Journal of Medical Imaging and Radiation Sciences 56 (2025) 101889

Journal of Medical Imaging and Radiation Sciences
Journal de l'imagerie médicale et des sciences de la radiation

www.elsevier.com/locate/jmir

Research Article

# Factors that impact upon interpretation efficacy of radiography advanced practitioners who interpret mammograms

Noelle Clerkin[a,*], Chantal Ski[b], Patrick Brennan[c] and Ruth Strudwick[a]

[a] *University of Suffolk, Waterfront Building, 19 Neptune Quay, Ipswich IP4 1QJ, UK*
[b] *University of Sydney, Camperdown NSW 2006, Australia*
[c] *School of Nursing and Midwifery, Queen's University Belfast, Belfast, UK*

## ABSTRACT

**Objectives:** In the United Kingdom (UK) Radiography advanced practitioners (RAPs) report mammographic images, however unlike other professional groupings who read mammograms, no data are available describing factors that impact reading performance. This preliminary study explores whether or not factors such as experience, mind set, access to prior images encountered by RAPs could impact upon their performance.

**Methods:** The performance of 15 RAPs interpreting a test set of 60 mammographic cases with known reading outcomes was assessed. Twenty of these 60 cases contained a cancer, whilst the remaining cases were normal or benign. Sensitivity, specificity, lesion sensitivity, receiver operating characteristic (ROC) and free response operating characteristic (AFROC) values were established for each RAP and Student-T and Mann Whitney tests were used to identify specific features that had a significant impact on accuracy.

**Results:** In terms of ROC values, higher values ($p = 0.0042$) were seen in those readers who: had less than [10 years experience] (0.93), compared to readers with greater than 10 years of experience (0.84); read greater than [100 cases/week] (0.93), compared to those who read less than 100 cases per week (0.87) ($p = 0.0358$) as well as readers who believed that emotional mind-set impacted their image interpretation (0.91) compared to those who did not (0.84) ($p = 0.0272$). Similar higher ROC values were noted in readers who consistently relied on [prior imaging](0.94), compared to those who occasionally relied on prior projections (0.89) ($p = 0.0231$).

**Conclusion:** This preliminary work suggests that factors may impact upon the diagnostic performance of RAPs when reading mammograms. These early results from a small sample size demonstrate that further explorations are required to optimise RAP reporting.

## RÉSUMÉ

**Objectifs :** Au Royaume-Uni, les radiographes en pratique avancée préparent des rapports sur les images mammographiques. Cependant, contrairement à d'autres groupes professionnels qui lisent les mammographies, aucune donnée n'est disponible décrivant les facteurs qui ont un impact sur les performances de lecture. Dans cette étude préliminaire, les auteurs explorent si les facteurs rencontrés par ces radiographes pourraient avoir un impact sur leurs performances.

**Méthodologies :** Nous avons évalué les performances de 15 radiographes en pratique avancée qui ont interprété un ensemble-test de 60 cas de mammographies dont la vérité était connue. Vingt de ces 60 cas contenaient des cancers, tandis que les autres étaient normaux ou bénins. Les valeurs de sensibilité, de spécificité, de sensibilité aux lésions, de la courbe ROC (receiver operating characteristic) et de la courbe AFROC (free response operating characteristic) ont été établies pour chaque radiographe et les tests T de Student et de Mann-Whitney ont été utilisés pour identifier les caractéristiques spécifiques ayant un impact significatif sur la précision.

**Résultats :** En termes de valeurs ROC, des valeurs plus élevées (p=0,0042) ont été observées chez les lecteurs qui avaient moins de 10 ans d'expérience (0,93), par rapport aux lecteurs ayant plus de 10 ans d'expérience (0,84); qui lisaient plus de 100 cas par semaine (0,93), par rapport à ceux qui lisaient moins de 100 cas par semaine (0,87) (p=0,0358); qui pensent que leur état émotionnel a un impact sur leur interprétation de l'image (0,91) par rapport à ceux qui ne

**Contributors**: All authors contributed to the conception or design of the work, the acquisition, analysis, or interpretation of the data. All authors were involved in drafting and commenting on the paper and have approved the final version.

**Funding**: This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Competing interests**: The authors declare no conflict of interest. Professor Patrick Brennan is CEO and Co-founder of DetectedX. Noelle Clerkin is Professor Patrick Brennan's niece.

**Ethical approval**: Ethics approval was granted from the University of Suffolk (paper number: RETH(P)21/006.

* Corresponding author.
*E-mail address:* noellefgeclerkin@gmail.com (N. Clerkin).


https://doi.org/10.1016/j.jmir.2025.101889

le pensent pas (0,84) ($p = 0{,}0272$); et qui se fient systématiquement aux projections antérieures (0,94) par rapport à ceux qui ne s'y fient qu'occasionnellement (0,89) ($p = 0{,}0231$).

**Conclusion :** Cette étude préliminaire suggère qu'une série de facteurs peut avoir un impact sur la performance diagnostique des radiographes en pratique avancée lors de la lecture des mammographies. Il s'agit toutefois de premiers résultats et des recherches supplémentaires sont nécessaires pour optimiser la production de rapports par les radiographes en pratique avancée.

*Keywords:* Image interpretation; Advanced practice; Cancer detection

## Introduction

The most prevalent type of global cancer in women is breast cancer [1]. In 2018, 2.09 million cases of breast cancer were recorded by the World Health Organisation with approximately 41,000 cases being based in the United Kingdom (UK). 10,000 women die from breast cancer in the UK each year [2] with the annual economic burden reaching £250m. When breast cancer is diagnosed at an early stage, 90% of women will survive for a minimum of five years, but this drops to 13% when the cancer is more advanced stage 3/4 [1]. In the UK, the National Health Service Breast Screening Programme (NHSBSP), available to women from 50-70 years old, aims to reduce mortality by detecting breast cancer at an early stage when the presence of the disease is pre-clinical and impalpable [3]. Mammography remains the gold standard in offering a non-invasive, inexpensive solution able to detect malignancies of less than 1 cm [4].

The efficacy of any breast cancer screening programme relies on efficient analysis of mammographic images. This success is reliant on an individual's abilities to recognise small changes that represent cancer whilst confidently declaring normal cases cancer-free [5]. Health care professionals that interpret and report on breast images, must have the knowledge to review the image competently, identify suspicious regions of interest and differentiate between normal, benign and malignant image appearances [6]. As a result, breast imaging readers have specific image interpretation training and experience to meet the high standards of radiology or radiography peers.

Since 1995, due to the increasing demand for mammography interpretation and the paucity of radiologists, radiographers in the UK have been trained to interpret screening mammograms [7]. To date, research has compared radiographers' abilities with that of radiologists' [8] with a number of papers demonstrating that the former can operate to the same standard [9]. However, whilst there is evidence on criteria and benchmarks that promote enhanced levels of radiologist performance [10,11], limited research on the factors that promote advanced radiographic interpretive excellence is available. Without this evidence, it would be premature to assume that the same factors apply to both populations when educational backgrounds, clinical experiences and demographic characteristics may be very different [12,13].

Identifying criteria that impact image interpretation for readers have been achieved with programmes such as Breast Screen Reader Assessment Strategy (BREAST) in Australia and Personal Performance in Mammographic Screening (PERFORMS) in the UK. However, the focus of both programmes and current literature has been on radiologist data, with analyses demonstrating an array of factors determining high diagnostic efficacy such as annual reading volume [14,15], sleeping patterns [16], time of day [17], training programs [18,13], types of lesions [12] social networking [19] and even gender [12]. Similar data for Radiography Advanced Practitioners (RAPs) remains unavailable [20,21].

This preliminary quantitative study will address this deficiency by focusing on factors that should impact RAP image interpretation. Exploring such data, will for the first time, identify factors that may impact upon RAP breast cancer diagnosis performance and highlight where future research may be directed.

## Methods

Ethics approval was granted from the University of Suffolk (paper number: RETH(P)21/006) and consent was obtained from each reader through completion of an informed consent form.

### *Study design*

This study applied a two-part quantitative approach to explore factors that may affect RAP image interpretation. Firstly, an image test-set was used where radiographers interpreted a collection of mammograms. Then a questionnaire was completed by each participant which documented demographic and other parameters (the details are documented below). Both the images and the questionnaire were embedded within the DetectedX on-line platform (DetectedX, Sydney). The DetectedX platform ensures that all images are displayed at full resolution, this supported the authors' decision to choose this design.

### *Study sample and recruitment*

The focus here was to invite advanced and consultant radiographers who report mammograms for the symptomatic and screening service within the National Health Service (NHS). For inclusion in the study, a qualification in breast image interpretation was essential. Fifteen RAPs were involved from regions across the UK, these areas included Norfolk, Suffolk, Derbyshire, Essex, Yorkshire, Cumbria, Hampshire, Somerset, Cornwall, Antrim, Armagh and Derry. The sample size described here compares well with a number of similar radiologist studies performed with participant numbers of five or lower [22,23].

Participants were successfully recruited through various methods: twelve participants were recruited through direct contact within their NHS Trust; three participants were recruited through presentation at a national conference. Once each participant had read the participant information leaflet and informed consent was obtained, an instruction leaflet and a demonstration video were presented to each participant. A video call or an in-person meeting was also made available to each participant and a 24/7 help line established. Participant information leaflet and informed consent form has been added to the study's appendix (Appendix 1,2).

*Study procedure*

An image test set containing 60 digital mammographic cases were presented to each study participant. Each case consisted of four images, a cranial caudal (CC) and a mediolateral oblique (MLO) projection of each breast. All cases were selected from those individuals attending the National Breast Screening programme in Australia between the ages of 50-74. All cases were verified to be typical mixed density screening cases by a senior radiologist responsible for training and quality assurance. The cases were considered by a senior reporting UK radiographer to ensure that cases represented a typical UK mix. From the 60 cases of mixed density selected for the study, 20 of these contained a biopsy proven cancer, the remainder were considered "normal" since they received a routine recall outcome, which was confirmed by two radiologists reading independently and a follow up negative screening mammogram result obtained two years later. Within these normal cases, some contained benign masses as well as 'clear' mammograms absent of any prominent findings. This presented a reflection of a generalised screening workload. The 20 biopsy proven cancer cases contained a variety of invasive and in situ malignancies. Eight cases demonstrated indeterminate micro calcifications representing ductal carcinoma in situ whilst the remaining 12 contained asymmetric densities (n=2) and irregular masses with indistinct margins (n=10). 55% of these detected malignancies were present in the left breast. This type of enriched test set has been shown to represent well the performance of breast reporting radiologists when actually dealing with much lower prevalence in the clinic [22].

RAPs were invited to review each mammogram and mark any area of concern. All typical post-processing options were available including contrast windowing, panning and magnification. Once a region of interest suspicious of a lesion was identified, the reader could rate the area. The presented scale was from two to five: two representing high confidence that the area was benign and five indicating a high probability of a definite malignancy; if nothing was marked, a score of one was automatically given. Whilst this represents an Australian scoring system, the approach aligns reasonably well with the UK's one-five presentation scale, one demonstrating no concern and five highlighting an almost definitive malignancy. The scoring system was explained to each study participant prior to commencing the study. Reader's markings within a specific pre-set radius were accepted as correctly identified with radii being set by expert radiologists with help from associated pathology reports. When a suspicious area was identified, the reviewer had the opportunity to describe whether the area under review was a mass, asymmetric density, architectural distortion, calcification or other finding. Within these choices, subsections were also available to select the shape of the mass, whether it was global or focal, whether any calcification was present and if so, was it amorphous, course heterogeneous, fine pleomorphic or linear branching. Finally, if a lesion was presumed as normal, variants such as a low-lying inframammary node, a superimposed nipple or a subcutaneous lesion could be also described. If the reader did not observe any areas of concern on the image, the RAP would move on to the next set of mammograms without any markings. Readers were under no time constraints and could complete the image test set in more than one sitting if required. They were able to revisit any case and amend or correct any previous decision. An instruction video on how to navigate the website was available to all participants. No information was provided to the reader on the number of cases with an abnormality.

To simulate a clinical reporting environment and as recommended by the NHSBSP [23], participants undertook the study on a 5MP reading workstation, in optimum lighting conditions suitable for image interpretation: maximum luminance was no lower than 450 cd/m2 and all rooms were darkened with no direct sunlight with a typical lux value of 10-20. As stated earlier the viewer enables full post-processing functionality; features that included: zoom, windowing and full native resolution were available to all readers at all times.

*Statistical analysis*

Performance metrics were calculated including: specificity, sensitivity, lesion sensitivity and the areas under (AUC) the response operating characteristic (ROC) and alternative free-response operating characteristic (AFROC) curves.

- Specificity is defined as the number of negative mammograms the RAP correctly identified divided by the number of true negative mammograms included in the test set [21];
- Sensitivity is the percentage of cases correctly identified with breast cancer over the number of total number of cases with breast cancer [22];
- Lesion Sensitivity is the ratio of the number of cancers the reader correctly identified versus the overall number of cancers.
- ROC curve is defined as a graph of test sensitivity (y coordinate) against the false positive rate (x coordi-

nate) [24], using the confidence grading as described above;
- AFROC curve is defined as a graph of the lesion localization fraction (LLF) along the y axis versus the non-lesion localisation fraction (NLF) along the x axis, where the denominators for the fractions are the total number of lesions and the total number of images, collectively [25].

Statistical tests including Student T test or Mann-Whitney test, analysed the impact of participants' age, gender, subjects undertaken at school, years qualified, type of qualification, work pattern, experience, other qualifications, extra-curricular activities, vision, and sleep pattern on the above five statistical metrics outlined above. A number of additional parameters and the resultant impact were also analysed. The factors included: the effect of the length of time the reader prefers to interpret mammograms, if regular breaks were taken during reads, the volume of reads, if readers felt that their image interpretation was impacted by emotional mindset, acoustic environment, or volume of interruptions. In terms of emotional mindset, the observer was simply asked, "did you feel that your emotional well-being on the day of reading could impact on your reading behaviour?". The type of service the reader interprets for, the pathology they find most challenging to interpret and if they rely on prior images/clinical history were also considered.

Descriptive data analyses yielded mean, median, standard deviation and interquartile range metrics. Since data normalisation tests demonstrated that sometimes data was normally distributed and at other times not, a parametric Student T test or a non-parametric Mann-Whitney test was used to examine differences between groups. This is shown in Tables 1 and 2. Power calculations demonstrated that the sample size recorded here could detect inter-group differences in for example AUC scores of 0.06 at 80% power and an alpha value of $p = 0.05$.

## Results

Self-reported demographic details such as age, number of years of qualified in image interpretation, the estimated volume of screening mammograms read per year and the time spent reading mammograms per week are demonstrated in Table 1. The mean age of the readers was 47.5 years (Range:30-69) and the mean number of years qualified in breast image interpretation was 22 years qualified (Range: 1-40 years). 59% of readers read greater than 5000 mammograms a year and all readers read a minimum of 50 mammograms a week.

### *Sensitivity*

The results of statistical analyses are presented in Table 1 and Table 2 with significant findings highlighted. In terms of sensitivity and lesion sensitivity, no significant findings were noted.

### *Specificity*

Lower specificity values ($p = 0.034$) were identified in the RAPs who stated that their emotional mind-set impacted their reading (88), compared to those who did not (66.5) (Table 2).

### *Receiver Operator Characteristic (ROC) values*

Statistically higher values ($p = 0.004$) were shown in those individuals who had less than (0.93), compared to readers with greater than 10 years of experience (0.84). Higher values ($p = 0.036$) were also seen in readers who: read greater than (0.93), compared to those who read less than 100 cases per week (0.87); thought that emotional mind-set impacted their image interpretation (0.91) compared to those who did not (0.84) ($p = 0.027$); consistently relied on (0.94), compared to those who occasionally relied on prior projections (0.89) ($p = 0.0231$).

### *Alternative free-response receiver operating characteristic (AFROC) values*

The AFROC analysis demonstrated that readers who undertook interpretation for the NHSBSP (0.87) compared to those who did not (0.80) exhibit significantly higher values ($p = 0.0486$).

## Discussion

In the UK, RAPs play a critical role in reporting mammograms. Variability in radiologist performance is well documented and utilised amongst all readers, but similar data on RAPs specifically are not currently available. This study's aim is to start a conversation around better understanding potential factors that may have an impact on RAP performance.

A predictor for improved diagnostic performance is the number of mammograms read in a year [6].

Guidance from the NHSBSP state that all readers interpreting mammograms are required to report a minimum of 5000 mammograms a year [26]. Reed (2010) demonstrates that readers who read between 2000 and 4999 mammograms a year had significantly higher performance than those who read less than 1000 [28], in keeping with this study's findings. We identified ROC values were higher in readers who read greater than 100 mammograms a week (equivalent to 5000 reads per year) compared to those who did not, supporting that optimum performance is achieved through reading over a certain threshold. This reading volume/performance relationship was further supported by the finding [6] that participants who undertook interpretation for the NHSBSP (and by default met the minimum 5000 threshold) achieved higher AFROC values compared to readers who did not read for the breast screening service and were likely to undergo less involvement with mammography.

Whilst reader volume is generally associated with improved performance against radiologists [26] and now RAPs, the impact of years of experience is less clear: one recent study from

Table 1
Clinical Background: Sensitivity, Specificity and Lesion Sensitivity is shown in the table with standard deviation or *interquartile values* in brackets.

| Parameter | Sensitivity | Specificity | Lesion Sensitivity |
|---|---|---|---|
| Reads for a screening programme (n=10) | 89.6 (7.8) | 89.0 (*76.2-93*) | 85 (*85-100*) |
| Does not read for a screening programme (n=5) | 89.2 (8.0) | 82.5 (*45.5-85.6*) | 85 (*80-100*) |
| Radiographer (n=11) | 88.5 (7.7) | 88 (*77.5-95*) | 87.7 (9.0) |
| Other Professional title (n=4) | 92.0 (7.6) | 80 (*41.5-88*) | 88.0 (5.7) |
| <10 years' experience (n=12) | 90.4 (8.3) | 84.2 (11.3) | 89.2 (8.9) |
| >10 years' experience (n=3) | 87.0 (5.7) | 64.4 (33.7) | 84.0 (4.2) |
| Specialised in breast disease (n=15) | 88.8 (*85-100*) | 85 (*77.6-95*) | 85.0 (*82.5-100*) |
| Not specialised in breast disease (n=0) | 100 (*100-100*) | 55 (*55-55*) | 100 (*100-100*) |
| <10 years working in this speciality (n=5) | 89.0 (10.3) | 85 (*77.5-93*) | 95 (80-100) |
| >10 years working in this speciality (n=10) | 89.6 (6.9) | 88 (*66.5-95*) | 85 (85-100) |
| <100 breast cases read in a week (n=7) | 89.4 (7.7) | 85 (*56.5-93*) | 87.8(7.9) |
| >100 breast cases read in a week (n=8) | 89.4 (8.1) | 88 (*77.5-95*) | 87.8(8.7) |
| <100 mammograms read in a week (n=10) | 89.2 (6.7) | 85 (*69-93*) | 88 (6.9) |
| >100 mammograms read in a week (n=5) | 90.0 (10.6) | 88 (*75-95*) | 87 (11.5) |
| >5000 mammograms read in 2019 (n=6) | 88.7 (7.9) | 85.5 (*76.3-95*) | 85.0 (*85-95*) |
| <5000 mammograms read in 2019 (n=9) | 90.0 (7.8) | 85.0 (*70.7-93*) | 87.5 (*80-100*) |
| DBT read weekly (n=4) | 87.0 (12.5) | 65.8 (35.2) | 85 (*75-95*) |
| DBT not read weekly (n=11) | 89.6 (7.2) | 83.7 (10.8) | 85 (*85-100*) |
| Morning reading session (n=13) | 87.5 (*85-95*) | 86.5 (*80-95*) | 85.0 (*85-100*) |
| Other time of the day (n=2) | 95.0 (*90-100*) | 66.5 (*58-75*) | 90.0 (*80-100*) |
| Regular intervals (n=14) | 90.0 (*85-98.8*) | 85.0 (*75-95*) | 85.0 (*81.3-100*) |
| Non regular intervals (n=1) | 87.5 (*85-90*) | 85.5 (*83-88*) | 87.5 (*85.0-90*) |
| Emotional mindset impacts reading (n=12) | 88.9 (7.6) | 88.0 (*80-95*)* | 88.2 (8.4) |
| Emotional mindset does not impact reading (n=3) | 91.2 (8.5) | 66.5 (*20.5-85*) | 86.3 (7.5) |
| Minimal noise (n=10) | 90.0 (7.3) | 85.0 (*78.5-93*) | 85 (*85.0-100*) |
| Background noise (n=5) | 88.8 (10.3) | 81.5 (*62.3-95*) | 85 (*72.5-95*) |
| Interruption impacts reading (n=11) | 88.6 (7.2) | 88 (*80-95*) | 87.5 (7.8) |
| Interruption does not impact reading (n=4) | 92.5 (9.6) | 75 (*62.3-85*) | 88.7 (10.3) |
| Reading session length <2Hours (n=8) | 87.5 (7.9) | 85 (*78.8-95*) | 87.5 (*80-100*) |
| Reading session length >2 Hours (n=7) | 91.8 (7.0) | 84 (*62.3-93*) | 85.0 (*85-100*) |
| Reads for symptomatic or screening (n=5) | 89.5 (6.9) | 85 (*75-95*) | 86.4 (9.4) |
| Reads for symptomatic and screening (n=10) | 89.3 (9.3) | 83 (*58-93*) | 88.6 (7.4) |
| Distortion most challenging to detect (n=11) | 88.9 (7.2) | 86.5 (*70.8-95*) | 87.1 (8.0) |
| Mass/MCC/ASD most challenging to detect (n=4) | 91.3 (10.3) | 82.5 (*76.3-93*) | 90.0 (9.1) |
| Reliance on prior images sometimes (n=9) | 87.5 (*85-98.8*) | 84.0 (*62.3-93*) | 89.2 (7) |
| Reliance on prior images all the time (n=6) | 90.0 (*78.8-96.3)* | 86.5 (*78.8-95*) | 85.0 (10) |
| Access to prior images (n=2) | 97.5 (3.6) | 77.5 (*75.0-80*) | 92.5 (*85.0-100*) |
| No access to prior images (n=13) | 88.4 (7.5) | 86.5 (*76.3-95*) | 85.0 (*81.3-100*) |
| Reliance on clinical history occasionally (n=14) | 89.1 (8.0) | 85.0 (76.3-95) | 85.0 (*81.3-100*) |
| Reliance on clinical history all the time (n=1) | 92.5 (3.6) | 50.5 (80.0-93) | 87.5 (*85.0-90*) |

(Asterisks indicates where a significant difference is shown, with a *= $p < 0.05$; ** $p < 0.01$; $p \leq 0.001$) Numbers of participants used for each analysis are shown by (n=XXX).

North America demonstrated that the cancer detection rate (CDR) for RAPs and radiologists was no different between those with more or less than ten years' experience [21] and an Australian study focusing on radiologist's CDR performance confirmed that years of mammographic reading had no impact on performance [27]. Conversely, in another paper from Australia, sensitivity was shown to improve with the experience of the reader [28]. Our findings showed no significant difference in sensitivity with varying levels of reading experience, although improved ROC values were noted amongst RAPs who had less than 10 years' experience compared with readers with more than 10 years was demonstrated.

This ROC result highlights a rather non-intuitive situation around experience (as defined by number of years) and performance. Further examination of our results, show that this finding is most likely linked to increased recalls. For example, specificity of the RAPs in this study showed that those with more than 10 years' experience had 31% lower specificity values than those with less than 10 years' experience. This has support from previous work where the literature identifies radiologists with greater years' experience demonstrate lower specificity [29-32]: in 2004 Barlow looking at 469512 mammograms interpreted by 124 radiologists analysed found that readers with less than 10 years' experience (1.27) had greater specificity than those with greater than or equal to 10 years' experience (1.0); Elmore (2009) reported higher specificity amongst readers who had less than ten years' (12.8) compared with clinicians with greater than twenty years' experience (7.1) across a total of 205 radiologists in seven US states; in 2006, Tan quantified the extent of reduction in specificity as years of experience increased amongst 1067 readers interpreting 27394 mammograms and showed that with every ten years increase in experience a breast

Table 2
Clinical Background: ROC, JAFROC values are shown in the table with standard deviation or interquartile values in brackets.

| Parameter | ROC | JAFROC |
|---|---|---|
| Reads for a screening programme (n=10) | 0.92 (0.04) | 0.87 (0.04) * |
| Does not read for a screening programme (n=5) | 0.88 (0.08) | 0.80 (0.09) |
| Clinical setting (n=11) | 0.92 (0.87-0.95) * | 0.86 (0.03) |
| Another test location (n=4) | 0.95 (0.84-0.96) | 0.85 (0.04) |
| <10 years' experience (n=12) | 0.93 (0.90-0.95) ** | 0.87 (0.04) |
| >10 years' experience (n=3) | 0.84 (0.76-0.89) | 0.85 (0.03) |
| Specialised in breast disease (n=15) | 0.92 (0.87-0.95) | 0.87 (0.85-0.92) |
| Not specialised in breast disease (n=0) | 0.85 (0.85-0.85) | 0.82 (0.82-0.82) |
| <10 years working in the speciality (n=5) | 0.91 (0.04) | 0.87 (0.04) |
| >10 years working in the speciality (n=10) | 0.89 (0.07) | 0.86 (0.03) |
| <100 breast cases in a week (n=7) | 0.87 (0.07) ** | 0.85 (0.03) |
| >100 breast cases in a week (n=8) | 0.93 (0.04) | 0.87 (0.04) |
| <100 mammograms read in a week (n=10) | 0.88 (0.07) | 0.85 (0.03) |
| >100 mammograms read in a week (n=5) | 0.94 (0.02) | 0.88 (0.02) |
| >5000 mammograms read in 2019 (n=6) | 0.89 (0.08) | 0.87 (0.03) |
| <5000 mammograms read in 2019 (n=9) | 0.90 (0.04) | 0.86 (0.04) |
| DBT read weekly (n=4) | 0.86 (0.09) | 0.87 (0.02) |
| DBT not read weekly (n=11) | 0.91 (0.04) | 0.86 (0.04) |
| Morning reading session (n=13) | 0.92 (0.87-0.96) | 0.86 (0.83-0.92) |
| Other time of the day (n=2) | 0.88 (0.80-0.95) | 0.88 (0.88-0.92) |
| Regular intervals (n=14) | 0.92 (0.87-0.96) | 0.86 (0.85-0.92) |
| Non regular intervals (n=1) | 0.90 (0.84-0.95) | 0.84 (0.81-0.88) |
| Emotional mindset impacts reading (n=12) | 0.91 (0.04) ** | 0.86 (0.04) |
| Emotional mindset does not impact reading (n=3) | 0.84 (0.09) | 0.85 (0.02) |
| Minimal noise (n=10) | 0.91 (0.86-0.95) | 0.85 (0.82-0.92 |
| Background noise (n=5) | 0.95 (0.84-0.96) | 0.88 (0.86-0.88) |
| Interruption impacts reading (n=11) | 0.92 (0.87-0.95) | 0.86 (0.03) |
| Interruption does not impact reading (n=4) | 0.91 (0.82-0.95) | 0.87 (0.04) |
| Reading session length <2Hours (n=8) | 0.898 (0.07) | 0.862 (0.04) |
| Reading session length >2 Hours (n=7) | 0.897 (0.04) | 0.861 (0.03) |
| Reads for symptomatic or screening (n=5) | 0.92 (0.04) | 0.87 (0.04) |
| Reads for symptomatic and screening (n=10) | 0.88 (0.07) | 0.86 (0.03) |
| Distortion most challenging to detect (n=11) | 0.91 (0.85-0.95) | 0.86 (0.034) |
| Mass/MCC/ASD most challenging to detect (n=4) | 0.92 (0.89-0.95) | 0.85 (0.032) |
| Reliance on prior Images sometimes (n=9) | 0.89 (0.85-0.95) * | 0.86 (0.033) |
| Reliance on prior Images all the time (n=6) | 0.94 (0.92-0.96) | 0.87 (0.034) |
| Access to prior images (n=2) | 0.94 (0.93-0.95) | 0.89 (0.86-0.92) |
| No access to prior images (n=13) | 0.93 (0.86-0.96) | 0.86 (0.82-0.92) |
| Reliance on clinical history occasionally (n=14) | 0.92 (0.87-0.96) | 0.86 (0.82-0.92) |
| Reliance on clinical history all the time (n=1) | 0.83 (0.72-0.93) | 0.85 (0.85-0.87) |

(Asterisks indicates where a significant difference shown, with a. *=$p < 0.05$; **$p < 0.01$; $p \leq 0.001$) Numbers of participants used for each analysis are shown by (n=XXX).

radiologist demonstrated an up to 12 % reduction in specificity [30–34].

Interestingly, findings around decreasing performance and increased years of experience are not unique to medical imaging: within the police force in the US a study that compared performance levels of newly recruited police officers with senior supervisors found that trying to predict job performance based on years of experience held no merit [35]; also, within the finance sector in Belgium, Chief Executive Officers (CEOs) demonstrated reduced productivity with increasing years [36].

But the question remains why this inverse performance/years of experience relationship might be the case, particularly regarding specificity levels. Confidence may be a factor. In medicine, experienced General Practitioners have demonstrated that confidence and decisiveness relating to definitive diagnoses decline with increasing experience [37]. This relationship however does not appear to be a linear one with the practitioners showing initial increasing confidence with time, however as years of experience increased, confidence levels reduce [38]. Previous workers have suggested a number of reasons including higher expectations of a negative outcome [39], litigation implications as a result of an error in judgement, anxiety to be seen to miss cancers amongst peers [40] as well as low motivation and interest in the task at hand [41].

Results demonstrate that having opinions around reader emotional mindset also result in lower specificity as well as ROC values. This has not been shown before in medical imaging, but in other domains emotion associated with prior experiences can impact decision making as well as interfering with, distracting from and interrupting tasks [42]. For example, stress is well-proven to have a negative impact on performance for nursing students, mathematicians, construction workers and bankers [43–45]. In radiology, we also know that previous psychological traumas can cause uncertainty in decision making [46] and a subsequent intolerance of risk averse decisions [47] in an effort to maximise sensitivity. This highlights the importance of gentle strategies such as REALM [48,49] or anonymised peer review to improve performance rather than harsher auditing measures. However, it should be acknowledged that the current results focussed on the opinion of whether emotional mindset played a part in determining performance levels not that it actually played a part.

## Limitations/recommendations

The sample size of 15 participants included in this study was less than ideal, however previous work with meaningful results has been published with participant numbers being 5 or lower [22]. Nonetheless, the authors fully recognise that the findings shown here are preliminary observations and simply highlight where in the future research could be directed with larger numbers of participants. The author has since developed a Special Interest Group affiliated with the Society and College of Radiographers which should aid and improve recruitment in the future. It is possible that the questionnaire presentation may have limited data exploration. For future recruitment, questionnaire redevelopment will be implemented and a wider range of activities and interests included.

## Conclusion

This paper, for the first time, identified possible factors that may impact upon the efficacy of RAP image interpretation. In particular, it was of interest to note that numbers of reads per year, RAPs' understanding of the importance of emotional agents and access to prior images may be important features that impact significantly on performance. Whilst these early data should facilitate future research so that RAP activities can be optimised, arguably and more importantly, there is now the identified need for a programme of work specific to radiography reporting rather than relying on data from radiologists. Future data from subsequent studies can be used as a benchmark to facilitate best practice in the diagnosis of breast cancer by RAPs.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jmir.2025.101889.

## References


[1] Public Health England. NHS breast screening: helping you decide [Internet]. GOV.UK. 2021. Available from: https://www.gov.uk/government/publications/breast-screening-helping-women-decide/nhs-breast-screening-helping-you-decide.

[2] Cancer Research UK (2018). *Breast cancer statistics*. [online] Cancer Research UK. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer.

[3] National Institute for Health and Care Excellence. *Familial Breast Cancer: Classification, Care and Managing Breast Cancer and Related Risks in People with a Family History of Breast Cancer*. London: National Institute for Health and Care Excellence (NICE); 2019 2019 Nov. (NICE Clinical Guidelines, No. 164.) Available from:. https://www.ncbi.nlm.nih.gov/books/NBK552606/.

[4] Gartlehner G, Thaler K, Chapman A, et al. Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk. *Cochrane Database Syst Rev*. 2013(4):CD009632.

[5] Kundel HL, Nodine CF, Carmody D. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Invest Radiol*. 1978 May 1;13(3):175–181.

[6] Rawashdeh MA, Lee WB, Bourne RM, et al. Markers of good performance in mammography depend on number of annual readings. *Radiology*. 2013 Oct;269(1):61–67.

[7] Culpan AM. Radiographer involvement in mammography image interpretation: a survey of United Kingdom practice. *Radiography*. 2016 Nov 1;22(4):306–312.

[8] Nair A, Gartland N, Barton B, et al. Comparing the performance of trained radiographers against experienced radiologists in the UK lung cancer screening (UKLS) trial. *Br J Radiol*. 2016 Oct;89(1066):20160301.

[9] Woznitza N, Piper K, Burke S, et al. Chest X-ray interpretation by radiographers is not inferior to radiologists: a multireader, multicase comparison using JAFROC (Jack-knife alternative free-response receiver operating characteristics) analysis. *Acad Radiol*. 2018 Dec 1;25(12):1556–1563.

[10] Mall S, Brennan PC, Mello-Thoms C. Modelling visual search behaviour of breast radiologists using a deep convolution neural network. *J Med Imaging*. 2018 Jul 1;5(3):035502.

[11] Geertse TD, Paap E, van der Waal D, et al. Utility of supplemental training to improve radiologist performance in breast cancer screening: a literature review. *J Am Coll Radiol*. 2019 Nov 1;16(11):1528–1546.

[12] Trieu PD, Lewis SJ, Li T, et al. Reader characteristics and mammogram features associated with breast imaging reporting scores. *Br J Radiol*. 2020 Oct 1;93(1114):20200363.

[13] Alakhras Maram, Al-Mousa DS, Brennan PJ, et al. *Proc. SPIE 11513, 15th International Workshop on Breast Imaging (IWBI2020)*. Impact of digital breast tomosynthesis on readers with different experience; 22 May 2020.

[14] Hoff SR, Myklebust TÅ, Lee CI, et al. Influence of mammography volume on radiologists' performance: results from BreastScreen Norway. *Radiology*. 2019 Aug;292(2):289–296.

[15] Gandomkar Z, Ekpo EU, Chen Z, et al. Expert radiologist performance does not appear to impact upon their capability in perceiving the gist of the abnormal on mammograms. *Proc. SPIE.* 15th International Workshop on Breast Imaging (IWBI2020*)*; 2020 May 22:156–162.

[16] Ganeshan D, Rosenkrantz AB, Bassett RL, et al. Burnout in Academic Radiologists in the United States. *Acad Radiol*. 2020 Sep;27(9):1274–1281.

[17] Alshabibi AS, Suleiman ME, Tapia KA, et al. Impact of time of day on radiology image interpretations. *Clin Radiol*. 2020 Oct;75(10):746–756.

[18] Ganesan A, Alakhras M, Brennan PC, et al. A review of factors influencing radiologists' visual search behaviour. *J Med Imaging Radiat Oncol*. 2018 Sep 10;62(6):747–757.

[19] Tavakoli Taba S, Hossain L, Heard R, et al. Personal and network dynamics in performance of knowledge workers: a study of Australian breast radiologists. *PLoS One*. 2016 Feb 26;11(2):e0150186.

[20] Clerkin N, Ski CF, Brennan PC, et al. Identification of factors associated with diagnostic performance variation in reporting of mammograms: A review. *Radiography*. 2023;29(2):340–346 [online]. doi:10.1016/j.radi.2023.01.004.

[21] Chen Y, James JJ, Michalopoulou E, et al. Performance of Radiologists and Radiographers in Double Reading Mammograms: The UK National Health Service Breast Screening Program. *Radiology*. 2023;306(1):102–109.

[22] The Society and College of Radiographers, 2020. *The role of the radiography workforce in image interpretation, film reading and clinical image reporting*. [online] Available from: <https://www.sor.org/getmedia

/761f4f0d-30e9-4402-ab97-dbf6a23a9f94/image_interpretation_info.pdf>[Accessed 23 May 2023].
[23] Qenam BA, Li T, Tapia K, et al. The roles of clinical audit and test sets in promoting the quality of breast screening: a scoping review. *Clin Radiol*. 2020;75(10) 794.e1-.e6.
[24] Soh BP, Lee WB, McEntee MF, et al. Mammography test sets: Reading location and prior images do not affect group performance. *Clin Radiol*. 2014;69(4):397–402.
[25] Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatr*. 2011;48(4):277–287.
[26] Chakraborty DP. Validation and statistical power comparison of methods for analyzing free-response observer performance studies. *Acad Radiol*. 2008;15(12):1554–1566.
[27] Miglioretti DL, Gard CC, Carney PA, et al. When Radiologists Perform Best: The Learning Curve in Screening Mammogram Interpretation. *Radiology*. 2009 Dec;253(3):632–640.
[28] Reed W, Lee W, Cawson JN, et al. Malignancy Detection in Digital Mammograms: Important Reader Characteristics and Required Case Numbers. *Acad Radiol*. 2010 Nov 1;17(11):1409–1413.
[29] Wong DJ, Gandomkar Z, Lewis S, et al. Do reader characteristics affect diagnostic efficacy in screening mammography? A systematic review. *Clin Breast Cancer*. 2023 Jan;23(2):e56–e67.
[30] Tan A, Freeman DH, Goodwin JS, et al. Variation in false-positive rates of mammography reading among 1067 radiologists: a population-based assessment. *Breast Cancer Res Treat*. 2006 Jul 4;100(3):309–318.
[31] Barlow WE, Chi C, Carney PA, et al. Accuracy of Screening Mammography Interpretation by Characteristics of Radiologists. *JNCI*. 2004 Dec 14;96(24):1840–1850.
[32] Elmore JG, Jackson SL, Abraham L, et al. Variability in Interpretive Performance at Screening Mammography and Radiologists' Characteristics Associated with Accuracy. *Radiology*. 2009 Dec;253(3):641–651.
[33] Miglioretti DL, Smith-Bindman R, Abraham L, et al. Radiologist Characteristics Associated With Interpretive Performance of Diagnostic Mammography. *JNCI*. 2007 Dec 19;99(24):1854–1863.
[34] Ericsson KA. The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance. In: The Cambridge Handbook of Expertise and Expert Performance [Internet]. *Psychology;*. 2006:683–704. Available from:. http://www.skillteam.se/wp-content/uploads/2011/12/Ericsson_delib_pract.pdf.
[35] Jacobs R, Hofmann DA, Kriska SD. Performance and Seniority. *Hum Perform*. 1990 Jun;3(2):107–121.
[36] Lallemand T, Rycx F. Are Older Workers Harmful for Firm Productivity? *De Economist [Internet].*. 2009 Sep;157(3):273–292. [cited 2019 Aug 13]Available from:. https://link.springer.com/content/pdf/10.1007%2Fs10645-009-9126-5.pdf.
[37] Serfling MA. CEO age and the riskiness of corporate policies. *J Corporate Finance*. 2014 Apr;25(25):251–273.
[38] Weber EU, Bockenholt U, Hilton DJ, et al. Confidence judgements as expressions of experienced decision conflict. *Risk Decision Policy*. 2000 Apr;5(1):69–100.
[39] Travado L, Grassi L, Gil F, et al. Physician-patient communication among Southern European cancer physicians: The influence of psychosocial orientation and burnout. *Psychooncology*. 2005;14(8):661–670.
[40] Lawthers AG, Localio AR, Laird NM, et al. Physicians' Perceptions of the Risk of Being Sued. *J Health Polit Policy Law*. 1992;17(3):463–482.
[41] Wayne SJ, Liden RC, Kraimer ML, et al. The role of human capital, motivation and supervisor sponsorship in predicting career success. *J Organ Behav*. 1999 Sep;20(5):577–595.
[42] Dolcos F, Wang L, Mather M. Current research and emerging directions in emotion-cognition interactions. *Front Integr Neurosci*. 2014 Nov 11;8(11:8).
[43] Beilock SL. Math Performance in Stressful Situations. *Curr Dir Psychol Sci*. 2008 Oct;17(5):339–343.
[44] Powl A, Skitmore M. Factors hindering the performance of construction project managers. *Constr Innov*. 2005 Mar;5(1):41–51.
[45] Siahaan E. Antecedents of employee performance and the influence on employee job satisfaction in banking service sector in Indonesia. *Banks Bank Syst*. 2017 Nov 30;12(4):75–89.
[46] Oducado RMF, Estoque H. Online learning in nursing education during the COVID-19 pandemic: stress, satisfaction, and academic performance. *J Nurs Pract*. 2021 Apr 1;4(2):143–153.
[47] Starcke K, Brand M. Decision making under stress: a selective review. *Neurosci Bio Behav Rev*. 2012 Apr;36(4):1228–1248. Available from:. https://www.sciencedirect.com/science/article/pii/S0149763412000218.
[48] Fergus TA, Wu KD. Searching for specificity between cognitive vulnerabilities and mood and anxiety symptoms. *J Psychopathol Behav Assess*. 2011 Jul 24;33(4):446–458.
[49] Standards for radiology events and learning meetings, The Royal College of Radiologists [Internet]. www.rcr.ac.uk. Available from: https://www.rcr.ac.uk/publication/standards-radiology-events-and-learning-meetings.