The Breast 84 (2025) 104621

Contents lists available at ScienceDirect

The Breast

journal homepage: www.journals.elsevier.com/the-breast

Evaluation of breast cancer screening programmes: Candidate performance indicators and their association with breast cancer mortality

Carlos Canelo-Aybar [a,b], Pablo Alonso-Coello [b,c,**], Sergei Muratov [d], Jean-Eric Tarride [d], Nadya Dimitrova [e,f], Francesco Giusti [e], Bettina Borisch [g], Xavier Castells [h], Stephen W. Duffy [i], Patricia Fitzpatrick [j,k], Markus Follmann [l], Livia Giordano [m], Solveig Hofvind [n], Annette Lebeau [o,p], Cecily Quinn [q,r], Alberto Torresin [s], Paolo Giorgi Rossi [t], Holger Schünemann [d], Lennarth Nyström [u], Mireille Broeders [v,w,*], on behalf of the ECIBC contributor group

[a] Iberoamerican Cochrane Center, Barcelona, Spain
[b] Institut de Recerca Sant Pau (IR Sant Pau), Barcelona, Spain
[c] CIBER Biomedical Research Center in Epidemiology and Public Health (CIBERESP), Health Institute Carlos III (ISCIII), Madrid, Spain
[d] Department of Health Research Methods, Evidence, and Impact, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada
[e] European Commission, Joint Research Centre, Via E. Fermi 2749 – TP 127, I-21027, Ispra, VA, Italy
[f] Public Health Officer, International Agency for Research on Cancer, Lyon, France
[g] Institute of Global Health, University of Geneva, Geneva, Switzerland
[h] IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
[i] Centre for Prevention, Detection and Diagnosis, Wolfson Institute of Population Health, Queen Mary University of London, London, UK
[j] National Screening Service, Dublin, Ireland
[k] UCD School of Public Health, Physiotherapy & Sports Science, Dublin, Ireland
[l] German Cancer Society, Berlin, Germany
[m] CPO-Piedmont - AOU Città della Salute e della Scienza, Torino, Italy
[n] Cancer Registry of Norway, Oslo, Norway
[o] University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[p] Private Group Practice for Pathology, Lübeck, Germany
[q] St. Vincent's University Hospital, Dublin, Ireland
[r] University College Dublin, Ireland
[s] Department of Physics, Università Statale di Milano (UNIMI), Milan, Italy
[t] AUSL Reggio Emilia, IRCCS, Reggio Emilia, Italy
[u] Department of Epidemiology and Global Health, Umeå University, Umeå, Sweden
[v] IQ Health Science department, Radboud University Medical Center, Nijmegen, the Netherlands
[w] Dutch Expert Centre for Screening, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Breast Neoplasms/diagnostic imaging*
Early detection of Cancer*/methods
Female
Mass screening/methods
Program evaluation
Quality indicators, health care/standards*

ABSTRACT

Aim: Evaluation of a breast cancer (BC) screening programme is necessary to ensure its quality. Performance measurements might be prioritized considering the association with outcomes related to BC mortality. We piloted an approach to explore the association of selected performance measurements with incidence-based BC mortality (IBM).

Methods: We performed an ecological analysis of aggregated data from regional or national population-based cancer registries and BC screening programmes in Europe, using 13 performance indicators. We built a panel data (longitudinal cross-sectional) regression model to estimate the association between screening performance measurements and IBM rates.

Results: We included data of 9 programmes and registries from Italy, Spain, Norway, Ireland and the Czech Republic. The number of screening years included in the dataset ranged from 5 to 20 years. In adjusted panel analyses, higher screening coverage, breast cancer detection rates (BCDR prevalent and subsequent rounds),

* Corresponding author. IQ Health science department, Radboud University Medical Center, PO Box 9101, 6500, HB, Nijmegen, the Netherlands.
** Corresponding author. Institut de Recerca Sant Pau (IR Sant Pau), Sant Quintí, 77-79 08041, Barcelona, Spain.
E-mail addresses: palonso@santpau.cat (P. Alonso-Coello), mireille.broeders@radboudumc.nl (M. Broeders).

https://doi.org/10.1016/j.breast.2025.104621
Received 13 January 2025; Received in revised form 17 October 2025; Accepted 21 October 2025
Available online 24 October 2025
0960-9776/© 2025 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

node-negative proportion, and episode sensitivity were associated with lower incidence-based mortality (IBM), whereas a higher interval cancer rate was associated with higher IBM. The association for recall rate in subsequent examinations was small and imprecise.

Conclusion: Our pilot approach suggests association of several performance indicators with IBM. These indicators were related to the implementation of the screening programme (screening coverage), sensitivity (BC detection rate), and efficiency (recall rate). Further studies with larger datasets and individual data may confirm these findings.

1. Introduction

Breast cancer (BC) screening programmes aim ultimately to reduce breast cancer mortality, which has been the primary endpoint in major screening trials and remains the key indicator of programme effectiveness (Tabar 2011). Because mortality reduction is observable only after long-term follow-up, surrogate indicators are commonly used for programme monitoring. Among available approaches, incidence-based breast cancer mortality (IBM) has been introduced as a robust metric that links screening exposure to mortality, thereby providing a more direct assessment of screening impact than all-cause or routine cause-specific mortality (Dibden 2020).

The implementation of population-based screening programmes should include a defined target population, specified screening intervals, testing and follow-up protocol along with a comprehensive clinical and technical quality assurance [1]. In European countries, the implementation of such programmes has steadily reduced the mortality from breast cancer (BC) [2]. To ensure the screening programmes perform as effective as expected requires continuous monitoring and evaluation of their quality [3]. Quality indicators (QI) are constructs that relate to the programme's structure, processes of care or health related outcomes, and are used to monitor, evaluate, and improve the standards of care [4]. We can measure QI through instruments describing the measurable elements (i.e. numerator and denominators) of practice performance (performance measure, PM) and using defined thresholds (performance indicator, PI) (see Box 1) [4].

To assess the impact of mammographic screening programmes on BC mortality demands the follow-up of screened and non-screened cohorts over an extended period of time (i.e. ten or more years of follow-up), and adequate links across data sources (e.g., medical records, pathology, death registries) which often are not readily available or feasible for timely decision-making. Thus, a frequently used alternative to assess BC screening programmes is to monitor performance indicators [5,6], as they reflect the quality of the activities in the continuum of the screening process without directly informing about the impact on BC mortality [7]. Few studies have attempted to document the association of performance indicators with BC mortality. One cross-sectional ecological study compared the BC mortality rates and performance indicators (recall rates, detection rate and screen-detection by stage) across two screening European programmes but it did not account for trends over time [8]. A systematic review including randomised controlled trials of mammography screening found an association between BC mortality with attendance rate and sensitivity [9].

In 2008, the European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis defined a set of common QI, standardised methods for measuring and settled performance indicators thresholds [7]. Later, the European Commission's Joint Research Committee (JRC) established the European Commission Initiative on Breast Cancer (ECIBC) to develop a quality assurance (QA) scheme for the full spectrum of BC services [10,11] and to provide evidence-based guidelines for screening and diagnosis. Then, the ECIBC commissioned a Monitoring and Evaluation subgroup (M&E) with the specific task of refining or developing new QI for monitoring and evaluation of screening activities [12].

In a previous study, we described the multi-stage process used by the M&E group to select 13 performance indicators. This process included a scoping review of the literature, development of a conceptual framework and a survey for the final selection using pre-determined criteria (relevance, measurability, accurateness, ethics and understandability) [12]. In this study, we piloted an approach to prioritize selected performance indicators by identifying which European screening programme's performance indicators are associated with incidence-based BC mortality (IBM).

2. Material and methods

We performed a longitudinal ecological analysis based on BC screening programmes' de-identified aggregated data of women between 50 and 69 years at screening (age group for which the ECIBC strongly recommends population screening) [13]. We requested yearly performance indicators statistics from BC screening programmes of the EU, and IBM rates collected from population-based cancer registry records at country or regional level accessible via the JRC database (European Cancer Information System - ECIS) [14]. Malignant BC incident

Box 1
Definitions of quality indicators in population-based breast cancer screening programmes

1. Quality indicators – QI (also quality markers/quality constructs/quality parameters) are constructs that relate to structures, processes or health status of care, and are used to monitor, evaluate, and improve the quality of care.
2. Performance measures (PM) are tools or instruments to quantify or describe measurable elements of practice performance allowing for rate-based measures (e.g., numerator and denominator) and can be utilized to assess the quality of programmes.
3. Performance indicators (PI) are quantifiable or measurable units or scores of practices. Thresholds, ranges or other targets permits the comparisons and evaluations of practice performance that is directly attributable to a healthcare organisation, a team, or single actor.

Example.
- QI: Participation in screening (related to process)
- PM: Proportion of people participating in screening (numerator) among all the eligible people for whom screening is recommended (denominator)
- PI: setting a threshold for an acceptable participation rate in screening at $>$ 50 %

cases were selected with topography codes C50 of the International Classification of Diseases for Oncology, third edition [15], while deaths due to malignant BC were selected either with code 174 of the International Classification of Diseases, Ninth Revision [16], or code C50 of the International Classification of Diseases, Tenth Revision [17]. Data quality for incidence data was evaluated with QI used to assess ECIS data [18].

2.1. Variables

Incidence-based breast cancer mortality (IBM) was defined as breast cancer deaths among women aged 50–69 years at the time of screening invitation, irrespective of their age at death. Thus, deaths occurring beyond age 69 were included if the tumour was diagnosed following an invitation in the 50–69 age range. A 5-year lag was applied between performance indicators and IBM to account for the expected time between screening and potential mortality impact.

We initially restricted the analysis to the 13 candidate screening indicators previously selected by the M&E group (described elsewhere [12]) (Appendix 1***)***. Through discussion, we designed a direct acyclic graph, describing the theoretical link of the 13 indicators between each other and with the final health outcomes (Appendix 2). Finally, we only included in the analysis those indicators with a plausible causal or intermediate link to IBM; thus, mastectomy rate (depends largely on women's preferences or value) and benign open biopsy rate (an adverse event by definition) were not included in the analysis (Njor 2012, Broeders 2005).

2.2. Inclusion criteria

We included data from screening programmes: 1) that covered the same regional or country territorial area of an existing ECIS cancer registry; 2) where the data from the cancer registry covered a period of 5 years or longer after implementation of the screening programme; 3) where information about the cause of death (i.e. BC related) was available in ECIS for the corresponding cancer registry of each screening programme. We confined our analysis to data from screening programmes represented in the M&E group in order to ensure better communication and interpretation of the results.

2.3. Statistical analysis

We built a panel data linear regression model, which allows the inclusion of data that span both in time and observation units. Our dependent variable was annual IBM expressed per 100,000 women-years, and the independent variables were all performance indicators. We considered a 5-year time lag between the performance indicator and IBM, i.e. for a given indicator with data at year 2000 the corresponding IBM was at year 2005. Our observation unit was the population covered by each ECIS cancer registry that overlapped with the population targeted by the screening programme, from which we derived the data on IBM. The units varied in population size and geographical area, as they are not homogeneously defined in each country or region in Europe.

Given that screening programmes began at different calendar periods, we considered for all the units, the year of inception of the programme as "year 1", and subsequently numbered the next years progressively for the analysis (i.e. for a programme starting in 2000, this will be the first year). We first estimated the regression coefficient of each of the 11 candidate performance indicators with IBM independently, across units and screening years, then we adjusted each analysis by age, and by the proportion of subsequent screening examinations for screening coverage, recall rate, and the three indicators related to breast cancer detection. We used the robust procedure to estimate standard errors and control for heteroskedasticity as correlation across units was foreseen at country or region level [19].

To model the data, we developed two different approaches in order to test the most appropriate structure: a) fixed effects model (FE), and b) random effects model (RE). The FE model uses the same unit at different time points as its own comparator to control for potential (unmeasured) biases which in practice removes the effect of time-invariant variables (i.e. country income level). The RE model uses all the information available (across units and between time points) to allow us to estimate the effects of time-invariant variables (which otherwise are absorbed in the FE model), but generally yielding larger standard errors.

After the two models were developed, we formally assessed which model fitted the cross-sectional data better. We used the Hausman test to assess the adequacy of fixed versus random effects models to fit the panel data [20,21]. Consequently, all results were reported based on the random effects model, capturing both within- and between-programme variation, allowing us to assess differences across programmes, which would be removed under a fixed-effects specification.

The selection of variables to be included in the models was guided by a directed acyclic graph (DAG, Appendix 2). For each performance measure, we reported one unadjusted model and a second model adjusted for age structure and first (prevalent) or subsequent screening examination.

We adopted an estimation-focused approach. Adjusted coefficients (b) are reported with 95 % confidence intervals and exact p-values; no pre-specified p-value threshold was used to accept or reject null hypotheses. Because sample size was determined by data availability across programmes, no formal power calculation was conducted. Interpretation prioritized the magnitude and precision of associations and their consistency across sensitivity analyses.

3. Results

We obtained data from nine screening programmes and their corresponding cancer registries: three national programmes (Ireland, Norway, Czech Republic) and six regional programmes, three from Italy (Bergamo, Reggio Emilia and Sondrio) and three from Spain (Girona, Pais Vasco and Tarragona). Initially, we included five screening programmes from Italy, but two programmes (Palermo, Viterbo) were later excluded due to inconsistent data reporting. Two screening programs provided longer periods of observations, 18 and 21 years, from the programmes of Norway and the regional program of Navarra (Spain), respectively. Across the time periods reported at each cancer registry, the annual rate of newly diagnosed BC cases among women 50–69 years ranged from 163 to 322 per 100,000 at the last year included in the analysis (Table 1).

Overall, we observed a decreasing trend in IBM over time across most programmes. We presented the bivariate plots for each screening performance indicator and IBM per year of screening activity in Appendix 5. There was a high heterogeneity in the trend between each indicator and IBM. However, for screening coverage and recall rate, there was a visual pattern of correlation with IBM (Appendix 5 and Table 2).

We described the estimates for each screening performance indicator per programme and year in Appendix 4**.** The unadjusted panel regression analysis suggested an association with IBM for five indicators: screening coverage (b= −0.26; 95 %CI -0.33 to −0.19), recall rate (b= 1.99; 95 %CI 1.03 to 2.95), breast cancer detection rate (b= 2.61; 95 % CI 0.72 to 4.49), invasive breast cancer detection rate (b= 3.09; 95 %CI 0.49 to 5.69), and invasive cancer >20 mm (b= 7.67; 95 %CI -0.27 to 15.62) (Appendix 3).

We found that the proportion of subsequent screening examinations might be associated to IBM in a negative direction (b= −0.16; 95 %CI -0.21 to −0.11), therefore we decided to include this variable in an adjusted model (or, if data was available, conduct a stratified analysis). The adjusted analysis found that increasing the performance indicators of screening coverage (b= −0.17; 95 %CI -0.33 to −0.02), breast cancer detection-prevalent examination (b= −0.59; 95 %CI -0.79 to −0.39), breast cancer detection-subsequent examination (b= −2.61; 95 %CI -4.66 to −0.56), lymph node status (b= -0.40; 95 %CI -0.73 to −0.07)

Table 1
Characteristic of the included screening programmes with biennial screening interval in the analysis[a].

	Czech Republic	Ireland	Bergamo (Italy)	Reggio Emilia (Italy)	Sondrio (Italy)	Norway	Girona (Spain)	Navarra (Spain)	Tarragona (Spain)
Start-up of the program	2002	2001	2001	1994	2001	1995–2005	2001	1990	1998
Target age	45+	50–69	45–74	45–74	45–74	50–69	50–69	50–69	50–69
IBM, years of available data	1999 to 2009	1997 to 2009	2007 to 2010	2003 to 2011	2003 to 2010	1991 to 2014	1996 to 2011	1987 to 2010	1995–2009
Indicators, years of available data	2002 to 2013	2001 to 2013	2006 to 2017	2006 to 2017	2006 to 2015	1996 to 2013	2002 to 2018	1990 to 2017	1998 to 2011
Years included in the analysis (n)	2002 to 2009 (8)	2001 to 2009 (9)	2007 to 2010 (5)	2006 to 2011 (6)	2006 to 2010 (5)	1996 to 2013 (18)	2002 to 2011 (10)	1990 to 2010 (21)	1998 to 2009 (12)
Population (last year included)	1413363	859,156	128,314	61,254	22,775	584,813	81,206	70,381	83,450
Annual BC incidence (last year included)[b,c]	217	163	268	322	303	284	192	230	229
Annual IBM (x 100,000) (last year included)	24.1	12.7	17.2	16.3	13.2	12.8	16.4	2.8	10.8

IBM = incidence-based breast cancer mortality.
[a] Data on population, breast cancer incidence and IBM include only the group from 50 to 69 years old.
[b] Mean incident cases per 100,000 women.
[c] The rates are estimated using data from population included in the European Cancer Information System (ECIS), depending on the implementation phase of the screening programme it may represent a lower geographical area, i.e. during 1995–2005, BC screening was not nationwide in Norway.

Table 2
Range of reported performance indicators over the time period included in the analysis, range[a].

Indicator, minimum and maximum values over the period	Regions (country) and years included in the analysis								
	Czech Republic 2002 to 2009	Ireland 2001 to 2009	Bergamo (Italy) 2007 to 2010	Reggio Emilia (Italy) 2006 to 2011	Sondrio (Italy) 2006 to 2010	Norway 1996 to 2013	Girona (Spain) 2002 to 2011	Navarra (Spain) 1990 to 2010	Tarragona (Spain) 1998 to 2009
Screening coverage (%)	1.1; 50.6	38.8; 68.1	57.7; 62.5	71.8; 80.6	61.4; 65.3	66.6; 73.9	43.3; 80.2	96.1; 98.4	8.1; 66.2
Participation rate (%)	NA	69.5; 78.3	59.0; 62.9	78.5; 81.1	59.2; 72.4	73.1; 80.5	51.8; 62.2	84.0; 89.5	49.8; 66.9
Recall rate (%)	10.0; 18.7	3.9; 4.6	3.4; 5.5	2.8; 3.9	3.1; 4.5	2.6; 4.3	7.4; 13.5	3.0; 19.0	0.4; 3.8
Breast cancer detection rate (per 1000)	5.1; 5.6	4.8; 8.0	3.4; 5.0	5.1; 6.0	2.9; 4.5	5.2; 6.8	4.0; 7.1	3.3; 7.0	2.0; 4.3
Invasive breast cancer detection rate (per 1000)	4.4; 5.0	4.0; 5.7	2.9; 4.3	4.2; 4.8	2.6; 4.5	4.0; 5.4	3.3; 4.8	2.1; 4.1	2.0; 3.7
Invasive cancer >20 mm (per 1000)	0.6; 1.7	0.5; 1.4	0.9; 1.0	0.5; 0.8	0.3; 0.7	0.7; 1.0	0.1; 0.8	1.2; 3.4	0.4; 1.2
Invasive cancer ≤10 mm (%)	23.5; 37.7	25.3; 30.1	30.2; 36.4	35.7; 50.4	25.0; 55.6	31.7; 38.5	14.1; 32.6	30.2; 56.3	17.9; 40.0
Lymph node status negative (%)	63.7; 70.6	67.1; 73.0	63.4; 71.1	63.2; 77.6	70.0; 100	72.2; 82.6	62.2; 76.3	61.0; 84.8	49.4; 84.3
Interval cancer (per 1000)	NA	1.3; 2.3	NA	NA	NA	1.6; 2.1	0.4; 1.7	0.7; 1.5	0.1; 1.4
Episode sensitivity (%)	NA	73.4; 84.0	NA	NA	NA	71.4; 79.7	75.4; 91.5	72.7; 91.7	71.4; 97.8

NA = not data available.
IBM=Incidence based mortality.
[a] The minimum and maximum values may have occurred over different years or without a common pattern across different registries. Please refer to appendix 4 for a detail report of performance indicators by screening years.

and episode sensitivity (*b*= -0.56; 95 %CI -0.86 to −0.25) may be associated to a reduction on IBM; while an increase on recall rate-subsequent examination, and interval cancer rate might be associated to an increase on IBM (Table 3).

4. Discussion

4.1. Main findings

Our analysis of aggregated data from organised screening programmes in the EU aimed to test an approach to identify performance measures potentially associated with IBM. After controlling by type of screening examination, we observed a potential association of IBM with measures of screening coverage, breast cancer detection-prevalent round, breast cancer detection-subsequent round, lymph node status, episode sensitivity, recall rate-subsequent round, and interval cancer rate. Organized screening programmes might consider this approach to prioritize performance measures for the quality assessment of programme implementation (i.e. stratifying performance measures by type of examination during assessment).

There is evidence from observational studies on the effectiveness of service screening programmes of breast cancer after their implementation [2]. For example, aggregated data at national level suggests a decrease of 19 %–24 % in breast cancer mortality among women invited to screening in the age group 50–64 years or 50–59 years respectively [6, 12]. A comprehensive review of 36 cohort studies, 17 case-control studies and 7 RCT showed a reduction in breast cancer mortality comparing attenders versus non-attenders to the organized screening programs ranging from 33 % to 43 % in Northern Europe, 43 %–45 % in Southern Europe and 12 %–58 % Western Europe [2].

A systematic review of nine RCTs reanalysed the impact of the attendance rate and quality indicators, i.e. proportional incidence rate and over-detection, on the outcomes of interest, i.e. advanced stage breast cancer and breast cancer mortality [9]. The authors estimated a

Table 3
Panel data estimates for multivariate analysis[a].

Indicator (regressor)[b,c]	Coefficient (*b*)	95 % CI	P value
Screening coverage (%)	*−0.17*	*−0.33 to -0.02*	*0.036*
Participation rate (%)	0.43	−0.36 to 1.21	0.220
Recall rate (%)- prevalent examinations	0.61	−1.22 to 2.44	0.464
Recall rate (%)- subsequent examinations	*0.07*	*−0.16 to 0.35*	*0.068*
*Breast cancer detection rate – prevalent examinations (*per *1000)*	*−0.59*	*−0.79 to -0.39*	*<0.001*
*Breast cancer detection rate – subsequent examinations (*per *1000)*	*−2.61*	*−4.66 to -0.56*	*0.020*
Invasive breast cancer detection rate (per 1000)	0.63	−2.81 to 4.07	0.670
Cancers >20 mm (per 1000)	−1.84	−12.15 to 8.49	0.679
Cancers ≤10 mm (%)	0.19	−0.27 to 0.65	0.345
Lymph node status (%)	*−0.40*	*−0.73 to -0.07*	*0.024*
*Interval cancer rate (*per *1000)*	*6.63*	*1.01 to 12.25*	*0.037*
Episode sensitivity (%)	*−0.56*	*−0.86 to -0.25*	*0.016*

[a] Each model included only age and proportion of subsequent screening in the adjusted analysis. In the final models (showed here), age was excluded as the p value was >0.4 for each performance measures.

[b] We did not have individual data for age, but the frequency of women for specific age groups (i.e. 50 to 55). We imputed a mean age group by taking a pondered mean assuming the middle age value for each reported age groups.

[c] The coefficient represents how much the dependent variable is expected to increase (if positive) or decrease (if negative) when the independent variable increases by one unit.

33 % (95 % CI: 24–42 %) and 13 % (95 % CI: 6–20 %) BC mortality reduction for the best scenario (90 % attendance rate and 95 % sensitivity) and the poorest scenario (30 % attendance rate and 55 % sensitivity), respectively. Those results also suggested that for ensuring the effectiveness of screening programmes minimum thresholds should be defined for indicators linked to participation or sensitivity (i.e. BC detection, interval cancer). Given the limitation of our data, we did not attempt to explore potential thresholds in our analysis at this stage.

Observational studies attempting to relate performance indicators to IBM are sparse. One study from Finland did not find an association between IBM and screening centres categorized at low, intermediate and high levels of recall rate [5], while results from the same screening programme from 1991 to 2003 reported a significant increase in the incidence of interval cancers by ascending recall rate categories [20]. Of note, studies of organised mammography screening from the Netherlands have reported low recall rates (0.7–1.3 %), reduced rates of advanced cancers, and similar impact on breast cancer mortality compared to studies from Finland [5,21]. The potential value of screening performance indicators in predicting the extent of reduction in BC mortality seems to be limited from these studies, although the studies themselves are limited by their cross-sectional methodology and inclusion of fewer programmes compared to our analysis.

4.2. Limitations and strengths

The use of aggregated programme-level data implies that individual-level confounding factors could not be accounted for, raising the risk of ecological fallacy and unmeasured confounding. This issue is particularly relevant when analysing time trends, as changes in risk factors, treatment protocols, or socioeconomic conditions may lead to spurious correlations. Moreover, information on breast density, cancer stage, and treatment—key determinants of breast cancer mortality—was not harmonised or consistently available across programmes. In this context, although we applied panel data analysis to incorporate observations over time and conducted multivariate analyses to account for collinearities among aggregate variables, the potential for residual confounding remains. These estimates should therefore be regarded as a first step towards exploring associations between performance indicators and breast cancer mortality. More generally, ecological designs provide a valid approach for hypothesis generation, but cannot yield definitive causal conclusions; further research with individual-level data from screening programmes will be required to validate these findings.

Mortality outcomes largely reflect data collected up to 15 years ago; while these historical datasets remain valid for studying associations, more recent evidence will be needed to confirm whether the same patterns persist today. In addition, some programmes lacked information on specific indicators, and not all population-based registries reported breast cancer mortality to ECIS, resulting in an unbalanced panel that may reduce the certainty of estimates.

Considerable heterogeneity was observed across countries and programmes, reflecting differences in implementation, data quality, treatment availability, and underlying population risk. Although we used panel regression models with random effects to address these issues, our estimates should be interpreted as exploratory, providing a first step towards identifying performance indicators that may be associated with breast cancer mortality.

The findings highlight which indicators could be useful for monitoring the performance of population-based programmes. However, it is important to note that causal links cannot be established for each individual indicator, and ecological fallacy may limit causal inference from aggregated data. Nevertheless, the proposed indicators are supported by evidence from large population-based trials. Importantly, this is the first study to combine data from breast cancer screening programmes across the EU with data from European population-based cancer registries. As such, the associations identified provide a preliminary basis for further exploration of acceptable and desirable levels for candidate performance indicators and can help prioritize their assessment in larger cohorts with individual-level data.

Finally, the generalisability of our findings is limited. The study is based on European population-based screening programmes, which share certain organisational models and quality assurance frameworks. Caution is needed when extrapolating these results to non-European settings, where health systems, resources, population risk profiles, and screening practices may differ substantially. Nevertheless, these are intrinsic limitations of monitoring indicators. The aim of this study was to test whether, despite these limitations, monitoring indicators were able to identify the impact of screening.

4.3. Implications for practice and research

The relation between the estimates of screening performance measures and IBM as a long-term outcome of mammography screening is not straightforward. While estimates of effectiveness reducing BC mortality remain as the best outcomes for evaluating the success of mammography screening programmes, the approach implemented in this study aimed to identify promising candidate performance indicators potentially associated to IBM. The findings suggest which indicators may be useful for monitoring the performance of population-based programmes, but it must be always considerate that causal link in each single case cannot be inferred and ecological fallacy alway affect causal inference from aggregated indicator results. Nevertheless, the proposed indicators grounds on the results of large population based trials.

Further research is needed to understand whether more complete, better harmonised, individual-level, and up-to-date data would overcome some of the limitations observed in this study particularly to disentangle contextual effects such as treatment and stage, and to establish acceptable and desirable levels for candidate indicators to optimise monitoring". Shared clinical and technical quality controls are

also essential to ensure uniform screening programme implementation methodology and to ensure greater uniformity of epidemiological information [10].

CRediT authorship contribution statement

Carlos Canelo-Aybar: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Pablo Alonso-Coello:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Sergei Muratov:** Writing – review & editing, Conceptualization. **Jean-Eric Tarride:** Writing – review & editing, Conceptualization. **Nadya Dimitrova:** Writing – review & editing, Writing – original draft, Project administration, Data curation, Conceptualization. **Francesco Giusti:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Bettina Borisch:** Writing – review & editing, Conceptualization. **Xavier Castells:** Writing – review & editing, Data curation, Conceptualization. **Stephen W. Duffy:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Patricia Fitzpatrick:** Writing – review & editing, Data curation, Conceptualization. **Markus Follmann:** Writing – review & editing, Conceptualization. **Livia Giordano:** Writing – review & editing, Conceptualization. **Solveig Hofvind:** Writing – review & editing, Data curation, Conceptualization. **Annette Lebeau:** Writing – review & editing, Conceptualization. **Cecily Quinn:** Writing – review & editing, Conceptualization. **Alberto Torresin:** Writing – review & editing, Conceptualization. **Paolo Giorgi Rossi:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Holger Schünemann:** Writing – review & editing, Methodology, Conceptualization. **Lennarth Nyström:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Mireille Broeders:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization.

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Ethical approval

Ethical approval for this study was not required.

Funding information

The work was funded and coordinated by the European Commission and its Joint Research Centre (JRC) in Ispra, Italy, within the scope of the project European Commission Initiative on Breast Cancer (ECIBC). Other than the scientific/technical contribution of individual JRC employees (Dr. Dimitrova, Dr. Giusti), the funding body did not have any other input into the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. This research did not receive any specific grant from other sources.

Declaration of competing interest

Nadya Dimitrova was staff of the Joint Research Centre during the period of the study and coordinated a working group of experts which did the study.

Stephen W. Duffy carries out occasional paid consultancy for the biotechnology company GRAIL and the pharmaceutical company Roche.

The remaining authors declare that they have no conflict of interest.

Acknowledgments

Our thanks go in particular to the colleagues from the European Commission's Joint Research Centre for their coordination and scientific support, including making available the data sources from the European Cancer Information System (ECIS).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.breast.2025.104621.

Data availability

The data analyzed in this study is subject to the following licenses/restrictions: requests to access the datasets should be directed to: JRC-CANCER-POLICY-SUPPORT@ ec.europa.eu.

References

[1] Lynge E, et al. Determinants of successful implementation of population-based cancer screening programmes. Eur J Cancer 2012;48(5):743–8.
[2] Zielonke N, et al. Evidence for reducing cancer-specific mortality due to screening for breast cancer in Europe: a systematic review. Eur J Cancer 2020;127:191–206.
[3] Broeders M, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. J Med Screen 2012;19 (Suppl 1):14–25.
[4] Parmelli E, et al. Guideline-based quality assurance: a conceptual framework for the definition of key elements. BMC Health Serv Res 2021;21(1):173.
[5] Sarkeala T, Heinavaara S, Anttila A. Organised mammography screening reduces breast cancer mortality: a cohort study from Finland. Int J Cancer 2008;122(3): 614–9.
[6] Broeders MJM, et al. The impact of mammography screening programmes on incidence of advanced breast cancer in Europe: a literature review. BMC Cancer 2018;18(1):860.
[7] Perry N, et al. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition–summary document. Ann Oncol 2008;19(4):614–22.
[8] Sarkeala T, et al. Validity of process indicators of screening for breast cancer to predict mortality reduction. J Med Screen 2005;12(1):33–7.
[9] Chen TH, et al. Clarifying the debate on population-based screening for breast cancer with mammography: a systematic review of randomized controlled trials on mammography with Bayesian meta-analysis and causal model. Medicine (Baltim) 2017;96(3):e5684.
[10] Deandrea S, et al. Italian guidelines for age range and test interval in breast cancer screening programmes: GRADE-ADOLOPMENT of the European guidelines. Radiol Med 2024;129(12):1778–89.
[11] EU. Committee note. Council Conclusions on reducing the burden of cancer (9636/08). 2008 1 November 2024]; Available from: http://register.consilium.europa.eu/doc/srv?l=EN&f=ST%209636%202008%20INIT.
[12] Muratov S, et al. Monitoring and evaluation of breast cancer screening programmes: selecting candidate performance indicators. BMC Cancer 2020;20(1): 795.
[13] Schunemann HJ, et al. Breast cancer screening and diagnosis: a synopsis of the European Breast Guidelines. Ann Intern Med 2020;172(1):46–56.
[14] ECIS, Source: ECIS - European Cancer Information System From https://ecis.jrc.ec.europa.eu, accessed on day/month/year © European Union, 2023. 2023.
[15] WHO, World Health Organization (WHO) (2013). International Classification of Diseases for Oncology. . 3rd ed. (ICD-O-3) ed. 2013, Geneva: World Health Organization.
[16] WHO, International classification of diseases : [9th] ninth revision, basic tabulation list with alphabetic index. 1978, Geneva: World Health Organization.
[17] WHO, ICD-10 : international statistical classification of diseases and related health problems : tenth revision. 2nd. edition ed. 2004, Geneva: World Health Organization.
[18] Giusti F, et al. Quality indicators: completeness, validity and timeliness of cancer registry data contributing to the European Cancer Information system. Front Oncol 2023;13:1219128.
[19] Hoechle D. Robust standard errors for panel regressions with cross-sectional dependence. STATA J 2007;7(3):281–312.
[20] Sarkeala T, et al. Episode sensitivity in association with process indicators in the Finnish breast cancer screening program. Int J Cancer 2006;118(1):174–9.
[21] Otto SJ, et al. Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. Lancet 2003;361(9367):1411–7.