European Journal of Radiology 183 (2025) 111938

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

# Breast imaging readers' performance in the PERFORMS test-set based assessment scheme within the MyPeBS international randomised study

Eleni Michalopoulou [a,1,*], Iain Darker [a,1], Valentina Iotti [b,1], Efrat Slonim [c,1], Harry J. de Koning [d,1], Rodrigo Alcantara Souza [e,1], Jean-Benoit Burrion [f], Sandrine De Montgolfier [g,1], Cécile Vissac-Sabatier [h,1], Michal Guindy [c,1], Pierpaolo Pattacini [b,1], Suzette Delaloge [i,1], Fiona J. Gilbert [j,1], Yan Chen [a,1], the MyPeBS Consortium

[a] Translational Medical Sciences, School of Medicine, University of Nottingham, City Hospital Campus, Hucknall Road, Nottingham NG5 1PB, United Kingdom
[b] Department of Diagnostic Imaging, Azienda Unità Sanitaria Locale - IRCCS di Reggio Emilia, Reggio Emilia, Italy
[c] Assuta Medical Centres, HaBarzel 20 St, Ramat Hahayal, Tel Aviv, Israel
[d] Department of Public Health, Erasmus University Medical Centre, Rotterdam, the Netherlands
[e] Department of Radiology, Hospital del Mar Medical Research Institute, Barcelona, Spain
[f] Institut Jules Bordet, Université Libre de Bruxelles, Rue Meylemeersch 1070, Brussels, Belgium
[g] Aix Marseille University, Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, ISSPAM, Marseille, France
[h] Unicancer, 101 rue de Tolbiac, 75013 Paris, France
[i] Department of Cancer Medicine, Institut Gustave-Roussy, Villejuif, France
[j] University of Cambridge, Department of Radiology, Clinical School, Cambridge Biomedical Research Centre, Cambridge CB2 0QQ, United Kingdom

ARTICLE INFO

*Keywords:*
Mammography
Quality assurance
Performance
Sensitivity and specificity
Cancer detection

ABSTRACT

*Purpose:* A survey conducted by the European Society of Breast Imaging (EUSOBI) in 2023 revealed significant variations in Quality Assurance (QA) practices across Europe. The UK encourages regular performance monitoring for screen readers. This study aimed to assess the variability in diagnostic performance among readers participating in a wider prospective randomised trial across multiple countries.
*Method:* In this retrospective multinational study, breast imaging readers from the MyPeBS clinical trial examined a test set of 40 challenging breast screening cases using the PERFORMS software, from March 2021 to February 2022. The challenging set, enriched with biopsy-proven cancers, aimed to differentiate readers by their level of diagnostic performance. Cancer detection and correct return to screen rates were calculated for each participant.
*Results:* A total of 110 readers from 6 countries completed the PERFORMS test set, while 88 also completed an accompanying questionnaire collecting information about their breast screening work and experience. The study revealed variability in cancer detection rates (M = 73.6 %, SD = 19.7 %, range 0.0 %–100.0 %) and correct return to screen rates (M = 79.7 %, SD = 10.5 %, range 46.4 %–100.0 %). Outliers with extremely low cancer detection (2.7 % of participants) and correct return to screen rates (1.8 % of participants) were also identified.
*Conclusions:* Breast imaging readers' performance in test set-based assessments like PERFORMS can reflect real-world screening proficiency. The presence of outlier readers with low diagnostic performance on the test highlights the need for double reading and for standardised QA protocols to ensure patient safety and service efficiency.

*Abbreviations:* EUSOBI, European Society of Breast Imaging; QA, Quality Assurance; MyPeBS, My Personal Breast Screening; PERFORMS, Personal Performance in Mammographic Screening; UK, United Kingdom.

* Corresponding author at: Translational Medical Sciences, School of Medicine, University of Nottingham, Clinical Science Building, City Hospital Campus, Hucknall Road, Nottingham NG5 1PB, United Kingdom.
*E-mail address:* Eleni.Michalopoulou@nottingham.ac.uk (E. Michalopoulou).

[1] All authors have equally contributed to the manuscript.

https://doi.org/10.1016/j.ejrad.2025.111938
Received 31 October 2024; Received in revised form 16 December 2024; Accepted 16 January 2025
Available online 19 January 2025

## 1. Introduction

Breast cancer is the most common cancer among European women. In 2020, there were an estimated 355,500 new cases diagnosed and 91,826 deaths across Europe, among a population of approximately 229 million women [1]. Regular mammography screening is the most commonly used method for detecting breast cancer at an early stage, in order to reduce mortality and reduce the need for more aggressive therapies [2]. Consequently, evaluating the diagnostic performance of mammography becomes paramount for healthcare professionals in ensuring accurate breast cancer detection.

The diagnostic performance of mammographic breast cancer screening can be summarised as the ability to detect a cancer in a mammogram when cancer is present (sensitivity), whilst avoiding false alarms (specificity). Within a breast screening programme, ideal diagnostic performance would be characterised by the detection of every cancer within the population screened, without incurring unnecessary recalls and clinical interventions [3]. In practice some cancers are too small to be detected on the mammogram, may be occult or may be missed with negative consequences for the patient resulting in delayed diagnosis and poor prognosis, whilst others without cancer are subjected to unnecessary follow-up investigations [4]. Thus, monitoring sensitivity and specificity within breast cancer screening programmes and among individual readers is recommended. Ideally individuals whose diagnostic performance is at a lower level may benefit from actions to improve their practice (education, further training, awareness of a specific blind spot). This helps achieve higher levels of patient safety and efficiency in a screening service and constitutes an effective measure of quality assurance [5].

In 2023, the European Society of Breast Imaging (EUSOBI) conducted a survey among its members to collect information on radiologists' preferences and practice regarding quality assurance (QA) measures in mammography image interpretation [6]. Only about half of respondents reported that there was some form of quality assurance in their workplace. Regular monitoring of real-life performance of a programme or specific performance testing allows the identification of individuals working in breast cancer screening with poor diagnostic performance and may avoid serious consequences for patient safety and service efficiency.

The 'My Personal Breast Screening' (MyPeBS) clinical study assesses a personalised risk-based screening approach against standard screening [7]. In order to further explore the concerns regarding the lack of performance monitoring and the absence of a harmonised international performance monitoring system, breast imaging readers participating in MyPeBS were invited to participate in a test set-based system called 'Personal Performance in Mammographic Screening' (PERFORMS). Participation in PERFORMS requires breast imaging readers to examine test sets of challenging breast screening cases on their mammographic workstation and record their decisions regarding overall breast classification and mammographic features on each case using the PERFORMS online reporting application software. The test sets include difficult normal, benign, and malignant cases, featuring subtle or interval cancers, bilateral appearances, multi-focal tumours, and more. These challenging cases aim to furnish participants with insights into their reading performance, whilst differentiating readers according to their level of diagnostic performance. Details of the PERFORMS scheme and its evaluation are described elsewhere [8].

The aim of this study was to investigate the mammographic interpretation skills of readers drawn from many countries and evaluate the variability in their diagnostic performance on a deliberately challenging mammography test-set.

## 2. Material and methods

This retrospective study was conducted within the framework of the Radiology Quality Control task of the MyPeBS clinical study (International Randomized Study Comparing personalized, Risk-Stratified to Standard Breast Cancer Screening in Women Aged 40–70, NCT03672331, sponsored by Unicancer). The study was carried out in accordance with local security and data protection policies. The requirement for ethical approval was waived after discussion with the organisational research and development team because this study was deemed to represent an audit of current practice (Ethics Reference No: 88-1223).

### 2.1. Study design

Breast imaging readers involved in the MyPeBS clinical trial from 6 countries (Belgium, France, Israel, Italy, Spain, and UK) were invited to participate in the study between March 2021 and February 2022. Participation required them to examine a set of anonymised 2D mammograms, which consisted of recent examples from routine breast screening in the United Kingdom (UK).

The test-set images were uploaded to mammography workstations at the hospitals where participants were based, allowing them to view the images. Participants recorded their findings on a password-protected website and rated each breast using the BI-RADS five-point scale: 1 for normal, 2 for benign, 3 for indeterminate, 4 for suspicious, and 5 for highly suspicious. A rating of 1 or 2 indicated a decision not to recall for further investigation separately for each breast, while a rating of 3 or higher indicated a decision to recall the breast for further investigation. (A clinical decision to recall or not is made at breast level for the purposes of the PERFORMS test set.)

Upon submitting their answers, participants received immediate feedback on each case, which included a comparison to pathology and the radiological opinion derived from a panel of expert radiologists. They also received a performance report containing comprehensive statistics on their cancer detection rate, correct return to screen rate, positive predictive value, and negative predictive value.

At the end of the test set, participants were also asked to complete an online questionnaire collecting information about their demographics, experience, and breast screening work.

### 2.2. Case set

The case set consisted of 38 challenging breast screening cases, each with a known outcome, along with 2 additional practice cases. These cases, collected from various sources, were anonymised. Each case comprised full-field 2D digital mammograms of the mediolateral oblique (MLO) view and the craniocaudal (CC) view of each breast. No additional clinical information or prior images were provided. The set contained biopsy-proven cancers with various radiological features such as architectural distortion, ill-defined masses, spiculated masses, and calcifications.

A panel of ten expert breast radiologists based in the UK had previously assessed the suitability of each case for training and assessment purposes and provided their radiological opinion on various aspects of each case. Each radiologist on the panel had more than 20 years of radiological experience and interpreted a minimum of 5000 screening or diagnostic mammograms per year, with a minimum of 1500 screening mammograms as a first reader. Where appropriate, each abnormal case had been followed up, and case pathology was available. Cases classified as normal had a three-year normal follow-up as per the UK screening procedure. Specifically, the 38 test cases were analysed as 76 breasts: 18 breasts with at least one biopsy-proven malignancy, which were used to calculate the cancer detection rates; 54 normal breasts and 2 breasts with biopsy-proven benign lesions that were not considered suspicious by the panel of expert radiologists, which were used to calculate the correct return to screen rate; and 2 breasts with biopsy-proven benign lesions that were considered suspicious by the panel of expert radiologists, which were not used in the performance metrics for the present analyses. Although participants were aware that the case set was heavily

enriched with biopsy-proven cancers, they were not informed about the ratio of cancerous to non-cancerous cases.

These cases were selected to present a sufficient level of difficulty, allowing for the differentiation of readers according to their level of diagnostic performance. This approach facilitated the identification of individuals with poor diagnostic performance compared to their peers. Thus, the test provided an estimate of the variability in sensitivity and specificity levels among this multinational sample of breast readers.

### 2.3. Performance metrics

Within the PERFORMS performance framework, two measures were calculated for each participant: cancer detection rate (sensitivity) and correct return to screen rate (specificity), both per breast. The cancer detection rate was calculated as the percentage of breasts with biopsy-proven malignant features that were recalled by the participant (i.e., given a rating of 3 or higher). The correct return to screen rate was calculated as the percentage of non-recallable breasts in the test set that were not recalled by the participants (i.e., given a rating of 1 or 2). All non-recallable breasts either have normal features with a normal, three-year follow-up mammogram, or biopsy-proven benign features deemed not suspicious by the panel of experts. Outlier participants on either measure were defined as those participants whose scores fell more than 1.5 times the interquartile range below the 25th percentile of the scores in the group.

Statistical calculations were performed using the IBM SPSS Statistics (version 28) statistical software (IBM SPSS Statistics for Windows, Version 28.0. Released 2021. Armonk, NY: IBM Corp.).

## 3. Results

A total of 267 breast imaging readers from 6 countries (Belgium, France, Israel, Italy, Spain and UK) were invited to take part in the PERFORMS scheme, of whom 110 (41.2 %) completed the test. The participation rates by country are presented in Table 1. The main reasons for non-participation were lack of time, clinical duties, and other commitments.

### 3.1. Characteristics of study sample

Of the 110 participants who completed the PERFORMS test, 88 also completed an accompanying questionnaire that collected information about their demographics, experience, and breast screening work. Most respondents (86.4 %) were women. The most commonly selected age group was between 41 and 50 years (38.6 %), followed by 51–60 years (34.1 %). Smaller proportions of respondents were aged 31–40 years (19.3 %) or 61–70 years (8 %).

These 88 readers had between 3 and 31 years of experience (M = 12.8, SD = 7.3, Mdn = 12, IQR = 6–16). They read between 1000 and 50,000 cases per year (M = 6979, SD = 6772, Mdn = 5000, IQR = 3000–8000), and read for between 3 and 40 h per week (M = 16.3, SD = 9.5, Mdn = 15, IQR = 8–22).

Approximately 62.5 % of the respondents worked in both screening and symptomatic settings with a further 36.4 % working exclusively in screening, and the remainder, 1.1 %, working exclusively in a symptomatic setting. The majority (69.3 %) performed independent double-reading in practice, with 17.0 % undertaking informed double-reading, and 13.6 % undertaking single reading. Only 12.5 % used CAD or AI in practice.

**Table 1**
Participation rates of invitees in the PERFORMS test by country of practice.

| Country | Invitees | Participants | Participation rate |
|---|---|---|---|
| Belgium | 41 | 24 | 58.5 % |
| France | 133 | 16 | 12.0 % |
| Israel | 13 | 13 | 100.0 % |
| Italy | 39 | 27 | 69.2 % |
| Spain | 6 | 2 | 33.3 % |
| UK | 35 | 28 | 80.0 % |
| **Total** | **267** | **110** | **41.2 %** |

### 3.2. Participants performance overview

Across the 110 participants, a broad range of cancer detection rates was observed, from 0.0 % to 100.0 % (M = 73.6 %, SD = 19.7 %, Mdn = 77.8 %, IQR = 61.1 %-88.9 %). For the correct return to screen rate, the range was from 46.4 % to 100.0 % (M = 79.7 %, SD = 10.5 %, Mdn = 80.4 %, IQR = 73.2 %-87.5 %). Table 2 presents a summary of these performance metrics across participants.

Poor-performing outliers were identified as those with a cancer detection rate 1.5 times the interquartile range below the 25th percentile, equating to a cancer detection rate below 19.4 %. On this basis, 3 of the 110 participants (2.7 %) were classified as statistical outliers in terms of their cancer detection rate (Fig. 1). The respective threshold for the correct return to screen rate was 51.8 %, resulting in 2 of the participants (1.8 %) being statistical outliers (Fig. 2). No participant was classified as an outlier on both measures.

Unfortunately, only one of the outlier participants had completed the accompanying study questionnaire. Consequently, it is not possible to evaluate whether the outlier participants differed from their peers in terms of factors such as experience or case reading volume.

## 4. Discussion

The median level of sensitivity (the cancer detection rate) amongst the study participants on this challenging set of cases was 77.8 % (inter-quartile range 61.1 %–88.9 %), whilst the median level of specificity (correct return to screen rate) was 80.4 % (inter-quartile range 73.2 %–87.5 %). A small percentage of participants had unusually low sensitivity or specificity scores compared to those of their peers: 2.7 % of participants had exceptionally low cancer detection rates, while in terms of specificity, 1.8 % had exceptionally low correct return-to-screen rates. These outliers were not simply those with the lowest performance in the group, as someone will always have the lowest level of performance in any group. Rather, these individuals were identified as outliers because their performance was more than one-and-a-half times the interquartile range below the 25th percentile, a common statistical method for identifying low-end outliers. The presence of outliers with exceptionally low sensitivity or specificity compared to their peers on the same cases is concerning and warrants attention.

Performance on test sets such as PERFORMS has been shown to correlate with performance in live breast cancer screening [8,9]. While unusually low scores do not necessarily indicate poor practice in live screening—potentially arising from challenging circumstances—they should still act as a flag for further investigation to ensure patient safety. Arguably, the consequences of missing a cancer are more serious than a false alarm, making it imperative to address diagnostic deficiencies [10]. Given that quality assurance in breast cancer screening is far from ubiquitous across Europe and beyond, the major concern is that outlier readers like those identified here could be missing a significant number of cancers in their practice.

These findings also suggest that implementing performance assessment schemes for readers participating in clinical studies evaluating novel approaches to breast cancer screening can be very valuable. Such schemes ensure that all participants maintain consistent, high-quality standards, reducing variability and safeguarding the integrity of trial outcomes. In the UK PROSPECTS trial [11], which is assessing the impact of using Digital Breast Tomosynthesis (DBT) in routine breast cancer screening, all trial readers are part of the NHS Breast Screening Programme, which mandates the annual completion of PERFORMS as part of reader appraisals. This helps to ensure image interpretation competency and promotes confidence in trial findings.

**Table 2**
Summary of cancer detection rate and correct return to screen rate across participants.

| Metric | N | Mean | SD | Percentiles | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 0 | 25 | 50 | 75 | 100 |
| Cancer detection rate | 110 | 73.64 % | 19.73 % | 0.00 % | 61.11 % | 77.78 % | 88.89 % | 100.00 % |
| Correct return to screen rate | 110 | 79.66 % | 10.46 % | 46.43 % | 73.21 % | 80.36 % | 87.50 % | 100.00 % |

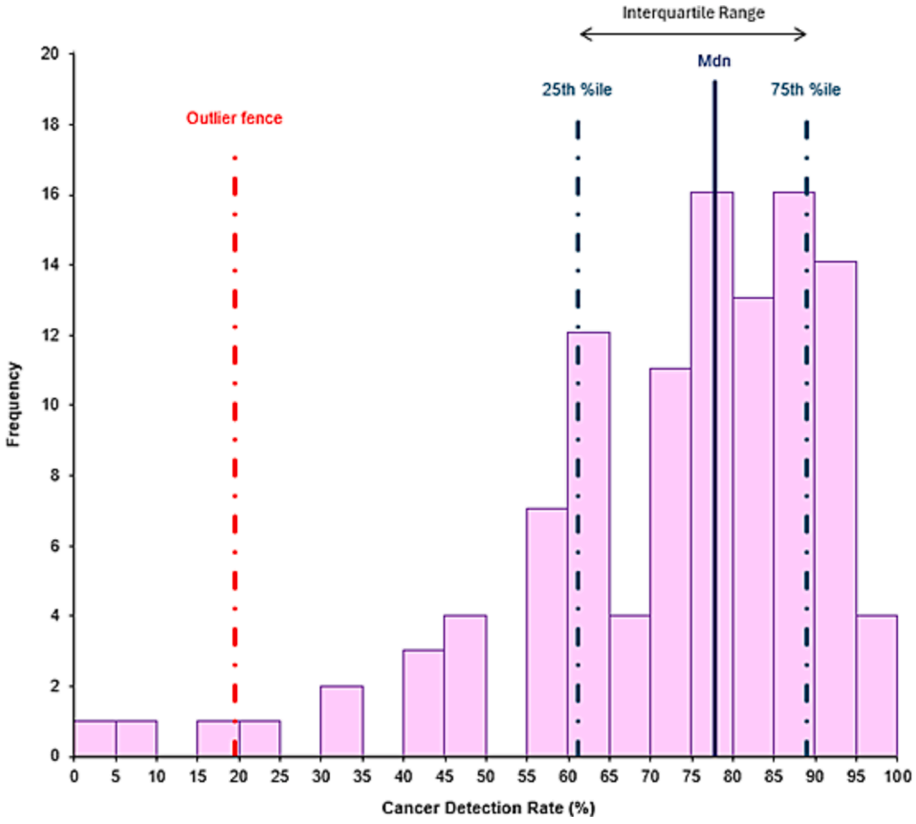


**Fig. 1.** Frequency distribution of cancer detection rates across participants with low performing outliers marked.

Reader performance is also affected by other parameters such as the number of mammograms read annually and years of experience. Previous studies have shown that high-volume readers have better performance outcomes in cancer detection rates, potentially due to their increased exposure to mammograms and subsequent optimised ability to discriminate between normal and malignant cases [12]. There is also a significant association between reading performance and readers' experience in mammography interpretation, with more experienced readers showing higher sensitivity during mammographic reading sessions than less experienced readers [13]. This may be because experienced readers are less susceptible to fatigue-related issues [14] and have a higher threshold for classifying mammograms as abnormal due to their ability to identify and avoid recalling benign lesions they have encountered in the past [15]. In our study, however, only one of the outlier participants provided information on reading volume and years of experience, so we cannot evaluate whether the outlier participants differed from their peers in terms of these parameters.

The UK Breast Screening Programme employs performance testing as a crucial component of its quality assurance framework [16]. As part of this, all NHS breast screening readers are required to participate in the PERFORMS test-set based assessment scheme [17], enabling the rapid evaluation of their diagnostic abilities using test datasets with confirmed clinical outcomes. The cases included in these test sets are deliberately challenging to differentiate readers based on diagnostic ability and hence, individuals with poor diagnostic performance relative to their peers can be identified. Additionally, the test sets contain a higher cancer prevalence than encountered in live screening, exposing readers to diverse radiographic cancer presentations at an accelerated rate. Within the PERFORMS quality assurance programme, individual readers with poor performance at outlier levels are contacted and offered targeted support and training to improve their practice [8]. By ensuring regular participation in performance testing, the programme helps maintain a high standard of image interpretation and enables timely identification and remediation of any performance issues.

This study has some potential limitations. First, only 41 % of those invited went on to complete the PERFORMS test set. There is no information on the characteristics or performance of those who did not participate, so this is an unquantifiable source of participation bias. Second, 20 % of the participants did not respond to the accompanying study questionnaire, preventing investigation into whether their performance relates to factors such as experience or case reading volume (these analyses were undertaken, but were underpowered). Finally, participants were aware that the PERFORMS case set was enriched with biopsy-proven cancers, and as a result their approach in undertaking the scheme could be different to that adopted in routine screening.

## 5. Conclusion

To our knowledge, this is the first study to evaluate the variability in diagnostic performance amongst a sample of breast imaging readers

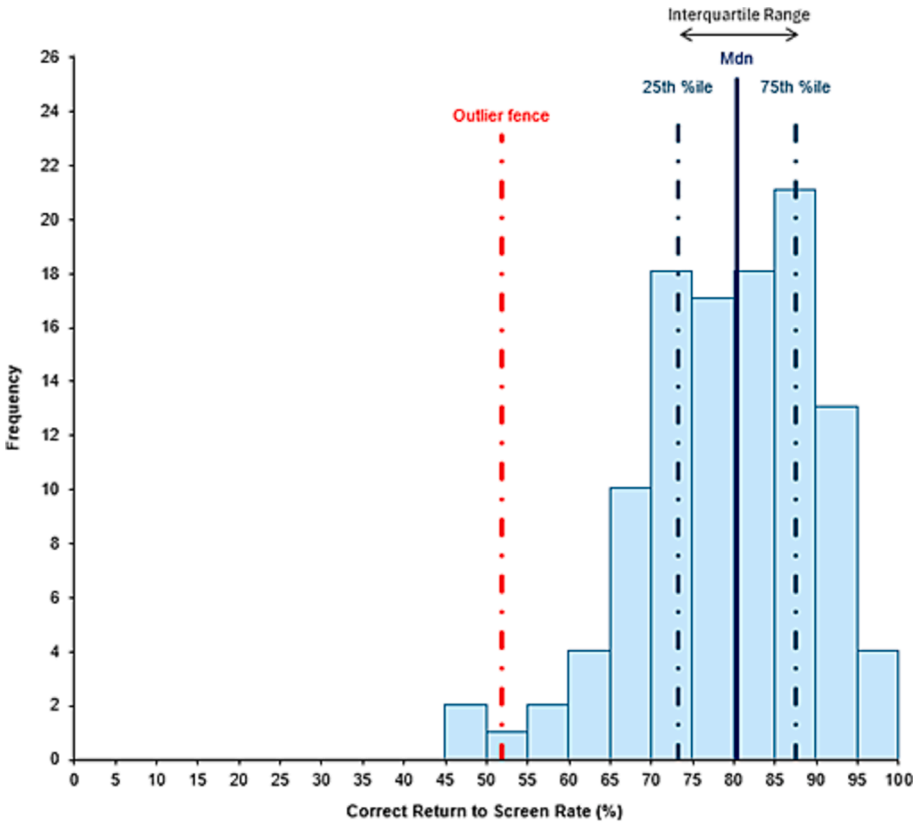


**Fig. 2.** Frequency distribution of correct return to screen rates across participants with low performing outliers marked.

drawn from multiple countries. The results underscore the vital importance of monitoring their diagnostic performance. The presence of outliers with especially low levels of sensitivity or specificity within this small sample highlights both the need for double reading to be maintained (which is recommended by European Guidelines [18] and provided to all participants in the MyPeBS trial) and the need for a broader and more comprehensive implementation of quality assurance in breast screening on an international scale.

## CRediT authorship contribution statement

**Eleni Michalopoulou:** Writing – review & editing, Writing – original draft, Project administration, Formal analysis, Data curation, Conceptualization. **Iain Darker:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Valentina Iotti:** Writing – review & editing, Visualization, Validation, Resources, Conceptualization. **Efrat Slonim:** Writing – review & editing, Visualization, Resources, Methodology, Conceptualization. **Harry J. de Koning:** Writing – review & editing, Visualization, Validation, Resources, Conceptualization. **Rodrigo Alcantara Souza:** Writing – review & editing, Visualization, Validation, Resources, Conceptualization. **Jean-Benoit Burrion:** Writing – review & editing, Validation, Resources, Conceptualization. **Sandrine De Montgolfier:** Writing – review & editing, Visualization, Validation, Resources, Conceptualization. **Cécile Vissac-Sabatier:** Writing – review & editing, Visualization, Validation, Methodology, Conceptualization. **Michal Guindy:** Writing – review & editing, Visualization, Validation, Resources, Conceptualization. **Pierpaolo Pattacini:** Writing – review & editing, Validation, Resources, Conceptualization. **Suzette Delaloge:** Writing – review & editing, Visualization, Project administration, Conceptualization. **Fiona J. Gilbert:** Writing – review & editing, Visualization, Methodology, Conceptualization. **Yan Chen:** Writing – review & editing, Visualization, Supervision, Methodology, Conceptualization.

## Funding

The current study is funded as part of the MyPeBS clinical trial, which is supported by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 755394. The funders had no role in in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the manuscript for publication.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yan Chen declares relevant support outside of this reported work. The PERFORMS scheme is funded by grant money paid to the author's institution by the NHS Breast Screening Programme through the National Health Service England. This financial relationship, however, is not related to the manuscript. All other authors of this work declare no relationships with any companies, whose products or services may be related to the subject matter of the article. The funding source however had no level of involvement in this study.

## Acknowledgements

We are deeply grateful to the radiologists who participated in this study. We would also like to thank the MyPeBS Consortium as follows: Stefan Michiels (Biostatistics and Epidemiology Service, Gustave Roussy, France), Paolo Giorgi Rossi (Epidemiology Unit, Azienda Unità Sanitaria Locale Reggio Emilia, Italy), Marta Roman (Epidemiology and Evaluation Department, Hospital del Mar Medical Research Institute, Spain), D Gareth Evans (Prevent Breast Cancer Centre, 9, Manchester University NHS Foundation Trust, UK), Livia Giordano (Piedmont Reference Centre for Epidemiology and Cancer Prevention, Italy), Paul

Pharoah (Department of Oncology, University of Cambridge, UK), Debbie Keatley (Independent Cancer Patients' Voice, UK), Nicolien Vanravesteyn (Department of Public Health, Erasmus University Medical Centre, The Netherlands) and Camille Baron (Unicancer, Institut Bergonié, France).

## References

[1] European Commission, European Cancer Information System, Breast cancer burden in EU-27, 2020. https://ecis.jrc.ec.europa.eu/factsheets.php (Accessed 27 January 2024).
[2] M. Løberg, M.L. Lousdal, M. Bretthauer, M. Kalager, Benefits and harms of mammography screening, Breast Cancer Res. 17 (2015) 63, https://doi.org/10.1186/s13058-015-0525-z.
[3] C.D. Lehman, R.F. Arao, B.L. Sprague, J.M. Lee, et al., National performance benchmarks for modern screening digital mammography: update from the breast cancer surveillance consortium, Radiology 283 (2016) 49–58, https://doi.org/10.1148/radiol.2016161174.
[4] C. Renzi, K.L. Whitaker, J. Wardle, Over-reassurance and under-support after a 'false alarm': a systematic review of the impact on subsequent cancer symptom attribution and help seeking, BMJ Open 5 (2015) e007002, https://doi.org/10.1136/bmjopen-2014-007002.
[5] N. Perry, M. Broeders, C. de Wolf, S. Törnberg, R. Holland, L. von Karsa L, European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis, fourth ed., European Communities, Belgium, 2006.
[6] E. Michalopoulou, P. Clauser, F.J. Gilbert, R.M. Pijnappel, R.M. Mann, P.A. T. Baltzer, Y. Chen, E.M. Fallenberg, A survey by the European Society of Breast Imaging on radiologists' preferences regarding quality assurance measures of image interpretation in screening and diagnostic mammography, Eur. Radiol. 33 (2023) 8103–8111, https://doi.org/10.1007/s00330-023-09973-7.
[7] MyPeBS Personalising Breast Screening, The Project, 2020. https://www.mypebs.eu/the-project/ (Accessed 27 June 2023).
[8] Y. Chen, J.J. James, E.J. Cornford, J. Jenkins, The Relationship between Mammography Readers' Real-Life Performance and Performance in a Test Set-based Assessment Scheme in a National Breast Screening Program, Radiol. Imaging Cancer. 25;2(5):e200016 (2020), https://doi.org/10.1148/rycan.2020200016.
[9] Y. Chen, J.J. James, E. Michalopoulou, I.T. Darker, J. Jenkins, The relationship between missed breast cancers on mammography in a test-set based assessment scheme and real-life performance in a National Breast Screening Programme, Eur. J. Radiol. 142 (2021) 109881, https://doi.org/10.1016/j.ejrad.2021.109881.
[10] E.L. Aaronson, G.R. Quinn, C.I. Wong, A.M. Murray, C.R. Petty, J. Einbinder, G. D. Schiff, Missed diagnosis of cancer in primary care: Insights from malpractice claims data, J. Healthc. Risk Manag. 39 (2019) 19–29, https://doi.org/10.1002/jhrm.21385.
[11] PROSPECTS Trial, Measuring the cost-effectiveness of DBT and 2DDM compared to standard 2DDM in breast cancer screening, 2022. https://medphys.royalsurrey.nhs.uk/prospects/ (Accessed 11 December 2024).
[12] D.J. Wong, Z. Gandomkar, S. Lewis, W. Reed, M. Suleiman, S. Siviengphanom, E. Ekpo, Do reader characteristics affect diagnostic efficacy in screening mammography? A systematic review, Clin. Breast Cancer. 23 (2023) e57–e65, https://doi.org/10.1016/j.clbc.2023.01.009.
[13] A.S. Alshabibi, M.E. Suleiman, S.M. Albeshan, R. Heard, P.C. Brennan, Variations in breast cancer detection rates during mammogram-reading sessions: does experience have an impact? Br. J. Radiol. 95 (2022) 20210895 https://doi.org/10.1259/bjr.20210895.
[14] N. Clerkin, C.F. Ski, P.C. Brennan, R. Strudwick, Identification of factors associated with diagnostic performance variation in reporting of mammograms: a review, Radiography 29 (2023) 340–346, https://doi.org/10.1016/j.radi.2023.01.004.
[15] D.L. Miglioretti, R. Smith-Bindman, L. Abraham, R.J. Brenner, P.A. Carney, E.J. A. Bowles, D.S.M. Buist, J.G. Elmore, Radiologist characteristics associated with interpretive performance of diagnostic mammography, J. Natl. Cancer Inst. 99 (2007) 1854–1863, https://doi.org/10.1093/jnci/djm238.
[16] A. Gale, Y. Chen, A review of the PERFORMS scheme in breast screening, Br. J. Radiol. 93 (2020) 20190908, https://doi.org/10.1259/bjr.20190908.
[17] GOV.UK, Breast screening: guidance for image reading. https://www.gov.uk/government/publications/breast-screening-guidance-for-image-reading/breast-screening-guidance-for-image-reading, 2024 (Accessed 9 December 2024).
[18] European Commission Initiative on Breast Cancer (ECIBC), European guidelines on breast cancer screening and diagnosis. https://cancer-screening-and-care.jrc.ec.europa.eu/sites/default/files/Guidelines/EtDs/ECIBC_GLs_EtD_double_reading.pdf, 2019 (Accessed 12 July 2024).