Computer-based training intervention to reduce unnecessary repeat images

Ann Mumby[1]; Mark Hartswood[2]; Patsy Whelehan[3]
[1]Scottish Mammography Education Centre, [2]University of Oxford, [3]University of Dundee

Introduction

With the transition from film-screen to digital mammography, increased technical repeat rates were noted in the NHS Breast Screening Programme[1] and internationally[2]. The risk of unnecessary repeats persists, arising from the ease and speed with which an image judged inadequate can be repeated, and from the desire of radiographic staff to produce the best image possible and avoid the need for women to be recalled by the reader. Reducing unnecessary repeats would save resources and reduce radiation dose.

Studies of computer-based training tools for radiology tasks in screening mammography have demonstrated improvements in reader performance and reduction in inter-observer variability[3]. Similar benefits may be possible for computer-assisted training of radiographers. Therefore, a computer-based training tool has been developed to improve radiographic decision-making on when to repeat a mammographic image.

The aim of this study was to obtain preliminary data on the validity and efficacy of this training intervention.

Materials and Methods

A set of mammograms, judged to have a range of imperfections which might challenge radiographers' judgements on whether the images were acceptable, was selected by the lead investigator. These cases were loaded into the computerised interface and subjected to review by three radiographers with national-level mammography training expertise, two of whom were qualified in mammography image interpretation. Observers indicated whether each image met specific criteria derived from NHSBSP guidance[4]. The purpose of this expert review was to set the reference standard for whether each image should be repeated or not. Each expert read all the cases twice and the reference standard for Accept or Reject was set as the majority decision of all six reads per case, with the lead investigator arbitrating when the expert classifications were tied at 3-3. The case set was then divided into "test cases" and "training cases".

Twenty-two participants with at least 2 years mammography experience were recruited from two breast screening centres in Scotland. They were then randomly divided into equally-sized control and intervention groups. Both groups read the test cases at the beginning and the end of the study process. Between the two test-set observations, only the intervention group also read the training set. They were given immediate feedback on what the "correct" Accept/Reject decision should have been, from the "expert" decision stored in the software.

All data were captured by the software tool and exported for analysis to Microsoft Excel and SPSS. Agreement with the reference standard opinion for each image was assessed before and after the training intervention using descriptive and kappa statistics.

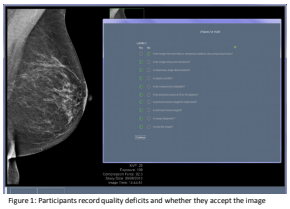

Figure 1: Participants record quality deficits and whether they accept the image

Figure 2: Overview screen with opportunity to compare images and review decisions

Figure 3: Feedback and rationale

Results

Nineteen participants completed the trial, 8 in the intervention arm and 11 controls. Two from the control arm were excluded for a protocol deviation, i.e. interval between before and after test-set reads < two weeks. Thus there were 17 datasets for analysis.

Mammography experience (years) in the final sample was similar between groups (Control - mean: 9.2, median: 7, range: 2-18; Intervention – mean: 10, median: 8, range: 4-18). In the control arm there were two film-readers, one stereo biopsy practitioner and one mammography trainer, and in the intervention arm, no film-readers, two stereo biopsy practitioners and one mammography trainer.

Mean, median and range of time intervals in days between first and second test-set reads were: Control - 83.4, 70, 28-189; Intervention - 122.9, 115.5, 75-160.

Percentage agreement with the reference standard increased in 2 of 9 controls and 4 of 8 members of the intervention group. It was unchanged in one of each group and decreased in 6 of 9 controls and 3 of 8 in the intervention group. Agreement with the gold standard according to Cohen's kappa was extremely low for both first and second reads of the test set in both groups, and most of the kappa scores were not statistically significant. (See Table 3.)

In the intervention group, 8 of 8 participants accepted higher numbers of images at their second read, compared with 3 of 9 controls.

Anecdotally, the tool was user-friendly and members of the intervention group considered it useful, particularly the provision of the experts' rationale for their decisions.

Participant	Percentage agreement with reference standard: first test	Percentage agreement with reference standard: second test	Agreement with reference standard: first test (kappa)	p-value*	Agreement with reference standard: second test (kappa)	p-value*
Control-1	81	72	0.02	0.85	-0.02	0.84
Control-2	77	73	0.24	0.05	0.12	0.30
Control-3	78	77	0.02	0.08	0.07	0.53
Control-4	77	78	0.33	0.01	0.16	0.14
Control-5	81	81	0.24	0.05	0.11	0.23
Control-6	78	83	0.11	0.23	0.19	0.06
Control-7	80	77	0.19	0.06	0.08	0.42
Control-8	83	80	0.19	0.06	0.08	0.42
Control-9	83	72	0.17	0.15	0.04	0.73
Intervention-1	78	81	0.27	0.02	0.20	0.08
Intervention-2	80	75	0.19	0.13	0.00	1.00
Intervention-3	81	86	0.33	0.01	0.28	0.03
Intervention-4	81	72	0.48	0.69	0.14	0.23
Intervention-5	63	75	0.00	1.00	0.08	0.51
Intervention-6	88	84	0.33	0.01	0.24	0.05
Intervention-7	80	80	0.30	0.01	0.13	0.25
Intervention-8	78	81	0.10	0.38	0.16	0.14

Table 1: Results: comparison of participants' reads with the reference standard

Discussion

This randomised, controlled, before-and-after trial aimed to produce preliminary evidence regarding validity and efficacy of a training intervention designed to promote good decision-making on image acceptability. A larger proportion of the intervention group compared with the controls achieved higher agreement with the reference standard on the post-test. However, this represents only very weak evidence of intervention efficacy and further work is required.

Development of the reference standard, reported elsewhere[5], was problematical, with agreement among the expert panel no better than moderate. Although we resolved this by setting the consensus (+/- arbitration) opinion as the reference for each image, the validity of the reference standard is still open to challenge. In the main trial, the kappa scores were not statistically significant, indicating insufficient power to assess levels of agreement above chance. One of the factors limiting the power of the study was the low prevalence of the Reject condition in the sample of images. Additional limitations likely to have led to underestimation of intervention effect include selection of cases where the decision was borderline, and the inclusion of experienced rather than trainee practitioners.

Conclusions

This computer-based training tool shows initial promise in developing skills to judge when an image is of acceptable quality.

Further work with a more balanced case set is required to quantify its efficacy in inexperienced practitioners.

The study provides pilot data to inform a power calculation for a definitive study.

Contact

Ann Mumby
Ann.mumby@ggc.scot.nhs.uk

Acknowledgements

Thanks to the College of Radiographers Industrial Partnership Scheme for funding the study, to the experts who contributed to the reference standard, and to the participants and their centres in Edinburgh and Glasgow.

References

1. 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)
2. Ripley O, Otto P. and 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/amos2/pdf/59-15807-20871-208.pdf (Accessed 24/01/2012)
3. Urban N., Longton G, Crowe A., Drucker M, Lehman C, Peacock S, K.Lowe, Zeliadt S., Gaul M. (2007) Computer-Assisted Mammography Feedback Program (CAMFP): An Electronic Tool For Continuing Medical Education Academic Radiology, 14(9), 1036-1042.
4. NHSBSP Publication 63.
5. Whelehan P, Hartswood M and Mumby A. Observer variability in accept/reject classification of clinical image quality in mammography. *ECR 2015*. Vienna: European Society of Radiology, 2015. http://dx.doi.org/10.1594/ecr2015/C-2432