening". IARC Handbooks of Cancer Prevention, vol. 7., IARC Press, Lyon, (2002)
13. American Cancer Society. Detailed Guide: Breast Cancer. http://www.cancer.org/docroot/cricri_2x.asp
14. Gale A.G., Roebuck, E.J., Riley, P.J. & Worthington, B.S., "Computer aided diagnosis in mammography". In Lemke, H.U., Rhodes, M.L., Jaffee, C.C. & Felix, R. (Eds.). *Computer Assisted Radiology*, (Berlin) Springer-Verlag, (1985).
15. Gale, A.G., Roebuck, E.J., Riley, P. & Worthington, B.S., "Computer aids to mammography diagnosis". *British Journal of Radiology*, 60, 887-891, (1987).

16. Gale A.G. & Walker G.E. , “Design for performance: Quality assessment in a national breast screening programme”. In Lovesay E. (Ed.) *Ergonomics - design for performance*, Taylor & Francis, London, (1991).
17. Gale A.G., “PERFORMS – a self assessment scheme for radiologists in breast screening”. *Seminars in Breast Disease: Improving and monitoring mammographic interpretative skills*, 6(3), 148-152, (2003).
18. Gale A.G., Scott H., “Measuring Radiology Performance in Breast Screening”. In M. Michell (Ed.) *Contemporary Issues in Cancer Imaging – Breast Cancer*, Cambridge University Press, Cambridge, (2010)
19. Gale A.G. & Scott H.J., “Keeping abreast of the times”. In P.D. Bust (Ed.) *Contemporary Ergonomics 2006*. Taylor & Francis, London, (2006).
20. Cowley H.C., Gale A.G. and Wilson A.R.M., Mammographic training sets for improving breast cancer detection. In H.L.Kundel (Ed.) *Medical Imaging 1996: Image Perception*, Proc. SPIE 2712, (1996).
21. Taylor-Phillips S., Wallis M.G. & Gale A.G., “Should previous mammograms be digitised in the transition to digital mammography?”, *Eur. rad.*, Vol. 19, No. 8, (2009).
22. Yan Chen, Alastair Gale, Hazel Scott, “Mammographic interpretation training: how useful is handheld technology?” In D. Manning and B Sahiner (Eds.). *Image Perception, Observer Performance, and Technology Assessment. Proceedings of SPIE 6917;12*, 1-10, (2008)
23. Maxwell A.J., Ridley N.T., Rubin G., Wallis M.G. Gilbert F.J. and Michell M.J., “The Royal College of Radiologists Breast Group breast imaging classification”, *Clinical Radiology*, Vol. 64 (6), pp 624-627, (2009).
24. Gale, A. G. & Scott, H.J., “Patient Safety in Radiology: An Example from Breast Screening”. In *Proceedings of Improving Patient Safety 2008 Conference*, Hignett S., Norris B., Catchpole K., Hutchinson A. & Tapley S. (Eds.), *IPS2008 Improving Patient Safety 2008: From Safe Design to Safe Practice*, Ergonomics Society, Loughborough, (2008).
25. Scott H.J., Evans A., Gale A.G., Murphy A. & Reed J., “The relationship between real life breast screening and an annual self assessment scheme”. In B. Sahiner & D.J. Manning (Eds.) *SPIE Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment*, Vol 7263, ppE1–E9, (2009)
26. Cowley H.C., Gale A.G. and Wilson A.R.M., “Mammographic training sets for improving breast cancer detection”. In H.L.Kundel (Ed.) *Medical Imaging 1996: Image Perception*, Proc. SPIE 2712, (1996).
27. Cowley, H C and Gale, A G., “Breast cancer screening: comparison of radiologists performance in a self-assessment scheme and in actual breast screening”. In *Medical Imaging 1999: Image and Performance*, E.A. Krupinski (ed), *Proceedings of SPIE Vol. 3663*, (1999).
28. Scott H.J., Gale A.G., & Wooding D.S., “Breast Screening Technologists: does real-life case volume affect performance?” In: *Image Perception, Observer Performance, and Technology Assessment*, D.P. Chakraborty & M.P. Eckstein (eds.) *Proceedings of SPIE Vol. 5372*, (2004).
29. Scott H.J., and Gale A.G., “Does mammographic practice affect film reading style – Breast Screening vs. Symptomatic Radiologists?” In *Image Perception, Observer Performance, and Technology Assessment*. MP. Eckstein and Y Jiang (Eds.) *Proceedings of SPIE 2006*
30. Esserman L., Cowley H., Eberle C., Kirkpatrick A., Chang S., Berbaum K., & Gale A.G., “Improving the Accuracy of Mammography: Volume and Outcome Relationships”. *J. Nat. Cancer Inst.*, Vol. 94, No. 5, 369-375, (2002).
31. Scott H.J., Gale A.G., & Wooding D.S., “European Breast Screening Performance: does case volume matter?” In: *Image Perception, Observer Performance, and Technology Assessment*, D.P. Chakraborty & M.P. Eckstein (eds.) *Proceedings of SPIE Vol. 5372*, (2004).
32. Scott H.J., and Gale A.G., “Breast screening: when is a difficult case truly difficult and for whom?” In *Image Perception, Observer Performance, and Technology Assessment*. MP. Eckstein and Y Jiang (eds.) *Proceedings of SPIE*, Vol. 5749, (2005)
33. Whatmough P., Gale A.G., and Wilson A.R.M., “Do radiologists agree on the importance of mammographic features?” In: K Doi, M L Giger, R M Nishikawa & R.A. Schmidt (Eds.), *Digital Mammography '96*. Elsevier, Amsterdam, (1996).
34. Savage, C.J., Gale, A.G., Pawley, E.M. & Wilson, A.R.M., “To err is human, to compute divine?” In: Gale, A.G., Astley, S., Dance, D. & Cairns, A.: (Co-eds.) *Digital Mammography*. (Amsterdam) Elsevier Science B.V. (1994).
35. Scott H.J., Gale A.G. & Hill S., “How are false negative cases perceived by mammographers? Which abnormalities are misinterpreted and which go undetected?” In D. Manning and B Sahiner (Eds.). *Image Perception, Observer Performance, and Technology Assessment. Proceedings of SPIE 6917;13*, 1-11, (2008)
36. Gale A.G., Walker G.E., Roebuck E.J. & Worthington B.S., “Visual search training in mammography”. In Luer G. & Lass U. (Eds.) *Fourth European Conference on Eye Movements Vol. 1*, (Gottingen) Hogrefe, (1987).

37. Gale, A.G., Savage, C.J., Pawley, E.F. & Wilson, A.R.M., "Breast screening: visual search and observer performance". In H.L. Kundel (Ed.), Medical Imaging 1994: Image Perception. Proc. SPIE 2166, (1994).
38. Mugglestone M.D., Gale A.G, Cowley H.C, Wilson A.R.M., "Diagnostic performance on briefly presented mammographic images", in H.L.Kundel (Ed.) Image Perception. Proc. SPIE Vol. 2436 pp 106-116, (1995).
39. Scott H.J. & Gale A.G., "Breast screening: PERFORMS identifies key mammographic training needs". British Journal of Radiology, 79: S127-S133, (2006).
40. Whatmough P., Gale A.G. and Wilson A.R.M., "Computer Assisted Education and Inter-Disciplinary Breast Cancer Diagnosis". In Yongmin Kim (Ed.) Medical Imaging 1996: Image Display, Proc. SPIE 2707, (1996).
41. Yap M.H., Gale A.G., Scott H.J. "Generic Infrastructure for Medical Informatics (GIMI): The Development of a Mammographic Training System", In: E.A. Krupinski (ed.): Digital Mammography, LNCS 5116, pp. 577-584, Springer-Verlag Berlin Heidelberg, (2008).