This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

The effect of digitising film prior mammograms on radiologists' performance in breast screening: a JAFROC study

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© 2009 Society of Photo-Optical Instrumentation Engineers

VERSION

VoR (Version of Record)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Taylor-Phillips, Sian, Matthew G. Wallis, Alison Duncan, and Alastair G. Gale. 2019. "The Effect of Digitising Film Prior Mammograms on Radiologists' Performance in Breast Screening: A JAFROC Study". figshare. <https://hdl.handle.net/2134/6297>.

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

The Effect of Digitising Film Prior Mammograms on Radiologists' Performance in Breast Screening: A JAFROC Study

Sian Taylor-Phillips^{*a}, Matthew G. Wallis^b, Alison Duncan^c, Alastair G. Gale^a

^aApplied Vision Research Centre, Loughborough University, UK

^bCambridge Breast Unit, Addenbrooke's Hospital, Cambridge, UK

^cWarwickshire, Solihull, and Coventry Breast Screening Service, University Hospital, Coventry, UK

ABSTRACT

After the introduction of digital mammography the film mammograms from the previous screening round (the prior mammograms) can be displayed in a variety of ways. This paper investigates the performance of radiologists reading digital screening mammograms with the prior mammograms displayed either as film or in digitised format. A set of 162 cases was assembled, each with two view digital mammograms and two view film prior mammograms. Of these cases 66 were malignant as proven by biopsy, and the others were normal or benign. The film prior mammograms were digitised at 75µm. Eight participants, with four to seventeen years experience of reading screening mammograms, each read the mammograms twice; once with the digitised prior mammograms displayed on the digital workstation, and once with the film prior mammograms displayed on an adjacent multi-viewer. The two viewings were at least one month apart. Participants marked the location of abnormalities on a paper copy of the mammograms and rated the probability of malignancy of each abnormality. Participants were video-taped whilst reading the cases to enable analysis of gross eye movements for information regarding the level of use of the prior mammograms. JAFROC analysis showed no difference in performance between the conditions.

Keywords: Observer performance evaluation, image display, prior mammograms, digital mammography, ROC methodology

1. INTRODUCTION

The introduction of digital mammography is underway throughout Europe, with many different implementations. Whilst standards are in place concerning the display of the current digital mammograms¹ there is little consensus about how the prior mammograms should be displayed. For the first round of digital screening these mammograms will be in film format. Roelofs et al.² provided evidence that these prior film mammograms should always be displayed, rather than requiring the radiologist to request them only when they feel these are necessary. This study presented a set of 160 mammograms twice, with and without the prior mammograms, and asked participants to identify the cases for which they would request to see the prior mammograms. It was found that performance was improved when the prior mammograms were always present rather than only present when requested by the radiologist. However, there is little evidence about whether the display medium of the prior mammograms affects radiologists' performance. The prior mammograms could be displayed in film format on a multi-viewer, or digitised and displayed on the digital workstation alongside the current mammograms. Two studies^{3,4} found no difference between detection performance between a hard copy display or soft copy display on a CRT or five mega-pixel LCD monitor, but these were digitally acquired images printed out, rather than screen-film images digitised. Nab et al.⁵ found no difference in performance between reading from original film screen mammograms on a light box and the same mammograms digitised and displayed in soft copy format; however, this study only employed two participants so may not have had sufficient power to identify any differences.

*s.phillips2@lboro.ac.uk

Some information is always lost in the digitisation of mammograms and therefore at certain resolutions digitising mammograms can reduce performance. Ruschin et al.⁶ found that the detection of the shape of microcalcifications was poorer on screen film mammograms digitised at over 100 μ m in comparison to 100 μ m or less. Chan et al.⁷ found an increase in detection accuracy with a pixel size of 35 μ m rather than 70 μ m but this was with a computerized model observer rather than human observers. The digital image quality can also be degraded by excessive levels of ambient light⁸ and therefore the presence of a bright light source to illuminate film prior mammograms could adversely affect the display of the current digital mammograms.

However, it is not only the image quality which affects radiologists' performance, but other image factors including breast density⁹, lesion location and lesion type¹⁰, observer factors such as experience and annual case volume¹¹, and workstation factors such as ambient light levels or the presence of other distractions.⁸ Taylor-Phillips et al.¹² found that participants used the prior mammograms for a greater proportion of cases when they were digitised rather than when displayed in film format. The aim of the current paper is to investigate whether this difference in behaviour, when considered alongside the other factors above, will translate to a difference in screening performance.

2. METHODS

2.1 Case Set

A set of 162 recent screening mammograms, of which 66 were malignant, was assembled. All cases had digital current mammograms taken between March 2005 and June 2007 as part of the Warwickshire, Solihull, and Coventry Breast Screening Programme, and film prior mammograms from three years previously. All malignant cases which fitted the criteria were considered for inclusion in the experiment (79 in total). A set of 100 normal/benign cases were chosen at random from the database of 'difficult' normal cases for the same time period. An expert radiologist with 20 years experience in breast screening used all available records to classify each case as malignant or benign. He also marked the locations of malignant lesions, characterised the lesion type, rated the difficulty of the case, and advised whether it should be included in the experiment. Cases were then paired according to diagnosis (malignant or normal), then by lesion type and difficulty rating. Some 17 cases were not appropriate for inclusion due to being mammographically occult, only having single view prior mammograms, technical problems, or not having an appropriate match. This gave two sets of 81 mammograms, matched for diagnosis, and where possible also for lesion type and case difficulty.

2.2 Participants

All eight screening observers at University Hospital (Coventry) took part in the study, four radiologists and four radiography advanced practitioners. Every participant was familiar with the equipment, but unfamiliar with the method of reporting on a confidence scale. Therefore before starting the experiment each participant was given a set of three practice cases to report, and an opportunity to ask questions about any aspect of the study.

2.3 Equipment

The digital mammograms were displayed using the MicroDose Mammography system (Sectra, Sweden) on twin five megapixel LCD screens. The film prior mammograms were digitised using an Array 2905 Laser Film Digitiser, (Array Corporation, New Hampshire, USA), set to 75 μ m standard resolution. Mammographic film display was on a Mammolux XL multi-viewer (Planilux, Germany), which was positioned adjacent and perpendicular to the digital workstation as shown in figure 1. Reading conditions were identical for both experimental conditions with the room darkened. For the soft copy reading and for the second experimental condition the only additional illumination being that from the hard copy mammo-illuminator, which participants could dim or turn off as necessary.

Figure 1 – Layout of both experimental workstations a. digital mammography workstation with digitised prior mammograms and b. digital mammography workstation with film prior mammograms.

2.4 Performance and Behaviour

To measure performance each participant read the 162 cases twice, once with film and once with digitised prior mammograms. The reading sessions were at least one month apart. Counterbalancing was applied for reader characteristics and order of presentation. In order to minimise the effects of fatigue the 162 cases were split into three sets, each of which were read at least a week apart. Each participant completed all of the reading sessions on the same day of the week and at the same time of day to minimise extraneous variables. The order of case presentation was randomised at the start of the experiment, and was the same for all participants.

Analysis using both Jackknife Free-Response Receiver Operating Characteristic (JAFROC) and Receiver Operating Characteristic (ROC) methodology was undertaken. JAFROC methodology was preferred as it takes into account the additional lesion location information, however due to recent questions about this methodology¹³ standard ROC analysis was also conducted. Each participant was asked to mark the location of any lesions on both the Cranio-Caudal and Medio-Lateral Oblique view and rate the probability of malignancy on a scale from 0 to 100 percent. Lesion location was marked on a 6cm x 5.5cm print out of each mammogram. For the JAFROC analysis, the marked lesion location was considered correct if this was within 2mm of the correct lesion boundary, as originally marked by an experienced radiologist with all of the clinical information available. If a participant marked the correct location on one view, but not on the other view the location was considered correct for the purposes of JAFROC.

Participants were videotaped whilst carrying out the experiment. Detailed analysis of participant behaviour, and in particular their eye movements will be conducted and then correlated to performance. However this analysis is beyond the scope of this paper and will be reported elsewhere.

3. RESULTS

ROC analysis was performed using SPSS 16.0.¹⁴ JAFROC analysis was performed using JAFROC1 v1.0.¹⁵ The data all met the criteria for maximum false positive fraction and effective number of bins as defined by the JAFROC software. There was one data point excluded from the analysis, due to a software fault which prevented one of the participants from viewing one of the cases. There was missing information in a further three instances. Two where the lesion location was marked correctly but no probability of malignancy indicated. In this instance the average probability of malignancy from cases where the lesion location was filled in correctly was taken for that participant. There was a further instance where the probability of malignancy was indicated but the lesion location not marked. In this case for the JAFROC analysis the lesion location was considered to be incorrect.

Both ROC and JAFROC methodologies found no difference between the two conditions. The ROC curves for each participant are shown in figures 2 and 3.

Fig. 2. ROC curves for each radiologist

Fig. 3. ROC curves for each radiography advanced practitioner

4. DISCUSSION

This study has found no difference in performance in reading digital mammograms whether the prior mammograms are digitised or displayed in film format. If this study is considered to have adequately modelled real world screening then it could be concluded that in the transition to digital breast screening there is no need to digitise the prior mammograms if these can be mounted on a multi-viewer. Roelofs et al.² provide compelling evidence that performance would be degraded if the prior mammograms were not displayed at all, but only available upon request, and therefore simply providing a light box for radiologists to hang the prior mammograms when they considered them necessary would not provide an acceptable solution. The evidence presented in this paper does not contradict that of Roelofs et al.² because here the prior mammograms were always displayed in both conditions.

Whether this study has accurately modeled real world behaviour is yet to be determined. Taylor-Phillips et al.¹² found in live screening that the screen readers used the prior mammograms for a greater proportion of cases when these were digitised in comparison to being displayed in film format on a multi-viewer. In the present study eye movement data have been recorded of how each participant visually examined the digital images or the digital and analogue images. Planned future analyses of these data will determine whether the same pattern of eye movements has been followed in the experimental scenario. If the same behavioural pattern has not been followed to that in the real world then it will limit the conclusions that can be made from these performance data.

In this study digitisation of the prior mammograms was at 75 μ m. Chan et al.⁷ suggest that performance would be improved if digitisation was at 35 μ m, however they use computerized observers rather than human observers, so the effect on the human observer is not known. There is always information lost in the process of digitisation of analogue films, and it is possible that this information loss degrades performance using the digitised prior mammograms.

Additionally, here the observers also had greater experience using film prior mammograms mounted on a multi-viewer than they had of digitised prior mammograms, as this was their chosen implementation at the study screening centre. It is not clear whether this affected their performance. Per Skaane et al.¹⁶ found considerable intra and inter observer variation in their examination of screen film reporting and full field digital mammography. They interpret their findings as indicating the need for proper training for soft copy reading.

The lack of an experimental effect could also be due to insufficient power in the experimental design. The case set was sufficiently large (162 cases) with 41% malignant cases which is comparable to many other published prospective studies, and in particular this is of the same order of magnitude as the 160 and 198 cases used in similar studies which did find an experimental effect.^{2, 17} The number of participants could be increased, but due to the characteristics of ROC methodology this would provide only a small increase in power.¹⁸ Therefore, as there is no significant effect, and indeed no trend at all, it can be concluded that the lack of experimental effect is unlikely to be due to the power of the study.

5. CONCLUSION

This study has found that digitising prior mammograms or displaying them in film format on a multi-viewer does not affect observer performance in detecting cancer for a difficult test set. Caution should be applied before generalising this to a population of screen readers as the data here refer to a specific experimental situation, and all of the participants were from the same breast screening centre. Should this finding extrapolate beyond the present study then this is important as in the UK widespread digital screening is set to be introduced over the next three years and consequently finding no performance differences implies that in the transition to digital the national screening performance should not be affected.

6. ACKNOWLEDGEMENT

We gratefully acknowledge the contribution of the participants at the Warwickshire, Solihull and Coventry Breast Screening Service who gave up their time to assist with this research, and the NHS Breast Screening Programme for their support. MGW receives research funding from the NIHR and the Cambridge Biomedical Research Group.

REFERENCES

- [1] Van Engen, R., Young, K., Bosmans, H., Thijssen, M. , [The European protocol for the quality control of the physical and technical aspects of mammography screening. Part B: Digital Mammography. In: European Guidelines for Breast Cancer Screening, 4th Edition], European Commission, Luxembourg, (2006).
- [2] Roelofs, A. A., Karssemeijer, N., Wedekind, N., Beck, C., van Woudenberg, S., Snoeren, P. R., Hendriks, J. H., Rosselli del Turco, M., Bjurstam, N., Junkermann, H., Beijerinck, D., Seradour, B., Evertsz, C. J. , "Importance of comparison of current and prior mammograms in breast cancer screening", *Radiology*, 242(1), 70-77 (2007).
- [3] Pisano, E. D., Cole, E. B., Kistner, E. O., Muller, K. E., Hemminger, B. M., Brown, M. L., Johnston, R. E., Kuzmiak, C. M., Braeuning, M. P., Freimanis, R. I., Soo, M. S., Baker, J. A., Walsh, R. , "Interpretation of digital mammograms: comparison of speed and accuracy of soft-copy versus printed-film display", *Radiology*, 223(2), 483-488 (2002).
- [4] Kamitani, T., Yabuuchi, H., Soeda, H., Matsuo, Y., Okafuji, T., Sakai, S., Furuya, A., Hatakenaka, M., Ishii, N., Honda, H. , "Detection of masses and microcalcifications of breast cancer on digital mammograms: comparison among hard-copy film, 3-megapixel liquid crystal display (LCD) monitors and 5-megapixel LCD monitors: an observer performance study", *Eur. Radiol.*, 17(5), 1365-1371 (2007).
- [5] Nab, H. W., Karssemeijer, N., van Erning, L., Hendriks, J. H. C. L. , "Comparison of Digital and Conventional Mammography: A ROC Study of 270 Mammograms", *Medical Informatics & The Internet in Medicine*, 17(2), 125-131 (1992).
- [6] Ruschin, M., Hemdal, B., Andersson, I., Borjesson, S., Hakansson, M., Bath, M., Grahn, A., Tingberg, A. , "Threshold pixel size for shape determination of microcalcifications in digital mammography: a pilot study", *Radiat. Prot. Dosimet.*, 114, 415-423 (2005).
- [7] Chan, H. P., Niklason, L. T., Ikeda, D. M., Lam, K. L., Adler, D. D. , "Digitization requirements in mammography: Effects on computer-aided detection of microcalcifications", *Med. Phys.*, 21(7), 1203 (1994).
- [8] Samei, E., Badano, A., Chakraborty, D., Compton, K., Cornelius, C., Corrigan, K., Flynn, M. J., Hemminger, B., Hangiandreou, N., Johnson, J. , "Assessment of display performance for medical imaging systems: Executive summary of AAPM TG18 report", *Med. Phys.*, 32(4), 1205 (2005). [9] Fajardo, L. L., Hillman, B. J., Frey, C. ,

- "Correlation between breast parenchymal patterns and mammographers' certainty of diagnosis", *Invest. Radiol.*, 23(7), 505-508 (1988).
- [10] Bird, R. E., Wallace, T. W., Yankaskas, B. C. , "Analysis of cancers missed at screening mammography", *Radiology*, 184(3), 613-617 (1992).
 - [11] Esserman, L., Cowley, H., Eberle, C., Kirkpatrick, A., Chang, S., Berbaum, K., Gale, A. , "Improving the Accuracy of Mammography: Volume and Outcome Relationships", *jnci*, 94(5), 369-375 (2002).
 - [12] Taylor-Phillips, S., Wallis, M. G., Gale, A. G., "Mammography workstation design: effect on mammographer behaviour and the risk of musculoskeletal disorders", *Proceedings of SPIE* 6917, 69171G (2008).
 - [13] Gur, D., Rockette, H. E. , "Performance Assessments of Diagnostic Systems Under the FROC Paradigm: Experimental, Analytical, and Results Interpretation Issues", *Academic Radiology*, 15(10), 1312-1315 (2008).
 - [14] SPSS Inc. , "SPSS 16.0 for Windows", 16.0.2 (2008).
 - [15] Chakraborty, D. P. , "Jackknife Free-Response Receiver Operating Characteristic", 1.0 (2008).
 - [16] Skaane, P., Diekmann, F., Balleyguier, C., Diekmann, S., Piguet, J. C., Young, K., Abdelnoor, M., Niklason, L. , "Observer variability in screen-film mammography versus full-field digital mammography with soft-copy reading", *Eur. Radiol.*, 18(6), 1134-1143 (2008).
 - [17] Varela, C., Karssemeijer, N., Hendriks, J. H. C. L., Holland, R. , "Use of prior mammograms in the classification of benign and malignant masses", *European Journal of Radiology*, 56(2), 248-255 (2005).
 - [18] Obuchowski, N. A. , "Sample Size Tables For Receiver Operating Characteristic Studies", *Am. J. Roentgenol.*, 175(3), 603-608 (2000).