

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.

## How much is enough: factors affecting the optimal interpretation of breast screening mammograms

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

SPIE

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Scott, Hazel J., and Alastair G. Gale. 2019. "How Much Is Enough: Factors Affecting the Optimal Interpretation of Breast Screening Mammograms". figshare. <https://hdl.handle.net/2134/3298>.

This item was submitted to Loughborough's Institutional Repository by the author and is made available under the following Creative Commons Licence conditions.



creative commons  
COMMONS DEED

**Attribution-NonCommercial-NoDerivs 2.5**

**You are free:**

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

**Under the following conditions:**

 **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/>

# HOW MUCH IS ENOUGH? FACTORS AFFECTING THE OPTIMAL INTERPRETATION OF BREAST SCREENING

## MAMMOGRAMS.

Hazel J. Scott & Alastair G. Gale

Applied Vision Research Centre, Loughborough University, Loughborough, UK

### SUMMARY

**PERFORMS** (Personal Performance in Mammographic Screening), a self-assessment scheme for mammographers is undertaken as an educational tool by film-readers involved in reading breast-screening films in the UK. The scheme has been running as a bi-annual exercise since its inception in 1991. In addition to completing the scheme each year, the majority of film-readers also choose to complete a questionnaire, administered as part of the scheme, indicating key aspects of their every-day reading practice. These key aspects include volume of cases read per week, time-on-task reading screening films, incidence and time of break periods as well as typical number of film-reading sessions per week. Previous recommendations on best screening practice (significantly the optimum time on task) were considered in the light of these film-readers' self-reports on a current PERFORMS case set. In addition we looked at performance accuracy of over 450 film-readers reading PERFORMS cases (60 difficult mammographic cases).

Performance on measures akin to True Positive (Correct Recall Percentages) and True Negative (Correct Return to Screen Percentages) decisions were investigated. Data presented demonstrate that individual behaviours in real life screening, for the interpretation of mammographic cases, affect film-reading accuracy on a test set of mammograms for specificity and sensitivity (namely volume of cases read per week and film-reading experience). The consequences, in light of previous research, for best screening practice in real life are considered.

### INTRODUCTION

The question as to the most efficient mammographic practice has been explored in recent years (Laming & Warren, 2000). Several recommendations were proffered including the optimum time-on-task period that any given film-readers should allow when reading breast screening mammograms (it was recommended that film-readers should not read for more than 30minutes at a time). In addition it was recommended that immediate feedback on radiological accuracy would improve performance. Such a mechanism exists as a self-assessment scheme in the UK, which allows all participating film-readers to evaluate their performance on a test (not real life) set of mammograms. This scheme, PERFORMS (PERSONAL PERFORMANCE in Mammographic Screening), established since 1991 is a free and anonymous exercise consisting of difficult screening cases and provides immediate and confidential feedback to all film readers on their respective performance based on a radiological "gold standard". The scheme functions as a self assessment whereby film readers have the opportunity to evaluate 120 difficult (60x2) cases bi-yearly. In addition, many film-readers also provide information, via questionnaire/self-report, on their regular mammographic reading practice. We compared these reading practices with performance on recent PERFORMS cases with a view to elucidating which practices were a) most common and b) were associated with optimum proficiency on PERFORMS. In this way we aimed to extend the research into best screening practice by exploring common reading practice (such as volume of cases read, time-on-task, breaks during task and number of film-reading sessions per week) with direct comparison to individual performance.

### METHOD

Sitting PERFORMS (see Figure 1) requires individuals to enter data about each case, into a tablet PC, whereupon they receive immediate and confidential feedback detailing their performance on each case compared to a radiological standard. The radiological standard is initially based on the decisions of a panel of very experienced readers as well as, where relevant, on case pathology. Specifically individuals receive detailed feedback on their number of 'Correct Recall' decisions (a measure of sensitivity), 'Correct Return to Screen' decisions (a measure of specificity) and percentage of correct malignancy detected.

**PERFORMS coverage is Nationwide**

**Figure 1. What is PERFORMS?**

**Radiologist sitting PERFORMS**

**60x2 OBL/CC views**

**Touch screen tablet PC interface**

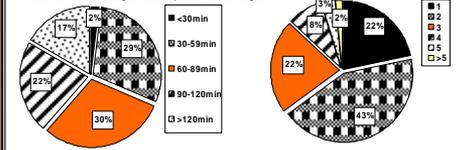
**Immediate personal feedback on the Tablet PC**

### DESIGN

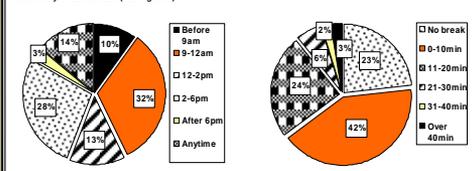
The anonymous data of over 450 UK breast-screening film-readers were compared on the most recent PERFORMS film set. Inclusion criteria for this study were those participants that completed not only the most recent round of PERFORMS (60 difficult cases) but who had also completed the computerised self-report, this was a cohort of the original group of participants who completed only the PERFORMS set (approximately 550). From their self-report we analysed participants regular screening practices, principally, the number/length of weekly sessions reading, time of day they read, details of task interruption, breaks on task, double reading practices, volume of cases read per week and years of screening experience. These were compared with accuracy measures on the PERFORMS film set based on a 'National Radiological Opinion' this is gleaned from pathology as well as from the majority opinion of all participating film-readers and allows us to include ROC measures such as  $d'$  and  $d''$  for pathology

### RESULTS

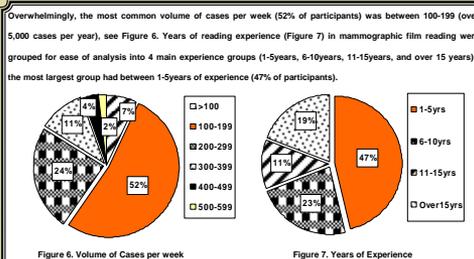
Film readers were asked to report on the most common session length, i.e. the amount of time that they would film-read before stopping. The most commonly reported session length was from 60 to 89 minutes (see Figure 2) with under 30 minutes being the least reported session length.



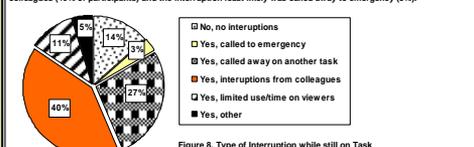
When reporting on the number of reading sessions per week the most common was two sessions (range 1- over 5) with 43% reporting this frequency (see Figure 3). The normal time of day (see Figure 4) for reading breast-screening films was most popularly between 9-12am (32% of participants) and 2-6pm (28%). Participants were asked to report on taking breaks whilst still on task, the type of break most commonly reported was 0-10 minutes followed by 11-20 minutes (see Figure 5).



Overwhelmingly, the most common volume of cases per week (52% of participants) was between 100-199 (over 5,000 cases per year), see Figure 6. Years of reading experience (Figure 7) in mammographic film reading were grouped for ease of analysis into 4 main experience groups (1-5years, 6-10years, 11-15years, and over 15 years). The most largest group had between 1-5years of experience (47% of participants).

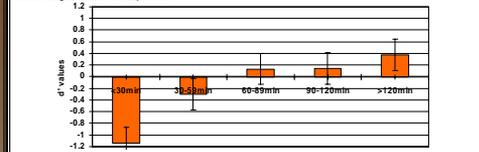


The kind of interruptions film-readers reported as the most common (see Figure 8) were interruptions from colleagues (40% of participants) and the interruption least likely was called away to emergency (3%).



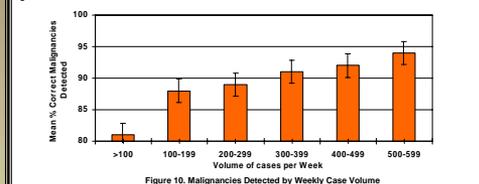
**A COMMON PROFILE OF TYPICAL READING BEHAVIOUR** emerged, namely; reading sessions lasting over an hour, reading in two weekly sessions, reading between 9-12pm with breaks of approximately 10 minutes. Interruptions were mostly from colleagues, case volume was typically between 100-199 cases per week (over 5,200 cases per year) and participants were more likely to have been reading screening mammograms from between 1-5 years.

Of these data, when compared with sensitivity, specificity and ROC measures, only three factors related to attainment on the PERFORMS scheme for the current year. Reading behaviour related to PERFORMS results were, session length, volume and experience.

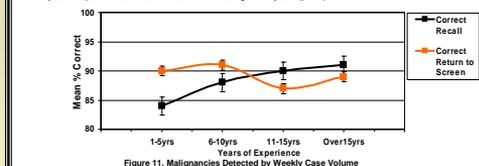


A one way ANOVA was performed on session length, with one IV (session length) and with DV's - all PERFORMS specificity and sensitivity measures. There were no significant differences for measures of specificity or for measures of sensitivity, but there were significant differences for overall ROC PERFORMS measures ( $d'$  [F(4,440) = 3.4, p<0.01] and  $d''$  for pathology [F(4,440) = 3.94, p<0.05]). Participants who reported reading for shorter sessions, 30 minutes or less, performed significantly less well compared with those who read for longer time periods (Figure 9).

For case volume, a one way Analysis of Variance identified significant differences for sensitivity measures Correct Recall [F(6,440) = 3.23, p<0.01] and Malignancies Detected[F(6,440) = 3.98, p<0.005] as well as for  $d'$  for pathology [F(6,440) = 2.45, p<0.05]. Student Newman Keuls, post hoc analysis revealed that those with a higher volumes (of over 400 cases per week) perform better than those who read less than 100 cases per week (see Figure 10) on all sensitivity measures (p<0.05), there were no significant group differences for  $d''$  pathology (p=n.s.). There were no significant differences for CS or for  $d'$ .



Years of reading experience was also related to performance on PERFORMS. A one-way ANOVA revealed significant differences for all PERFORMS measures (p<0.05). However, post hoc Student Newman Keuls identified that for all sensitivity measures, experience groups of 11 years and over performed significantly better than groups of 1-10 years (p<0.05), which in turn performed significantly better than the lowest experience group of 1 to 5 years (p<0.05). This trend was contrary to that found for specificity measures however, where the low experience groups (1-5 years and 6 to 10 years) performed significantly better than higher experienced groups of 11-15 years (p<0.05), see Figure 11. For overall ROC measures Student Newman Keuls post-hoc tests showed that groups 6-10years and over 15 years outperformed those who had been reading for 1-5years (p<0.05)



In order to extricate which of the three factors (session time but more notable years of experience and case volume) had more of an effect on performance measures, all three were entered into a multiple regression analysis (enter method). A weak but significant model emerged in which volume was not a significant predictor of any performance measures over years of reading experience (where session length was also entered). Contrastingly, using the same model, years of reading experience was a significant predictor of all sensitivity, specificity and ROC measures. In addition, length of reading session was only a significant predictor of malignancies detected,  $d'$  and  $d''$  Pathology.

### CONCLUSION

To the question of best screening practice, in terms of what is common (from participants' self-reports) individuals read in sessions twice as long as the Laming and Warren (2000) report recommends (30 minutes). Furthermore, from overall  $d'$  values, those individuals that report typically reading for under 30 minutes, performed less well on PERFORMS than those who read for longer. This may suggest that as well as being a recommended maximum time for film-reading sessions, there may also be a minimum. However, the main factors affecting performance were shown to be volume and years experience. Those with more years of experience and volume of cases attain greater sensitivity levels. From multiple regression analysis, years of screening experience, rather than any particular reading habit or 'style', affects performance on the present PERFORMS case set.