Spatial Patterns and Healthcare Access in Early-Onset Breast Cancer Diagnosis in Lagos, Nigeria: A Bayesian Multilevel Analysis

Olopha Paul Omoh and Elias Temidayo Mayowa

Department of Statistics, Federal University of Technology, Akure

Abstract

Early-onset breast cancer (diagnosis before age 40) is an emerging public health concern in Nigeria, yet its spatial distribution and determinants remain poorly understood, particularly in urban settings with unequal access to diagnostic services. This study analysed retrospective breast cancer registry data from the Lagos University Teaching Hospital (LUTH), with early-onset diagnosis defined as a binary outcome. A Bayesian structured additive logistic regression model was used to assess socio-demographic, clinical, and spatial effects. Nonlinear effects of age and distance to LUTH were modelled using smooth functions, while spatial heterogeneity across Local Government Areas (LGAs) was captured using a Gaussian Markov Random Field. A three-level hierarchical structure accounted for clustering, and estimation was performed using Integrated Nested Laplace Approximation (INLA). The unadjusted model showed geographic variation, but this was substantially attenuated after adjustment, with no strong evidence of elevated risk across LGAs. Distance to LUTH showed a decreasing nonlinear relationship with early-onset diagnosis, indicating higher probabilities among individuals living closer to the facility. Age also exhibited a declining nonlinear association. Overall, spatial variation is modest after adjustment and largely reflects diagnostic access rather than true geographic clustering.

Keywords: Bayesian modelling; Structured additive regression; INLA; Spatial epidemiology; Early-onset breast cancer; Healthcare accessibility

1.0 Introduction

Breast cancer remains the most commonly diagnosed cancer among women globally and a leading cause of cancer-related mortality (Bray et al., 2018; Sung et al., 2021). While the disease predominantly affects older women in high-income countries, a substantial proportion of cases in low- and middle-income countries (LMICs), particularly in sub-Saharan Africa, occur at younger ages, often referred to as early onset breast cancer (Jedy-Agba et al., 2016; Vanderpuye et al., 2017). Early onset breast cancer is typically defined as breast cancer diagnosed before the age of 40 or 45 years and is frequently associated with aggressive tumour biology, delayed diagnosis, and poorer survival outcomes (Azim et al., 2014; Paluch-Shimon et al., 2020).

In sub-Saharan Africa, breast cancer incidence has been rising steadily over recent decades, driven by rapid urbanisation, lifestyle transitions, reproductive pattern changes, and population ageing (Bray et al., 2018; Jedy-Agba et al., 2020). Nigeria, the most populous country in Africa, bears a particularly high burden, accounting for a large share of breast cancer cases and deaths in the region (Ferlay et al., 2020; Sung et al., 2021). Evidence consistently shows that Nigerian women tend to be diagnosed at younger ages and more advanced stages compared with women in high-income settings, reflecting both biological and systemic factors (Jedy-Agba et al., 2012; Ogundiran et al., 2013).

Early onset breast cancer poses distinct clinical and public health challenges. Younger women are more likely to present with high-grade tumours, hormone receptor–negative disease, and rapid disease progression, all of which complicate treatment and worsen prognosis (Azim et al., 2014;

ISSN NUMBER: 1116-249X

Anders et al., 2009). In LMICs, these biological challenges are compounded by limited awareness, weak screening infrastructure, diagnostic delays, and financial barriers to care, leading to advanced-stage presentation and poor outcomes (Anderson et al., 2011; Ginsburg et al., 2017; Unger-Saldaña, 2014). In Nigeria, over 70% of breast cancer cases are diagnosed at Stage III or IV, with younger women disproportionately affected (Jedy-Agba et al., 2016; Sighoko et al., 2018). Socio-demographic and socioeconomic factors play a critical role in shaping breast cancer risk, detection, and outcomes. Ethnicity, marital status, employment, and socioeconomic position have all been shown to influence reproductive behaviours, lifestyle exposures, health-seeking behaviour, and access to diagnostic services (Aizer & Currie, 2014; Diez-Roux, 2000; Merlo et al., 2005). In Nigeria, ethnic and regional disparities in breast cancer incidence and stage at diagnosis have been reported, reflecting complex interactions between culture, migration, urban residence, and healthcare access (Jedy-Agba et al., 2012; Ogundiran et al., 2013). However, evidence specifically focusing on early onset breast cancer and its multilevel determinants remains limited.

Geographic context further shapes breast cancer risk and diagnosis. Spatial inequalities in healthcare infrastructure, distance to tertiary facilities, and urban–peri-urban gradients can result in substantial geographic variation in observed disease patterns (Jones et al., 2010; Wakefield, 2007). In LMIC settings, cancer diagnostic and treatment services are often concentrated in a small number of urban tertiary hospitals, creating strong distance-related barriers to timely diagnosis (Ginsburg et al., 2018). Failure to account for these spatial processes may lead to biased inference and mask important place-based determinants of disease (Banerjee et al., 2015; Lawson, 2018).

Recent advances in Bayesian spatial epidemiology provide powerful tools for addressing these challenges. Structured additive regression models allow for the simultaneous modelling of linear effects, nonlinear covariate effects, and spatial random effects, offering greater flexibility than traditional regression approaches (Fahrmeir et al., 2013; Wood, 2017). Bayesian spatial models are particularly well suited to cancer epidemiology in LMICs, where data sparsity, hierarchical structures, and unmeasured contextual effects are common (Lawson, 2018; Wakefield, 2007). Despite their advantages, such models have been infrequently applied to early onset breast cancer in sub-Saharan Africa, and evidence from Nigeria remains scarce.

Lagos State, Nigeria's largest metropolitan area, presents a unique setting for investigating early onset breast cancer. The state is characterised by extreme population density, rapid urban expansion, socioeconomic inequality, and spatially uneven distribution of healthcare services, including specialised oncology facilities. These features make Lagos an important context for understanding how individual, clinical, and spatial factors interact to shape early onset breast cancer risk and diagnosis in an urban African setting.

Against this background, the present study investigates the socio-demographic, socioeconomic, clinical, and spatial determinants of early onset breast cancer among women in Lagos State using a Bayesian structured additive regression framework. Specifically, the study examines (i) individual- and household-level factors associated with early onset breast cancer, (ii) nonlinear effects of age and geographic accessibility to tertiary care, and (iii) residual spatial variation across Local Government Areas after adjusting for observed covariates. By adopting a multilevel spatial modelling approach, this study aims to generate evidence that can inform targeted early detection strategies, reduce diagnostic inequities, and improve breast cancer outcomes among younger women in Lagos State and similar urban LMIC contexts.

ISSN NUMBER: 1116-249X

2.0 Methodology

2.1 Study location

Figure 1: Study Area Map

Map of Lagos State showing the 20 Local Government Areas (LGAs) used as spatial units in the analysis. The map provides the geographic context for examining spatial variation in early-onset breast cancer diagnosis.

This study was conducted in Lagos State, Nigeria, the country's largest metropolitan region and economic hub. Lagos comprises 20 Local Government Areas (LGAs), which serve as the spatial units of analysis. The state is characterized by high population density, rapid urbanization, and substantial socioeconomic and infrastructural inequalities. While Lagos hosts major tertiary healthcare facilities, including the Lagos University Teaching Hospital (LUTH), access to cancer diagnostic services remains uneven across LGAs. These spatial disparities make Lagos an appropriate setting for examining geographic variation in breast cancer diagnosis patterns.

2.2 Data Source

The study utilized retrospective breast cancer registry data obtained from the oncology centre and cancer registry of Lagos University Teaching Hospital (LUTH). The dataset includes individual-level clinical and socio-demographic information for women diagnosed with breast cancer. As a hospital-based dataset, the sample reflects patients who successfully accessed diagnostic services at LUTH and does not represent the general population.

2.3 Outcome Variable

The outcome variable is a binary indicator of early-onset breast cancer diagnosis, defined as:

$$Y_{\mathrm{IJK}} = \begin{cases} 1 & if\ age\ at\ diagnosis < 40 \\ 0 & \text{otherwise} \end{cases}$$

Let $p_{ijk} = Pr(Y_{ijk}=1)$ denote the probability that individual i, residing in household j and LGA k, is diagnosed before age 40.
Accordingly, the outcome is modeled as:

$$Y_{ijk} \sim Bernoulli(p_{ijk})$$

This study models the conditional probability of early-onset diagnosis given that breast cancer has already been diagnosed. Therefore, the outcome reflects diagnostic timing among hospital-presenting cases rather than population-level incidence risk.

ISSN NUMBER: 1116-249X

2.4 Covariates

Explanatory variables include socio-demographic and clinical characteristics. Socio-demographic variables comprise current age, ethnicity, marital status (partnered/unpartnered), employment status, socioeconomic status (SES), and religion. Clinical variables include cancer stage at diagnosis and comorbidity status. Geographic accessibility is captured through distance from the patient's residence to LUTH, measured as a continuous variable.

2.5 Statistical Modelling Framework

2.5.1 Baseline Logistic Model

A baseline logistic regression model was first specified as:

$$logit(p_{ijk}) = \beta_0 + \sum_{r=1}^{R} \beta_r X_{rijk}$$

where β_0 is the intercept, X_{rijk} represents covariates, and β_r are regression coefficients.

2.5.2 Structured Additive Model

To allow for nonlinear covariate effects and spatial dependence, a Bayesian structured additive regression model was employed:

$$logit(p_{ijk}) = \beta_0 + \sum_{r=1}^{R} \beta_r X_{rijk} + \sum_{m=1}^{p} f_m(Z_{m,ijk}) + u_j + v_k$$

Where:

- β_0 is the intercept
- $\sum_{r=1}^{R} \beta_r X_{rijk}$: linear covariate effects
- $\sum_{m=1}^{p} f_m(Z_{m,ijk})$: nonlinear smooth effects
- u_j : household-level random effect
- v_k: LGA-level spatial effect

2.5.3 Spatial Specification (BYM2 Model)

The LGA-level spatial effect γ_k was modeled using the Besag–York–Mollié 2 (BYM2) specification:

$$\gamma_k = \frac{1}{\sqrt{\tau}}\left(\sqrt{\phi}\, s_k + \sqrt{1-\phi}\, v_k\right)$$

where:

- $s_k$is the structured spatial effect following an intrinsic conditional autoregressive (ICAR) prior,
- $v_k \sim N(0,1)$ is the unstructured spatial effect,
- τ is a precision parameter controlling overall variability,
- $\phi \in [0,1]$ is a mixing parameter representing the proportion of spatially structured variation.

ISSN NUMBER: 1116-249X

The structured component captures spatial correlation between neighbouring LGAs, while the unstructured component accounts for independent area-level heterogeneity.

2.5.4 Hierarchical Structure and Rationale

A multilevel structure was specified with individuals nested within households, while LGAs were incorporated as a spatial contextual level. Specifically:

- Level 1: Individuals
- Level 2: Households
- Contextual level: LGAs (modeled via spatial effects)

Household-level random effects u_j capture shared characteristics such as health-seeking behaviour, financial constraints, and referral patterns that influence diagnostic timing. LGA-level effects are modeled using a spatial specification (BYM2), capturing geographic variation arising from differences in accessibility to diagnostic services, healthcare infrastructure, and other unobserved contextual factors. This formulation accounts for both intra-household correlation and spatial dependence, improving the validity and interpretability of model estimates.

2.5.5 Final Model for Inference

The final model used for inference corresponds to the structured additive specification in Section 2.5.2 and is implemented as follows:

$$logit(p_{ijk}) = \beta_0 + \sum_{r=1}^{R} \beta_r X_{rijk} + f_1(Age_{ijk}) + f_2(Distance_{ijk}) + u_j + v_k$$

where:

- $f_1(\cdot)$ and $f_2(\cdot)$ are smooth nonlinear functions of age and distance to LUTH, respectively,
- u_j is a household-level random effect,
- v_k represents the LGA-level spatial effect.

2.5.6 Prior Specification

A Bayesian framework was adopted with the following prior specifications. Fixed effects were assigned weakly informative Gaussian priors. Nonlinear effects were modeled using second-order random walk (RW2) priors, which impose smoothness while allowing flexible functional forms for continuous covariates. Household-level random effects were assigned Gaussian priors to capture within-household correlation. For the spatial effects, the Besag–York–Mollié 2 (BYM2) specification was used, with penalized complexity (PC) priors assigned to both the precision parameter and the mixing parameter. These PC priors provide principled regularization by shrinking toward a base model of no spatial variation, thereby reducing overfitting and improving interpretability of spatial effects.

2.5.7 Sensitivity Analysis and Extreme Estimates

Sensitivity analyses were conducted by varying hyperparameters of the prior distributions. Posterior estimates were found to be stable across reasonable prior choices, indicating robustness of the results. Extreme odds ratios observed for certain categorical variables were investigated and found to be

ISSN NUMBER: 1116-249X

primarily driven by sparse data in some subgroups and quasi-separation in the binary outcome. These effects are well documented in logistic regression and were partially mitigated through Bayesian regularization.

2.5.8 Estimation Procedure

Model estimation was performed using Integrated Nested Laplace Approximation (INLA), implemented in the R-INLA framework. INLA provides accurate and computationally efficient approximation for Bayesian inference in latent Gaussian models, making it particularly suitable for structured additive and spatial hierarchical models.

2.6 Model Building Strategy

A sequential model-building strategy was adopted to assess the contribution of spatial, compositional, nonlinear, and hierarchical effects. Model M1 included only spatial effects. Model M2 added fixed covariates. Model M3 further incorporated nonlinear effects for continuous variables, while Model M4 extended M3 by including household-level random effects. This progression allows evaluation of the incremental contribution of each model component.

Let $\eta_{ijk} = logit(p_{ijk})$,

Model Specifications

Model M1 (Spatial-only model): $\eta_{ijk} = \beta_0 + v_k$

Captures crude geographic variation only.

Model M2 (Spatial + Fixed effects) : $\eta_{ijk} = \beta_0 + \sum_{r=1}^{R} x_{ijkr}\,\beta_r + v_k$

Adds covariates effects (socio-demographic + clinical variables).

Model M3 (Add Nonlinear Effects)

$$\eta_{ijk} = \beta_0 + \sum_{r=1}^{R} x_{ijkr}\,\beta_r + f_1(Age_{ijk}) + f_2(Distance_{ijk}) + v_k$$

Now captures nonlinear relationships.

Model M4 (Final model): Add Household Random Effects

$$\eta_{ijk} = \beta_0 + \sum_{r=1}^{R} x_{ijkr}\,\beta_r + f_1(Age_{ijk}) + f_2(Distance_{ijk}) + u_j + v_k$$

Model Comparison

Model performance was evaluated using the Deviance Information Criterion (DIC) and the Widely Applicable Information Criterion (WAIC), with lower values indicating better fit.

ISSN NUMBER: 1116-249X

3.0 Results

3.1 Descriptive Statistics

A total of 835 breast cancer patients were analysed. The study population was predominantly Christian (738; 88.4%), with a smaller proportion identifying as Muslim (99; 11.9%). Most patients belonged to the medium socioeconomic group (489; 58.6%), followed by those in the low socioeconomic category (279; 33.4%), while only a small fraction were in the high socioeconomic group (67; 8.0%). In terms of ethnicity, Yoruba women constituted the largest group (425; 50.9%), followed by Igbo women (303; 36.3%); smaller proportions were observed among Edo (49; 5.9%), other ethnic groups (47; 5.6%), Cross-River (6; 0.7%), and Hausa (5; 0.6%). Regarding marital status, the majority of patients were partnered at diagnosis (653; 78.2%), while 182 (21.8%) were unpartnered. Occupationally, most respondents were economically active, with self-employed women forming the largest group (390; 46.7%), followed by those in paid employment (345; 41.3%), and a smaller proportion being unemployed (97; 11.6%).

3.2 Model Comparison

Table 1. Early Onset Logistic Models Diagnostics

	Early Onset Models	
Model	DIC	WAIC
M1: Spatial only	839.126	839.601
M2: M1+ fixed	838.341	838.782
M3: M2 + non_linear	235.469	319.775
M4: M3 + community	-96.579	165.679

Model comparison was conducted using the Deviance Information Criterion (DIC) and the Widely Applicable Information Criterion (WAIC), both of which balance model fit and complexity. Lower values indicate improved explanatory performance (Spiegelhalter et al., 2002; Watanabe, 2010). Inference was conducted using Integrated Nested Laplace Approximation (INLA), which provides accurate deterministic approximation for latent Gaussian models (Rue et al., 2009). The spatial-only model (M1) yielded the highest DIC and WAIC values, indicating poor fit. Inclusion of fixed effects (M2) resulted in modest improvement, whereas the addition of nonlinear smooth terms (M3) substantially reduced model deviance. The full multilevel spatial model (M4) achieved the lowest DIC and WAIC values, demonstrating the importance of jointly modelling nonlinear, hierarchical, and spatial dependence structures. All models were fitted using identical likelihood specifications and sample sizes to ensure comparability of information criteria. Inference is therefore based on the fully adjusted model (M4), which provides the most reliable representation of the data structure.

3.3 Fixed Effects

Posterior odds ratios (ORs) were obtained from the fully adjusted Bayesian model (M4). Statistical significance was assessed using 95% credible intervals. Intervals excluding 1 indicate meaningful associations. The results indicate substantial heterogeneity across socio-demographic and clinical variables. Ethnicity, partnership status, socioeconomic position, and cancer stage were strongly associated with early-onset breast cancer. These estimates represent associations with the probability of early diagnosis among observed breast cancer cases, not the population-level risk of developing early-onset breast cancer.

ISSN NUMBER: 1116-249X

Table 2. Odds Ratios for Factors Associated with Early Onset Breast Cancer Diagnosis in Lagos State

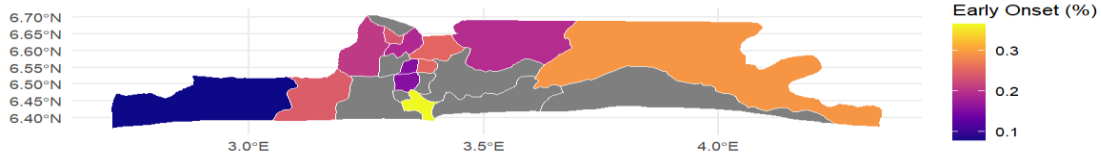
Variable (Reference category)	OR	95% CrI	Interpretation
Religion (Christianity)			
Islam	3.45	0.72 – 16.64	Higher odds, but not statistically significant
Ethnicity (Others)			
Igbo	13.59	1.32 – 142.40	Significantly higher odds
Hausa	≈ 0.00	≈ 0.00 – 0.09	Markedly lower odds
Cross-River	≈ 0.00	≈ 0.00 – ≈ 0.00	Markedly lower odds
Edo	4.34	0.26 – 73.70	Higher odds, but not statistically significant
Yoruba	1.95	0.19 – 20.70	Higher odds, but not statistically significant
Partnership (Unpartnered)			
Partnered	6.25	1.57 – 25.00	Significantly higher odds
Employment status (Employed)			
Self-employed	0.21	0.07 – 0.67	Significantly lower odds
Unemployed	0.69	0.03 – 14.90	Not statistically significant
Socioeconomic status (High)			
Medium	0.33	0.04 – 2.90	Not statistically significant
Low	0.08	0.01 – 0.80	Significantly lower odds
Health status (Non-comorbidity)			
Comorbidity	1.19	0.40 – 3.53	Higher odds but no significant association
Cancer stage (Stage I)			
Stage II	48.10	1.74 – 1331	Strong association
Stage III	9.15	0.36 – 226.00	Uncertain association
Stage IV	224.40	9.80 – 5134	Very strong association
Unstaged	24.00	1.09 – 557	Strong association

Interpretation

Ethnicity showed substantial heterogeneity, with Igbo women exhibiting significantly higher odds of early-onset diagnosis (diagnosis before age 40) compared with the reference group However, extremely large or near-zero odds ratios (e.g., Hausa and Cross-River categories) likely reflect sparse data and should be interpreted cautiously rather than as true biological effects. Partnered women had higher odds of early diagnosis compared with unpartnered women, which may reflect differences in healthcare utilisation and support systems. Self-employed women and those in lower socioeconomic categories exhibited lower odds of early diagnosis, which is more plausibly explained by diagnostic access limitations and under-detection, rather than reduced biological risk. Clinical stage showed the strongest association, with advanced stages exhibiting markedly higher odds. This likely reflects delayed diagnosis and aggressive disease presentation in younger women.

ISSN NUMBER: 1116-249X

3.4 Spatial Results



. .

Figure 3.1: Map of the proportion of early-onset breast cancer across Lagos State LGAs.

The spatial distribution of early-onset breast cancer across LGAs shows variation in the proportion of cases diagnosed before age 40.

Higher proportions are observed in Eti-Osa, Ibeju-Lekki, and Epe, while lower proportions are seen in Badagry and Ojo. This pattern reflects differences in diagnostic capture and healthcare accessibility, rather than true differences in underlying disease risk. Central LGAs such as Lagos Mainland and Surulere show intermediate proportions, likely reflecting better diagnostic infrastructure and more balanced age detection patterns. The observed spatial pattern suggests an access-driven gradient rather than a true epidemiological hotspot structure.

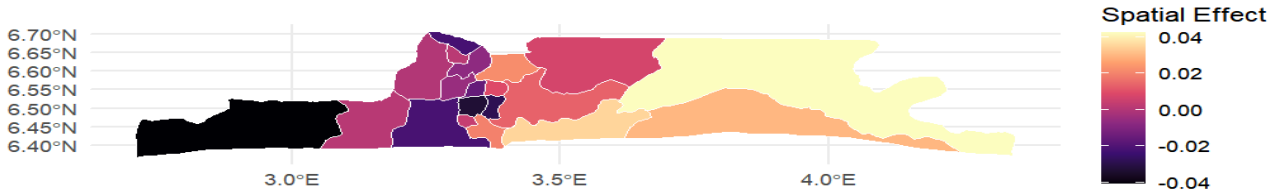


Figure 3.2: Unadjusted Spatial Effects (Model M1)

Figure 3.2 displays the spatial distribution of unadjusted (crude) spatial effects on the log-odds of early-onset breast cancer diagnosis across Local Government Areas (LGAs) in Lagos State, derived from the spatial-only model (M1). Positive values indicate higher log-odds relative to the state average, while negative values indicate lower log-odds. The map reveals clear geographic variation across LGAs. Western LGAs, including Badagry and Ojo, are characterised by negative spatial effects. Central LGAs such as Lagos Mainland, Surulere, Mushin, and Shomolu exhibit values close

ISSN NUMBER: 1116-249X

to zero. In contrast, eastern and peri-urban LGAs - including Ikorodu, Epe, Ibeju-Lekki, and parts of Eti-Osa - display positive spatial effects, indicating relatively higher log-odds of early-onset diagnosis.

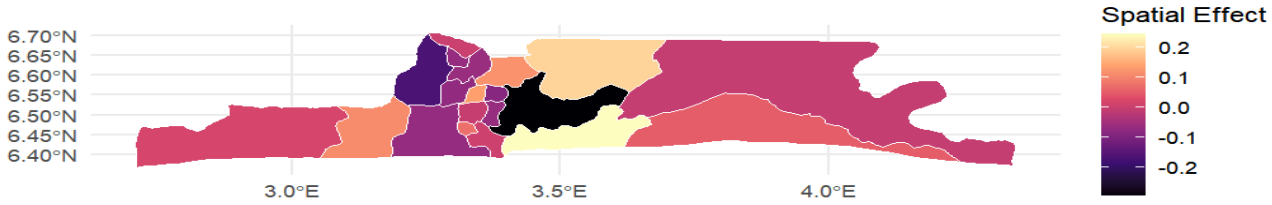


Figure 3.3: Adjusted Spatial Effects (Model M4)

Posterior mean spatial effects (log-odds scale) from the fully adjusted Bayesian structured additive model (M4), accounting for covariates and hierarchical structure.

After adjustment, spatial variation across LGAs is reduced. The distribution shows modest deviations around the state average, with no clear pattern of pronounced clustering. A small number of LGAs exhibit relatively higher or lower effects; however, the magnitude of these differences is limited.

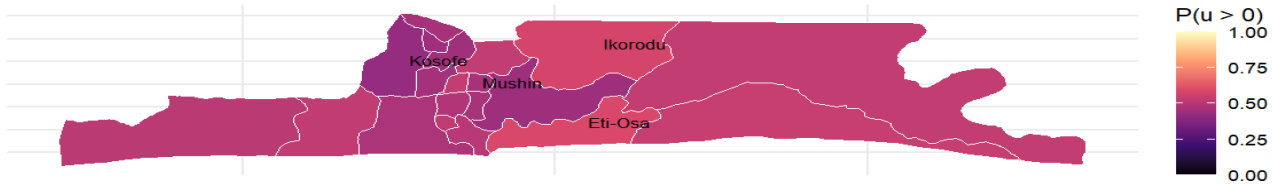


Figure 3.4: Exceedance Probability Map

Figure 3.4 presents the exceedance probabilities, defined as the posterior probability that the spatial effect exceeds a specified threshold, across Local Government Areas (LGAs) in Lagos State. Higher values indicate stronger statistical evidence that the probability of early-onset breast cancer diagnosis is above the overall mean. The exceedance probability map provides a formal assessment of spatial

ISSN NUMBER: 1116-249X

variation in the early breast cancer diagnosis. No LGA attains exceedance probabilities at or above the conventional threshold of 0.8, indicating an absence of statistically strong evidence for elevated spatial effects after adjusting for covariates. A small number of LGAs, including Eti-Osa and Ikorodu, exhibit moderately higher exceedance probabilities; however, these remain below thresholds typically used to identify statistically significant hotspots. Overall, the pattern suggests limited spatial heterogeneity, with residual spatial variation appearing weak rather than indicative of pronounced geographic clustering.

3.5 Nonlinear Effects

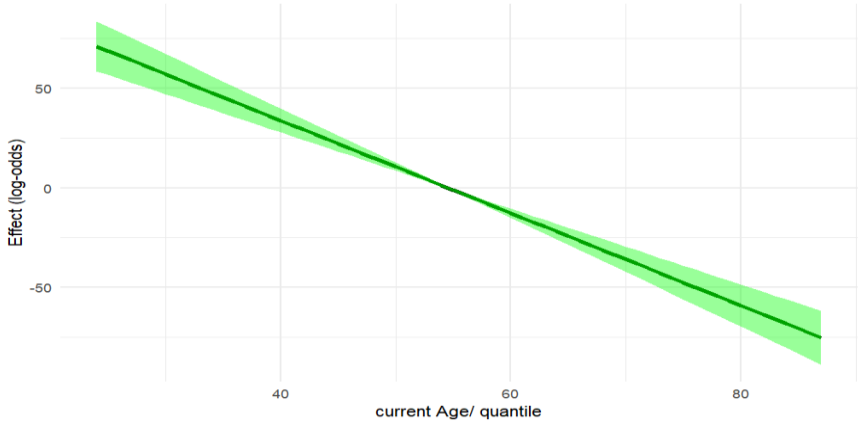


Figure 3.5: Nonlinear effect of current age on the log-odds of early onset breast cancer diagnosis

The figure shows the estimated nonlinear effect of current age on the log-odds of early-onset breast cancer diagnosis, with the shaded region representing the 95% credible interval. The effect exhibits a strong negative pattern, with higher log-odds at younger ages that decline progressively as age increases. The curve indicates a steep decrease in the log-odds within the younger age range, followed by a more gradual decline at older ages. The credible intervals are relatively narrow in the central age range and widen slightly at the extremes, indicating increased uncertainty at lower and higher age values.

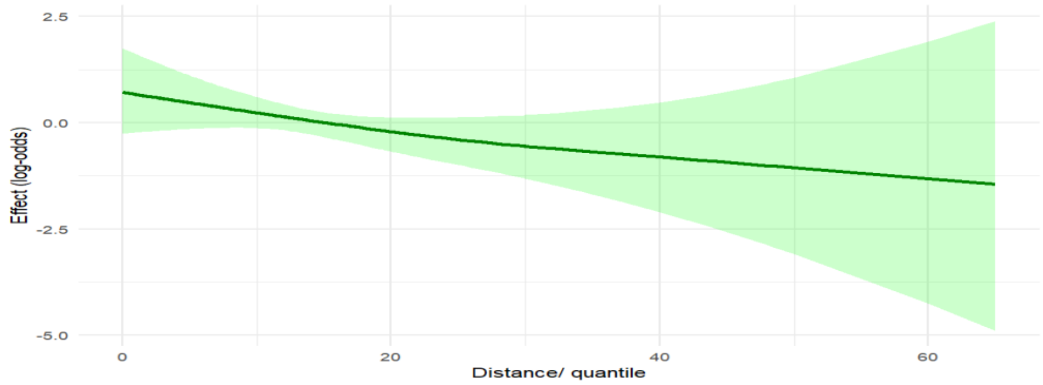


Figure 3.6: Nonlinear Effect of Distance to LUTH

ISSN NUMBER: 1116-249X

Figure 3.6 presents the estimated nonlinear effect of distance to Lagos University Teaching Hospital (LUTH) on the log-odds of early-onset breast cancer diagnosis, with the shaded region representing the 95% credible interval. The results indicate a monotonic decreasing relationship between distance and the likelihood of early-onset diagnosis. Individuals residing closer to LUTH have higher log-odds of being diagnosed before age 40, while the log-odds decrease progressively with increasing distance. The uncertainty associated with the estimates increases at larger distances, as reflected by the widening credible intervals, suggesting greater variability or fewer observations in those regions.

3.6 Discussion

This study provides comprehensive evidence on the socio-demographic, clinical, accessibility-related, and spatial determinants of early-onset breast cancer diagnosis in Lagos State using a Bayesian structured additive regression framework. By jointly modelling fixed effects, nonlinear covariate effects, and structured spatial dependence through the BYM2 model, the analysis advances both methodological and applied understanding of early-onset breast cancer in an urban low- and middle-income country (LMIC) context. Importantly, the findings should be interpreted as determinants of the probability of early-onset diagnosis among observed breast cancer cases, rather than population-level risk. This distinction reflects the hospital-based nature of the data and is essential for interpreting accessibility-related effects.

Socio-demographic and socioeconomic determinants

Ethnicity emerged as an important factor, with Igbo women exhibiting significantly higher odds of early-onset diagnosis compared with the reference group. Ethnic disparities in breast cancer diagnosis patterns have been reported in Nigeria and other sub-Saharan African settings, reflecting differences in cultural practices, migration patterns, and healthcare utilisation (Jedy-Agba et al., 2012; Ogundiran et al., 2013). However, extreme estimates observed for smaller ethnic groups such as Hausa and Cross-River likely reflect sparse data and should be interpreted cautiously.

Partnership status was positively associated with early-onset diagnosis, potentially reflecting stronger social support and improved healthcare-seeking behaviour (Aizer & Currie, 2014; Ginsburg et al., 2017). Socioeconomic status and employment showed inverse associations, particularly among self-employed and lower socioeconomic groups. Rather than indicating biological protection, this pattern likely reflects disparities in access to screening and timely diagnosis (Unger-Saldaña, 2014; Ginsburg et al., 2018), consistent with broader evidence from Nigeria (Jedy-Agba et al., 2020).

Clinical determinants

Clinical stage at diagnosis showed the strongest association with early-onset diagnosis. The higher odds observed for advanced-stage disease likely reflect delayed presentation and aggressive tumour progression among younger women (Azim et al., 2014; Anders et al., 2009; Jedy-Agba et al., 2016). The substantial proportion of unstaged cases further highlights systemic limitations in cancer documentation in resource-constrained settings (Sighoko et al., 2018).

Spatial patterns and contextual influences

Spatial analysis revealed geographic variation in early-onset diagnosis across LGAs. While unadjusted models showed clear differences, adjustment for individual, household, and clinical factors substantially reduced spatial variation. Higher probabilities were observed in eastern and peri-urban LGAs such as Eti-Osa, Ibeju-Lekki, Epe, and Ikorodu, while lower probabilities were

ISSN NUMBER: 1116-249X

observed in Badagry and Ojo. However, exceedance probability analysis provided no strong evidence of spatial clustering after adjustment, indicating limited residual spatial heterogeneity.

Remaining spatial variation likely reflects unmeasured contextual factors such as healthcare infrastructure, referral pathways, and differential access to diagnostic services. Importantly, these spatial effects represent variation in diagnostic processes rather than differences in underlying disease incidence.

Nonlinear effects of age and distance

Age showed a clear monotonic decline in the probability of early-onset diagnosis with increasing age, consistent with the outcome definition and supporting the use of flexible modelling approaches (Fahrmeir et al., 2013; Wood, 2017).

Distance to Lagos University Teaching Hospital exhibited a decreasing relationship with early-onset diagnosis, consistent with distance-decay effects in healthcare utilisation (Jones et al., 2010; Ginsburg et al., 2018). This effect reflects healthcare access and diagnostic delay rather than a causal influence on disease onset, reinforcing the importance of distinguishing biological risk from observed diagnostic patterns in hospital-based data.

4.0 Conclusion

This study examined determinants of early-onset breast cancer diagnosis in Lagos State using a Bayesian structured additive regression framework. Socio-demographic characteristics, socioeconomic position, and clinical stage were significantly associated with the probability of early-onset diagnosis among observed breast cancer cases. Spatial analysis revealed limited geographic variation across Local Government Areas, with residual differences persisting after covariate adjustment but no statistically meaningful evidence of spatial clustering. These patterns are consistent with variation in healthcare access and diagnostic practices rather than differences in underlying disease incidence. The nonlinear effect of distance to tertiary care highlights the importance of geographic accessibility to diagnostic services, particularly the Lagos University Teaching Hospital, in shaping observed diagnostic patterns. Overall, the findings underscore the need to improve equitable access to early diagnostic services and demonstrate the utility of Bayesian spatial modelling for analysing complex health data in resource-constrained settings.

Limitations of the Study

This study is based on facility-derived, cross-sectional data, which may introduce selection bias and limits causal inference. Some covariate categories, particularly certain ethnic subgroups, had small sample sizes, resulting in unstable posterior estimates and wide credible intervals. Spatial effects were modelled at the Local Government Area level, which may mask finer-scale heterogeneity due to the modifiable areal unit problem (MAUP). Additionally, key factors such as tumour molecular subtype, genetic predisposition, and detailed reproductive histories were unavailable and could not be accounted for in the analysis.

2.5 **Ethical Considerations**

Ethical clearance was granted by University of Lagos Teaching hospital health research ethics committee. All data were anonymized in compliance with GDPR (2018) and NHREC (2022) guidelines. A waiver of informed consent was obtained due to the retrospective nature of the study.

ISSN NUMBER: 1116-249X

References

Adebamowo, C. A., & Ajayi, O. O. (2000). Breast cancer in Nigeria. *West African Journal of Medicine, 19*(3), 179–191.

Aizer, A., & Currie, J. (2014). The intergenerational transmission of inequality: Maternal disadvantage and health at birth. *Science, 344*(6186), 856–861. https://doi.org/10.1126/science.1251872

Anders, C. K., Johnson, R., Litton, J., Phillips, M., & Bleyer, A. (2009). Breast cancer before age 40 years. *Seminars in Oncology, 36*(3), 237–249. https://doi.org/10.1053/j.seminoncol.2009.03.001

Anderson, B. O., Yip, C. H., Smith, R. A., et al. (2011). Guideline implementation for breast healthcare in low-income and middle-income countries. *The Lancet Oncology, 12*(4), 387–398. https://doi.org/10.1016/S1470-2045(11)70083-4

Azim, H. A., Partridge, A. H., & Gelber, S. (2014). Biology of breast cancer in young women. *Breast Cancer Research, 16*(4), 427. https://doi.org/10.1186/s13058-014-0427-5

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2015). *Hierarchical modeling and analysis for spatial data* (2nd ed.). CRC Press.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates. *CA: A Cancer Journal for Clinicians, 68*(6), 394–424. https://doi.org/10.3322/caac.21492

Diez-Roux, A. V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health, 21*, 171–192.

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression: Models, methods and applications*. Springer.

Ferlay, J., Colombet, M., Soerjomataram, I., et al. (2020). Cancer statistics for the year 2020. *International Journal of Cancer, 149*(4), 778–789.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC Press.

Ginsburg, O., Bray, F., Coleman, M. P., et al. (2017). The global burden of women's cancers. *The Lancet, 389*(10071), 847–860. https://doi.org/10.1016/S0140-6736(16)31392-7

Ginsburg, O., Yip, C. H., Brooks, A., et al. (2018). Breast cancer early detection: A phased approach to implementation. *Cancer, 124*(Suppl 18), 3548–3557. https://doi.org/10.1002/cncr.31598

Jedy-Agba, E., Curado, M. P., Ogunbiyi, O., et al. (2012). Cancer incidence in Nigeria. *Cancer Epidemiology, 36*(5), e271–e278. https://doi.org/10.1016/j.canep.2012.04.001

Jedy-Agba, E., McCormack, V., Adebamowo, C., & dos-Santos-Silva, I. (2016). Stage at diagnosis of breast cancer in sub-Saharan Africa. *The Lancet Global Health, 4*(12), e923–e935. https://doi.org/10.1016/S2214-109X(16)30259-5

ISSN NUMBER: 1116-249X

Jedy-Agba, E., Oga, E., Odutola, M., et al. (2020). Developing national cancer registration in Nigeria. *Cancer Epidemiology, 67*, 101763.

Jones, A. P., Haynes, R., Sauerzapf, V., Crawford, S. M., & Zhao, H. (2010). Travel time to hospital and treatment for breast cancer. *European Journal of Cancer, 44*(7), 992–999.

Lawson, A. B. (2018). *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology* (3rd ed.). CRC Press.

Merlo, J., Chaix, B., Ohlsson, H., et al. (2005). A brief conceptual tutorial of multilevel analysis in social epidemiology. *Journal of Epidemiology & Community Health, 59*(6), 443–449.

Ogundiran, T. O., Huo, D., Adenipekun, A., et al. (2013). Case–control study of body size and breast cancer risk in Nigeria. *British Journal of Cancer, 108*(8), 1653–1658.

Paluch-Shimon, S., Cardoso, F., Partridge, A. H., et al. (2020). ESO–ESMO consensus guidelines for breast cancer in young women. *Annals of Oncology, 31*(6), 674–696.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using INLA. *Journal of the Royal Statistical Society: Series B, 71*(2), 319–392.

Sighoko, D., Forae, G. D., & Nwosu, S. O. (2018). Cancer staging challenges in resource-limited settings. *ecancermedicalscience, 12*, 867.

Sung, H., Ferlay, J., Siegel, R. L., et al. (2021). Global cancer statistics 2020. *CA: A Cancer Journal for Clinicians, 71*(3), 209–249.

Unger-Saldaña, K. (2014). Challenges to early diagnosis of breast cancer in developing countries. *World Journal of Clinical Oncology, 5*(3), 465–477.

Wakefield, J. (2007). Disease mapping and spatial regression. *Annual Review of Public Health, 28*, 75–100.

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). CRC Press

ISSN NUMBER: 1116-249X