

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.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

BY: **Attribution.** You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

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

The time course of cancer detection performance

Sian Taylor-Phillips^{*a}, Aileen Clarke^a, Matthew Wallis^b, Margot Wheaton^c, Alison Duncan^c, and Alastair G Gale^d

^aWarwick Medical School, The University of Warwick, Coventry, CV4 7AL, UK;

^bCambridge Breast Unit, Addenbrookes Hospital, Cambridge, CB2 2QQ, UK;

^cWarwickshire, Solihull, and Coventry Breast Screening Service, University Hospital, Coventry, CV2 2DX, UK

^dApplied Vision Centre, Loughborough University, Leicestershire, LE11 3TU, UK.

ABSTRACT

The purpose of this study was to measure how mammography readers' performance varies with time of day and time spent reading. This was investigated in screening practice and when reading an enriched case set. In screening practice records of time and date that each case was read, along with outcome (whether the woman was recalled for further tests, and biopsy results where performed) was extracted from records from one breast screening centre in UK (4 readers). Patterns of performance with time spent reading was also measured using an enriched test set (160 cases, 41% malignant, read three times by eight radiologists). Recall rates varied with time of day, with different patterns for each reader. Recall rates decreased as the reading session progressed both when reading the enriched test set and in screening practice. Further work is needed to expand this work to a greater number of breast screening centres, and to determine whether these patterns of performance over time can be used to optimize overall performance.

Keywords: Performance, mammography, mass screening, fatigue, radiologist

1. INTRODUCTION

In the UK National Health Service Breast Screening Programme qualified readers are required to read more than 5000 cases per year [1]. These readers are either radiologists or radiography advanced practitioners (radiographers trained to take mammograms, who are re-trained to read and interpret those mammograms alongside the radiologists). This is a monotonous task, which requires high levels of skill and vigilance to detect the sometimes subtle indications of malignancy. Understanding the patterns of performance of these readers could potentially lead to methods to optimize that performance, and therefore detect a greater number of cancers whilst recalling fewer women for further tests, which are both invasive and can cause anxiety and distress [2]. Furthermore, any patterns of performance variation with time spent reading a case set or time of day may be of methodological interest, as they may affect the results of enriched case set Receiver Operating Characteristic (ROC) studies unless controlled for across conditions. This study presents some pilot data in the field.

There is limited research on changes in mammography performance with time spent reading, however Cowley and Gale [3] have reported an increase in specificity and concurrent decrease in specificity with time spent reading the PERFORMS [4] test sets of mammograms in the UK. This is accordant with a theory of threshold increase over time. Changes in performance with time spent working have been detected for many tasks which have similarities to reading screening mammograms. A vigilance decrement of performance decrease over time has been observed in a wide range of tasks, most notably in repetitive visual tasks where the target is difficult to detect, [5] (circumstances similar to those found in mammography). There is also a prevalence effect which manifests itself in a decrease in detection rate over time when searching for a rare target such as a cancer in breast screening. This has been studied for the similar task of airport baggage screening. [6] The "2 to 4 hour shift" [7] is a dip in performance detected with time on shift for lorry, train and car drivers, and may also be relevant as driving is a safety-critical repetitive visual task similar to breast screening, although it is conceivable that such a dip may occur sooner in breast screening since it is a more intense visual task.

*s.taylor-phillips@warwick.ac.uk; phone +44(0)7725000262

Variation of performance in radiology with time of day has been investigated in several studies. Gale et al. [8] report a drop in sensitivity searching for lung nodules after lunch, in comparison to reading in the morning and evening. Two more recent studies have found a drop in radiology performance as the working day progresses. A retrospective analysis of reading performance of mammography test sets showed slightly improved performance (ROC A_z) in the morning in comparison to the afternoon, [9] but the retrospective study design may have some biases. A prospective analysis of diagnostic accuracy reading bone fractures found a decrease in performance (ROC AUC) after a working day in comparison to before. [10] The authors attribute this effect to oculomotor fatigue. However no performance variation with time of day has been found for radiologists reading the PERFORMS test sets, [3] although there is a trend towards a slight threshold increase in the evening. [11]

2. METHODOLOGY

Two data sources were analysed: Data extracted from records in the UK National Health Service (NHS) Breast Screening Programme (representing performance in screening practice), and data collected from participants readings an enriched test set of mammograms under test conditions.

2.1 Screening practice

In The UK Breast Screening Programme readers interpret the mammograms and input their decision to the National Breast Screening Service [12] database either via barcode reading (if they still use film screen mammography), or directly into NBSS via the digital mammography workstation if they have this technology available (fig 1). The NBSS database records the exact date and time at which each decision is entered. In this study a Crystal [13] report was developed to extract the time and date that each decision was entered, whether the decision was to recall the woman for further tests or not, and if the woman had further tests what the outcome was. For analysis of performance with time of day a report was used to extract data for 4 readers at one centre, between 1st January 2009 and 30th September 2010, in total over 20,000 cases. For the analysis of performance with time spent reading, data for seven readers was extracted for the time period 1st January 2007 to 31st December 2009, in total over 170,000 cases. For the time of day analysis the time and date stamp from the records was used to group cases into the time slot within which they were read. For the analysis of time spent reading the time and date stamps were used to calculate when breaks of over one hour were taken, and therefore for each case at what time it was read in relation to those breaks.

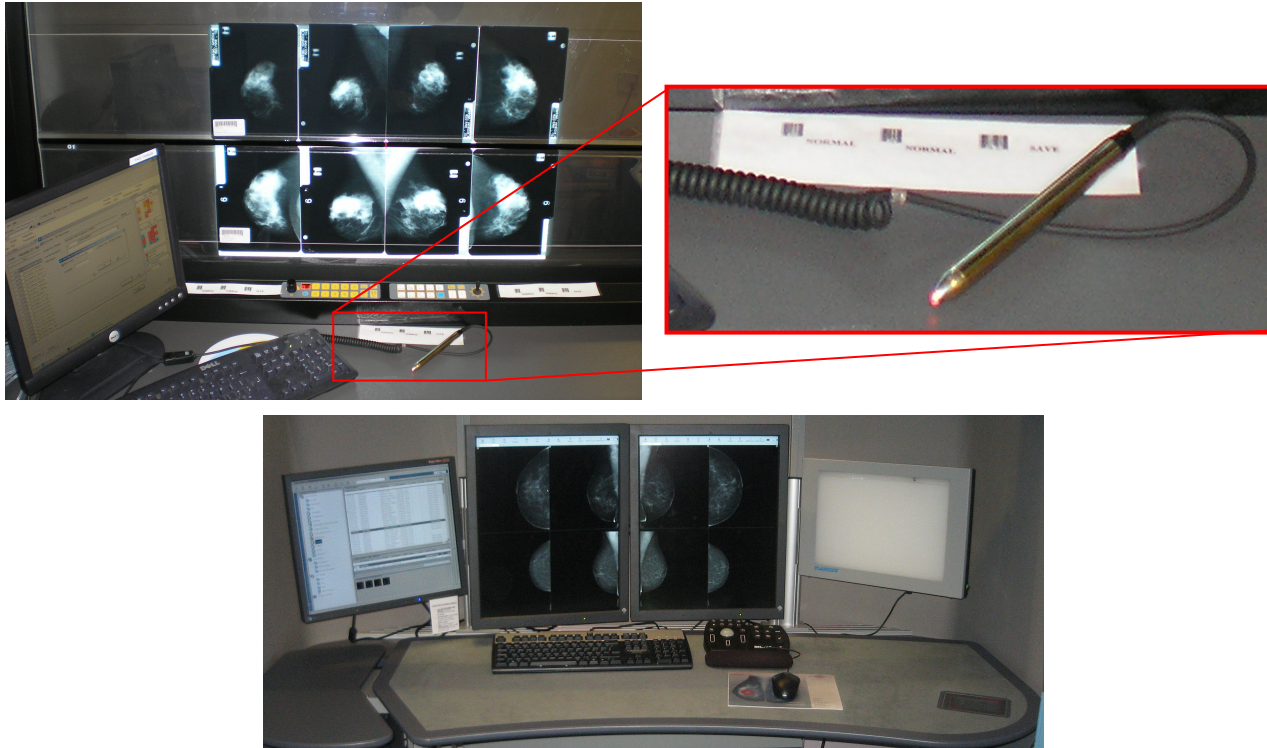


Figure 1. Typical mammography workstations used in the UK. A film workstation (top) where mammograms are mounted on a multi-viewer and results are inputted by barcode scanning. A digital workstation (bottom) where results are inputted using the mouse or control pad. In both cases the exact time and date of results input is recorded on the database system NBSS.

2.2 Enriched set of test cases

Recall rates with time spent reading were analysed for eight participants reading an enriched test set of 160 mammograms three times, over a total of nine sessions each.

3. RESULTS

3.1 Variation with Time of Day

The volume of cases read varied with time of day, with the most popular reading times mid-morning or mid-afternoon (fig 2). None of the readers in this study had set reading times, they read batches of mammograms whenever they got the opportunity.

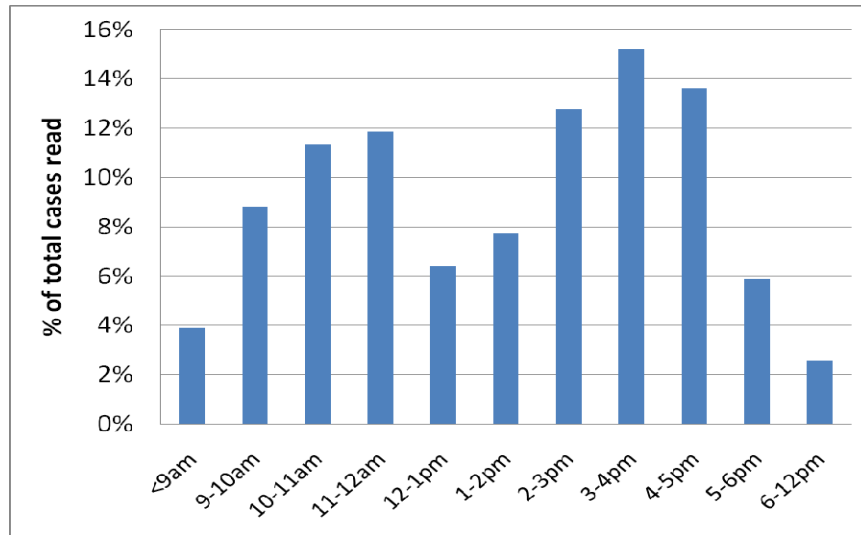


Figure 2. The proportion of cases that were read at each time of day. The majority of cases were read between 10am and 12pm in the morning, and between 2pm and 5pm in the afternoon.

Recall rate for mammography readers appears to vary with time of day, but not in the same manner for all readers (fig 3 and 4). Statistical tests for patterns were not possible due to the small number of readers and diversity in patterns of performance. Data for a larger number of readers is being collected, and these will be analysed to determine whether the readers can be characterized into types with similar patterns.



Figure 3. Variation of recall rates with time of day for two radiologists in the NHS Breast Screening Programme, both with over 5 years experience reading screening mammograms. Radiologists 1 (top) appears to recall fewer women in the afternoon than the morning reading session.

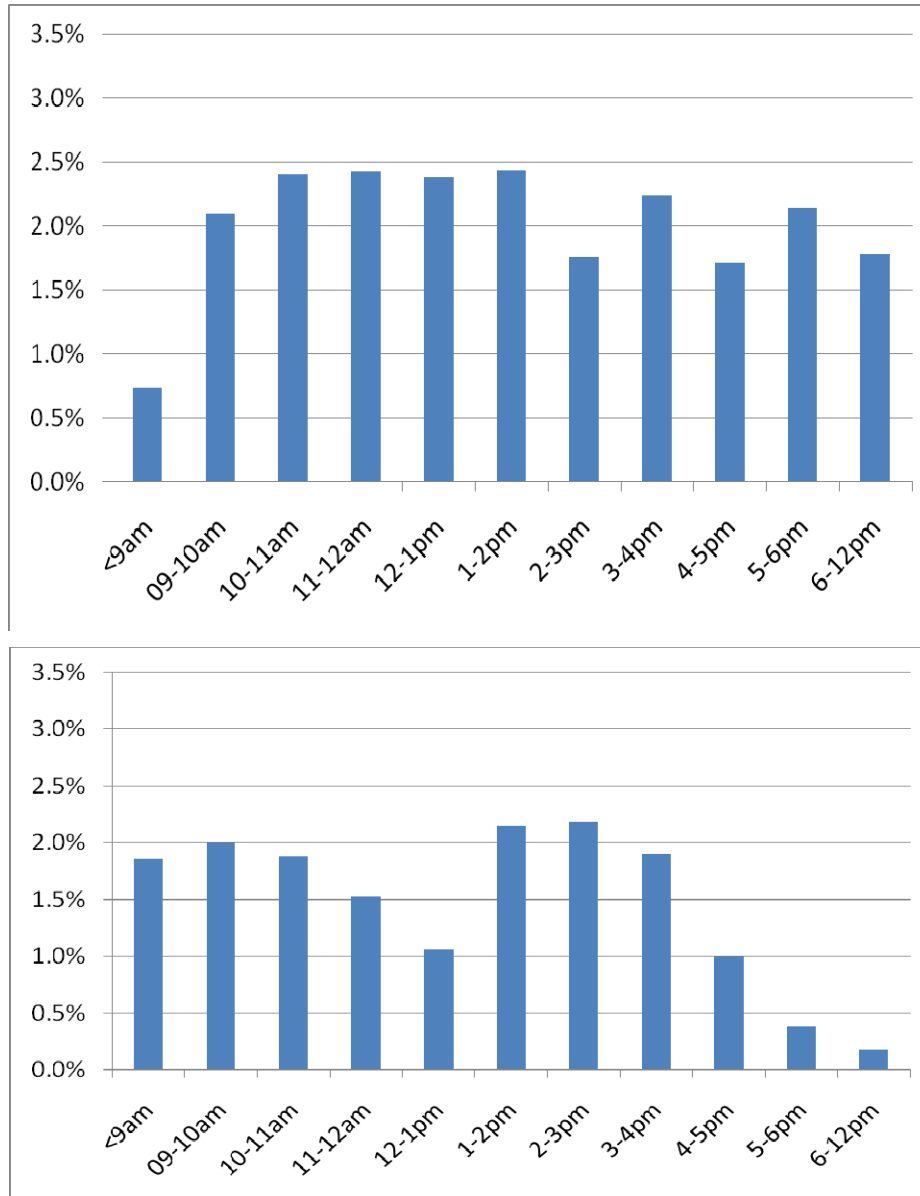


Figure 4. Variation of recall rates with time of day for two radiography advanced practitioners in the NHS Breast Screening Programme, both with over 3 years experience reading screening mammograms. Participant 4 (bottom) appears to recall a lower percentage of women around lunch time and at the end of the day, but no patterns are observed in participant 3 (top).

3.2 Variation with time spent reading

In screening practice recall rate was 2.3% higher for the first ten cases than the second ten read since a break ($t(6)=4.6$, $p=.004$, 95%CI of difference = 1.0% to 3.5%). In the enriched test set a small trend towards the same pattern of declining recall rate with time spent reading was observed ($t(7)=2.2$, $p=.06$). There was a large amount of noise in the data which is due to variation in case difficulty and the small number of cases. An example of the trend of recall rate over time for one reader is given in figure 5.

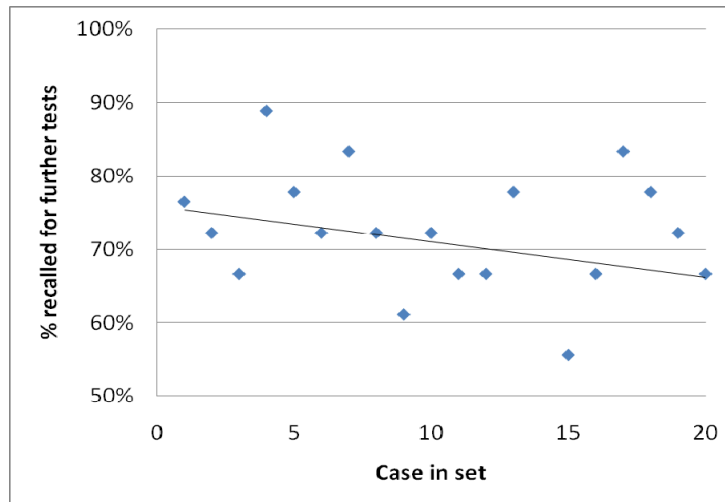


Figure 5. Percentage recalled over time in the enriched case set study for reader 1. The high levels of noise in the data are due to variation in case difficulty in the small case set (160 cases).

4. CONCLUSIONS

Performance reading mammograms may vary with time of day, time spent working and break scheduling. Further research is necessary in this field, both using experimental studies with test cases and observational studies re-analysing the results of cases read in screening practice.

5. ACKNOWLEDGEMENTS

Sian Taylor-Phillips is supported by a Wellcome Trust Value in People fellowship.

REFERENCES

- [1] [Quality assurance guidelines for breast cancer screening radiology] NHS Cancer Screening Programmes, (2005).
- [2] Aro, A. R., Pilvikki Absetz, S., van Elderen, T. M. *et al.*, "False-positive findings in mammography screening induces short-term distress - breast cancer-specific concern prevails longer," *Eur J Cancer*, 36(9), 1089-97 (2000).
- [3] Cowley, H. C., and Gale, A. G., "Time of day effects on mammographic film reading performance," *Proc. SPIE*, 3036, 212-221 (1997).
- [4] Gale, A. G., "Performs: a self-assessment scheme for radiologists in breast screening," *Seminars in Breast Disease*, 6(3), 148-152 (2003).
- [5] See, J. E., Howe, S. R., Warm, J. S. *et al.*, "Metaanalysis of the Sensitivity Decrement in Vigilance," *Psychological Bulletin*, 117(2), 230-249 (1995).
- [6] Wolfe, J. M., Horowitz, T. S., Van Wert, M. J. *et al.*, "Low target prevalence is a stubborn source of errors in visual search tasks," *J Exp Psychol Gen*, 136(4), 623-38 (2007).
- [7] Folkard, S., "Black times: temporal determinants of transport safety," *Accident Analysis & Prevention*, 29(4), 417-30 (1997).
- [8] Gale, A. G., Murray, D., Millar, K. *et al.*, [Circadian Variation in Radiology] North-Holland, (1984).
- [9] Krupinski, E., "Reader Fatigue Interpreting Mammograms," *Lecture Notes in Computer Science*, 6136, 312-318 (2010).
- [10] Krupinski, E. A., Berbaum, K. S., Caldwell, R. T. *et al.*, "Long Radiology Workdays Reduce Detection and Accommodation Accuracy," *Journal of the American College of Radiology : JACR*, 7(9), 698-704 (2010).
- [11] Gale, A., and Scott, H., [Measuring radiology performance in breast screening] Cambridge University Press, Cambridge(2010).
- [12] [National Breast Screening Service (NBSS) database] Bespoke Services, Temenos, London.
- [13] [Crystal Report] Systems Applications and Products in Data Processing (SAP), Walldorf, Germany(2010).