Economic Sciences
https://economic-sciences.com
ES(2026) 22, 2 (2) , 146-154 | ISSN:1505-4683

Birth Sex-Ratio Imbalance and Global Breast Cancer Mortality: A Robustness-Focused Ecological Panel Analysis

Anikaa Sharma
Independent Researcher, Los Angeles, USA
Sharma4anikaa@gmail.com

Abstract

This study evaluates whether birth sex-ratio imbalance has an independent ecological association with breast cancer mortality in country-year panel data. Country-year data for 1990-2023 were assembled from GBD/IHME, World Bank, and WHO Global Health Observatory sources. The manuscript framing and literature audit used a multi-perspective workflow: question generation, source-grounded evidence curation, outline synthesis, and blind-spot review. The empirical analysis used a descriptive nested ordinary least squares model ladder, country-clustered standard errors, country/year fixed-effects robustness checks, lagged predictors, registry-quality sensitivity analysis, and leave-one-country-out validation. Adding SRB imbalance produced a very small incremental gain beyond incidence and standard ecological adjustment variables (R^2 gain = 0.000783). SRB was statistically significant under heteroskedasticity-robust standard errors, but not under country-clustered inference ($p = 0.160548$). SRB also remained non-significant after observed adjustment covariates, two-way country/year fixed effects, lagged temporal checks, and alternative natural-baseline assumptions of 1.03, 1.05, and 1.07. Overall, SRB provided at most a small, non-robust ecological signal and should not be interpreted as an independent breast cancer mortality determinant.

Keywords: *breast cancer, mortality, sex ratio at birth, machine learning*

I. INTRODUCTION

Breast cancer mortality varies substantially across countries. Prior studies link these differences to incidence, development level, stage at diagnosis, screening access, registry quality, treatment availability, and broader health-system capacity [3], [9]. Mortality is not determined by incidence alone: countries with similar incidence can have different mortality because of differences in early detection, diagnostic capacity, treatment access, affordability, continuity of care, and completeness of mortality registration. The WHO Global Breast Cancer Initiative similarly emphasizes early detection, timely diagnosis, and treatment completion as major levers for reducing breast cancer mortality [10].

Birth sex-ratio imbalance is usually studied as a demographic or contextual indicator, not as a cancer-specific risk factor. It may reflect social, biological, environmental, or population-level conditions, including sex-selective practices, population composition, environmental exposure patterns, or broader gender-related structures [4], [5]. These mechanisms do not imply that SRB directly affects breast cancer mortality. Instead, SRB may function as a country-level contextual marker correlated with unmeasured health-system, surveillance, or socioeconomic factors.

The novelty of this study is therefore deliberately cautious. It does not claim that SRB causes breast cancer mortality or that SRB is a major epidemiologic mechanism. Rather, it tests whether an apparent SRB association survives progressively stricter adjustment for incidence, socioeconomic conditions, health-system context, repeated country observations, fixed effects, and temporal ordering. To avoid a single-angle interpretation, the paper was revised using a multi-perspective research workflow that explicitly considered epidemiologic, demographic, health-system, econometric, data-quality, and skeptical/null-result perspectives before finalizing the outline and interpretation. In this framing, a weak or non-robust SRB result is informative because it distinguishes a superficial ecological correlation from a stable independent association.

Economic Sciences
https://economic-sciences.com
ES(2026) 22, 2 (2) , 146-154 | ISSN:1505-4683

II. DATA AND VARIBLE

This study used public country-year aggregate data from GBD/IHME, World Bank, and WHO Global Health Observatory sources for 1990-2023. Breast cancer mortality and incidence were obtained from GBD/IHME. Socioeconomic and demographic adjustment variables included GDP per capita, female life expectancy, fertility rate, urbanization, and sex ratio at birth from the World Bank. SRB imbalance was defined as |SRB - 1.05|, where 1.05 represents the approximate expected natural sex ratio at birth.

Observed health-system indicators were added only when public source values matched the same ISO3 country code and year. This included health expenditure per capita, physicians per 1,000 people, hospital beds per 1,000 people, out-of-pocket expenditure as a percentage of current health expenditure, UHC service coverage, and nurses/midwives per 1,000 people. No K-nearest-neighbor filling, interpolation, forward-fill, backward-fill, or synthetic value generation was used.

Several cancer-specific WHO indicators were available but sparse in exact country-year panel form, so they were retained for coverage transparency and sensitivity/data-quality checks rather than main adjustment. In the active augmented dataset, radiotherapy unit density was observed for 349 rows across 172 countries, national cancer-registry existence for 738 rows across 189 countries, and population-based registry existence for 191 rows across 187 countries.

Cancer-registry indicators were also used as an external data-quality screen. Countries with any observed national or population-based cancer-registry indicator equal to one were classified as registry-confirmed for sensitivity analysis; countries with only observed zero values or no registry-quality observation were treated as weaker or unknown registry-data settings.

The primary full explanatory model used 6,763 complete country-year observations. In the active compact Q2 dataset, the core primary variables used in Models 1-3 had no excluded rows, so the incidence-only, socioeconomic, and SRB model ladder was estimated on the same country-year sample. A model with observed hospital beds, out-of-pocket expenditure, UHC, and nurses/midwives retained 2,232 observations across 170 countries. The broad health-system robustness model retained 2,057 observations across 167 countries. This design separates breadth from specificity: the broad model tests whether SRB appears at maximum coverage, while health-system models test whether the signal persists in observed-data subsets with stronger adjustment.

III. METHODS

All data were publicly available country-level aggregates and did not involve individual-level human-subject data. The analysis used complete-case regression for each specified model. Reproducible scripts generate the augmented dataset, regression outputs, validation summaries, tables, and figures from the source CSV files.

A. MULTI-PERSPECTIVE LITERATURE AND FRAMING WORKFLOW

Before final manuscript revision, the literature framing and interpretation were reorganized using a multi-perspective workflow. Topic coverage was broadened by identifying multiple relevant perspectives, asking source-grounded questions from each perspective, curating the resulting evidence, synthesizing an outline, and checking for missing angles, source bias, and unsupported associations.

This study adapted that process as a manuscript-development and evidence-audit step rather than as an automated source-generation system. Six perspectives were used to stress-test the paper's framing: breast cancer epidemiology, SRB/demography, health-system access and quality, ecological/panel econometrics, data-quality and registry surveillance, and a skeptical null-results perspective. For each perspective, the revision asked what assumptions that expert lens would challenge, what evidence would be needed, and whether the paper's current wording overstated SRB as a causal or independent determinant.

Economic Sciences
https://economic-sciences.com
ES(2026) 22, 2 (2) , 146-154 | ISSN:1505-4683

The multi-perspective audit led to three manuscript decisions. First, known breast cancer mortality determinants were treated as adjustment context rather than competing focal discoveries. Second, SRB was framed as a contextual demographic marker, not as a biological cancer-risk factor. Third, the interpretation emphasized robustness, clustered inference, fixed effects, registry-quality sensitivity, and temporal checks instead of relying on a single significant pooled OLS coefficient. This workflow improved breadth and organization, but the reported empirical findings come only from the reproducible statistical pipeline described below.

This study was designed as a descriptive ecological panel analysis, not a causal identification study. No instrumental variable, natural experiment, difference-in-differences design, or regression-discontinuity strategy was available, so all estimates are interpreted as adjusted associations rather than causal effects or policy effects.

The descriptive model ladder used nested OLS regression:

$$Mortality_{i,t} = \beta_0 + \beta_1 (SRB\ Imbalance_{i,t}) + \beta_2(Incidence_{i,t}) + \beta_3 (GDP\ per\ Capita_{i,t}) + \beta_4 (Life\ Expectancy_{i,t}) + \beta_5 (Fertility\ Rate_{i,t}) + \beta_6(Urbanization_{i,t}) + \varepsilon_{i,t}$$

The model ladder compared an incidence-only baseline, a socioeconomic model, a socioeconomic model plus SRB, and health-system robustness models. The model sequence and sample-size pattern are shown in Figure 1 and detailed in the Results section. Because the dataset contains repeated annual observations for countries, heteroskedasticity-robust standard errors were treated as preliminary; country-clustered standard errors provided the more conservative SRB inference [8].

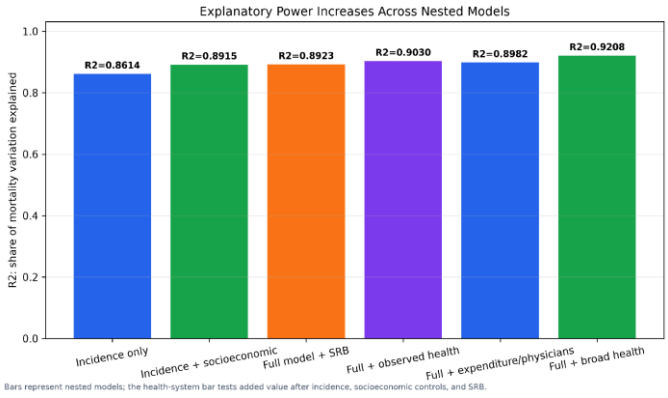

Figure 1. Nested model R2 comparison

Robustness checks assessed whether SRB persisted after health-system adjustment, registry-quality restriction, country/year fixed effects, lagged predictors, alternative SRB baselines, GBD uncertainty bounds, and first-difference sensitivity. Country fixed effects absorb stable country characteristics; year fixed effects absorb global period shocks. Lagged predictors improve temporal ordering but do not establish causality.

Missingness and coverage checks were used to avoid overstating sparse cancer-service variables as complete patient-level controls.

Supplementary panel diagnostics evaluated residual dependence and time-series persistence, including serial correlation, cross-sectional dependence, unit-root behavior, and within-country coefficient stability. Spatial Moran/SAR and Arellano-Bond GMM were retained as code templates only because the current dataset lacks spatial weights and the local environment does not include the required dynamic-panel package. Standardized coefficient ranking is shown in Figure 2 as a descriptive comparison; conservative SRB inference is based on clustered and fixed-effect robustness checks.

Economic Sciences
https://economic-sciences.com
ES(2026) 22, 2 (2) , 146-154 | ISSN:1505-4683

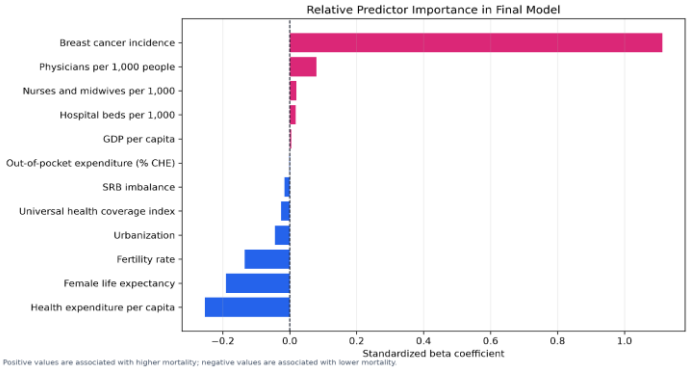

Figure 2. Standardized coefficient ranking

Leave-one-country-out validation trained the explanatory model on all countries except one and predicted mortality for the excluded country during 2016-2023. CatBoost and XGBoost were included only as supplementary forecasting models because they used lagged mortality and therefore primarily test temporal predictability rather than causal explanation.

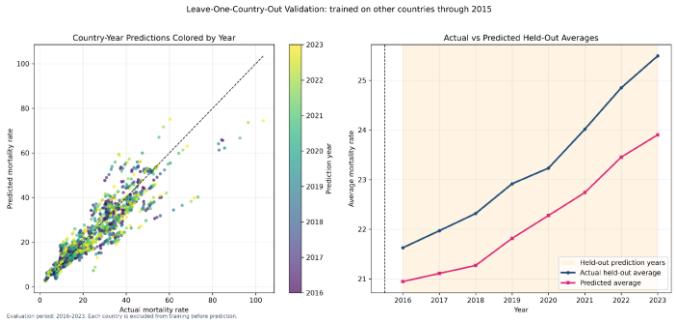

Figure 3. Leave-one-country-out actual vs predicted mortality

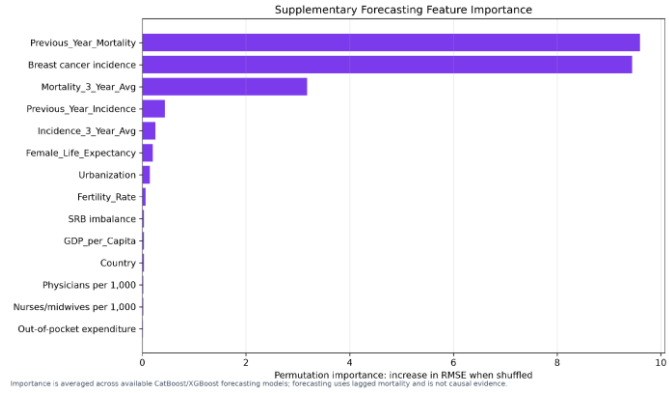

Figure 4. Supplementary forecasting permutation importance

IV. RESULTS

A. EXPLANATORY MODEL LADDER

The model ladder was used to assess whether the SRB association persisted under progressively broader ecological adjustment. SRB added only a very small increment in the broad baseline model, and health-system models served as robustness checks in smaller observed-data subsets.

Economic Sciences
https://economic-sciences.com
ES(2026) 22, 2 (2) , 146-154 | ISSN:1505-4683

MODEL	ROWS	R^2	ADJ. R^2	GAIN
M1: Incidence only	6,763	0.861435	0.861414	NA
M2: + socioeconomic controls	6,763	0.891490	0.891410	0.030056
M3: + SRB	6,763	0.892274	0.892178	0.000783
M4: beds + OOP + UHC + nurses	2,232	0.903025	0.902588	NA
M5: health exp, physicians	2,875	0.898233	0.897949	NA
M6: all broad health-system controls	2,057	0.920799	0.920334	NA

B. SRB ATTENUATION

The central result is the attenuation of SRB across increasingly conservative model stages. SRB is detectable only in the simpler robust-OLS specification and disappears after country clustering, observed capacity/affordability controls, broader health-system adjustment, fixed effects, or lagged temporal checks.

ROBUSTNESS CHECK	ROWS	R^2	SRB TERM TESTED	SRB P-VALUE
Primary full model, HC3 SE	6,763	0.892274	SRB imbalance	6.51e-12
Country-clustered SE	6,763	0.892274	SRB imbalance	0.160548
Beds + OOP + UHC + nurses	2,232	0.903025	SRB imbalance	0.761905
Health expenditure + physicians	2,875	0.898233	SRB imbalance	0.905731
All broad health-system controls	2,057	0.920799	SRB imbalance	0.499744
Country fixed effects	6,763	0.981691	SRB imbalance	0.099296
Health + country fixed effects	2,875	0.990479	SRB imbalance	0.089034
Country + year fixed effects	6,763	0.982892	SRB imbalance	0.175718
Broad health + country + year fixed effects	2,057	0.992450	SRB imbalance	0.548640
Lagged + year fixed effects	2,685	0.896496	Lagged SRB imbalance	0.718952

In realistic units, the broad-model SRB coefficient implies that a 0.01 increase in SRB imbalance is associated with approximately +0.266 mortality-rate units before clustering or health-system adjustment. This effect is small and not stable under stricter inference.

As a data-quality sensitivity check, the core model was re-estimated after excluding countries without registry-confirmed data quality. In registry-confirmed countries only, the core model retained 5,049 rows across 148 countries and SRB was not significant under country-clustered inference (p =

Economic Sciences
https://economic-sciences.com
ES(2026) 22, 2 (2) , 146-154 | ISSN:1505-4683

0.763615). In the health-system robustness subset, registry-confirmed countries retained 1,993 rows across 138 countries and SRB also remained non-significant (p = 0.501792).

The missingness comparison showed that all 6,763 rows in the active compact Q2 dataset were complete for the core primary model variables. SRB baseline sensitivity also supported the same conclusion: SRB was not significant under country-clustered inference when the imbalance baseline was set to 1.03, 1.05, or 1.07.

SRB BASELINE	ROWS	R^2	SRB P-VALUE
1.03	6,763	0.891863	0.333803
1.05	6,763	0.892274	0.160548
1.07	6,763	0.891768	0.372357

C. SUPPLEMENTARY PANEL DIAGNOSTIC

Additional panel diagnostics supported the conservative interpretation. The Wooldridge-style residual AR(1) check indicated within-country serial correlation, and the Pesaran CD check indicated cross-sectional residual dependence. The Fisher-combined ADF check did not reject the unit-root null for mortality in the tested country series. These diagnostics strengthen the rationale for country-clustered inference, fixed-effects checks, and cautious non-causal interpretation rather than strengthening the SRB association.

DIAGNOSTIC	STATISTIC	P-VALUE	INTERPRETATION
Wooldridge-style residual AR(1)	226.846	< 0.001	Within-country serial correlation present
Pesaran CD residual check	22.592	< 0.001	Cross-sectional residual dependence present
Fisher-combined ADF panel check	173.407	1.000	Unit-root null not rejected

The Mundlak-style comparison produced a non-significant SRB within-country estimate, consistent with the broader conclusion that SRB is not a robust independent ecological signal. Spatial Moran/SAR and Arellano-Bond GMM were not treated as completed analyses because they require spatial weights or dynamic-panel tooling not available in the current reproducible environment.

D. PARAMETRIC EFFECTS

Standardized coefficients show that mortality is dominated by incidence and health-system/socioeconomic conditions, not SRB. In the broad health-system robustness model, SRB had a much smaller standardized effect than incidence, health expenditure, life expectancy, fertility, physician density, urbanization, and UHC service coverage.

RANK	PREDICTOR	STD. BETA	DIRECTION	P-VALUE
1	BC incidence	1.113150	Positive	< 0.001
2	Health expenditure pc	-0.253873	Negative	< 0.001
3	Female life expectancy	-0.191340	Negative	< 0.001
4	Fertility	-0.134245	Negative	< 0.001

Economic Sciences
https://economic-sciences.com
ES(2026) 22, 2 (2) , 146-154 | ISSN:1505-4683

5	Physicians per 1,000	0.079532	Positive	< 0.001
6	Urbanization	-0.044377	Negative	< 0.001
7	UHC service coverage	-0.026022	Negative	0.040299
10	SRB imbalance	-0.016132	Negative	0.007220

The parameter table uses HC3 standard errors for coefficient ranking; conservative SRB inference is based on the country-clustered robustness table above.

VIF diagnostics suggested multicollinearity among development-related variables, especially health expenditure per capita and GDP per capita, but SRB had low VIF, indicating that its weak performance was not simply due to duplication with other adjustment variables. Full VIF and coefficient tables are provided as supplementary outputs.

E. VALIDATION AND FORCASTING

Leave-one-country-out validation showed reasonable international generalizability with the updated feature set: 196 held-out countries, 1,556 held-out rows, overall R^2 = 0.835353, MAE = 3.852548, RMSE = 5.779625, and median country MAE = 2.656541. The largest errors occurred in Palau, Monaco, Andorra, Nauru, Portugal, Uruguay, Tuvalu, Georgia, Norway, and Fiji, likely reflecting country-specific health systems, small-population instability, screening/treatment differences, or reporting differences.

Supplementary forecasting models using lagged mortality achieved high predictive performance (CatBoost R^2 = 0.986756; XGBoost R^2 = 0.986484) [6], [7]. Permutation importance was added as a supplementary forecasting importance check; these results confirm temporal persistence in mortality rates but do not provide causal evidence.

V. DISCUSSION

The discussion centers on one question: whether SRB imbalance has a stable ecological association with breast cancer mortality. The answer is negative. SRB provides at most a small contextual demographic signal and does not behave like a stable independent ecological marker.

The multi-perspective framing is important because it keeps the interpretation from collapsing into one disciplinary angle. From a breast cancer epidemiology perspective, incidence, stage, detection, treatment, and survival are more proximate to mortality than SRB. From a demographic perspective, SRB can be meaningful as a social or population marker without being a cancer mechanism. From a health-system and surveillance perspective, SRB may cover reporting completeness, vital registration, diagnostic access, or treatment capacity. From an econometric perspective, repeated country observations, stable country differences, time trends, and residual dependence can make weak ecological signals appear stronger than they are. These perspectives all point toward the same cautious interpretation: SRB should be evaluated as a contextual marker and stress-tested heavily before being treated as independently meaningful.

The SRB result is best interpreted as a negative robustness finding. In the broad model, SRB is statistically significant under heteroskedasticity-robust standard errors, but the incremental R^2 gain is only 0.000783. Once repeated country observations are handled through country-clustered inference, SRB is no longer significant. The association weakens further under adjustment covariates, registry-quality restriction, country fixed effects, two-way country/year fixed effects, and lagged temporal checks. Thus, the apparent SRB signal is more consistent with ecological covariation, surveillance context, stable country characteristics, or global time patterns than with an independent SRB-mortality relationship.

This finding is still useful. It shows that a demographic variable that appears significant in a broad cross-national model can disappear under more appropriate ecological panel adjustment. The

Economic Sciences
https://economic-sciences.com
ES(2026) 22, 2 (2) , 146-154 | ISSN:1505-4683

paper therefore contributes less by elevating SRB as a risk factor and more by demonstrating why SRB should be interpreted cautiously as a contextual marker. The results should not be used to infer causal policy effects of SRB imbalance on breast cancer mortality.

The validation and forecasting results are supplementary checks only. They help confirm that the SRB-centered ecological interpretation is not driven by one narrow split of the data, but they are not separate claims about prediction, health-system performance, or causal mechanisms.

VI. LIMITATIONS

This study has several limitations. First, it is ecological and uses country-year aggregate data. Results should not be interpreted as individual breast cancer risk estimates and are subject to ecological fallacy.

Second, breast cancer incidence and mortality came from GBD/IHME modeled estimates. Because both were drawn from the same modeled system, the strong incidence-mortality association may partly reflect shared modeling assumptions, source-data structure, and correction procedures rather than fully independent observed epidemiologic measurement. The reproducible pipeline includes GBD lower/point/upper target sensitivity when lower and upper mortality columns are present; the current compact Q2 dataset contains the point estimate only, so lower/upper uncertainty-bound sensitivity is reported as unavailable rather than imputed.

Third, adjustment variables were incomplete in some robustness subsets. This is transparent but may introduce selection bias if missingness is related to income, healthcare infrastructure, or reporting capacity.

Fourth, cancer-specific variables such as screening, mammography, radiotherapy, and registry quality were too sparse for main full-panel adjustment. They were documented for coverage transparency and registry-quality exclusion sensitivity.

Finally, the study does not include a causal identification strategy. Even with country-clustered standard errors, country/year fixed effects, lagged predictors, panel diagnostics, and registry-quality restriction, residual ecological patterning may remain. The panel diagnostics also indicated serial correlation, cross-sectional dependence, and possible mortality-series persistence, so the SRB results should be interpreted as descriptive ecological association evidence rather than causal or structural identification.

The multi-perspective review process is also a limitation. It improved perspective coverage and helped identify overclaiming risks, but it was not a formal systematic review or meta-analysis. Source selection and final synthesis still required author judgment.

VII. CONCLUSION

SRB imbalance does not show a stable independent ecological association with breast cancer mortality. It appears statistically detectable only in the simpler broad model, adds very little association information, and is not robust to country-clustered inference, adjustment covariates, country/year fixed effects, or lagged temporal checks. SRB should therefore be interpreted as a minor ecological contextual marker rather than an independent mortality determinant.

VIII. REFERENCES

1 Institute for Health Metrics and Evaluation, "Global Burden of Disease Results Tool," IHME, University of Washington, Seattle, WA, 2024. Available: https://ghoapi.azureedge.net/gbd-results-tool
2 World Bank, "World Development Indicators," World Bank DataBank, Washington, DC, 2024. Available: https://databank.worldbank.org/source/world-development-indicators
3 Allemani et al., "Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): Analysis of individual records for 37,513,025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries," Lancet, vol. 391, no. 10125, pp. 1023-1075, 2018.
4 Chao, P. Gerland, A. R. Cook, and L. Alkema, "Systematic assessment of the sex ratio at birth for all countries and estimation of national imbalances and regional reference levels," Proc. Natl. Acad. Sci. USA, vol. 116, no. 19, pp. 9303-9311, 2019.

Economic Sciences
https://economic-sciences.com
ES(2026) 22, 2 (2) , 146-154 | ISSN:1505-4683

5 M. L. Terrell, K. P. Hartnett, and M. Marcus, "Can environmental or occupational hazards alter the sex ratio at birth? A systematic review," Emerging Health Threats J., vol. 4, no. 1, 2011.
6 T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785-794.
7 L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in Advances in Neural Information Processing Systems, vol. 31, 2018.
8 J. M. Wooldridge, Econometric Analysis of Cross Section and Panel Data, 2nd ed. Cambridge, MA: MIT Press, 2010.
9 Bray et al., "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA Cancer J. Clin., vol. 74, no. 3, pp. 229-263, 2024.
10 World Health Organization, Global Breast Cancer Initiative Implementation Framework: Assessing, Strengthening and Scaling Up Services for the Early Detection and Management of Breast Cancer. Geneva: World Health Organization, 2023.
11 Supplementary materials available on request: full coefficient tables, sparse cancer-specific coverage, registry-quality exclusion sensitivity, VIF diagnostics, residual diagnostics, panel diagnostic tables, country-level error rankings, SRB interaction tables, country-clustered standard errors, country/year fixed-effects tables, GBD uncertainty sensitivity status, lagged year fixed-effects outputs, and forecasting plots.