European Journal of Radiology 168 (2023) 111117

Contents lists available at ScienceDirect

European Journal of Radiology

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

A comparison of breast screening performance based on a standardised test, during and outside of the COVID-19 lockdown period

Eleni Michalopoulou *, Padma Priya Pugalenthi, Iain Darker, Yan Chen

University of Nottingham, Translational Medical Sciences, School of Medicine, City Hospital Campus, Hucknall Road, Nottingham NG5 1PB, United Kingdom

ARTICLE INFO

Keywords:
Mammography
COVID-19 pandemic
Performance
Decision making

ABSTRACT

Purpose: In the UK, the COVID-19 pandemic and the resulting lockdown significantly impacted routine breast screening and led to reduced case volumes for breast screening readers. The aim of the study was to evaluate whether breast screening performance on a test-set based assessment scheme (PERFORMS) varied over time, in association with the lockdown period.

Method: In this retrospective study, performance data were obtained for all breast cancer screening readers in England who completed the PERFORMS schemes based on digital mammography (from 2015 to 2022). Cancer detection (sensitivity), correct return to screen (specificity) and the difference between the sensitivity and specificity were calculated for each breast screener in each PERFORMS scheme.

Results: During the study period, 4906 readers participated in eight PERFORMS schemes. Both the cancer detection ($H(7) = 775.56$, $p < .00001$) and correct return to screen rate ($H(7) = 401.13$, $p < .00001$) varied significantly by scheme. The difference between cancer detection and correct return to screen rate in the scheme administered during the lockdown period was significantly higher than in all other schemes ($p < .05$, Bonferroni correction applied).

Conclusions: The performance of breast screeners on the PERFORMS test-set based assessment scheme which was administered during the lockdown period, was characterised by markedly high sensitivity and low specificity when compared to their performance on other tests taken pre-COVID and following the lockdown.

1. Introduction

Each year more than 2 million women have breast cancer screening in the UK [1]. The National Health Service Breast Screening Programme (NHSBSP) invites women between the ages of 50 and 71 for a screening mammogram at three yearly intervals and women with abnormal mammograms are recalled for further assessment [2]. However, the COVID-19 national lockdown that was announced in the UK on the 23rd of March 2020 led to the halting of the breast cancer screening services by the end of the month [3]. All 78 NHS breast screening units decided to pause screening for approximately 3 months in order to allow staff to be redeployed to respond to COVID-19, and to protect patients and staff from the virus [4].

During that time, new invitations to the screening program were not sent as services were reconfigured [3]. Screening focused on following up with existing screen positive women and women with high familial risk [3]. The programme resumed sending out invitations for routine screening between April to September 2020 [5]. However, infection control measures, workforce sickness and fewer women presenting for breast cancer screening continued to affect routine screening numbers [5].

Performance in the UK breast screening programme is heavily audited as part of the quality assurance process since the screening programme began in 1988. A pivotal part of this process is the PERFORMS (Personal Performance in Mammographic Screening) test-set based assessment scheme. For over 30 years, all breast screening readers in the NHSBSP are required to participate in this annual, standardised test of breast screening ability based on a carefully collated set of mammograms, each of known outcome. The PERFORMS scheme serves as a quality assurance and training mechanism for the NHSBSP to

Abbreviations: COVID-19, Coronavirus disease 2019; UK, United Kingdom; NHSBSP, National Health Service Breast Screening Programme; PERFORMS, Personal Performance in Mammographic Screening.

* Corresponding author at: The University of Nottingham Clinical Science Building Nottingham City Hospital, Hucknall Road, Nottingham NG5 1PB, United Kingdom.

E-mail address: Eleni.Michalopoulou@nottingham.ac.uk (E. Michalopoulou).

https://doi.org/10.1016/j.ejrad.2023.111117
Received 8 June 2023; Received in revised form 20 September 2023; Accepted 26 September 2023
Available online 28 September 2023

evaluate each breast screener's ability to detect challenging breast screening cases and identify potential problems at the earliest opportunity to allow interventions to change practice [6].

From 2019−2020, 2.12 million women were screened by NHSBSP in England [7]. However, during the pandemic, there was a 44.1% decrease with only 1.19 million women being screened [7]. There is speculation regarding how the reduction in case reading volumes for breast screeners has affected their screening ability, with consequences for the early detection of breast cancer in the women of the screening population. The PERFORMS scheme, which was administered throughout the lockdown period of the COVID-19 pandemic, offers an opportunity to evaluate whether breast screening performance on this standardised test varied over time, in association with the lockdown period. The aim of the study is to evaluate whether performance on the PERFORMS test taken during the COVID-19 lockdown period differed from that on the tests taken at other times.

2. Material and methods

The retrospective 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.

2.1. The PERFORMS scheme

In the UK, where double reading of mammograms is the standard of care, the longstanding shortage of radiologists has meant that non-radiologists have routinely been involved in the reading of screening mammograms since the mid-1990s. Non-radiology readers are either radiographer advanced practitioners or breast clinicians trained specifically in mammographic interpretation. All national breast screening personnel (radiologists, radiographers and breast clinicians) involved in reporting screening mammograms take part in PERFORMS annually by reading test sets of 60 challenging cases in their clinical workstation and report their decisions about each case using the PERFORMS reporting software. These cases are recent examples of mammograms arising from the routine breast screening in the UK. They are double read by two independent readers and any discrepancies between readers are dealt with by either arbitration using an independent third reader or consensus depending on local screening centre protocols.

In order for a case set to be included in the PERFORMS scheme, a panel of ten UK expert breast radiologists (with more than 20 years of experience each) also assesses the suitability of each case for training and assessment purposes and provides their radiological opinion on various aspects of each case. All malignant and benign cases are pathology proven while all normal cases have a normal three year follow up screening outcome.

The test-set images are uploaded to the Picture Archiving and Communication System (PACS) at each screening centre where they can be viewed. Participants' findings are recorded on a password protected website and they receive immediate feedback on each case at the end of the set, compared to pathology and an opinion derived from the panel of expert radiologists. Once PERFORMS is completed by all participants, comprehensive performance statistics are produced providing an individual with a comparison with their peers nationally. Details of the PERFORMS scheme and its evaluation are described elsewhere [8]. Given the nature of PERFORMS, each scheme presented different sets of cases with differing numbers of malignant, benign and normal breasts, and potentially differing levels of difficulty. This allowed PERFORMS to remain up to date with new cases and aided in providing a reliable quality assurance platform.

2.2. Study sample and case sets

All participants (radiologists, radiographers and breast clinicians) were readers in the NHSBSP at the time of completing the schemes with different levels of experience. For this study, data from all those participants (n = 4906) who completed the PERFORMS schemes based on digital mammography (annual test sets from 2015 to 2022) were included in the analysis. Older PERFORMS schemes not based on digital mammography and current ongoing PERFORMS schemes have been excluded from the analysis. All included tests follow the same format, with the same total number of cases and similar ratios of malignant to non-malignant cases per set. The composition of each PERFORMS scheme is detailed in Table 1.

Specifically, the 2020 case set was administered during the lockdown period of the COVID-19 pandemic. The present analysis compares breast screening performance on the 2020 case set test with that on seven other PERFORMS tests (2015–2019 and 2021–2022 – all the other PERFORMS case sets that are based on digital mammography, and for which participation in England was completed at the time of writing). However, the 2020 case set that coincided with the COVID lockdown periods was not selected especially for the lockdown period, indeed it was collated before the lockdowns were known about. These other case sets were administered before and after the COVID-19 pandemic lockdown period. The aim of the analyses was to evaluate whether performance on the test taken during the COVID-19 lockdown period differed from that on the tests taken at other times.

2.3. Performance metrics

Within the PERFORMS performance framework, two measures are calculated for each participant: cancer detection rate (sensitivity) and correct return to screen (specificity). A panel of expert radiologists examine each breast of each case in the PERFORMS case set and determines which breasts should be recalled (analysis of performance is taken at individual breast level rather than at case level). The cancer detection rate is expressed as the percentage of breasts with biopsy-proven malignant features that were recalled by the individual screener. The correct return to screen rate is calculated for each breast screener by the percentage of non-recallable breasts not recalled by the screener. 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.

2.4. Statistical analysis

For the purposes of this analysis where performance across eight different schemes was compared, each scheme was treated as a separate

Table 1
The composition of each PERFORMS test set.

Scheme Year	Recallable breasts			Number of non-recallable breasts	Total number of breasts
	with malignant features	with benign or normal features*	Total		
2015	22	6	28	92	120
2016	32	3	35	85	120
2017	34	7	41	79	120
2018	39	1	40	80	120
2019	41	3	44	76	120
2020	29	4	33	87	120
2021	28	5	33	87	120
2022	16	0	16	104	120
Total	**241**	**29**	**270**	**690**	**960**

* benign or normal features that are suspicious and indistinguishable from a malignant feature on the basis of inspection of the mammogram alone.

pool of participants in a between-participants design. The performance metrics were not normally distributed, and absolute differences between schemes are of little value due to the differing case sets; consequently, the data were analysed using non-parametric statistical techniques. The alpha-level for statistical significance was set at 0.05 for all analyses. Statistical calculations were performed using statistical software (SPSS Statistics version 27.0; IBM).

3. Results

There were 4906 readers (radiologists, radiographers and breast clinicians) who completed the PERFORMS test sets in England and hence included in the study. Summary statistics for cancer detection and correct return to screen measures are given in Table 2.

A Kruskal-Wallis test indicated that both the cancer detection and correct return to screen rate varied significantly by scheme (cancer detection: $H(7) = 775.56$, $p < .001$, correct return: $H(7) = 401.13$, $p < .001$).

In pairwise post-hoc tests of the critical difference in ranks (Bonferroni correction applied), the cancer detection rate on the 2020 case set was significantly higher than in all other schemes ($p < .05$), while the correct return to screen rate on the 2020 case set was significantly lower than in all other schemes ($p < .05$) except the 2019 case set.

Relatively high levels of sensitivity on the 2020 case set, coupled with relatively low levels of specificity, indicate that breast screeners were "over-recalling" when reading the 2020 case set, i.e., recalling cases more readily than in other case sets. This possibility was examined directly by looking at the difference between the sensitivity and specificity for each reader: the cancer detection rate minus correct return to screen rate was calculated for each breast screener in each PERFORMS scheme (Table2).

The difference between the cancer detection rate and the correct return to screen rate varied significantly by scheme ($H(7) = 365.11$, $p < .001$); specifically, the difference between cancer detection rate and correct return to screen rate on the 2020 case set was significantly higher than in all other schemes ($p < .05$, Bonferroni correction applied).

4. Discussion

This study was undertaken to determine whether reduced case volumes in the NHS Breast Screening Programme in England during the COVID-19 lockdown period were associated with changes in the screening performance of breast screeners, as assessed using the standardised PERFORMS test of breast screening. The performance of breast screeners on the 2020 case set, which was administered during the COVID-19 lockdown period, was characterised by markedly high sensitivity and low specificity when compared to their performance on PERFORMS tests administered at other times before and after the lockdown period.

In particular, the gap between the sensitivity and specificity of individual readers was largest for this case set, indicating that breast screeners were "over-recalling" breasts from the 2020 case set. Individual readers tended to detect a greater percentage of the malignancies in the 2020 case set, but without an accompanying increase in specificity, resulting in a greater percentage of unnecessary recalls. The extreme degree of over-recalling is peculiar to the test administered during the lockdown period. This over-recalling may have been associated with the changes to live screening over the lockdown period. The effect was temporary, and the same degree of over-recalling was not observed before or after the lockdown period.

A recent audit study that evaluated the real-life performance indicators of a screening service in the public health system in Sao Paulo State, Brazil showed that the recall rate was significantly lower ($p < 0.001$) during the pandemic compared to the pre COVID-19 period, with increased specificity for screening and diagnostic mammograms [9]. Similarly, performance data from 66 facilities of the Breast Cancer Surveillance Consortium, USA showed a significant decrease in cancer detection (23% fewer cancers, $p < .001$) during the pandemic compared with the same pre COVID-19 period [10]. However, in both studies, authors explained that the reduction in recall rate and cancer detection were due to the reduction in number of women screened during the pandemic. These results contrast with our findings. Yet, our study focused on readers' performance as assessed using a standardised test-based scheme, where screeners' decisions do not have a direct impact on patient care, rather than their real-life performance. The noted effect of over-recalling in the test may reflect a temporary change in breast screeners' decision making and cancer detection ability associated with their reduced volume of live screening practice at that time.

This study does have some limitations. PERFORMS is a standardised test of breast screening performance, which, although undertaken using real cases from the NHSBSP, is not live. As such, breast screeners may interpret mammograms differently during a test set compared to real-life practice as their decisions on the test do not have consequences for the women whose cases contribute to the test [11].

A further limitation of the study is that PERFORMS uses a relatively small case set (60 cases, 120 breasts) as the basis of its performance metrics when compared to real-life screening. Inevitably, the case set varies in difficulty year-by-year, with associated differences in performance on the test. Consequently, variation in the case sets could contribute to the differences in sensitivity and specificity observed each year, in addition to any differences that might be attributed to the COVID-19 lockdown period. However, it is argued that the extreme degree of over-recalling observed for the test administered during the lockdown is, at least in part, a consequence of temporary changes in breast screeners' performance associated with their reduced volume of live screening practice at that time. Other significant differences were noted between case sets. However, the study focused on the fact that sensitivity was highest, specificity was second lowest, and the difference

Table 2
Summary statistics for cancer detection and correct return to screen measures of each PERFORMS test set.

Scheme Year	Cancer detection			Correct return to screen			Cancer detection minus Correct return to screen		
	M	SD	SE	M	SD	SE	M	SD	SE
2015	91.3%	7.8%	0.3%	83.3%	8.9%	0.4%	8.0%	13.6%	0.6%
2016	86.0%	10.4%	0.4%	84.0%	7.2%	0.3%	2.0%	15.1%	0.6%
2017	86.3%	10.3%	0.4%	83.9%	9.3%	0.4%	2.4%	16.2%	0.6%
2018	92.2%	7.8%	0.3%	86.0%	9.5%	0.4%	6.2%	14.4%	0.6%
2019	85.0%	9.3%	0.4%	78.8%	10.8%	0.4%	6.2%	17.5%	0.7%
2020	94.9%	6.8%	0.3%	78.9%	10.8%	0.4%	16.0%	14.6%	0.6%
2021	89.8%	8.5%	0.3%	82.3%	10.4%	0.4%	7.5%	15.8%	0.6%
2022	92.1%	9.5%	0.5%	86.2%	9.4%	0.5%	6.0%	15.0%	0.7%
Overall	89.6%	9.5%	0.1%	82.7%	10.0%	0.1%	6.9%	15.9%	0.2%

between sensitivity and specificity for individual readers was highest for the COVID-19 set. These are the only differences that can be associated specifically with the 2020 COVID-19 lockdown year.

The PERFORMS test itself, and the bases of the performance metrics (the number and nature of the cases), have not been affected directly by changes in screening volumes and attendance during the lockdown period. The test set administered during the lockdown period was collated prior to the COVID-19 related lockdowns, following the same procedure and criteria used for all other case sets in the present study.

In addition, the findings of our study are only likely to be generalisable to screening programs that use test-set based assessment schemes. Our study only provides evidence to support that the aforementioned changes in the screening performance of breast screeners, as assessed using the PERFORMS test sets, was a consequence of the changes to live screening over the lockdown period.

5. Conclusion

To our knowledge, this is the first study carried out that evaluates whether performance on a test-set based assessment scheme taken during the COVID-19 lockdown period differed from that on the tests taken at other times. The results demonstrate that although breast screeners were "over-recalling" breasts from the PERFORMS case set that was administered during the COVID-19 lockdown period, the same effect was not observed before or after the lockdown period.

Funding

The PERFORMS test scheme was funded by Public Health England, Breast Screening Programme. The funding source had no level of involvement in this study.

CRediT authorship contribution statement

Eleni Michalopoulou: Conceptualization, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Padma Priya Pugalenthi:** Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Iain Darker:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Yan Chen:** Conceptualization, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dr Yan Chen declares activities related to the present work: disclosed grant money paid to author's institution by the National Health Service Breast Screening Programme through Public Health England. The funding source however had no level of involvement in this study. 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.

References

[1] Department of Health and Social Care, New breast cancer screening units to speed up diagnosis - Press release, 2023. https://www.gov.uk/government/news/new-breast-cancer-screening-units-to-speed-up-diagnosis#:~:text=Continued%20investment%20in%20mobile%20breast,cancer%20screening%20in%20the%20UK (accessed 11 April 2023).

[2] E. Cornford, S. Cheung, M. Press, O. Kearins, S. Taylor-Phillips, Optimum screening mammography reading volumes: evidence from the NHS Breast Screening Programme, Eur. Radiol. 31 (2021) 6909–6915, https://doi.org/10.1007/s00330-021-07754-8.

[3] S.W. Duffy, F. Seedat, O. Kearins, M. Press, J. Walton, J. Myles, D. Vulkan, N. Sharma, A. Mackie, The projected impact of the COVID-19 lockdown on breast cancer deaths in England due to the cessation of population screening: a national estimation, Brit. J. Cancer 126 (2022) 1355–1361, https://doi.org/10.1038/s41416-022-01714-9.

[4] Office for Health Improvement and Disparities, NHS screening programmes in England: 2020 to 2021 - Corporate report, 2023. https://www.gov.uk/government/publications/nhs-screening-programmes-annual-report/nhs-screening-programmes-in-england-2020-to-2021#:~:text=In%20March%202020%2C%20all%2078,and%20staff%20from%20the%20virus (accessed 11 April 2023).

[5] COVID-19 impact during 2020-21, NHS Digital, 2022. https://digital.nhs.uk/data-and-information/publications/statistical/breast-screening-programme/england—2020-21/covid-impact—programme-summary-2020-21 (accessed 11 April 2023).

[6] Y. Chen, J.J. James, E. 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 2 (5) (2020) e200016.

[7] Breast Screening Programme, England 2019-20, NHS Digital, 2021. https://digital.nhs.uk/data-and-information/publications/statistical/breast-screening-programme/england—2019-20 (accessed 13 April 2023).

[8] A. Gale, Y. Chen, A review of the PERFORMS scheme in breast screening, Brit. J. Radiol. 93 (2020) 20190908, https://doi.org/10.1259/bjr.20190908.

[9] E.M. Solla Negrao, C. Cabello, L. Conz, E.C. Mauad, L.C. Zeferino, D.B. Vale, The impact of the COVID-19 pandemic on breast cancer screening and diagnosis in a Brazilian metropolitan area, J. Med Screen. 30 (2023) 42–46, https://doi.org/10.1177/09691413221122055.

[10] K.P. Lowry, M.C.S. Bissell, D.L. Miglioretti, K. Kerlikowske, N. Alsheik, T. Macarol, E.J.A. Bowles, D.S.M. Buist, A.N.A. Tosteson, L. Henderson, S.D. Herschorn, K. J. Wernli, D.L. Weaver, N.K. Stout, B.L. Sprague, Breast biopsy recommendations and breast cancers diagnosed during the COVID-19 pandemic, Radiology 303 (2023) 287–294, https://doi.org/10.1148/radiol.2021211808.

[11] D. Gur, A.I. Bandos, C.S. Cohen, C.M. Hakim, L.A. Hardesty, M.A. Ganott, R.L. Perrin, W.R. Poller, R. Shah, J.H. Sumkin, L.P. Wallace, H.E. Rockette, The "laboratory" effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography.