International Journal of Women's Health

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Performance Metrics of Mammography Screening Programmes in Primary Health Care Centres in Bahrain

Ghufran Ahmed Jassim[1], Muna Ahmed AlMohri[2], Hala Yusuf AlAsoomi[2], Ameera AlNooh[3], Shahla Fahad AlDoseri[2]

[1]Department of Family Medicine, Royal College of Surgeons in Ireland – Medical University of Bahrain, Busaiteen, Muharraq, Bahrain; [2]Department of Medical Services, Primary HealthCare Centres, Manama, Bahrain; [3]Department of Non Communicable Disease, Public Health Directorate, Ministry of Health, Manama, Bahrain

Correspondence: Ghufran Ahmed Jassim, Department of Family Medicine, Royal College of Surgeons in Ireland – Medical University of Bahrain, PO Box 15503, Adliya, Bahrain, Tel +973 3968 0009, Email gjassim@rcsi-mub.com

Background: Mammography is the cornerstone of breast cancer screening. Its diagnostic performance, however, is influenced by population demographics such as age and breast density.

Purpose: The aim of this study was to establish contemporary performance benchmarks for mammography screening in Bahrain's primary health-care centres (PHCs) and to identify areas for quality improvement.

Methods: A cross-sectional retrospective analysis was performed on mammograms from asymptomatic women aged ≥40 years who were screened in 2020 at primary health care centres. Screening outcomes were cross-referenced with subsequent breast cancer diagnoses recorded in the Bahrain Cancer Registry (2021–2022). Mammographic findings were categorised using the Breast Imaging Reporting and Data System (BI-RADS), and performance metrics including the cancer detection rate (CDR), sensitivity, and specificity were calculated.

Results: A total of 2196 screening mammograms were included, with a mean patient age of 56.1 years. The cancer detection rate was 12.3 per 1000 screens, and the recall rate was 30.6%. Sensitivity and specificity were 69.2% and 71.2%, respectively, and the interval cancer rate was 5.4 per 1000. The most common breast density category was "scattered fibroglandular" (61.8%). Recall rates were significantly associated with breast density ($p < 0.001$). Among the 39 patients diagnosed with breast cancer, invasive ductal carcinoma was the most prevalent subtype (71.8%).

Conclusion: The findings highlight the moderate utility of mammography screening in Bahrain, characterised by a relatively high recall rate and a need to optimise reporting standards and recall criteria. Tailored strategies such as supplemental imaging for women with dense or high-risk breasts, strict adherence to BI-RADS guidelines, and implementing double reading or single reading with computer-aided detection could improve screening outcomes. These results establish important regional performance benchmarks and can inform policies to enhance breast cancer detection and management in the Arabian Gulf region.

Keywords: breast imaging, mammogram, breast cancer screening, benchmark, Arabian Gulf

Introduction

Mammography, despite its limitations, remains the only breast cancer screening test endorsed by major health bodies such as the US Preventive Services Task Force and the American Cancer Society, and it is the most effective means of reducing breast cancer mortality.[1,2] To improve the quality of mammography practice, the American College of Radiology (ACR) introduced the Breast Imaging Reporting and Data System (BI-RADS) in the 1980s.[3]

Established performance metrics for screening mammography such as the cancer detection rate, sensitivity, and specificity are defined by the ACR.[4,5] These metrics were originally designed to evaluate radiologists' interpretive performance, but they

Received: 10 May 2025
Accepted: 19 July 2025
Published: 18 August 2025

are also used to inform patients, healthcare providers, and policymakers about the benefits, harms, and limitations of mammography screening.[6]

Bahrain has the highest age-standardised breast cancer incidence among Gulf Cooperation Council countries.[7] Moreover, as in many Middle Eastern countries, breast cancer tends to present at a younger age in Bahrain.[8,9] The prevalence of dense breast tissue in these younger patients can make mammographic interpretation more challenging.[10,11]

Currently, mammography in Bahrain is offered every two years to women aged 40 years and above at five of the 27 primary healthcare (PHC) centres equipped with breast imaging facilities. Screening is opportunistic, meaning that women are not invited for screening but may receive it during patient–doctor encounters. In the last couple of years, a new healthcare system called "Choose Your Doctor" has been introduced in select health centres. This system allows patients to select their family physician, enabling the physician to actively invite eligible women for screening. The service is provided free of charge to Bahraini nationals in the public health care sector.

Previous Gulf-region mammography reports are methodologically heterogeneous and often involve estimation of the uptake and barriers to breast cancer screening.[8,12] The scarce mammogram metric performance studies in the region are single-centre, enrolling small sample size, with no linkage to subsequent clinical outcomes.[13,14] One local study reported a recall rate of 32% and a cancer detection rate of 0.5% in 2018,[14] but other performance metrics were not assessed and the study did not track interval cancers.[14] Evaluating these metrics can help policymakers establish benchmarks for local practice and identify opportunities for improvement. Moreover, given the similarities in breast cancer epidemiology and risk factors across the region, our findings can serve as a reference point for other regional studies of mammography performance. The purpose of this study was to establish the first national performance benchmarks for screening mammography in Bahrain's PHC system and to assess screening performance trends over time. Our design provides the first regional estimates of programme metrics based on standardized reporting and longitudinal verification, offering a more robust benchmark for future quality-improvement efforts.

Methods

Study Design

This cross-sectional observational study was conducted in two stages. In the first stage, we retrospectively reviewed mammography reports from two primary health care centres (Naim Health Centre and Hamad Kanoo Health Centre) between 1 January 2020 and 31 December 2020. At that time, these two centres were the only PHCs with breast imaging facilities receiving screening referrals from across Bahrain. In the second stage, we obtained data on subsequent breast cancer diagnoses for the same cohort from the Bahrain Cancer Registry for the period 1 January 2021 to 31 December 2022, corresponding to a two-year post-screening interval. Performance measures were defined according to the ACR BI-RADS 5th edition[3] and were calculated using standard BI-RADS formulae. All study procedures were conducted in accordance with relevant guidelines and regulations.

Inclusion Criteria

All digital screening mammograms performed at the two study centres in 2020 for asymptomatic women aged ≥40 years (between 1 January and 31 December 2020) were included.

Exclusion Criteria

We excluded (1) screening examinations in women with a personal history of breast cancer, mastectomy, or breast augmentation (populations in which screening performance metrics are known to differ significantly); and (2) any mammogram with a BI-RADS category 6 (known breast cancer).

Data Collection Procedures and Definitions

All included mammograms were part of the opportunistic screening programme at PHCs, where patients from any PHC in Bahrain were referred to the two designated centres. Demographic and clinical information (such as age, nationality, and family history) was collected from the medical records and radiology reports. All mammograms were interpreted by

a single radiologist. Each report included the BI-RADS assessment category assigned to the mammogram, according to the following standard definitions:

- **BI-RADS 0**: Requires additional imaging evaluation
- **BI-RADS 1**: Negative (no findings)
- **BI-RADS 2**: Benign findings
- **BI-RADS 3**: Probably benign; short-interval follow-up suggested
- **BI-RADS 4**: Suspicious for malignancy
- **BI-RADS 5**: Highly suggestive of malignancy
- **BI-RADS 6**: Known malignancy (biopsy-proven prior to definitive therapy)

If a mammogram resulted in a recommendation to recall the patient for further assessment, we retrieved the subsequent diagnostic radiology and pathology reports from the patient's records. The radiology reports also documented breast density according to the ACR BI-RADS Atlas categories: (a) almost entirely fatty, (b) scattered fibroglandular densities, (c) heterogeneously dense, and (d) extremely dense (5). Finally, we cross-checked the Bahrain Cancer Registry records for 2021 and 2022 to identify any breast cancer diagnoses among the women screened in 2020.

Outcome Measurements and Statistical Analysis

Performance metrics (eg recall rate, CDR, sensitivity, specificity, and likelihood ratios) were defined in accordance with the ACR BI-RADS 5th edition.[3] Data were analysed using SPSS version 29. Descriptive statistics (frequencies, percentages, means, and medians) are reported for the participants' demographic and clinical characteristics (age, breast density, family history, nationality) and for tumour characteristics of the cancers detected. The performance metrics were computed as defined above. Associations between age and mammography findings or breast density were tested using one-way ANOVA, and associations between mammography findings, density category, follow-up recommendation, and patient recall were tested using chi-square tests. A p-value < 0.05 was considered statistically significant. All data generated or analysed during this study are included in this published article.

Results

Demographics and Clinical Characteristics

We analysed 2196 screening mammograms from asymptomatic women. The mean age of the patients was 56.1 years (SD 8.2), with a median of 56 and a range of 43–93 years. The vast majority of women were Bahraini nationals (97.2%), with only 2.8% being non-Bahraini (Table 1). First-degree family history of breast cancer was recorded as "Unknown" in 98% of cases, limiting the usefulness of this variable.

The majority of mammograms were reported as either "negative" (68.8%) or "incomplete, needs additional evaluation" (28.5%). Findings suspicious for, or highly suggestive of, malignancy were very uncommon (0.2% of examinations for each category). The most common breast density category was "scattered fibroglandular" (61.8%), followed by "heterogeneously dense" (26.9%); only 1.6% of women had "extremely dense" breasts. Consistent with these findings, 67.9% of patients were recommended to return for routine screening (reflecting the high proportion of negative results), whereas 31.2% were advised to undergo further imaging based on indeterminate or concerning findings.

Associations Between Variables

We examined the relationships between patient age, mammography findings, and breast density. There were statistically significant differences in mean age across the breast density categories ($p < 0.001$, ANOVA). Women with almost entirely fatty breasts tended to be older on average, whereas those with extremely dense breasts were younger on average. By contrast, mean age did not differ significantly across the mammography report categories ($p = 0.325$) (Table 2).

Table 1 Clinical and Demographic Characteristics of the Screening Population (N = 2196)

Variable	Number (%)
Age	Mean (SD): 56.12 (8.18) Median: 56 Range: 50 (min 43, max 93)
Nationality	
Bahraini	2134 (97.2)
Non-Bahraini	62 (2.8)
1st degree Family history of breast cancer	
Yes	21 (1)
No	23 (1)
Unknown	2152 (98)
Mammogram interpretation	
Incomplete need additional evaluation	625 (28.5)
Negative	1510 (68.8)
Benign	37 (1.7)
Most likely, Benign	15 (0.7)
Suspicious	5 (0.2)
Highly suggestive of malignancy	4 (0.2)
Density classification	
Almost entirely fatty	143 (6.5)
Scattered fibroglandular	1357 (61.8)
Heterogeneously dense	591 (26.9)
Extremely dense	35 (1.6)
Unknown	70 (3.2)

Breast density was also significantly associated with the mammography assessment outcome ($p < 0.001$, chi-square), suggesting that certain density types correspond to specific mammographic findings (Table 3). For instance, women with almost entirely fatty breasts were more likely to have benign or negative mammogram results.

Table 3 also shows that mammography outcome, breast density, recommended follow-up, and recall status were all interrelated. In general, higher breast density was associated with a greater likelihood of needing additional imaging or intervention. Nearly all patients with an "Incomplete" assessment were appropriately recalled for further evaluation (95%). However, only 80% of those with a "Suspicious" finding and 75% of those with a "Highly suggestive of malignancy" finding were recalled. The failure to recall some patients with clearly abnormal findings appears to have resulted from inconsistencies in radiology reporting and non-adherence to recommended follow-up.

Breast Cancer Diagnoses

Within two years of the 2020 screening, 39 out of 2196 women (1.8%) were diagnosed with breast cancer. The mean age at diagnosis was 60.6 years (SD 9.8). Among these cases, 53.8% involved the left breast, 41.0% the right breast, and 5.1% were

Table 2 Association Between Age, Mammography Report Category, and Breast Density (N = 2196)

Variable	N	Mean (SD)	95% CI	P value
Mammogram report				
Incomplete: need additional imaging evaluation	625	55.90 (8.32)	55.25–56.55	0.325
Negative	1510	56.21 (8.03)	55.80–56.61	
Benign	37	54.72 (10.95)	51.07–58.38	
Most likely, benign	15	56.33 (6.88)	52.52–60.14	
Suspicious	5	60.60 (12.46)	45.12–76.07	
Highly suggestive of malignancy	4	62.75 (11.95)	43.72–81.77	
Density categories				
Almost entirely fat	143	59.44 (8.01)	58.12–60.77	<0.001
Scattered fibroglandular density	1357	57.17 (8.22)	56.74–57.61	
Heterogeneously dense	591	53.27 (7.37)	52.67–53.86	
Extremely dense	35	51.71 (6.11)	49.61–53.81	
Unknown	70	55.10 (7.86)	53.22–56.97	

Table 3 Association Between Mammography Findings, Breast Density, Follow-up Recommendation, and Recall Status (N = 2196)

Result of Mammogram reports	Incomplete: Need Additional Imaging Evaluation n(%)	Negative n(%)	Benign n(%)	Most Likely, Benign n(%)	Suspicious n(%)	Highly Suggestive Of Malignancy N(%)	P- value
Density categories							
Almost entirely fat	28 (4.5%)	115 (7.6%)	0	0	0	0	<0.001
Scattered fibroglandular density	328 (52.5%)	989 (65.5%)	24 (64.9%)	9 (60%)	4 (80%)	3 (75%)	
Heterogeneously dense	193 (30.9%)	379 (25.1%)	12 (32.4%)	6 (40%)	0	1 (25%)	
Extremely dense	20 (3.2%)	13 (0.9%)	1 (2.7%)	0	1 (20%)	0	
Unknown	56 (9.0%)	14 (0.9%)	0	0	0	0	
Follow up recommendations by the radiologist							
Return for routine screening	18 (2.9%)	1436 (95.1%)	24 (64.9%)	12 (80%)	1 (20%)	1 (25%)	<0.001
Short-interval follow-up imaging at 6 months	3 (0.5%)	6 (0.4%)	1 (2.7%)	0	0	0	
Recommendation for tissue biopsy	2 (0.3%)	4 (0.3%)	0	1 (6.7%)	0	2 (50%)	
Further imaging is needed immediately	602 (96.3%)	64 (4.2%)	12 (32.4%)	2 (13.3%)	4 (80%)	1 (25%)	
Recall of patients							
Yes	592 (95%)	59 (3.9%)	12 (32.4%)	3 (20%)	4 (80%)	3 (75%)	<0.001
No	31 (5%)	1451 (96.1%)	25 (67.6%)	12 (80%)	1 (20%)	1 (25%)	

bilateral. Regarding stage at diagnosis, 17.9% of the cancers were in situ, 43.6% were localised disease, 23.1% showed direct extension, and 15.4% had an unknown stage. In terms of tumour grade, 43.6% were Grade 2, 23.1% were Grade 3, and 15.4% had an unknown grade. Histologically, invasive ductal carcinoma was the most common type, accounting for 71.8% of the diagnosed cases.

Recall and Interval Cancers

A total of 673 women (30.6% of those screened) were recalled for further imaging. Among those recalled, 24 (3.6%) were ultimately diagnosed with breast cancer and 649 (96.4%) were not, reflecting a high rate of false-positive recalls. Conversely, 15 women who were not recalled after screening were later diagnosed with breast cancer. Of these 15 "missed" cases, 12 had normal mammogram findings at the time of screening and were diagnosed within the subsequent two-year interval, qualifying them as interval cancers. The remaining 3 cases had discordant initial readings: the mammogram interpretation noted a suspicious lesion requiring further imaging (a positive finding), but the final recommendation was mistakenly recorded as routine follow-up, and thus these patients were not recalled. All three of these errors occurred in women with heterogeneously dense or extremely dense breasts. Excluding these reporting errors, the adjusted interval cancer rate was 5.4 cases per 1000 screens.

Among the 12 true interval cancers (cancers that were not detected at screening but presented within two years), 58.3% were in the left breast, 54.5% arose in heterogeneously dense breasts, 41.7% were localised Grade 2 tumours, 16.7% were stage 3 with direct extension, and 41.7% had unknown grade or stage. Considering all 39 cancers diagnosed (24 in the recalled group and 15 in the not-recalled group), the majority (61.5%) were detected among those who had been recalled. However, because 649 of the 673 recalls did not result in a cancer diagnosis, the recall process yielded a substantial number of false positives.

Among the patients who were recalled, follow-up investigations included repeat mammography in 85 cases (12.6%), ultrasound in 323 cases (48.0%), and both ultrasound and mammography in 250 cases (37.1%); for the remaining 15 cases (2.2%), the type of follow-up was not recorded.

Performance Metrics

Figure 1 summarizes the clinical flow and Table 4 summarises the key performance metrics of the screening programme. The positive mammogram rate (also known as the abnormal interpretation rate, comprising BI-RADS categories 0, 3, 4,

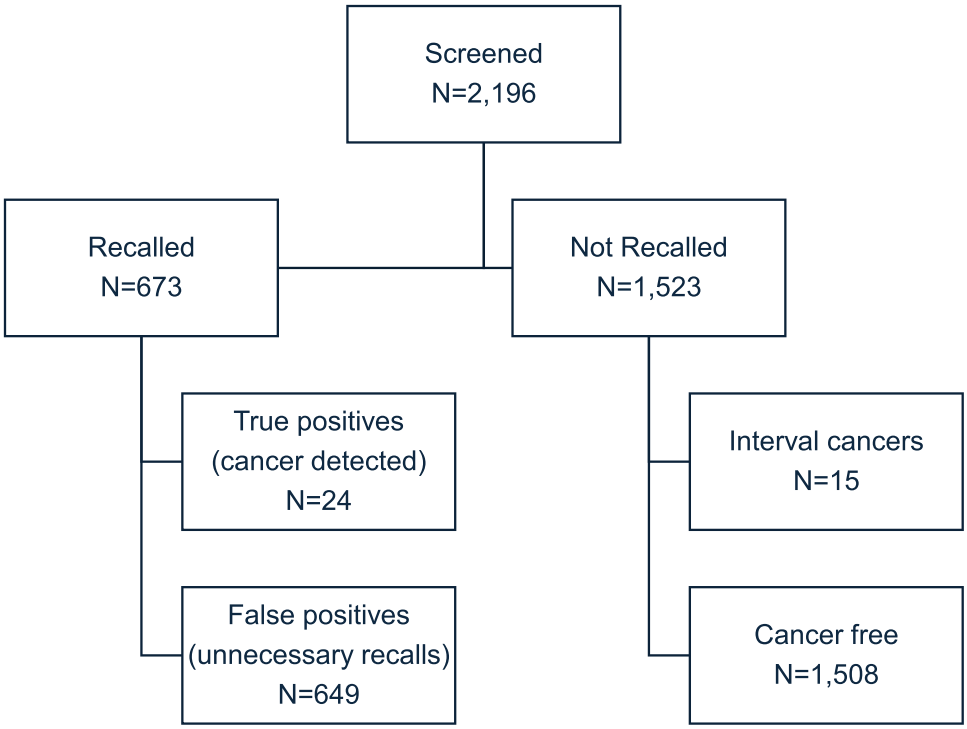

Figure 1 Summary of the clinical flow from initial screening to interpretation.

Table 4 Screening Performance Metrics for Mammography (N = 2196)

Mammogram Result	Breast Cancer Diagnosis		Total
	Yes	No	
Positive mammogram	27 (4.2%)	622 (95.8%)	649
Negative mammogram	12 (0.8%)	1535 (99.2%)	1547
Total	39	2157	2196

or 5) was 29.6% (649 out of 2196 screenings). The interval cancer rate, defined as the number of cancers diagnosed within 12–24 months after a negative screening result, was 5.4 per 1000 screens. The screening sensitivity was 69.2%; 95% CI (52.4% to 83.0%) and specificity was 71.2%; 95% CI (69.2% to 73.1%), corresponding to a positive likelihood ratio of 2.4 and a negative likelihood ratio of 0.43. The cancer detection rate (true positive cases per 1000 screenings) was 12.3 per 1000.

Discussion

This study evaluated the performance of a mammography screening programme in a primary health care setting in Bahrain, analysing 2196 screening mammograms. The screened population had a median age in the mid-50s. Approximately 30% of the mammograms were classified as heterogeneously dense or extremely dense, while about 60% were classified as having scattered fibroglandular densities. It is important to acknowledge that the BI-RADS breast density classification is based on visual assessment by radiologists and can suffer from inter- and intra-observer variability. Nevertheless, our observed density distribution is in line with previous reports that Middle Eastern women tend to have higher breast density compared to Western populations.[14–16]

Overall, our results indicate that the screening mammography programme has a modest screening utility on the basis of the given likelihood ratios and recall rates. The sensitivity was modest, and the recall rate was high, yielding only moderate likelihood ratios for test performance. The sensitivity of 69.2% means that roughly 69% of all breast cancers in the screened population were detected on the initial mammogram, while about 31% were missed (false negatives). The specificity of 71.2% corresponds to a false-positive rate of about 28.8%, meaning a considerable number of women without cancer received an abnormal screening result (622 out of 2157 cancer-free women). These false positives can lead to anxiety, unnecessary follow-up procedures (such as biopsies), and higher healthcare costs.

The positive likelihood ratio of 2.4 further reflects the limited discriminative power of the screening test. A positive mammogram made the probability of breast cancer only 2.4 times higher than it would be if the test were negative. Similarly, the negative likelihood ratio of 0.43 indicates that a negative mammogram does reduce the probability of cancer, but not enough to rule it out entirely, especially in women at higher risk.

Root cause review of the three patients who were not recalled despite suspicious findings identified two administrative scheduling errors (patients were placed on the follow-up list but appointments were never generated) and one instance of patient refusal to return. Because the screening pathway failed to trigger timely diagnostic work-up, we have treated all three cases as false-negative outcomes in the performance calculations and described them as manifestations of a system failure in the recall logistics.

Our findings are generally consistent with the range of results reported in the literature, though certain metrics differ in magnitude. A comparison between our results and international benchmark can be found in Supplementary Table 1. Published studies have shown considerable variability in mammography performance depending on the technology used, population demographics, and clinical setting.[17] For example, a systematic review of digital mammography reported CDRs of 4–6.6 per 1000 and recall rates of 3–10%.[18] Another review focusing on 2D mammography found CDRs of 3.2–7.1 per 1000 and recall rates of 2.6–17.5%.[19] Both of these reviews indicate much lower recall rates than observed in our study (which had a CDR of 12 per 1000 and a 30% recall rate). In a separate study, a 2D mammography programme

had an abnormal interpretation rate of 10.4%,[20] compared to our AIR of 29%. These discrepancies may be partly explained by the high prevalence of dense breasts in our cohort and our use of single-reader image interpretation; both factors have been shown to increase recall rates and reduce specificity.[21–23] Other attributable factors could be first round screening effect, conservative interpretive threshold of the single reader and absence of standardised protocol for BIRADS 3 pathway.[24] Moreover, while a single reader approach is not ideal, it was the prevailing situation at the time of data collection. As our study is retrospective in nature, we presented the data as they occurred in real clinical practice which might have led to increased classification bias.

Another systematic review reported average sensitivity and specificity for digital mammography of about 73% and 88%, respectively.[25] Our sensitivity (69.2%) and specificity (71.2%) were lower. One likely explanation is that our screening programme targets a younger age group (starting at 40 years) with a higher proportion of dense breasts, compared to Western populations where screening often starts at age 50.[22] Dense breast tissue in younger women can mask cancers on mammography, lowering both sensitivity and specificity. The lack of comparable national or regional data also makes it difficult to benchmark our results against other Gulf populations.

Variations in screening technology and practice patterns may further contribute to performance gaps. Digital breast tomosynthesis (3D mammography), for instance, can improve cancer detection in women with dense breasts by reducing the problem of overlapping tissue, as evidenced by increased sensitivity in some studies.[26,27] However, tomosynthesis is not yet widely available in Bahrain and comes with higher costs, limiting its use. Additionally, differences in screening intervals and radiologist expertise might influence outcomes; for example, shorter intervals between screenings can catch more interval cancers (potentially improving overall sensitivity) but might also inflate the recall rate by detecting indolent findings.[26,27] Radiologist credentials were not systematically captured in our dataset and hence future studies should prospectively document radiologist training and experience to evaluate their potential impact on performance metrics.

This study has several limitations that may affect the interpretation of the findings. First, the screening population's relatively younger age and greater breast density likely reduced the observed sensitivity and specificity of mammography, potentially underestimating the true cancer detection rate. Second, we did not include women who were screened at secondary care hospitals, which may have introduced selection bias since hospital screenings often involve higher-risk or symptomatic patients. Third, the study period overlapped with the COVID-19 pandemic, during which healthcare utilisation dropped substantially; this likely resulted in fewer women being screened and some cancers being diagnosed later, influencing the performance metrics particularly the interval cancer rate. Finally, several key clinical details (such as tumour stage, grade, and receptor status) were frequently missing from records, limiting our ability to analyse correlations between screening outcomes and cancer severity. In addition, incomplete documentation of risk factors (eg family history or prior biopsies) hindered risk stratification of the screened population.

Despite these limitations, our findings provide valuable insights that can inform improvements in the breast screening programme. One clear priority is to improve data documentation and integration ensuring that risk factor information is systematically recorded and that a comprehensive national database captures all screening and cancer outcomes. Given that only a small fraction of recalls resulted in a cancer diagnosis (and that some patients were recalled despite negative findings), there is a need to refine recall criteria to avoid unnecessary follow-ups and undue anxiety for patients. Early detection strategies should balance the psychological and logistical burdens of high recall rates with the benefit of detecting cancers early. To improve accuracy and reduce unnecessary recalls, we recommend adopting a double-reading system for mammograms, as recommended by the European screening guidelines,[28] or using computer-aided detection to assist a single reader (as is common in the United States). We observed that inconsistent application of BI-RADS assessment and follow-up recommendations contributed to inappropriate recall decisions; thus, standardising radiology reporting practices according to BI-RADS guidelines is essential.

It may also be beneficial to incorporate supplemental screening modalities for women at higher risk or with dense breasts, for example, routine 2-view digital breast tomosynthesis (DBT), targeted ultrasound for BI-RADS c/d, incorporate MRI for extremely dense (d) with risk factors. Further, standardized management of BI-RADS 3, for example, adopting short-interval (6-month) follow-up rather than immediate recall, aligning with BI-RADS guidance. Educating patients about the possibility of false-positive results and establishing clear, evidence-based follow-up protocols can help

mitigate the psychological impact of recalls. Tailoring screening and follow-up recommendations based on individual risk profiles (considering factors like family history and genetic predisposition) will further enhance the effectiveness of the programme. For women at average risk, personalised counselling is important to ensure informed decision-making regarding screening.

Implementing double-reading or computer-aided detection (CAD) in Bahrain's national breast-screening programme is feasible but will require thoughtful resource planning. The principal operational barrier is workforce and capital outlay for a server based CAD solution. Several facilitators offset these hurdles. First, a phased roll-out, initial double reading limited to BI-RADS 0 cases and CAD triage for dense-breast screens, targeting the groups that drive 70% of unnecessary recalls. Second, strong public–private partnerships (eg, corporate sponsorships that already subsidise screening costs) provide a realistic avenue for co-funding.

Conclusion

In summary, the performance of the mammography screening programme in Bahrain's primary health care centres was moderate. While most cancers in the study were eventually detected through the recall process, the high recall rate and modest accuracy indicate room for improvement and warrant a critical review of the referral, follow-up, and radiological reading protocols. By enhancing data quality, optimising recall protocols, utilising double-reading or advanced imaging technologies, and integrating supplemental screening for appropriate cases, the programme's effectiveness can be improved. Future research should explore the implementation of these strategies and evaluate outcomes in comparable populations, thereby guiding evidence-based enhancements in breast cancer screening practice.

AI Language-Editing disclosure

Portions of this manuscript's wording were refined for grammar and style with the assistance of *ChatGPT* (OpenAI, GPT-4 o model; accessed March 2025). The tool was applied only after the authors had drafted the original text, and solely to improve readability; it did not generate, alter, or interpret any scientific content, data, analyses, or conclusions. All AI-suggested edits were critically reviewed, revised as necessary, and approved by the human authors, who accept full responsibility for the integrity and accuracy of the final work. ChatGPT is not listed as an author, consistent with ICMJE and WAME guidance that AI tools do not meet authorship criteria and must be transparently acknowledged.

Ethics Approval and Consent to Participate

We obtained ethics approval from the ethics committee at RCSI Bahrain, Primary Care and the Public Health Directorate (RCSI Bahrain REC Approval Number: REC/2025/262/09-Mar-2025). The institutional ethics committee granted a waiver of informed consent due to the retrospective design of the study. Data were collected from existing medical records. In accordance with institutional policy, all patients undergoing mammography or consulting a physician—particularly during their initial visit—sign a declaration permitting the use of their anonymized medical data for purposes including audit, quality assurance, and research. Therefore, the use of these data complies with institutional ethical standards and the principles outlined in the Declaration of Helsinki.

The anonymized data were saved in a password protected document that is accessible only to the principal investigators (PIs). Stored data will be discarded 3 years after publication. The data were analysed and presented collectively, and no patients were identified.

Acknowledgment

We would like to thank two RCSI Bahrain medical students (Nathania Keny and Lulwa Khalil Alqallaf) for their assistance with data entry.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically

reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

The authors declare that they have no competing interests in this work.

References

1. DeAngelis CD, Fontanarosa PB. US preventive services task force and breast cancer screening. *JAMA*. 2010;303(2):172–173. doi:10.1001/jama.2009.1990
2. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast cancer screening for women at average risk: 2015 guideline update from the American cancer society. *JAMA*. 2015;314(15):1599–1614. doi:10.1001/jama.2015.12783
3. Atlas AB-R. Breast imaging reporting and data system. *Ame College Radiol*. 2013;2013:1.
4. Sickles EA. ACR BI-RADS® Atlas, Breast imaging reporting and data system. *Ame College Radiol*. 2013;2013:1.
5. Sprague BL, Miglioretti DL, Lee CI, Perry H, Tosteson AAN, Kerlikowske K. New mammography screening performance metrics based on the entire screening episode. *Cancer*. 2020;126(14):3289–3296. doi:10.1002/cncr.32939
6. Siu AL. Screening for breast cancer: U.S. preventive services task force recommendation statement. *Ann Intern Med*. 2016;164(4):279–296. doi:10.7326/m15-2886
7. Ferlay JEM, Lam F, Laversanne M, et al. Global Cancer Observatory: Cancer Today. Available from: https://gco.iarc.who.int/today. Accessed October 27, 2024.
8. Al-Shamsi HO, Abdelwahed N, Al-Awadhi A, et al. Breast cancer in the United Arab Emirates. *JCO Global Oncol*. 2023;(9):e2200247. doi:10.1200/go.22.00247
9. Jassim GA, Whitford DL. Quality of life of Bahraini women with breast cancer: a cross sectional study. *BMC Cancer*. 2013;13:212. doi:10.1186/1471-2407-13-212
10. Brown AL, Vijapura C, Patel M, De La Cruz A, Wahab R. Breast cancer in dense breasts: detection challenges and supplemental screening opportunities. *RadioGraphics*. 2023;43(10):e230024. doi:10.1148/rg.230024
11. Gordon PB. The impact of dense breasts on the stage of breast cancer at diagnosis: a review and options for supplemental screening. *Curr Oncol*. 2022;29(5):3595–3636. doi:10.3390/curroncol29050291
12. Alshammari SA, Alhazmi AM, Alenazi HA, Alshammari HS, Alshahrani AM. Mammography uptake among the female staff of King Saud University. *J Family Med Prim Care*. 2020;9(1):221–228. doi:10.4103/jfmpc.jfmpc_706_19
13. Taif S, Tufail F, Alnuaimi AS. Mammography performance in Oman: review of factors influencing cancer yield and positive predictive value. *Asia Pac J Clin Oncol*. 2016;12(2):e250–8. doi:10.1111/ajco.12181
14. Almohri, Almohri M, Al Jerdabi F, et al. Outcome of screening mammography for women attending primary healthcare centers: one-year review. *J Bahrain Med Soc*. 2023;35(2):1–7. doi:10.26715/jbms.35_2_1
15. AlSaleh N, AlRammah T, Alatabani A, et al. Mammographic density in relationships with relevant contributing factors: a multicentric study from Riyadh, Saudi Arabia. *Gland Surg*. 2024;13(6):844–851. doi:10.21037/gs-23-374
16. Salem C, Atallah D, Safi J, et al. Breast density and breast cancer incidence in the Lebanese population: results from a retrospective multicenter study. *Biomed Res Int*. 2017;2017:7594953. doi:10.1155/2017/7594953
17. Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology*. 2009;253(3):641–651. doi:10.1148/radiol.2533082308
18. Zeng B, Yu K, Gao L, Zeng X, Zhou Q. Breast cancer screening using synthesized two-dimensional mammography: a systematic review and meta-analysis. *Breast*. 2021;59:270–278. doi:10.1016/j.breast.2021.07.016
19. Marinovich ML, Hunter KE, Macaskill P, Houssami N. Breast cancer screening using tomosynthesis or mammography: a meta-analysis of cancer detection and recall. *JNCI J Nat Cancer Inst*. 2018;110(9):942–949. doi:10.1093/jnci/djy121
20. Bahl M, Mercaldo S, Vijapura CA, McCarthy AM, Lehman CD. Comparison of performance metrics with digital 2D versus tomosynthesis mammography in the diagnostic setting. *Eur Radiol*. 2019;29(2):477–484. doi:10.1007/s00330-018-5596-7
21. Hofvind S, Geller BM, Skelly J, Vacek PM. Sensitivity and specificity of mammographic screening as practised in Vermont and Norway. *Br J Radiol*. 2012;85(1020):e1226–32. doi:10.1259/bjr/15168178
22. Larsen M, Lynge E, Lee CI, Lång K, Hofvind S. Mammographic density and interval cancers in mammographic screening: moving towards more personalized screening. *Breast*. 2023;69:306–311. doi:10.1016/j.breast.2023.03.010
23. Mann RM, Athanasiou A, Baltzer PAT, et al. Breast cancer screening in women with extremely dense breasts recommendations of the European Society of Breast Imaging (EUSOBI). *Eur Radiol*. 2022;32(6):4036–4045. doi:10.1007/s00330-022-08617-6
24. Carney PA, Sickles EA, Monsees BS, et al. Identifying minimally acceptable interpretive performance criteria for screening mammography. *Radiology*. 2010;255(2):354–361. doi:10.1148/radiol.10091636
25. Abdullah P, Alabousi M, Ramadan S, et al. Synthetic 2D mammography versus standard 2d digital mammography: a diagnostic test accuracy systematic review and meta-analysis. *AJR Am J Roentgenol*. 2021;217(2):314–325. doi:10.2214/ajr.20.24204
26. Abdelattef SAA, Ibrahim SF, Abdelhamid WR, Mahmoud FM. Three-dimensional tomosynthesis versus two-dimensional mammography in detection and characterization of different breast lesions. *Egypt J Radiol Nucl Med*. 2021;52(1):268. doi:10.1186/s43055-021-00648-1
27. Giampietro RR, Cabral MVG, Lima SAM, Weber SAT, Dos Santos Nunes-Nogueira V. Accuracy and effectiveness of mammography versus mammography and tomosynthesis for population-based breast cancer screening: a systematic review and meta-analysis. *Sci Rep*. 2020;10(1):7991. doi:10.1038/s41598-020-64802-x
28. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition--summary document. *Ann Oncol*. 2008;19(4):614–622. doi:10.1093/annonc/mdm481