BENIN JOURNAL OF
STATISTICS
ISSN 2682-5767
Vol. 6, pp. 44– 58 (2023)

Covariate Effect in Spatio-temporal Bayesian Model with Two-level Spatial Structure

U. Abdullahi[*,1] , **A. Isah** [2] **, I. K. Olayemi** [3] **and R. A. Adeyemi** [4]

[1, 2, 4]*Department of Statistics, Federal University of Technology, Minna, Nigeria.*

[3]*Department of Animal Biology, Federal University of Technology, Minna, Nigeria.*

(*Received: 16 November 2022; Accepted: 03 April 2023*)

Abstract. Spatio-temporal models suffer from comparability problems of relative risks (RRs) based on the removal of the covariate effect as a confounding factor on the risk estimate of the study population through distribution of standardized mortality ratios (SMRs). Two spatio-temporal models with two-level spatial structure with different approaches are considered in this study for comparison. The first model followed SMRs procedure by removal of the effect of the confounding factors on the risk estimate in the study population through distribution standardization., while the second model included covariate effect as confounding factors on the risk estimate in the study population. The two models were fitted within a hierarchical Bayesian framework with integrated nested Laplace approximation (INLA) estimation procedures. The objectives of this study are to compare both models in terms of their performance and identify the age-group(s) of women with significant higher risk due to breast cancer disease. The models are applied to female breast cancer mortality data in Nigeria.

Keywords: Spatio-temporal modeling, Standardized mortality ratios, deviance information, criterion, covariate effect.

Published by: Department of Statistics, University of Benin, Nigeria

1. Introduction

Research into spatial and spatio-temporal disease mapping has been carried out within a hierarchical Bayesian framework, with generalized linear mixed models (GLMM) playing a major role. A variety of spatio-temporal models dealing with single-level and two-level spatial random effects have been developed for the analysis of spatio-temporal areal data Librero *et al.* (2017)). Spatio-temporal models with a single level of spatial grouping are the parametric models with linear time trend proposed by Bernardinelli *et al.* (1995b)

*Corresponding author. Email: u.abdullahi@futminna.edu.ng

and the non-parametric model including different types of space-time interactions between the spatial and temporal main effects described by Knorr-Held (2000). Duncan *et al.* (2016) developed Bayesian models for analysing spatio-temporal data in order to understand the temporal patterns of mammography screening service utilisation in Brisbane. Melkamu *et al.* (2018) fitted different spatio-temporal models within Bayesian hierarchical framework allowing different space-time interaction for mortality mapping with integrated nested Laplace approximations to analyse mortality data extracted from the health and demographic surveillance system in Kersa District in Hararege, Oromia Region, Ethiopia. Nurul *et al.* (2019) used generalized linear mixed models for spatio-temporal study, the models incorporated spatially correlated random effects as well as temporal effects. They used two different spatial random effects and compared them. The first model was based on Leroux spatial model, while the second model was based on the stochastic partial differential equation (SPDE) approach. Naresh *et al.* (2021) recently studied the spatio-temporal variation in annual Lyme disease cases in Virginia from 2001-2016 and modeled the disease with a spatio-temporal hierarchical Bayesian model using observed ecological and environmental covariates.

The first study dealing with a two-level spatial random effect in spatio-temporal disease mapping was proposed by Schrödle *et al.* (2011) to analyse reported cases of bovine viral diarrhea in Switzerland. Ugarte *et al.* (2016) proposed a new family of spatio-temporal models where the spatial effect has a two-level structured to analyse brain cancer mortality data in the municipalities of Navarre and the Basque Country. In addition, Ugarte *et al.* (2017) used one-dimensional, two-dimensional, and three-dimensional B-splines to specify space-time interactions in Bayesian disease mapping: model fitting and model identifiability. Librero *et al.* (2017) developed hierarchical Bayesian spatio-temporal models in capturing two spatial structures in order to explain hospital risk variations using three different disease conditions: Percutaneous Coronary Intervention (PCI), Colectomy in Colorectal Cancer (CCC) and Chronic Obstructive Pulmonary Disease (COPD). *et al.* (2019) proposed a novel two-stage approach to estimate and map disease risk in the presence of such local discontinuities and clusters. They proposed approaches in both spatial and spatio-temporal domains, where for the latter the clusters can either be fixed or allowed to vary over time. Win *et al.* (2020) reviewed the types and applications of fully Bayesian (FB) spatial-temporal models and covariates used to study cancer incidence and mortality This review highlighted the need for Bayesian spatial-temporal models to incorporate patient-level prognostic characteristics through the multi-level framework and forecast future cancer incidence and outcomes for cancer prevention and control strategies. These models adjusted for covariates at the patient, area or temporal level, and through standardization procedure. In other study, Sujit and Dankmar (2021) proposed a two-stage hierarchical Bayesian model as a joint bivariate model for the number of cases and Covid-19 death observed weekly for the different local authority administrative regions in England. Besides that, in different study, Nushrat *et al.* (2022) conducted a systematic literature search of spatial studies of COVID-19 published in English from Embase, Scopus, Medline, and Web of Science databases. In this study, the methodological approaches used to identify the spatial and spatio-temporal variations of

COVID-19 and the socioeconomic, demographic and climatic drivers of such variations were reviewed. This review highlighted the need for more local-level advanced Bayesian spatio-temporal modelling through the multi-level framework for COVID-19 prevention and control strategies.

However, these spatio-temporal models suffer from the comparability problems of relative risks (RRs) as a result of the removal of the covariate effect as a confounding factor on the risk estimate of the study population through distribution of standardized mortality ratios (SMRs). This paper is aimed at fitting a spatio-temporal Bayesian model with two-level spatial structure that includes covariate effect as an alternative to the traditional use of SMRs in these types of models. We shall evaluate the effect of covariate in terms of model performance and detect the female age-group with higher risk on breast cancer.

2. Materials and Method

2.1 *Description of Data*

Over the past 3 decades, the Institute of Human Virology, Nigeria, the Nigerian National System of Cancer Registries (NSCR) and the Nigerian Federal Ministry of Health have collaborated on training and streamlining the activities of cancer registries in Nigeria. In view of this, the Nigerian National System of Cancer Registries (NSCR) since 2009 has been involved in coordinating cancer registration in Nigeria (Elima *et al.* 2015). During this period, 21 institutions have been trained and provided support for cancer registration. Of these, nineteen (19) met the definition of Hospital-Based Cancer Registries (HBCRs) and two (2) met the criteria for Population-Based Cancer Registries (PBCRs).

The NSCR requested for data from the director of the cancer registries for the period under review, 2009 to 2016. Most, 16 of the 19 HBCRs (84%) responded. The NSCR received data from 16 HBCRs namely; University of Nigeria Teaching Hospital, Enugu (UNTH), University of Ilorin Teaching Hospital, Ilorin (UITH), University of Port Harcourt Teaching Hospital, Port Harcourt (UPTH), Federal Medical Centre, Ekiti (FMC Ido-Ekiti), University of Calabar Teaching Hospital, Calabar (UCTH), Obafemi Awolowo University Teaching Hospital, Ile-Ife, (OAUTH) Nnamdi Azikiwe University Teaching Hospital, Nnewi (NAUTH), Lagos State University Teaching Hospital, Ikeja(LASUTH), Lagos University Teaching Hospital, Suru-Lere (LUTH), Jos University Teaching Hospital, Jos (JUTH), Aminu Kano Teaching Hospital, Kano (AKTH), Federal Medical Centre, Gombe, University of Benin Teaching Hospital, Benin (UBTH) University of Abuja Teaching Hospital Gwagwalada (UATH), and the Ahmadu Bello University Teaching Hospital, Zaria (ABUTH). However, 3 registries did not respond to the call for data despite several reminders namely: Federal Medical Centre, Keffi, Federal Medical Centre, Lokoja and University of Uyo Teaching Hospital, Uyo (UUTH).

All these registries used CanReg5 software by the International Agency for Research on Cancer (IARC) for data entry and management and International Classification of Disease for Oncology, third edition (ICD-O-3) for coding the site. Data abstracted by the HBCRs included information on name, age, morphology and topography of tumor, tribe, address, treatment, education level,

marital status, religion and cause of death. However, the data for this study was extracted from the NSCR from 2009 to 2016. The data consist of female breast cancer mortality from 16 hospital-based cancer registries (HBCRs). A total of 4,437 female breast cancer deaths were recorded throughout the study period. The data for each case on individual-level characteristics (that is, age, occupation, marital status, and religion at the time of diagnosis) were obtained at individual hospital-based cancer registry.

Individual patient characteristics of interest

(i) **Age group**(AGE) — Age at time of diagnosis was collapsed into fifteen-year age groups from: (0-14 years) as Puberty age group and (15 – 29 years), (30-49 years), (50 – 69 years) and (70 and above) as adult's age groups, resulting in five age groups.
(ii) **Occupation**(OCCUP) —Three categories of occupation are skilled, semi-skilled and unskilled.
(iii) **Marital status** (MARITAL) — Women's marital status includes single/never married, married, widowed, and separated.
(iv) **Region**(RIG) — Women's region includes south east, south west south south, north east, north south and north center.

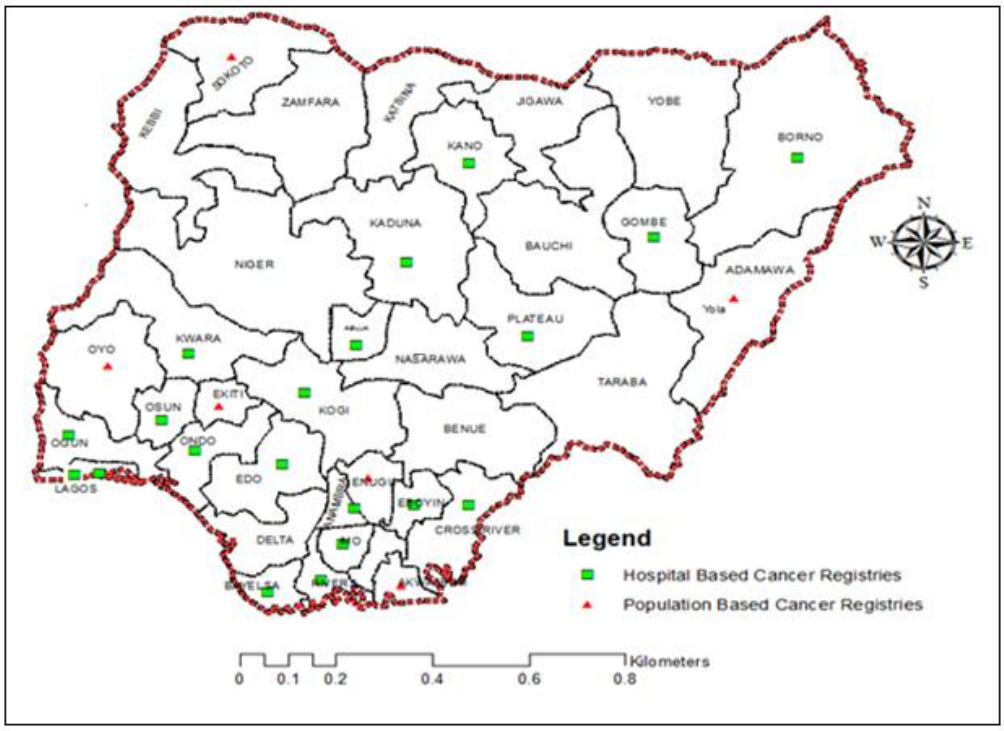

Figure 1: Map of the $n = 36$ States of Nigeria showing distribution of cancer registries.

2.2 *Statistical Methods*

Two spatio-temporal models with two-level spatial structure with different approaches are considered in this paper for comparison. The first model followed standardised mortality ratios (SMRs) procedure by removal of the effect of the

confounding factors on the risk estimate in the study population through distribution standardization., while the second model included covariate effect as confounding factors on the risk estimate in the study population. Two-level spatial structure modeled by Ugarte *et al.* (2016) formed the basis of the proposed model. The model has a State-level spatial random effect as first-level areas (FLAs) and a health area-level spatial effect as second-level areas (SLAs). In addition to covariate effect, spatial random effects, temporal random effect and space-time interaction effect were studied. These two models are built in the form of hierarchy and involve Gaussian Markov random fields (GMRF) model with the integrated nested Laplace approximation (INLA) for estimation (Rue *et al.*, 2009) which is computationally more efficient than the well-known Markov chain Monte Carlo approach. (Schrodle Held (2011); Rue *et al.* (2009) and Blangiardo (2013)).

2.3 *Model I: Spatio-temporal model with two-level spatial structure using standardised mortality ratios (SMRs) procedure*

Suppose we have a region with non-overlapping small areas divided into q first-level areas (FLAs) labeled as $i = 1, 2, ...q$ that can be aggregated into p second-level areas (SLAs) labeled as $j = 1, 2, ..p$ where $p < q$ For each area i, data are available for different time periods labeled by $t = 1, 2, ..T$. A spatio-temporal model that accounts for two-level spatial structure of dependence of the SLAs which are in FLAs and for the temporal dependence assumes that, conditional on the underlying relative risk r_{it} the number of deaths counts in each area and time period, it, follows a Poisson distribution with mean the number of deaths counts in each area and time period, y_{it} it, follows a Poisson distribution with mean $\mu_{it} = r_{it}E_{it}$ for area i and time t written as:

$$y_{it}|r_{it} \sim Poisson(\mu_{it} = r_{it}E_{it}) \tag{1}$$

$$log(\mu_{it}) = log(r_{it}) + log(E_{it}) \tag{2}$$

Here, $log(E_{it})$ is an offset and depends on the specification of $log(r_{it})$ In this study, the log-risk (r_{it}) is modeled taking into account the need of distinguishing between space organised in two-level structure (that is, FLAs and SLAs) and time components, and including interaction in space and time.
Using direct or indirect 'age' standardisation mortality ratio procedures, the number of expected deaths E_{it} in Equation (2), an be performed using Equations (3) and (4) respectively:

$$E_{it} = \sum_{i=1}^{L} N_{igt} \frac{\sum_{i=1}^{q} y_{igt}}{\sum_{i=1}^{q} N_{igt}} \quad i = 1, 2, ..., q \quad t = 1, 2, ..., T \tag{3}$$

where y_g and N_g are respectively the observed deaths and the population size in "age" groups $g \in \{1, ...G\}$

$$E_{it} = \sum_{g=1}^{G} \left(N_{igt} \frac{y_g}{N_g} \right) \quad i = 1, 2, ..., q \quad t = 1, 2, ..., T \tag{4}$$

so that,

$$y_g = \sum_{i}^{q} \sum_{t}^{T} y_{igt} \;\; N_g = \sum_{i}^{q} \sum_{t}^{T} N_{igt} \tag{5}$$

then E_{it} represents the number of deaths we would expect if the area i in time point t behaves as the whole region during the studied period. The log-risk can be modelled as:

$$log(r_{it}) = b_0 + \alpha_i + \beta_{j(i)} + \eta_t + \delta_{j(t)} \tag{6}$$

where b_0 is an overall risk level, α_i represents the spatial level for the i^{th}State areas,$\beta_{j(i)}$ represents the spatial level for the j^{th} Health areas which are in State i^{th} area, η_i represents temporal structured effect, and $\delta_{j(t)}$ are space-time interaction effects.

2.4 *Model II: Spatio-temporal models with two-level spatial structure that includes covariate effect*

Suppose that the region under study is divided into q first-level areas (FLAs) labeled as$i = 1, 2, .., q$ that can be aggregated into p second-level areas (SLAs) labeled as $j = 1, 2, ..., p$ where $p < q$ For each area i data are available for different time periods $t = 1, 2, ..., T$ let y_{ikt} denote the number of counts of the k_{th} individual in the FLA i_{th} region at the t_{th} time interval, with individual characteristics of interest $x_{it} = \left(x_{ik}^{(1)}, ..., x_{ik}^{(r)} \right)$ where $i = 1, 2, ..., n$, $k = 1, 2, ..., K$ and $t = 1, 2, ..., T$. Here, y_{ikt} is the number of counts that follows a Poisson distribution. This model assumed that, the mean μ_{it} is defined in terms of a rate r_{it} and the patient age Pa_{it} as $\mu_{it} = r_{it}Pa_{it}$ for area i and time t, which can be written as:

$$y_{ikt}|r_{itk} \sim Poisson(\mu_{itk} = r_{itk}Pa_{itk}) \tag{7}$$

$$log(\mu_{ikt}) = log(r_{itk}) + log(Pa_{itk}) \tag{8}$$

The individual risk is modelled via a regression model with inclusion of patient's age, spatial and temporal random effects and spatio-temporal interaction

term. The log-risk is modelled as:

$$log(r_{it}) = b_0 + \sum_r \omega_r x_{ik}^r + \alpha_i + \beta_{j(i)} + \eta_t + \delta_{jt} \tag{9}$$

where b_0 is an overall risk level, ω_r is the coefficients of covariates on the response, x_{it} is the value of the r_{th} covariate for the i_{th} area,α_i represents the FLA for the i_{th} area,$\beta_{j(i)}$ represents the SLA for the j^{th} Heath areas which are in FLAs i^{th} area, η_t denotes temporal effects and δ_{jt} is the SLA space-time interaction effect.

2.5 *Modeling the prior distributions for covariate, spatial, temporal and spatio-temporal effects*

2.5.1 Covariate effect

In modeling the covariate effect, univariate normal priors with mean zero and precision σ_ω are assigned to each of the regression effects $\omega = (\omega_1 + \omega_2 +, , , +\omega_r)$ The precision parameter σ_ω is assigned a Gamma distribution $\sigma_\omega \sim Gamma\,(a_\omega, b_\omega)$ with a and b representing the shape and scale of the Gamma distribution (Bernardinelli *et al.* 1995a).

2.5.2 Two-level spatial effects

For the spatial random effects, the conditional autoregressive (CAR) prior by Leroux *et al.* (1999) (LCAR) was adopted for FLA spatial random effect given by:

$$\alpha = (\alpha_1 + \alpha_{q=37})^{'} \sim N\left(0, \left[\sigma_\alpha\left(\lambda_\alpha R_\alpha + (1-\lambda_\alpha)\right) I_{q=37}\right]^{-1}\right) \tag{10}$$

where λ_α is a spatial smoothing parameter taking values between 0 and 1, I_q is an identity matrix of dimension 37×37, and R_α is the 37×37 spatial neighborhood matrix with diagonal elements equal to the number of neighbours of each State and non-diagonal element:

$$(R_\alpha) = \begin{cases} -1, if\ \ State\ i\ and\ j\ \ are\ \ neigbours \\ 0, if\ \ otherwise \end{cases} \tag{11}$$

Here, two States are considered as neighbours if they share a common border. The SLA spatial random effect is given by:

$$\beta = (\beta_1 + \beta_{q=16})^{'} \sim N\left(0, \left[\sigma_\beta\left(\lambda_\beta R_\beta + \left(1-\lambda_\beta\right)\right) I_{q=16}\right]^{-1}\right) \tag{12}$$

where R_β is the 16×16 spatial neighbourhood matrix of the SLAs and I_q is an identity matrix of dimension 16×16. This means that in space, each HA may have its own risk, but all HA within a State region share a common spatial effect.

2.5.3 Temporal effect

For structured temporal effects $\eta = (\eta_1, ..., \eta_T)^{'}$ a random walk of first order (RW1) is considered

$$RW1 = \eta_t|\eta_{t-1} \sim N\left(\eta_{t-1}, \sigma^2\right) \tag{13}$$

$$\eta = (\eta_1, ..., \eta_T)^{'} \sim N\left(0, [\sigma_\eta, R_\eta]^{-1}\right) \tag{14}$$

R_η denotes the temporal structure matrix of a RW1 and the symbol "-"denotes the Moore-Penrose generalised inverse. That is, in time each year has two neighbors, the previous point and the following one, except for the first and last year, which only depends on one.

2.5.4 Spatio-temporal effects

A completely structured interaction terms $\delta = (\delta_{11}, ..., \delta_{nT})^{'}$by Knorr-Hel (2000) are adopted and assumed to be distributed normally for the FLA and SLA interactions in Equation (9) given by:
$\delta_{it} \sim N\left(0, \sigma_\delta \left(R_\alpha \otimes R_\eta\right)^{-}\right)$ and $\delta_{jt} \sim N\left(0, \sigma_\delta \left(R_\beta \otimes R_\eta\right)^{-}\right)$ The parameters of interest are thus $\theta = (b_0, \omega, \alpha, \beta, \eta, \delta)$ with hyper-parameters represented by:

$$\varphi = \left(\sigma_\alpha, \omega_i, \lambda_\alpha, \beta_\beta.\lambda_\beta, \sigma_\eta, \sigma_\delta\right) \tag{15}$$

Then, to estimate the marginal posterior distribution of all parameters in Equation (15), the method of approximation, INLA is used. In the INLA approach, the deviance information criterion (DIC) can be evaluated for the best model selection.

2.6 *Model selection*

The best model is chosen based on the lowest DIC value. According to Spiegelhalter et al. (2002), DIC is the summation of the deviance posterior mean, and the effective parameters number, pD. The deviance posterior mean is a measure for model fit while the effective parameters number is a measure for model complexity. The lowest DIC value provides a balance between model fit and model complexity given as:

$$DIC = \bar{D} + pD \tag{16}$$

3. Results and Discussion for model I and II.

3.1 *Results*

The first step in these analyses is the preliminary analyses of the four covariate effects using combination method $\binom{n}{r}$ to fit the possible combinations of the four covariates . That is, all the combinations of four covariates $\binom{4}{4} = 1$ model, three covariates $\binom{4}{3} = 4$ models, two covariates $\binom{4}{2} = 6$ models, and one covariate $\binom{4}{1} = 4$ models. A total of 15 models were fitted to the female breast cancer data. The spatial and temporal effects were not included in the model in these preliminary analyses.

Table 1: DIC values of 15 model combinations for the selection of the covariate effect.

	Model	AGE	DIC	OCCU	DIC	MARIT	DIC	RIG	DIC
4 covariates	1	*	544.1	*	548.3	*	547.7	*	550.2
3 covariates	1	*	582.6	*	588	*	586		
	2	*	660.4	*	659.7			*	662.7
	3	*	693.3			*	681.4	*	685.5
	4			*	699.5	*	711.2	*	710.4
2 covariates	1	*	708.1	*	716.7				
	2	*	729.5	*	723.4				
	3	*	741.1						
	4							*	734.5
	5					*	736.7	*	742.2
	6			*	738	*	743.9	*	744.1
1 covariate	1	*	527.8						
	2								
	3					*	534.5		
	4							*	541.6

*This model has the smallest DIC value and parameter estimates that do not contain zero in the 97.5% credible interval.

The combinations of the covariates are shown in Table 1. The DIC values were examined and were in the range (527.80, 744.16) The model with one (1) covariate AGE appeared to have the smallest DIC value. The difference between the smallest DIC value and the second smallest DIC value is 6.67, which suggested a significant improvement in the model with AGE. The study proceeded with further analyses using the model in Equation (9) as,

$$log(r_{it}) = b_0 + \sum_r \omega_r AGE_{it} + \alpha_i + \beta_{j(i)} + \eta_t + \delta_{j(t)} \tag{17}$$

The second step in the analyses is by fitting regression model of the covariate effect (patient's age) as a factor categorized into five age-groups as in Section 2.1. The result of the regression analysis is presented in Table 2.

Table 2: Result of Regression Model of the Covariate Effect (Patients' age group).

Coefficients:				
	Estimate	Std. Error	t-value	$Pr(>\|t\|)$
(Intercept) Age group in years	0.03834	0.2177	0.48	0.638848
(0 - 14)	-0.09882	0.23359	-0.423	0.678686
(15 - 29)	0.94255	0.17799	5.296	0.000113
(20 - 29)	1.03678	0.02997	34.598	$5.81e^{-15}$
(50 - 69)	0.97809	0.02369	41.289	$5.00e^{-16}$
≥ 70	0.97112	0.05	19.423	$1.60e^{-11}$

Multiple R-squared: 0.9999, Adjusted R-squared: 0.9999, F-statistic: 3.643e+04 on 5 and 14, DF $p - value$:<2.2e-16

The result of fitted regression model on covariate effect (patients age) in Table 2 shows that, the Puberty age group had no positive effect on cancer mortality. While, the adult age groups all had positive effects on cancer mortality. Their p-values were significant and explained 99% of variability in the regression model. Examining the $p - value$ of each age-group, the most significant age group is (50 – 69 years) with the smallest $p - value$ ($5.00e^{-16} < 0.05$); followed by (30 – 49 years) with $p - value$ ($5.81e^{-15} < 0.05$), and then (70years +) with p-value ($1.60e^{-11}$). The significant age group was (15-29 years) with $p - value$ (0.000113). The multiple R-squared implies that 99.99% of variation in patients age was well explained by the fitted model and the regression model fit the sample data. Adjusted R-squared (:9999), implies 99.99% of goodness of fit of the model.
The model in Equation (16) can be reformulated as:

$$log(r_{itk}) = b_0 + \sum_i \omega_i Adultsage_{it} + \alpha_i + \beta_{j(i)} + \eta_t + \delta_{jt} \tag{18}$$

The third step in the analyses is by fitting the model in Equations (6) and (18) based on SMRs and patient's covariate as an alternative to the traditional method.

Table 3: Posterior estimates and DIC for two-level model without and with covariate:

Parameters	**Mean**	**SD**	**2.50%**	**50%**	**97.50%**
$log(r_{it}) = b_0 + \alpha_i + \beta_{j(i)} + \eta_t + \delta_{j(t)}$ Without covariate DIC= 1516.507					
b_0	0.0371	0.3597	0.0324	0. 0587	0.0734
σ_α	12.3924	4.4137	3.846	12.3252	21.3264
λ_α	0.9667	0.3641	0.262	0.961	1.7044
σ_β	69.3504	10.0996	49.8092	69.188	89.8324
λ_β	2.1298	0.3625	1.4285	2.124	5.8647
σ_η	77.0241	20.1195	35.1063	73.7112	114.7685
σ_δ	106.4027	74.5801	38.1419	105.346	257.0629
$log(r_{itk}) = b_0 + \sum_i \omega_i Adultsage_{it} + \alpha_i + \beta_{j(i)} + \eta_t + \delta_{jt}$ With covariate DIC= 1516.493					
b_0	0.0383	0.2177	0.3084	0.0597	0.0766
$\omega_{(15-29)years}$	0.2264	-0.3832	-0.8113	-0.0709	0.7043
$\omega_{(30-49)years}$	0.6041	0.3659	0.6029	1.3054	2.0522
$\omega_{(50-69)years}$	0.8286	0.3798	-0.6387	0.0938	0.8637
$\omega_{(70+)years}$	0.5902	0.372	-0.1146	0.6008	1.3579
σ_α	12.2461	4.2674	3.7003	12.1789	21.1901
λ_α	0.9988	0.3962	0.2941	0.9931	1.7365
σ_β	69.3599	10.1091	49.8187	69.1975	89.8419
λ_β	2.1307	0.3634	1.4294	2.1249	2.8656
σ_η	76.813	19.9084	35.2574	74.0669	116.0051
σ_δ	106.2911	74.1643	37.1219	104.5275	256.0429

Table 3 shows the estimated posterior means and quantiles intervals for the model parameters implemented with SMRs and with covariate effect. The fixed effects b_0, $\omega_{(15-29)years}$, $\omega_{(30-49)years}$, $\omega_{(50-69)year}$ and $\omega_{(70year+)}$ estimated as relative risks: an increase of age in the adult's age groups is associated respectively with an increase of around 22.64%, 60.41%, 82.86%, and 59.02% in the risk of female breast cancer mortality. It is observed that, age group $(15-29)\,years$, $(30-49)\,years$, $(50-69)\,year$, $(70year+)$ is the most significant with higher risk of breast cancer mortality. This means that, there is evidence of breast cancer effectiveness for women from 50 years of age.
Considering the DIC values presented in Table 3 as a criterion for evaluating the fit of the model, suggested that the model with covariate effect is better suited for the data (with a DIC of 1516.507 against 1516.493 obtained). Though the two DIC values are almost identical, and their posterior estimates for the parameters do not change considerably between the two models.

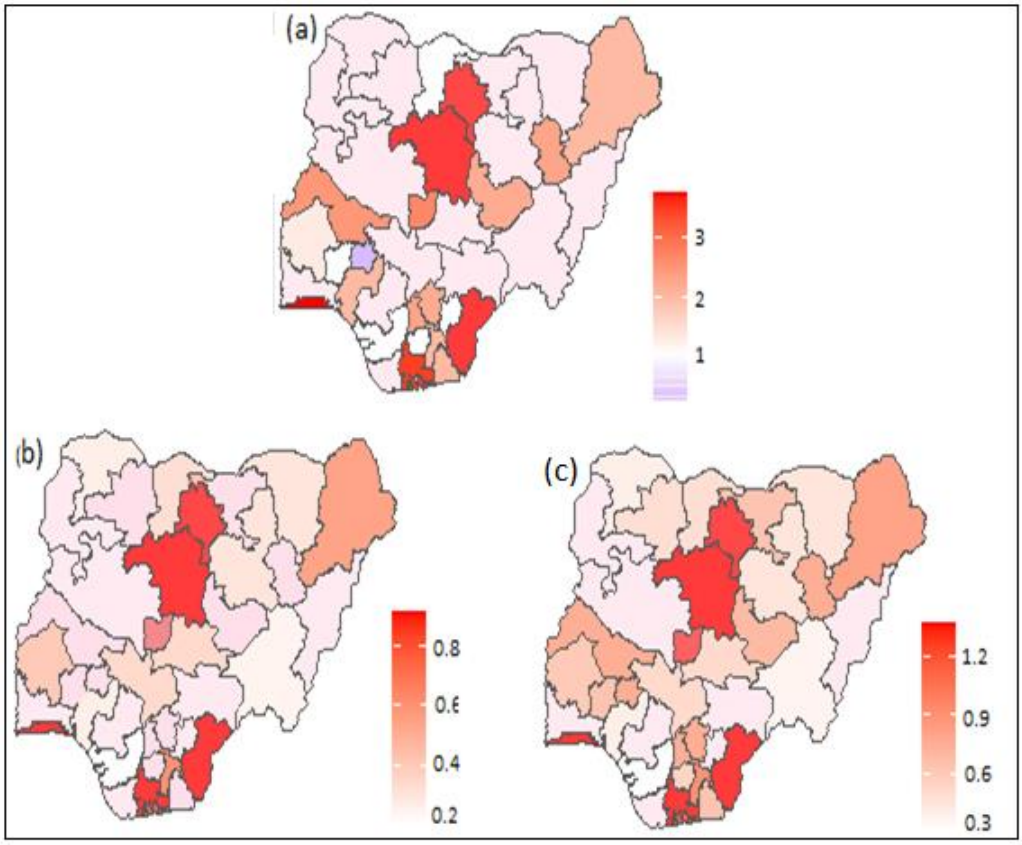

Figure 2: Map of the spatial pattern of mortality risks and Map of posterior probabilities of the spatial pattern of mortality risk using (b) model I and (c) model II.

Figures 2 shows the spatial patterns of mortality risk due to breast cancer at each hospital-based cancer registry and their posterior probabilities that the spatial risk is greater than one (1) using the two methods. In spatio-temporal disease mapping studies, usually the regions with probabilities above 0.8 and 0.9 are considered as high-risk regions, similar to Ugarte *et al.* (2009a) suggested. In this study, a reference threshold equal to 1 and a cut-off value of 0.8 is considered. From Figures (b) and (c) it can be used to detect the health areas where the posterior probability of breast cancer mortality is exceeding 0.8 and hence, the focus should be given to those health areas.

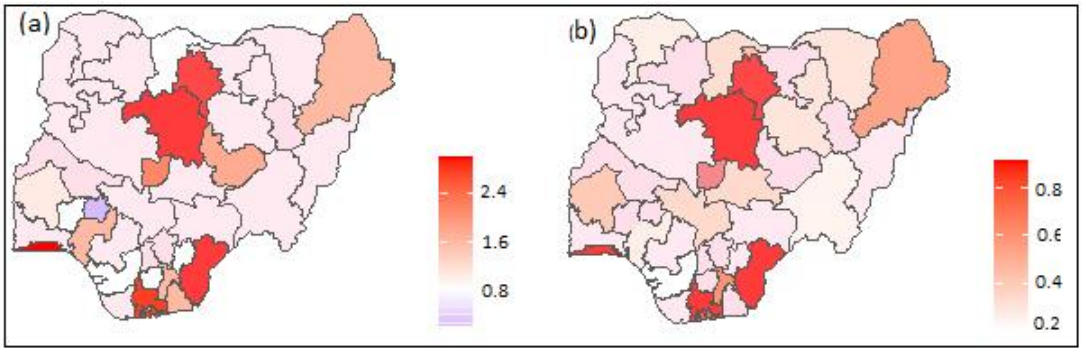

Figure 3: (a) Map of the spatial pattern of mortality risks without covariate effect (b) with covariate effect.

Figures 3 further shows the map of posterior means of the spatial pattern of

mortality risks of the two-level model without and with covariate effect. From both Figures (3b) and (3b), it can be seen that the health areas showing red colour are Kaduna, Kano, Lagos, Cross River and Rivers as the health areas of high risk of mortality due to breast cancer disease. Besides that, health areas that having tendency to become a high value of risk are health areas in Abuja, Jos, Gombe, Ile Ife and Enugu showing in light red colour. Other health areas having low risk are health areas in Ado Ekiti, Nnewi, Benin, Kwara and Borno.

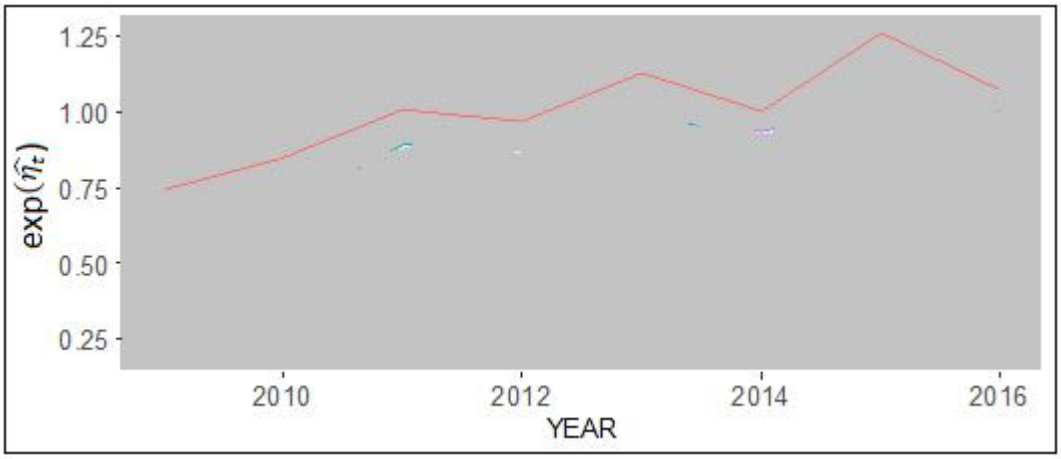

Figure 4: Map of temporal trend of breast cancer mortality relative risk $exp\,(\hat{\eta}_t)$

Figure 4 is the temporal pattern common to all heath areas and it can be shown that these times increase in the female breast cancer mortality throughout the period.

Table 4: Deviance information criterion (DIC) for the two-level models.

Two-level Model	$\bar{D}$	pD	DIC
Without covariate	1506.646	10.3404	1516.507
With covariate	1509.407	11.1017	1516.493

Table 4 shows summary of the posterior mean of the deviance, the number of effective parameters and the deviance information criterion (DIC) as a measure of trade-off between model fit and complexity for the two models. This also confirmed the result obtained in Table 3 suggesting that two-level model with covariate effect is the best model and displayed the best fit for the data.

3.2 *Discussion*

This study fitted an alternative spatio-temporal Bayesian model with two-level spatial structure that included covariate effect to evaluate the effect of covariate in terms of model performance and estimation through integrated nested Laplace approximation (INLA) procedures.
The result of fitted regression model on covariate effect suggested that, adult age groups all had positive effects on cancer mortality. Their $p - values$ were significant and explained 99 of variability in the regression model. The study found that, the posterior summary statistics of the model parameters implemented with and without covariate effect do not change considerably between

the two models and their DIC values are almost identical, though the model implemented with covariate effect produced the smallest DIC value of (1516.493 against 1516.507). In the model with covariate effect, the fixed effects estimated as relative risks revealed that adult's age group (50 – 69 years) are those with high risk. The study also showed that there is no substantial difference between the map of posterior means of the spatial pattern of mortality risk (a) without covariate effect (b) with covariate effect.

4. Conclusion

The study has found that, model II does not suffer from the comparability problems of RRs based on the SMRs. Though the DIC values for both models are almost identical, and their posterior estimates for the parameters do not change considerably between the two models. It can be concluded that both models are very beneficial for spatio-temporal study evolution. However, the most suitable model only can be chosen and determined by the nature of the data and also based on the objective of the study.
Besides that, the regression analysis of the covariate effect (patients' age group) female breast cancer mortality in Nigeria shows that the gap of the relative risks between the age groups under study is identified. Women of 50-69 years age group is identified as most significant high risk of breast cancer mortality. Hence, primary health care workers should teach self-examination to women aged 30 years and above. In addition, the national primary health care should also, establish policy for more than 80 of women aged 50-69 years to be screened for mammography programmes once every 2-3years.

Acknowledgement

The authors of this study acknowledged the entire staff of National Cancer Control Program and the Institute of Human Virology, Federal Ministry of Health of Nigeria.

References

Adin, A., Lee, D. Goicoa, T. and Ugarte, M. D. (2019). A two-stage approach to estimate spatial and spatio-temporal disease risks in the presence of local discontinuities and clusters. Statistical Methods in Medical Research, 28(9), 2595-2613.

Bernardinelli, L., Clayton, D., and Montomoli, C. (1995a). Bayesian estimates of disease maps: how important are priors? Statistics in Medicine, 14(21-22), 2411-2431

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M. (1995b). Bayesian analysis of space-time variation in disease risk. Statistics in Medicine, 14(21-22),2433-2443.

Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013). Spatial and spatiotemporal models with R-INLA. Spatial and Spatio-Temporal Epidemiology, 7, 39-55.

Duncan, E. W., White, N. M., and Mengersen, K. (2016). Bayesian spatio-temporal modelling for identifying unusual and unstable trends in mammography utilisation. BMJ Open, 6 (5), e010253. doi: 10:1136/bmjopen-2015-010253.

Elima, E. J. A., Emmanuel, A. O., Michael, O., Yusuf, M. A., Abiodun, P., Peter, A., Enoch, A., Adekunbiola, A. F. B., Ima-Obong, E., Olagoke, E., Emmanuel, E., Festus, I.,

Christopher, O., Olufemi, O., Abidemi, O., Clement, O., Cornelius, U., Patience, O., Ramatu, H., William, B., Patrick, D. and Clement, A. A. (2015). Developing national cancer registration in developing countries. Case study of the Nigerian National system of cancer registries. Frontiers in Public Health, 3(186).

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. Statistics in Medicine, 19(17-18), 2555-2567.

Leroux, B. G., Lei, X., and Breslow, N. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In Halloran, M. and Berry, D., editors, Statistical Models in Epidemiology, the Environment, and Clinical Trials, pages 179-191. Springer-Verlag: New York.

Librero, J., Ibañez, B., MartõÂnez-Lizaga, N., PeiroÂ, S. and Bernal-Delgado, E. (2017). Applying spatio-temporal models to assess variations across health care areas and regions: Lessons from the decentralized Spanish National Health System. PLoS ONE 12(2), e0170480.

Melkamu, D., Henry, M., Sileshi, F. and Nega, A. (2018). Spatio-temporal mapping and detection of mortality cluster due to cardiovascular disease with Bayesian hierarchical framework using integrated nested Laplace approximation: A discussion of suitable statistic applications in Kersa, Oromia, Ethiopia. Geospatial Health, 13, 681.

Naresh, N. , Goldbloom-Helzner, A. and Ali, A. (2021). Spatio-temporal modeling for confirmed cases of lyme disease in Virginia. PMID: 34555712, DOI: 10.1016/j.ttbdis.2021.101822

Nurul, S. A, Nuzlinda, A. and Fatin, A. M. F. (2019). A spatial–temporal study of dengue in Peninsular Malaysia for the year 2017 in two different space–time model, Journal of Applied Statistics, 164-8391

Nushrat, N., Zahid, A. B., Melanie, L. B., Wang-Choi, T., Hibah, S. and Jane, L. (2022). Methods Used in the Spatial and Spatiotemporal Analysis of COVID-19 Epidemiology: A Systematic Review. Int J Environ Res Public Health, 6;19(14):8267; PMID: 35886114; PMCID: PMC9324591 doi: 10.3390/ijerph19148267.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society, Series B, 71, 319-392.

Schrödle, B. and Held, L. (2011). Spatio-temporal disease mapping using INLA. Environmetrics, 22(6), 725-734.

Schrödle, B., Held, L., and Riebler, A. (2011). Using INLA for the Evaluation of Veterinary Surveillance Data from Switzerland: a case study. J Roy Stat Soc C, 60, 261-279.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(4), 583-639.

Sujit, K. S. and Dankmar, B. (2021) Bayesian spatio-temporal joint disease mapping of Covid-19 cases and deaths in local authorities of England; PMID: 33996424; PMCID: PMC8114675; doi: 10.1016/j.spasta.2021.100519

Ugarte, M. D., Adin, A., and Goicoa, T. (2016). Two-level spatially structured models in spatio-temporal disease mapping. Statistical Methods in Medical Research, 25(4), 1080-1100.

Ugarte, M. D., Adin, A., and Goicoa, T. (2017). One-dimensional, two-dimensional, and three-dimensional B-splines to specify space-time interactions in Bayesian disease mapping: model fitting and model identifiability. Spatial Statistics (accepted for publication, DOI: 10.1016/j.spasta.2017.04.002).

Ugarte, M. D., Goicoa, T., Ibáñez, B., and Militino, A. F. (2009a). Evaluating the performance of spatio-temporal Bayesian models in disease mapping. Environmetrics, 20(6), 647-665.

Win, W., Susannah, A. and Arul, E. (2020). A systematic review of Bayesian spatial-temporal models on cancer incidence and mortality. International Journal of Public Health, 65(5), 673-682, doi: 10.1007/s00038-020-01384-5.