

# **Can a computer-based educational tool improve radiographers' decision-making on when to perform repeat mammography views on quality grounds?**

## **Background**

Quality assurance issues are of the utmost importance in delivering high quality and reliable screening programmes. Image quality impacts on cancer detection, whereas technical repeats impact on the radiation dose a woman receives, as well as lengthening appointment time which increases screening cost. Although there is a shortage of published evidence on the frequency of unnecessary repeats in mammography, Dunn and Rogers have suggested that in general, as many as 50% of original images repeated for positional errors would have been found diagnostically acceptable by radiologists (Dunn and Rogers, 1997). Furthermore, routine NHSBSP audit procedures have identified spikes in technical repeat rates concurrent with the introduction of digital mammography. While there have been favourable reports of digital mammography reducing the need for appointments for technical recalls (i.e. repeats at a later date) (TRs), these are countered by evidence suggesting this may be at the expense of additional, possibly unnecessary, technical repeats (TPs) at the initial screening appointment (NHSBSP Equipment Report 0601). One study suggests that initial repeat rates for full field digital mammography may be more than five percentage points higher than for film-based mammography (Ripley et al, 2012). Instant accessibility of the image by the radiographer may have negative as well as positive consequences. Although it enables technically inadequate images to be spotted immediately and repeats taken without undue inconvenience to the women screened, conversely, the emphasis on producing a technically perfect film the first time may be reduced, and the radiographer may be tempted to retake a suboptimal but diagnostically acceptable image, to satisfy professional pride by producing a "perfect" image (NHSBP Equipment report 0601). In addition, radiographers need to be aware of subtle changes to image quality assessment. For example, digital mammography exaggerates the appearance of some types of technical flaw, for example skin folds, which do not necessarily require the image to be retaken. Conversely, there are currently some concerns that blur on images may not always be recognised at immediate quality assessment, coming to light only upon reading in optimal diagnostic conditions.

Maximising the benefit of digital mammography depends on training mammographers to promote awareness of these issues, and to support accurate grading of digital image quality and appropriate decision-making on technical repeats. Studies of computer-based training tools for radiology tasks in screening mammography have demonstrated improvements in reader performance and reduction in inter-observer variability (Urban et al, 2007). This study will evaluate a computer-based training tool to assess if its use can improve the quality and the consistency of image assessments made by mammographers by comparing their decisions with a robust gold standard and providing feedback. The tool was developing in a prior collaboration between the Scottish Mammography Education Centre, Manchester and Edinburgh Universities and University College London as a flexible environment for image-based assessment and learning tasks. Validating the tool would pave the way for the software's adoption as a training and assessment aid within core mammography qualifications (for example, NHSBSP trainers envisage using the software to replace ad-hoc arrangements currently in

place for delivering objective structured practical examinations), as well as for CPD.

## **Aim**

### **Primary research questions:**

1. How reliable and valid are the “gold standard” image acceptance decisions built into this training tool?
2. Does training with this tool improve image quality assessment decisions among participants, as measured by agreement between participants and the gold standard?
3. Does training with this tool increase the consistency of image quality acceptance decisions between participants?

The main aim of this project is to explore whether similar benefits are possible for computer-assisted training of radiographers. National mammography training and quality assurance groups have raised the issue of inconsistency between radiographers and centres in making decisions on when images are acceptable. In order to assess image quality there are various methods in use across the UK. Currently the predominant system used is P,G,M,& I (perfect, good, moderate and inadequate). This method is extremely subjective and inter-observer variability is known to exist, with Moreira et al (2005) demonstrating its poor reliability and validity. A national working group has been set up to develop modified assessment criteria that can be utilised for QA and for training. The criteria used within our training intervention are being refined in consultation with the national working group.

This project therefore aims to assess the validity and efficacy of this tool as a method of providing training to improve quality and consistency in mammographic image quality assessment.

## **Methods:**

### **Intervention refinement:**

We will pilot use of the training intervention with three radiographers to resolve any usability and technical issues with the software. A standard laboratory usability approach will be used where system use for predefined tasks is videotaped and usability problems identified. This will enable us to fine-tune the tool configuration and evaluation protocol.

### **Establishing validity through testing and refining the “gold standard”:**

A group of four experts at national mammography trainer level, including at least one trained in mammography image interpretation, who have not been involved in developing the training tool, will assess the existing **training case set** based on nationally agreed criteria. Intra-rater reliability for each expert will be assessed by comparison with a repeat observation a minimum of two weeks later. Inter-rater reliability between the four experts will also be assessed. Intraclass correlation coefficients (ICC) will be used to measure agreement and an ICC of at least 0.8 will be required to establish the gold standard. Consensus discussion and re-assessment of the training set will take place, revising the set if required, until this level of agreement is achieved.

A second set of cases, the **test set**, will be defined and the image quality assessed by the expert panel in the same way as for the training set, to set the gold standard opinion for testing outcomes the efficacy of the training tool.

## **Design:**

A quasi-experimental design will be used in the form of a before-and-after assessment. Although participants do in a sense act as their own controls in a before-and-after design, we consider a control group to be worthwhile as a way of distinguishing effects of the training tool from natural effects over time.

**Pre-intervention assessment:**

Twenty radiographers from two Scottish breast screening units will undertake the **test** set. Participating radiographers will have a minimum of two years' mammography experience and will be fully trained and competent in digital mammography, to reduce the risk of natural learning curves over time confounding the detection of effects of the training tool. Each participant's level of agreement with the gold standard classification will be assessed using the kappa statistic. Consistency between radiographers will also be measured, using the intraclass correlation co-efficient.

**Intervention:**

Half of the radiographers (n=10) will then undertake the training intervention, where they will assess the **training** images and receive immediate feedback on the correctness of their responses in comparison with the established gold standard.

**Post-intervention assessment:**

Following an interval of at least two weeks, to minimise recall bias, all 20 radiographers will re-asses the **test** set and their levels of agreement will again be compared with the gold standard. Differences in agreement with the gold standard within participants between the pre- and post-intervention tests will be assessed. Consistency between the radiographers will also be re-tested after the intervention. Thus, effects of the training exercise on the appropriateness of practitioners' image quality judgements, as well as on variability between practitioners' judgements will be detected. We have calculated that with a test set consisting of 21 cases, one third of which will have been deemed of acceptable quality, one third deemed acceptable but wrongly rejected in real life, and one third correctly rejected in real life, 10 participants and 10 controls should be sufficient to detect an effect of the training tool on image quality decision making by the means explained above. This is with a degree of caution given the early stage of this investigation and the hitherto unknown baseline size of the problems of poor decisions and inconsistency.

**Impact**

This study will yield validated training and test sets of mammograms for the purposes of improving image quality assessments, particularly with respect to overall acceptability of images. It will also provide data on the efficacy of our software-based training tool for improving the standards and consistency of radiographer decision making on image acceptance. The data could be used to design and power a larger-scale evaluation of the tool's effectiveness and cost-effectiveness, including effects on repeat rates in practice. We believe the training tool could be most valuable if used in tandem with the on-going implementation and early years of digital mammography, where there is a recognised problem around decision-making on image quality acceptance. Even if the problems seen with image acceptance decisions around digital implementation turn out to be short-lived, there is a clear demand for an effective training aid in this area. A number of lead educators from the UK's mammography training centres have indicated that they would be keen to utilise this package as soon as possible. We therefore consider it urgent to establish its validity and efficacy.

## 11. References:

Dunn M. and Rogers A. (1998) X-ray Film Reject Analysis as a Quality Indicator. Radiography, 4, 29-31.

NHSBSP Equipment Report 0601 (2006) Evaluation and Clinical Assessment of Digital Mammography Screening Using a Sectra Microdose Full Field Digital X-Ray Unit.

<http://www.cancerscreening.nhs.uk/breastscreen/publications/nhsbsp-equipment-report-0601.pdf>

(accessed 24/1/2012)

Urban N., Longton G, Crowe A., Drucker M. Lehman C, Peacock S, Lowe K, Zeliadt S., Gaul M. (2007) Computer-Assisted Mammography Feedback Program (CAMFP): An Electronic Tool For Continuing Medical Education Academic Radiology, 14(9), 1036-1042

Moreira C, Svoboda K, Poulos A, Taylor R, Page A, Rickard M (2005) Comparison of the validity and reliability of two image classification systems for the assessment of mammogram quality. J Med Screen 12: 38-42

Medical Research Council (2008), Developing and evaluating complex interventions: new guidance

<http://www.mrc.ac.uk/utilitis/documentrecord/index.htm?d=MRC--4871>

(accessed 30/1/2012)

Quality Improvement Scotland (2008) Determining the most clinically and cost-effective way of implementing digital mammography services for Breast Screening in NHS Scotland. Health Technology Assessment Advice 10.

[http://www.healthcareimprovementscotland.org/previous\\_resources/hta\\_report/hta/hta\\_10.aspx](http://www.healthcareimprovementscotland.org/previous_resources/hta_report/hta/hta_10.aspx)

(accessed 31/3/2010)

Riple O, Otto P, Hatab R (2011) Repeat Rates and Their Effect On Dose in Transitioning From Screen/Film to Full-Field Digital Mammography. Annual Meeting of the American Association of Physicists in Medicine. July.

<http://www.aapm.org/meetings/amox2/pdf/59-15807-20871-208.pdf>

(accessed 24/1/2012)