Spatial heterogeneity and determinants of childhood anaemia in Nigeria

Chigozie Louisa Jane Ugwu (✉ 217075063@stu.ukzn.ac.za)
University of KwaZulu-Natal - Westville Campus https://orcid.org/0000-0001-5013-2253
Temesgen Zewotir
University of KwaZulu-Natal Westville Campus

Research article

Keywords: Empirical Bayesian method, Childhood anaemia, Markov random field models, Structured additive, Spatial epidemiology

Posted Date: March 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-15535/v1

Spatial heterogeneity and determinants of childhood anaemia in Nigeria

Chigozie Louisa J. Ugwu, Temesgen Zewotir

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal Westville Campus, Durban, South Africa

Abstract

Background: Childhood anaemia is highly prevalent in Nigeria. According to the 2015 Nigeria Malaria Indicator Survey (NMIS) report, more than 68% of children aged 6-59 months were found to be anaemic. This estimate is far above the World Health Organization's 40% cut off point which classifies anaemia a severe public health challenge in the country. Identifying environmental, health, socioeconomic and demographic influential factors and mapping the prevalence of anaemia can help guide geographically targeted intervention programmes to reduce the risk of anaemia associated morbidity among vulnerable children in Nigeria.

Methods: Geographically linked national level datasets obtained from the 2015 Nigeria Demographic and Health Survey (NDHS) and Malaria Indicator Survey (NMIS) programmes were used for this study. For the analysis, a binary structured additive regression (BSAR) model was explored. This model incorporated a Markov Random Field (MRF) prior, with posterior parameters estimated via Bayesian framework.

Results: After accounting for the spatial heterogeneity, we found a strong negative association between the odds of anaemia and the child's demographic variables in terms of increasing age and being female. Increased odds of anaemia were also associated with child's malaria and fever status, living in a rural environment, lower household wealth quintile, having younger and illiterate mothers. We also found that a decreased distance to vegetated areas was associated with increased childhood anaemia risk, while the odds of anaemia decreased as cluster altitude increased at 95% credible interval. The maps of the posterior means of the spatial random effects revealed evidence of spatial variation in the odds of childhood anaemia, while accounting for the model covariates (Fig. 4). Greater risk of anaemia was observed for children who resided in Adamawa, Ebonyi, Edo, Cross river, FCT, Jigawa, Kaduna, Kano and Kebbi states in Nigeria.

Conclusions: In this study, a binary structured additive regression model was utilized, allowing for a flexible semiparametric predictor that accounted for the effects of different types of covariates, while simultaneously incorporating spatial variables directly. Our results revealed significant spatial variability of childhood anaemia, suggesting that spatially targeted interventions could result in efficiency gains for anaemia control in Nigeria.

Keywords: Empirical Bayesian method, Childhood anaemia, Markov random field models, Structured additive, Spatial epidemiology

Background

Anaemia is considered a severe public health problem in sub-Saharan African countries, including Nigeria, where it primarily affects preschool children under the age of five years (Under-5s) as well as pregnant women [1]. Anaemia is challenging to address due to the endemicity of malnutrition and malaria in most of these countries [1]. For children under 5 years of age (under-5s), anaemia disease is a major cause of morbidity and mortality [1]. According to the 2015 Malaria Indicator Survey report, the prevalence of anaemia is extremely high in Nigeria, about 68% of the children aged 6-59 months were found to be anaemic [2]. This estimate is far above the World Health Organization's 40% cutoff point, which classifies anaemia as a severe public health problem in the country [1, 2]. Anaemia has continued to threaten the health and wellbeing of millions of Nigerian children, inducing heavy economic burden to the nation [1-5]. Among the well documented consequences of childhood anaemia include impaired cognitive development, impaired immune functions, increased susceptibility to infectious diseases, impaired motor development, poor performance in pre-school, and short or long term mortality in acute severe cases of anaemia [1, 6-12].

The causes of anaemia are multifactorial and interlinked, over 45% of anaemia cases are due to iron deficiency, although, the proportion varies among population, location as well the prevailing local conditions [1, 6, 11]. The non-communicable (e.g., malnutrition) and communicable diseases (e.g., malaria and neglected tropical diseases) are also main known contributors [1, 4, 13-18]. In Uganda, population based cross-sectional studies revealed that low-level of iron intake and malaria were the two major causes of childhood anaemia [9, 19]. Studies from other African countries have also demonstrated that the likely cause of childhood anaemia varies depending on the area of the world in which the child lives, although infectious diseases

such as malaria, helminth infections, HIV, and tuberculosis are also important causes of anaemia [3, 8, 9, 16, 20, 21]. It was revealed that the prevalence of anaemia is highly associated with changes in malaria transmission, that malaria and anaemia are inextricably related, so children who are anaemic are at a great risk of death, and greater risk if it is combined with malaria [4, 15, 22-24]. A combination of malnutrition [19, 21, 25, 26] and other infectious diseases such as diarrhea, fever and intestine helminth [23, 24, 27-32], as well differences in socioeconomic, environmental and genetic factors have been found to influence anaemia and its etiology [4, 25, 28, 33]. Several authors have also established a link between the childhood anaemia outcome and household socioeconomic status, low maternal education attainment, less nurturing child-rearing environments, younger maternal age, non-exposure to mass media as well as number of household members [3-5, 8, 10, 20, 28, 30, 33-36]. Studies have consistently reported that children who reside in rural areas, belonging to poorer family and whose mothers have no formal education are at a greater risk of anaemia [4, 10, 25, 28, 35].

Severe anaemia cases can be influenced not only by individual and household factors, but also by environmental determinants [25, 28]. The exposure to many of these conditions may be determined by anthropogenic environmental factors (i.e., human impacted habitat conditions) such as sanitation, urbanisation, and overcrowding as well natural environmental factors such as elevation, rainfall, enhanced vegetation index, cluster altitude, temperature, and water availability [25]. Anaemia indicators are likely to vary between locations given the geographical variability of its major determinants. This property instigates a challenge to decisions around efficient allocation of public health services and resources to reduce the burden of disease in high risk populations [33]. The application of spatial epidemiological approaches has been useful for identifying areas where the risk of diseases is very high and for highlighting areas where interventions must be

enhanced. Spatial epidemiological approaches have been utilized to investigate childhood anaemia prevalence in many of the sub-Sharan African countries [10, 14, 25, 28, 32-34].

Mapping of anaemia prevalence based on Bayesian spatial models that include environmental drivers of the disease has not been widely studied in Nigeria. Previous studies in Nigeria have well documented the nonspatial factors affecting childhood anemia, the studies were limited to small sample data using hospital-based data and community-based surveys, which do not generally represent national level information on childhood anaemia prevalence [15, 23, 37-39], others were majorly on the prevalence of anaemia among pregnant women [29, 40-46]. However, very few studies that did venture into spatial analysis using national level data did not explore the impacts of geospatial-environmental covariates on childhood anaemia prevalence [4, 5]. Despite the public health significance of anaemia in Nigerian children, its broader study and important preventable risk factors remain largely unexplored using nationally representative data.

To gain a more comprehensive understanding of the distribution of childhood anaemia in Nigeria and to examine its important determinants, nationally representative data must be used. It is imperative to assess the combined effect of geographic, environmental, health and sociodemographic factors that may explain the distribution of anaemia using geographical information system (GIS) and model-based Bayesian geostatistical methods [47-49]. Given the relationships between environmental factors in malaria transmission, along with the indirect associations of many environmental factors with anemia through malaria transmission, anaemia can as well be strongly influenced by environmental conditions [17, 25, 28].

In this study, we used nationally representative NMIS and NDHS datasets to-quantify the geographical disparities in childhood anaemia prevalence in Nigeria and to investigate the role of environmental, health, socioeconomic and demographic factors on the anaemia prevalence among

children aged under 5 years. Geospatial-environmental covariates such as population density, enhanced vegetation index (EVI) and proximity to the permanent waters (coast or large lakes), were included in our study, specifically due to the relative lack of literature that explores their impact on anemia prevalence. We employed a structured additive regression (STAR) models for the anaemia binary response, that allowed for a flexible semiparametric predictor in order to account for the effects of different types of covariates. Using this model, the study simultaneously accounted for the nonlinear effects of continuous covariates, spatial effects of state of child's residence as well as usual linear effects of the covariates in a unified model framework. The outcomes of this research can help inform the design intervention and control programmes that may be required to address not only anaemia prevalence, but also malaria prevalence across all the states of Nigeria.

Methods

Study area and data

Nigeria is a very large landlocked country located in sub-Saharan Africa, comprising of 37 geographical states with Abuja as the Federal Capital Territory (Fig.1). These states are within the six main geographical regions (Fig 1). According to the last census (2006), Nigeria has a total population of 140,431,790 inhabitants, and an area of 923, 768 square kilometers with a population density of 226 inhabit./km^2 [2, 50]. The country has a tropical climate of wet and dry seasons driven by the movement of the two dominant-winds; the rain-bearing southwesterly winds and the cold, dry, and dusty northeasterly winds, normally called the Harmattan [2]. The country is characterized with a diverse climate and topography, encompassing uplands of 600m - 1,300m in the North Central Zone, east highlands, and lowlands $< 20m$ in the coastal areas, additional lowlands extend through the Sokoto plains to the Borno plains in the north, the coastal lowlands

in western Nigeria, and the Cross-River basin in the east. The highland areas of the country include, the Jos, and Adamawa highlands in the north, which expand down to the Obudu Plateau and Oban Hills in the South-South Zone [2]. Other topographic features include the Niger-Benue Trough and Chad Basin.

The 2015 Nigeria Malaria Indicator Survey (NMIS)

The anaemia data on children aged 6-59 months used in this study were collected from the 2015 Nigeria Malaria Indicator Survey (NMIS) programme within the Demographic and Health Survey (DHS) programme [2, 50]. The 2015 NMIS is the latest national level survey that was designed to provide information on basic indicators of child health, prevalence of anaemia as well as malaria [2]. The enumeration areas (EAs) based on the 2006 population census of the Federal Republic of Nigeria, were the primary sampling units (PSUs) for the survey. The PSUs were designed to provide separate estimates for the key indicators for each of the 37 states in Nigeria including the Federal Capital, Abuja (Fig.1). The survey adopted a stratified two-stage cluster design, where a total of 333 clusters (including 138 clusters in urban and 195 clusters in the rural areas) was chosen in stage 1 [2]. In the second stage of sampling, a systematic sample of 25 households (HHs) was selected on the average for each cluster [2]. The survey provided geocoded data for individual survey clusters (enumeration areas (EAs)) and using the GPS coordinates for the DHS survey clusters, the observed anaemia prevalence for each cluster was subsequently linked to the geographical information provided by a map object in boundary [50]. But, for the respondent's privacy, the coordinates of the 333 Nigeria's DHS clusters were randomly displaced by 5km in rural areas and 2km in urban areas [2, 50]. Detailed information about the NMI and DHS survey can be found online (http://www.measuredhs.com).

Study variables

Response variable

The response variable is the anaemia status of children under the age of 5 years in Nigeria extracted from the 2015 Nigeria Malaria Indicator Survey (NMIS) [2]. During the Survey, all children aged 6-59 months were eligible for anaemia test [2]. Blood samples were collected by a child's finger or heel-prick to estimate the level of hemoglobin level via the HaemoCue system and the results obtained after the test was recorded in the **Biomarker Questionnaire** [2]. The World Health Organization (WHO) recommended specific Hb levels at which a child is classified as having anaemia, children aged 6 to 59 months old are considered anaemic if the Hemoglobin (Hb) concentration levels are below 11.0 g/dL [1]. For the purpose of this research, the anaemia outcome was re-coded as a binary variable where 1 signifies positive outcome and 0 otherwise for $n = 6065$ children aged 6-59 months [2]. More detailed information on the anaemia testing data can be obtained from (http://www.measuredhs.com).

Explanatory variables

Health, socioeconomic and demographic covariates

The NMI survey used a **Household Questionnaire** to obtain basic demographic information on the characteristics of each person listed in the household, including age, sex, education, and relationship to the head of the household [2]. Data on age and sex were used to identify women who were eligible for the individual interview. Other information collected under this questionnaire were the socioeconomic, health information and characteristics of the household possessions, including household wealth index, type of toilet facilities, housing material, availability of mosquito net, use of mosquito nets and many more. Additionally, the **Woman's Questionnaire** obtained information on all women aged 15-49 [2]. The information under this category comprised information on the following main topics (1) education, (2) child malaria

status (3) child fever status and (4) information on malaria treatment and many more [2]. Due to the strong association between malaria and anaemia, the 2015 NMIS included malaria testing for children aged 6 to 59 months using finger or heel prick blood samples, the information was also recorded using **Biomarker Questionnaire** [2].

The literature review helped us to choose typical individual-level child and maternal demographic factors, socioeconomic status, health status of the children and other household level covariates used as predictor variables in our analysis. These variables include: (1) Child-specific factors such as: sex of the children (male=Ref or female), the age of children in months (continuous), child has malaria (yes=Ref or no) and the child's fever status (yes=Ref, or no). For maternal and household characteristics, we included: Mother's education attainment coded as (no education=Ref, primary, secondary or higher education), the mother's age (continuous), ethnicity (Hausa/Fulani, Yoruba, Igbo and Others= Ref), type of place of residence (urban=Ref, or rural) and household wealth index. We adopted the DHS standard household wealth index factor obtained through principal component analysis (PCA) and categorized into Quintiles as (lowest=Ref, second, middle, fourth and highest). More information on the 2015 NMIS survey data may be obtained from [2, 50].

Environmental data (the DHS's geospatial covariates)

We examined effects of environmental factors on the distribution of childhood anaemia in Nigeria using geospatial covariates data obtained from the 2015 NDHS Spatial Analysis repository [50], these are a standardized set of datasets extracted from publicly available remote sensing sources [50, 51]. From the literature, there is an indirect association of the environmental factors with anemia through the malaria transmission and a relative scarcity of literature examining such covariates in the context of anemia prevalence in Nigeria. For example, rainfall creates conditions

that allow enough surface water for mosquito breeding and is often studied for its effects on the transmission of malaria [52]. However, research on direct interactions between rainfall and anemia is limited. Like rainfall, vegetation cover linked to population data has been identified as a predictor for malarial transmission, studies have shown that the risk of dense vegetation cover near households as a facilitates malaria transmission [52, 53]. Moreover, intestinal parasites and helminth infections thrive under good environmental conditions such as temperature and adequate soil moisture, these infections cause blood loss and development of iron deficiency anaemia in children if not properly treated [25, 28]. Based on the combined results from a wide range of literature [14, 24, 25, 28, 33, 34], we felt that it is necessary to include geospatial environmental covariates as part of our analysis and these covariates included the DHS Program's rainfall covariate, defined as the average rainfall of the cells whose centroid falls within a radius of 10 km (for rural points) or 2 km (for urban points), Enhanced Vegetation Index (EVI) covariate, Cluster altitude, land surface temperature (LST) and proximity to water bodies (rivers, coast or large lakes). Table 1 shows information on the data sources and the description of the selected variables, further definitions for the covariates can be found in the DHS geospatial covariate report [54].

Statistical methods

Model description

This study investigates the effects of covariates on the childhood anaemia prevalence and describes the spatial patterns of the childhood anaemia, given that the anaemia observations were clustered in a connected geographical state where each of the children reside.

Let y_i be the anaemia status of the child i at geographical location $s_i, s = 1,2,3....,37$. We define the response $y_i = 1$ if a child's anaemia outcome is positive and $y_i = 0$ otherwise. Thus,

y_i is a Bernoulli outcome with an expected probability of positive anaemia outcome as p_i. This outcome can be modelled using the logistic regression model of the form;

$$y_i \mid \eta_i \sim Bernoulli(p_i) \tag{1}$$

where, the probability of positive anaemia outcome $p_i = p\{y_i = 1 \mid \eta_i\} = \left\{\frac{\exp(\eta_i)}{1+\exp(\eta_i)}\right\}$ and η_i represents the predictors.

However, to model the effects of different types of covariates, we employed a structured additive regression model to extend the common linear predictors in Equation (1) to a more general and flexible, semiparametric predictor [49, 55]. This allows us to account for the effects of different types of covariates, such as plausible nonlinear effect of continuous covariates, the usual linear effects as well as the spatial effects in this form [55];

$$\eta_i = \log\left\{\frac{p_i}{1-p_i}\right\} = \alpha_0 + \nu_i'\gamma + \sum_{j=1}^{q} f_j(\vartheta_{ij}) + f_{spat}(s_i), \tag{2}$$

where, α_0 denotes the overall intercept, γ represent the vector of fixed effects covariates, including: child's malaria RDT outcome, child's fever status, type of toilet facility, land surface temperature, wealth index, mothers' level of education, ethnicity and type of place of residence. These variables correspond to the usual parametric linear part of the predictors $\nu_i' = (\nu_{i1}, \nu_{i2}, \ldots\ldots., \nu_{ip})'$. The f_j components represent possible smooth functions for the nonlinear effects of the continuous covariates ϑ_{ij} which includes: the child's age, the mother's age, the cluster altitude, proximity to water bodies, population density, enhanced vegetation index and rainfall, modelled nonparametrically using P-splines, while f_{spat} represents the possible nonlinear spatial effect parameter used to capture the unobserved spatial heterogeneity in the geographical location s_i. These spatial components can be viewed as proxies for many important unobserved

influential covariates, some of them have strong spatial structures (correlated) and some may only be present locally within the states (uncorrelated) [56]. The spatial effect f_{spat} can be decomposed into spatially structured (correlated) and spatially unstructured (uncorrelated) random effects known as convolution-prior in disease mapping literature, and can be incorporated simultaneously in the model as $f_{spat}(s_i) = f_{spatstr}(s_i) + f_{spatunstr}(s_i)$ [55-57]. The spatially structured effect $f_{spatstr}$ aims at capturing the spatially dependent heterogeneity and can be modelled by a Markov random field (MRF), while the spatially unstructured spatial effect $f_{spatunstr}$ aims at capturing the local effects and can be modelled using a random effect term for the states in Nigeria [56].

The fundamental idea of the structured additive regression model is to approximate any of the nonlinear effects using the linear combination of basic functions (B-spline), such that;

$$f(\vartheta) = \sum_{l=1}^{L} \beta_l B_l(\vartheta) \tag{3}$$

where $B_l(\vartheta)$ represent the basis functions and β_l is the corresponding basis coefficients. For better implementation of the model, we cast all the effects into a generic form such that all the unknown functions can be appropriately expressed as a product of regression coefficients [58]. This was realized by rewriting the structured predictor in (2) in the form of;

$$\eta = v\gamma + \vartheta_1\beta_1 + \vartheta_2\beta_2 + ... + \vartheta_l\beta_l + ...\vartheta_{spat}\beta_{spat} \tag{4}$$

which reduces to $\eta = P\phi$, where $P = (\gamma, \vartheta_1, \vartheta_2, ..., \vartheta_{spat})$ represent suitable design matrices for each of the fixed effects, nonlinear effects as well as the spatial effects, while $\phi = (\alpha, \beta_1, \beta_2, ..., \beta_{spat})$ is a high dimensional vector of coefficients [55]. The components $\vartheta_1, \vartheta_2, ...\vartheta_l, ..., \vartheta_{spat}$ as well as $\beta_1, \beta_2, ...\beta_l, ..., \beta_{spat}$ are defined such that $f_l = \vartheta_l\beta_l$. Similarly, we rewrite the spatial component $f_{spat}(s_i)$ to represent $f_{spat}(s_i) = \vartheta_{spatstr}\beta_{spatstr} + \vartheta_{spatunstr}\beta_{spatunstr}$ [58].

Prior assumptions on functions and model parameters

For the fixed effect parameter γ, we assumed an independent diffuse prior $p(\gamma) \propto const.$

For the continuous covariates, we employed a Bayesian P-splines introduced by [55, 59], as a Bayesian analogue of the P-splines [60]. A P-spline is basically a polynomial spline approximation of smoothing functions [60]. For each of the continuous covariates ϑ_l, an x degree P-splines approximation is given by;

$$f_l(\vartheta_l) = \sum_{c=1}^{c_l} \beta_{lc} B_{lc}(\vartheta_l) \tag{5}$$

where, $\vartheta_{l,\min} = \chi_{l0} < \chi_{l1} < ...\chi_{l,k-1} < \chi_{lk} = \vartheta_{l,\max}$ are equally spaced knots of ϑ_l domain. Following [59] specification, such P-splines can be presented as a linear combination of the $C_l = k + x$ basis functions B_c, such that the estimation of the function f_l can be reduced to the estimation of vector $\boldsymbol{\beta}_l = (\beta_{l1}, ..., \beta_l c_l)'$ of unknown regression parameters [61]. In this study, we employed a cubic spline with 29 knots combined with the second order random walk (RW2) model for estimating the regression coefficients $\boldsymbol{\beta}_l = (\beta_{l1}, ..., \beta_l c_l)'$. The second order random walk can be expressed as:

$$\beta_{lc} = 2\beta_{l,c-1} - \beta_{l,c-2} + m_{lc} \tag{6}$$

where , $m_{lc} \sim N(0, \tau_l^2)$ for $c > 0$ and the diffuse priors β_{l1} and β_{l2} are used as initial values, respectively .

For the spatial effect $f_{spatstr}(s_i)$, a Gaussian Markov Random Field (MRF) prior was adopted [61-63]. The specification of MRF assumes that two geographical states s_i and s_j are neighbors if they share a common boundary [64]. Defining $\beta_{spat} = f_{spat}(s_i), s = 1, 2, ...S$, the MRF prior for the evaluation of the structured spatial effect can be defined as;

$$\beta_{spatstr}(s_i) \mid \beta_{spatstr}(s_j), s_i \neq s_j = \beta_{spatstr}(s_i) \mid \beta_{spatstr}(s_j), s_j \in \partial_{s_i} \sim N\left(\sum_{s_j \in \partial_{s_i}} \frac{\beta_{spatstr}(s_j)}{N_{s_i}}, \frac{\tau^2_{spatstr}}{N_{s_i}} \right) \quad (7)$$

where, $\tau^2_{spatstr}$ is the spatial dispersion parameter, N_{s_i} is the number of neighboring states, and $s_j \in \partial_{s_i}$ denotes that state s_j is a neighbor of state s_i [56]. For the unstructured heterogeneity, an exchangeable Gaussian prior was assumed, where $\beta_{spatunstr} \sim N(0, \tau^2_{spatunstr})$ [56, 61].

General structure of the priors

To establish specific estimation of the functional properties of the model such as smoothness function in the case of nonlinear and spatial effects, we adopted a Bayesian approach where suitable prior distributions are assigned to the vectors of the regression parameters in (4). Generally, this can be realized in the form of global smoothness prior for β_l in this form;

$$p(\beta_l \mid \tau_l^2) \propto \exp\left(-\frac{1}{2\tau_l^2} \beta_l^{'} K_l \beta_l \right) \quad (8)$$

where, τ_l^2 represents the prior variance that acts as an inverse smoothing parameter and determines the tradeoff between flexibility and smoothness, while K_l is the prior precision or penalty matrix that shrinks the parameters towards zero. The penalty matrix also penalizes large abrupt jumps between neighboring coefficients of β_l. In some cases, the penalty matrix K_l may be rank insufficient. The prior distribution of the vector β_l will be partially improper, but for the empirical Bayesian inference, the prior variance τ_l^2 will be considered to be an unknown constant which may be determined using penalized likelihood estimate or restricted maximum likelihood (REML) [65].

Empirical Bayesian inference

For our analysis, we relied on the empirical Bayesian approach that treats the smoothness variance τ_l^2 as a fixed unknown parameter and to be computed from the corresponding marginal posterior [58]. Estimation was thus inherently reduced to an iteration between optimizing a penalized logistic likelihood, including variable selection [55, 58, 61, 66, 67] and numerical determination of the smoothing variances by an approximate restricted maximum likelihood estimation (REML) [55, 58, 68]. The penalized likelihood estimation of the structured additive regression coefficients was implemented in R2BayesX-in R software, a public domain software for Bayesian analysis [55, 58, 67, 69] and the computation of conditional confidence bands was based on the Markov chain Monte Carlo (MCMC) algorithm subsequent to variable selection procedure [58].

The model comparison using the goodness of fit criterion

To decide whether the spatial effects are needed, and which models fitted our data better, two different models were constructed and compared. Model I included all the selected covariates with no spatial effect and Model II accommodated all the selected covariates with both spatially structured and unstructured random effects (convolution model). The models were compared using two goodness of fit criterion, including the Akaike's Information Criterion (AIC) and the Generalized Cross-Validation (GIC) model based on the residual sum of squares, where smaller values of AIC or GIC signifies a better model [55, 61, 70, 71].

Results

Sociodemographic Information:

Table 2 summarizes the sociodemographic characteristics of all the 6065 Nigerian children aged 6-59 months included in the study. The mean age of the children was 33.0 months with a standard deviation (SD) of ±15.49. Both genders were equally presented with 3079 (50.7%) being boys and 2992 (49.3%) being girls. Most of the children were above 25 months of age, and more than

half of them were living in the rural areas at the time of the survey. In terms of mother's educational level, almost half of the mothers had no previous education. The household income distribution showed that 1244 (20.5%) of the children belonged to lowest wealth Quintiles and in overall, 4150 (68.4%) children aged 6-59 months of the 6065 children were found to be anaemic.

Result of the model comparison

Table 3 presents the results of model comparison. Evidently, based on the Akaike's Information Criterion (AIC) and the Generalized Cross-Validation (GIC) model, Model I have AIC = 2973.41 and GIC = 1.11302, while Model II have AIC = 2960.69 and GIC = 1.10913, suggesting that the combined effects of covariates with both spatially structured and unstructured random effects was the best-fitting model that has explained the risk of childhood anaemia prevalence in Nigeria better than the model without any spatial effects. Thus, subsequent results were based on the best fitted model (Model II) and we used ArcMap software in ArcGIS 10.6.1 (ESRI, Redlands, CA) to generate the maps of the posterior means of both spatially structured and unstructured random effects from Model II.

Linear effects

Linear effects

Table 4 presents the results of the fixed effects obtained using the spatial semiparametric regression model on the association between childhood anemia and possible predictors. The statistical significance of the predictors using the Model II was scrutinized using the means and the 95% credible intervals (95% Crl), which represents the range of values containing the true value with a probability of 95%. From the results, we found that the prevalence of anaemia was strongly and linearly associated with land surface temperature at 95% credible interval (OR=1.101; 95% CrI: 1.019, 0.074). The coefficients of malaria status (OR=3.577; 95% CrI: 1.047, 1.501) and

fever (OR=1.260; 95% CrI: 0.047, 0.426) were positive, which means that both malaria disease and fever were significantly and positively associated with childhood anaemia risk in Nigeria. Children who have a malaria infection were 3 times more likely to be anaemic compared to their counterpart who did not have malaria. The results also revealed that female children were less likely to be anaemic (OR=0.718; 95% CrI: -514, -150), as compared to their male counterpart. This result implies that anaemia risk decreased for a child being female.

Regarding household socioeconomic status (HSES), increased HSES status was associated with reduced risk of childhood anaemia. Children residing in the highest wealth quintile households (richest households), fourth and middle quintiles (rich and middle range), respectively, were found to be less vulnerable to anaemia more than those residing in the lowest and second quintiles (poorest and poorer) households, respectively. The odds of anaemia increased among children living in the rural areas (OR=1.424; 95% CrI: 0.102, 0.594), compared to those in the urban areas. The use of improved toilet facility was found to be a protect factor against childhood anaemia. The odds of anaemia increased among children residing in households with unimproved toilet facility (OR=1.240; 95% CrI: 0.106, 0.421), unimproved toilet facilities including no facilities, pit latrine or open defecation around a household has a significant positive effect on childhood anaemia risk.

Regarding maternal characteristics, the analysis revealed a negative association between higher maternal education status and the odds of childhood anaemia. The risk of anaemia reduced for an increased level of mother's educational attainment. Children from mothers who had completed primary (OR= 0.657; 95% CrI: -0.717, -0.122), secondary education (OR=0.628; 95% CrI: -0.754, -0.172), more than secondary education (OR=0.463; 95% CrI: -1.155, -0.382), respectively, were less likely to be anaemic than children from mothers who had no formal

education, which implies that the mother's higher level of education reduces the risk of childhood anaemia at 95% credible interval.

Nonlinear effects

The nonlinear effects are shown in Fig. 2 and Fig. 3. Fig. 2a and Fig 2b showed the posterior model estimates of the child's age and mother's age as well as the 80% and 95% pointwise credible intervals. We found that childhood anaemia outcome decreased nonlinearly with increasing age of the children (Fig. 2a), anaemia prevalence was highest among children under 15 months and decreased thereafter as the child's age increased. There was a strong nonlinear effect of mother's age on the odds of childhood anaemia as shown in (Fig. 2b). We observed that the risk of anaemia decreased as the mother's age increased from 15 years up to 24 years, and then start to increase again from age 30 to 35 years and then decreased thereafter from 36 years. This shows that lower maternal age increased the risk of childhood anaemia, however, the figure has revealed that the mothers between ages 30 to 36 years are likely to have anaemic children more than their counterpart.

Regarding the continuous environmental factors, we found that the environmental factors that were significantly and nonlinearly associated with increased risk of childhood anaemia are: the decreased distance to the nearest water bodies (Fig. 3c), the decreased distance to vegetated areas (Fig 3b) and higher population density (Fig. 3d). It was also found that increased rainfall decreased the odds of anaemia (Fig. 3e) and the effect of cluster altitude also indicated a decrease in childhood anaemia probability as altitude increased (Fig. 3a).

Spatial effects

The spatial random effects were divided into spatially structured random effect (Fig. 4A) and spatially unstructured random effect (Fig. 4B). The structured spatial effect plot (Fig. 4A) revealed

that Ebonyi, Edo, Cross river, FCT, Jigawa, Kaduna and Kebbi states in Nigeria are higher risk areas of childhood anaemia. This implies that children who reside in the above-mentioned states are at a higher risk of having anaemia than their counterpart in other states. As expected, the map of the posterior means of the unstructured spatial random effects (Fig. 4B) confirmed that childhood anaemia was randomly distributed, although, the map showed similar anaemia spread with structured spatial effects except for Adamawa state. The result has shown evidence of possible local effects (state specific) on the prevalence of childhood anaemia within Adamawa, Ebonyi, Edo, FCT, Jigawa, Ondo, Kaduna and Kano states. Therefore, more detailed studies to ascertain the possible causes of local influence on childhood anaemia prevalence, especially in Adamawa state may be required.

Discussion

The 2015 Nigeria Demographic and Health Survey (NDHS) and Malaria Indicator Survey (NMIS) programme provided extensive information on the childhood anaemia prevalence and the geographical coordinates of location clusters, which has allowed us to determine the important influencing factors and areas of greater risk. This study employed a binary structured additive regression (BSAR) model, allowing for a flexible semiparametric predictor that accounted for the effects of different types of covariates, while simultaneously incorporating the spatial random effects directly. The model outputs demonstrated that the odds of childhood anaemia predominantly exist in spatial clusters of the neighboring states, as spatial clustering is an intrinsic characteristic of an epidemiological data. It is critical to incorporate this phenomenon in a statistical analysis to avoid violation of the assumption that anaemia observations are independent [33]. Moreover, the identification of childhood anaemia clusters can provide a public health tool to explore reasons for the spatial clustering [10].

The study witnessed a strong significant association between childhood anaemia and child's infection status, such as malaria and fever. Despite malaria prevention efforts, we found that the likelihood of anemia in a child increased with positive malaria and fever status. The risk of anaemia, increased in children who were infected with malaria parasite, and those who had fever two weeks prior to the survey, the results are in accordance findings from the published works of [4, 8, 27, 28, 32, 72-74], but contrary to the findings of [33]. The possible significance of the results is to be expected because malaria infection can cause both outcomes, malaria parasite attacks and destroys the red blood cells, thereby increasing the risk of anaemia among those children infected with the malaria parasite [4, 10, 23, 24]. Fever may not be a specific indication of malaria, although the illness can lead to an indication of the presence of other hidden childhood infections that may cause anaemia, like the presence of bacteremia in a child [4, 10, 11].

We found that the odds of anaemia among the under-5 children decreased with increasing age of the children, similar to the findings of [4, 14, 35, 73]. The findings suggest that, children between 6-15 months needs high iron requirements for rapid growth, while older children can consume several diets rich in nutrient aside breastfeeding and that may have contributed to the reduced risk of anaemia problems among older children [1, 73]. Regarding the effect of gender, our analysis revealed that male children have a higher risk of anaemia than their female counterpart of the same age group, consistent with the findings of [19, 27, 28, 33, 35]. The results confirmed that both sexes have different physiological needs for iron stores, infant iron stores are lower in males than in females during the first year of life [18]. Therefore, since both sexes have different hemoglobin level thresholds, males appear more likely to have a higher growth rate resulting in greater need of iron by their body, not supplied by the diet [1, 18, 35]. Moreover, the African

practice of gendered feeding priority may have also contributed to differences in the odds of anemia between the male and female children [12, 75].

The maternal characteristics such as mother's education and age play a crucial contextual role as a protective factor against childhood anaemia, especially for children born in rural areas and poor households. It was revealed that low mother's educational attainment is linked to higher odds of the child's anaemia outcome. The children of more educated mothers were significantly less likely to be anaemic. Children of the same age group with mothers who had no formal education were more likely to be anaemic than children whose mothers had secondary or more than secondary education. Our findings are in accordance with the results obtained in Kenya and Togo [28, 73], Uganda [19, 36] and in Rwanda [32], which revealed that children from illiterate mothers were 3 times more likely to be anaemic as compared to their counterpart. It is evident from our results that education aids the capacity of a woman to grasp the knowledge needed for adequate childcare and effectively understand information concerning the nutritional needs of her children, which may enable the prevention other diseases. Mothers with higher levels of education have improved health knowledge and greater control over food choices for their children. Therefore, promoting anaemia education programs that will target illiterate younger and rural mothers may help control childhood anaemia prevalence in Nigeria. Similarly, we also found a statistically significant nonlinear association between mother age and childhood anaemia risk. Analysis revealed that the risk of childhood anaemia decreased for mothers aged between 24-30 and 36-49 years. Children whose mothers are < 24 years and between 30-36 years are of greater risk of having anaemic children, similar to the results of [8, 14, 36, 76]. The findings suggest that, women who are ≥ 36 years may have given birth to two or more children, therefore, may be more

exposed and experienced in childcare and good nutritional practices as compared to younger mothers [8].

The household socioeconomic status is a well-known significant predictor of the child's health status for many of the childhood health outcomes, including anaemia [4]. Our findings indicated that the reduction in the prevalence of anaemia varies according to the household wealth-status, the odds of anaemia reduced greatly among children living in the wealthiest quintile (higher quintile). Children who resided in the households with little household income (lower quintile families) were more susceptible to anaemia and other consequences [17, 28] and children in this age group who reside in households with the highest wealth quintile (richer) were less affected by anaemia than those living in the lowest wealth quintile [5, 20]. In Nigeria, for example, children in the most deprived socioeconomic quintile of the states have been shown to have more than 30% higher rate of mortality due to anaemia than children in the higher socioeconomic quintile [2]. These findings suggest that higher household socioeconomic status is associated with better nutrition, better education and general well-being. Thus, children residing in the highest quintile households are likely to afford adequate and diversified food requirements needed for optimum growth and development.

Furthermore, the rural type of place of residence was found to influence the risk of childhood anaemia. Children living in rural households are usually more likely to be exposed to malaria and anaemia diseases [4, 12, 27]. Similarly, children residing in households that use the unimproved toilet facility were found to be at higher risk of anaemia. For example, helminthiasis has been shown to significantly contribute to the problem of anaemia, higher prevalence of helminth infection, which causes anaemia is associated with household activities that involves access to open defecation, open latrines and the manipulation of untreated human faeces in

agricultural farms [1, 25]. Even though, it is difficult for under-5 children living in rural areas to eat food rich in iron, as such food might be very difficult to preserve in a rural setting, the inaccessibility of a good toilet facility is also a strong influencing factor to anaemia prevalence due to helminthiasis. Our findings reinforce the need to prioritize anaemia intervention programmes in rural areas, especially among the rural poor households [31]. There was non-significant associations between ethnicity and childhood anaemia prevalence, contrary to the findings of [77]. Our result suggests that, the dietary habit of specific ethnic groups of the children's families does not really matter, children are better nourished if they live in an urban environment where there is access to food varieties. Therefore, adequate iron intake, general child's living conditions and access to quality health care services are expected to significantly reduce the proportion of anemia prevalence in a population.

This study contributes towards understanding the role of environmental factors in the spatial distribution of childhood anaemia in Nigeria. We found that the risk of anaemia was nonlinearly and negatively associated with cluster altitude and rainfall, but positively associated with distance to permanent water bodies, population density and enhanced vegetation index (EVI). The analysis revealed that children who reside at a higher altitudes have lower odds of anaemia than those living at lower altitudes, consistent with the findings of [25, 28, 72]. This effect may be attributable to the malaria-altitude relationship. Regarding rainfall, we found a nonlinear negative association between rainfall and the odds of childhood anaemia. The analysis revealed that increased rainfall decreased the risk of anaemia. This result is not surprising because heavy rainfall flushes away the breeding larvae, thereby reducing the number of mosquitoes and the risk of malaria, a well-known common cause of anaemia in children [53, 80].

The analysis revealed that living in a more densely populated areas was significantly associated with greater odds of a child being anaemic, this implies that the higher the population density (expressed as people/km$^{2)}$, the greater the chance of anaemia. High human population act as a source of many infections, which could be a representative of other important determinants of anaemia, such as poor environmental condition (impoverished area) with low socioeconomic status and lack of a diverse range of iron rich foods. Moreover, a higher risk of anaemia was observed for a shorter distance to permanent waters such as rivers, lakes, etc. According to the literature, closeness of household to permanent water bodies enhances the risk of malaria, hookworms and helminth, which are some of the major causes of anaemia, this may be because the presence of permanent waters enhances the development of mosquito larvae, hookworms and helminth [10, 25, 31]. The transmission of hookworms around the water bodies could be facilitated by open defecation or open fecal disposal because the permanent waters attract open fecal disposals from fishermen and other water users, thereby influencing the transmission of hookworm and helminth causes of anaemia [16, 25]. The odds of anaemia were significantly higher for shorter vegetation areas and increased with the increased vegetation area. The EVI is an important factor that induces a suitable condition for the availability of agricultural propagation, which enhances the spread of malaria, an important anaemia risk factor [16, 28, 30]. Nevertheless, the effect of EVI appears non-monotonic, as seen in Fig. 3e, we observed a sharp decline in EVI values above 3000, increased again at values up to 4000, presumably the highest values of EVI that reflect very bushy vegetation. The very high values of EVI may be assumed to be very thick and bushy, therefore, farmers clearing vegetation to gain more productive land can often inadvertently create breeding sites for mosquitoes [78]. In addition, increased agricultural development and irrigation, such as rice farming, can increase the transmission of malaria due to the creation of large stagnant

pools of water that provide breeding habitats favorable to *Anopheles gambiae* that causes anaemia [78, 79]. Intestinal parasites and helminth infections thrive under good environmental conditions such as land surface temperature and adequate soil moisture, and we found a linear relationship between the odds of anaemia and the land surface temperature. This result is evident that ineffective treatment of such infection as well as hookworm can lead to blood loss and development of iron-deficiently anaemia in children [14, 25]. Notably, it is clear that all the selected environmental factors included in our analysis were closely tied to adequate conditions for propagation and spread of malaria, helminth infections and intestinal parasites, which are strong risk factors for the development of anemia in children [10, 16, 28, 30].

This study found that there exists a geographical heterogeneity for childhood anaemia prevalence in Nigeria, after accounting for the model covariates [4, 5]. The spatial maps (Fig. 4), demonstrated that childhood anaemia was clustered around the neighboring states and higher clusters were found in Adamawa, Ebonyi, Edo, Cross river, Jigawa, Kaduna, Kano and Kebbi states. The result confirms that the odd of childhood anaemia varies depending on the geographical area in which a child resides, children residing in the above-mentioned states were found to be at the highest risk of anaemia. These states are in areas characterized by high population density as well as agriculture practices, hence more susceptible to several infectious diseases like anaemia. However, a study in Malawi found that the spatial variation in childhood anaemia was related to variation in socioeconomic conditions within the households where a child reside [14, 25]. The observed state-level prevalence of childhood anaemia highlight the potential benefit of identifying anaemia hotspot targets for effective public health interventions. Some areas in Nigeria may have had a higher burden of anaemia due to high prevalence of multiple risk factors which are unobserved. This variability may also be significantly associated with underlying differences in

state-level socioeconomic status, knowledge about anaemia and environmental conditions. The generated maps of both structured and unstructured spatial distribution of childhood anaemia which showed similar pattern could easily support policy makers about the states of needed concern for household-level intervention targeting, state-level or beyond state-level intervention targeting. However, we recommend that the identified higher risk states should be prioritized to obtain financial support, economic and infrastructure development, in-house or community health care workforce training and health education on childhood anaemia.

Few limitations need to be considered when interpreting the results of this study. First, the study utilized a cross-sectional data from the most recent NMIS survey, the nature of the survey did not allow any causal relationship to be established. Second, apart from the factors that were found to be associated with childhood anaemia in this study, there are a variety of other factors that may increase the risk of anaemia in children, but they were not included in our analysis due to lack of information on them. Such variables include; sickle cell disease and α-thalassemia, infectious diseases such as HIV/AIDS status, soil-transmitted helminths caused by hookworm, schistosomiasis, bacteremia and many more. Similarly, information on iron intake is also a major limiting factor in this research as we were unable to obtain data on receipt of iron supplement and other nutritional indices. However, the strength of this study lies in the ability to generalize the findings to the whole country having used nationally representative survey data that is internationally recognized, and being able to incorporate some selected environmental factors [2, 50]. Despite the limitations of this study, the research adds evidence to the literature concerning the spatial analysis and the identification of important risk factors of childhood anaemia using a flexible semiparametric predictor that accounted for the effects of different types covariates, while also accounting for the spatial structure of the data. We believe that this study could be more useful

if other missing variables were considered. Therefore, future studies should attempt to include other important factors not included in this study, which may bias results.

Conclusion

Due to the increasing availability of complex, geographically linked survey data, flexile semiparametric model considered in this study is useful and therefore, essential in empirical research. The utilized model allowed us to account for the effects of different types of covariates in determining the role of environmental, health, socioeconomic and demographic factors on childhood anaemia prevalence, and to quantify the geographical heterogeneity in childhood anaemia risk unaccounted by these measured factors. Findings suggest that the geographical variation may be reflection of variations in socioeconomic status, malaria endemicity and environmental conditions. The results have generated important anaemia epidemiological resource, highlighting national priority areas for childhood anaemia morbidity control in Nigeria, and provided public health authorities in the country with useful geographical information for developing effective prevention and intervention strategies. Additionally, the study contributes towards efforts to understand the role of environmental factors in determining the spatial distribution of anaemia and to establish environmental risk assessment of childhood anaemia in Nigeria.

Abbreviation

AIC: Akaike's information criterion; CAR: Conditional Autoregressive Model; BIC: Bayesian Information Criterion; EA: Enumeration Area; EVI: Enhanced Vegetation Index; FCT: Federal Capital Territory; GCV: Generalized Cross Validation; MRF: Markov Random Field; NMIS: Nigeria Malaria Indicator Survey; NDHS: Nigeria Demographic and Health Survey; OR: Odds Ratio; PSUs: Primary Sampling Unit; RW: Random walk; REML: Restricted Maximum Likelihood; SES: Socioeconomic Status; WHO: World Health Organization

Acknowledgements

The authors thank the permission granted by MEASURE DHS to use the NDHS and NMIS datasets in this study. The first author appreciates the study leave and the opportunity granted by the University of Nigeria Nsukka, Nigeria. The first author also thanks the Department of Biostatistics, University of Washington, United States, for the training and the opportunity given to her during the 10th Annual Summer Institute in Statistics and Modeling in Infectious Diseases (SISMID).

Author's contributions

CLJU obtained permission to use the 2015 NMIS and NDHS geospatial data sets. CLJU and TZ conceptualized the modeling idea and CLJU performed the analysis. Both CLJU and TZ jointly drafted and revised the manuscript. The manuscript is part of CLJU's PhD work. All authors read and approved the final version of the manuscript before submission.

Funding

No funding

Availability of data and materials

The analyzed dataset is freely available upon request from the Measure Demographic Health Survey (DHS) websites: www.dhsprogram.com/data/dataset/Nigeria. These covariate datasets are available for download in two places: the DHS Program website and the Spatial data repository www.spatialdata.dhsprogram.com/covariates/.

Ethics approval and consent to participate

The 2015 Nigeria Malaria Indicator survey (NMIS)-protocols were ethically cleared by the Nigeria Health Research Ethics Committee of the Federal Ministry of Health (NHREC) and the Internal Review Board of the ICF International in Calverton (USA). The study was based on a publicly available data obtained upon request through MEASURE DHS http://www.measuredhs.com and the consent to participate was not applicable, hence, an informed consent was provided by all the surveyed participants through their caregiver or parents prior to malaria test and the administration of questionnaires.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

1. WHO, The global prevalence of anaemia in 2011. World Health Organization, 2015. https://www.who.int/nutrition/publications/ micronutrients/global_prevalence -anaemia_2011/en/Accessed 02 Jan 2020.
2. Programme NMC. Nigeria MIS final report. Abuja: Federal Republic of Nigeria; 2015. NMIS, https://dhsprogram.com/pubs/pdf/MIS20/MIS20.pdf. Accessed 02 Nov 2019.

3. Fleming AF, Werblinska B. Anaemia in childhood in the guinea savanna of Nigeria. Ann Trop Paediatr. 1982; **2**(4):161-73. https://doi.org/10.1080/02724936.1982.11748250.
4. Adebayo SB, Gayawan E, Heumann C, Seiler C. Joint modeling of Anaemia and Malaria in children under five in Nigeria. Spat Spatiotemporal Epidemiol. 2016; **17:105**-115. https://doi.org/10.1016/j.sste.2016.04.011.
5. Gayawan E, Arogundade ED, Adebayo SB. Possible determinants and spatial patterns of anaemia among young children in Nigeria: a Bayesian semi-parametric modelling. Int Health. 2014; **6**(1):35-45. https://doi.org/10.1093/inthealth/iht034
6. P Phiri KS, Calis JCJ, Faragher B, Nkhoma E, Ng'oma K, Mangochi B, Molyneux ME, van Hensbroek MB. Long term outcome of severe anaemia in Malawian children. PloS one. 2008; **3**(8): e2903. doi: 10.1371/journal.pone.0002903.
7. Bobonis GJ, Miguel E, Puri-Sharma C. Anemia and school participation. J Hum Resour. 2006; **41**(4): 692-721.
8. Khan JR, Awan N, Misu F. Determinants of anemia among 6-59 months aged children in Bangladesh: evidence from nationally representative data. BMC pediatrics. 2016;**16**(1):3. DOI 10.1186/s12887-015-0536-z.
9. Kuziga F, Adoke Y, Wanyenze RK. Prevalence and factors associated with anaemia among children aged 6 to 59 months in Namutumba district, Uganda: a cross-sectional study. BMC pediatrics. 2017; **17**(1): 25. DOI 10.1186/s12887-017-0782-3.
10. Magalhaes RJS, Clements ACA. Mapping the risk of anaemia in preschool-age children: the contribution of malnutrition, malaria, and helminth infections in West Africa. PLoS medicine. 2011; **8**(6): e1000438. doi: 10.1371/journal.pmed.1000438.
11. Konstantyner T, Roma O, Thais C, de Aguiar Carrazedo Taddei, JA. Risk factors for anemia among Brazilian infants from the 2006 National Demographic Health Survey. Anemia. 2012; **2012**. https://doi.org/10.1155/2012/850681.
12. Halliday KE, Karanja P, Turner EL, Okello G, Njagi K, Dubeck MM, Allen E, Jukes MCH, Brooker SJ. Plasmodium falciparum, anaemia and cognitive and educational performance among school children in an area of moderate malaria transmission: baseline results of a cluster randomized trial on the coast of Kenya. Trop Med Int Health. 2012; **17**(5): 532-49. https://doi.org/10.1111/j.1365-3156.2012.02971.x
13. Gayawan E, Arogundade ED, Adebayo SB. A Bayesian multinomial modeling of spatial pattern of co-morbidity of malaria and non-malarial febrile illness among young children in Nigeria. Trans R Soc Trop Med Hyg. 2014; **108**(7): 415-24. https://doi.org/10.1093/trstmh/tru068
14. Ngwira A, Kazembe LN. Bayesian random effects modelling with application to childhood anaemia in Malawi. BMC public health. 2015; **15**(1): 161. DOI 10.1186/s12889-015-1494-y.
15. Osazuwa F, Ayo OM. Contribution of malnutrition and malaria to anemia in children in rural communities of Edo state, Nigeria. North Am J Med Sci. 2010; **2**(11): 532.
16 Foote EM, Sullivan KM, Ruth LJ, Oremo J, Sadumah I, Williams TN, Suchdev PS. Determinants of anemia among preschool children in rural, western Kenya. Am J Trop Med Hyg. 2013; **88**(4): 757- 64. DOI: https://doi.org/10.4269/ajtmh.12-0560
17 Ngnie-Teta I, Receveur O, Kuate-Defo B. Risk factors for moderate to severe anemia among children in Benin and Mali: insights from a multilevel analysis. Food Nutrition Bull. 2007; **28**(1): 76-89.
18. W.H.O., Nutritional anaemias: tools for effective prevention and control. 2017.

19 Legason, I.D., et al., Prevalence of anaemia and associated risk factors among children in North-western Uganda: a cross sectional study. BMC hematology, 2017. **17**(1): p. 10.
20. Balarajan, Y., et al., Anaemia in low-income and middle-income countries. The Lancet, 2011. **378**(9809): p. 2123--2135.
21. Zuffo, C.R.K., et al., Prevalence and risk factors of anemia in children. Jornal de Pediatria, 2016. **92**(4): p. 353--360.
22. W.H.O., World malaria report 2018. World Health Organization Geneva, Switzerland. 2018: WHO https://www.who.int/malaria/publications/world-malaria-report-2018/en/.
23. Oladeinde, B.H., et al., Malaria and anemia among children in a low resource setting in Nigeria. Iranian journal of parasitology, 2012. **7**(3): p. 31.
24. ONG'ECHA, J.M., et al., Parasitemia, anemia, and malarial anemia in infants and young children in a rural holoendemic Plasmodium falciparum transmission area. The American journal of tropical medicine and hygiene, 2006. **74**(3): p. 376--385. https://doi.org/10.1177/156482650702800109
25. Ngwira A, Kazembe LN. Analysis of severity of childhood anemia in Malawi: a Bayesian ordered categories model. Open Access Med Stat. 2016; **6**(9):20.
26. Zhao A, Zhang Y, Peng Y, Li J, Yang T, Liu Z, Lv Y, Wang P. Prevalence of anemia and its risk factors among children 6--36 months old in Burma. Am J Trop Med Hyg. 2012; **87**(2): 306-11. DOI: https://doi.org/10.4269/ajtmh.2012.11-0660.
27. Ronald LA, Kenny SL, Klinkenberg E, Akoto AO, Boakye I, Barnish G, Donnelly MJ. Malaria and anaemia among children in two communities of Kumasi, Ghana: a cross-sectional survey. Malaria J. 2006; **5**(1): 105. Doi:10.1186/1475-2875-5-105
28. Roberts DJ, Zewotir T. District Effect Appraisal in East Sub-Saharan Africa: Combating Childhood Anaemia. Anemia; 2019: 2019.
29. Oladeinde BH, Omoregie R, Odia I, Oladeinde OB. Prevalence of malaria and anemia among pregnant women attending a traditional birth home in Benin City, Nigeria. Oman med J. 2012; **27**(3): 232. DOI 10. 5001/omj.2012.52.
30. Menon MP, Yoon SS. Prevalence and factors associated with anemia among children under 5 years of age—Uganda, 2009. Am J Trop Med Hyg. 2015; **93**(3): 521-526. DOI: https://doi.org/10.4269/ajtmh.15-0102.
31. Brooker S, Peshu N, Warn PA, Mosobo M, Guyatt HL, Marsh K, Snow RW. The epidemiology of hookworm infection and its contribution to anaemia among pre-school children on the Kenyan coast. Trans R Soc Trop Med Hyg. 1999; **93**(3): 240-246.
32. Habyarimana F, Zewotir T, Ramroop S. Structured additive quantile regression for assessing the determinants of childhood anemia in Rwanda. Int J Environ Res Public health. 2017; **14**(6): 652.
33. Ejigu BA, Wencheko E, Berhane K. Spatial pattern and determinants of anaemia in Ethiopia. PloS one. 2018; **13**(5): e0197171.
34. Crawley J. Reducing the burden of anemia in infants and young children in malaria-endemic countries of Africa: from evidence to action. Am J Trop Med Hyg. 2004. **71**(2): 25-34. DOI: https://doi.org/10.4269/ajtmh.2004.71.25
35. Ngesa O, Mwambi H. Prevalence and risk factors of anaemia among children aged between 6 months and 14 years in Kenya. PLoS One. 2014; **9**(11): e113756.
36. Kikafunda JK, Lukwago FB, Turyashemererwa F. Anaemia and associated factors among under-fives and their mothers in Bushenyi district, Western Uganda. Public health nutrition. 2009; **12**(12): 2302-08.

37. Muoneke VU, Ibekwe RC, Nebe-Agumadu HU, Ibe BC. Factors associated with mortality in under-five children with severe anemia in Ebonyi, Nigeria. Indian pediatrics. 2012 **49**(2): 119-23. https://doi.org/10.1007/s13312-012-0026-4.
38. Bamidele JO, Abodunrin OL, Olajide FO, Oke YF. Prevalence and determinants of anemia among primary school pupils of a peri-urban community in Osun State, Nigeria. Int J Adolesc Med health. 2010; **22**(4): 461-68.
39. Akinbo FO, Omoregie R, Mordi R, Okaka CE. Prevalence of malaria and Anemia among young children in a tertiary Hospital in Benin City, Edo state, Nigeria. FJHS. 2009; **1**(2): 81-84. https://doi.org/10.1016/S1877-8607(10)60004-5.
40. Olubukola A, Odunayo A, Adesina A. Anemia in pregnancy at two levels of health care in Ibadan, south west Nigeria. Ann Afr Med. 2011; **10**(4).
41. Dim CC, Onah HE. The prevalence of anemia among pregnant women at booking in Enugu, South Eastern Nigeria. Med Gen Med. 2007; **9**(3):11.
42. Agan TU, Ekabua JE, Udoh AE, Ekanem EI, Efiok EE, Mgbekem MA. Prevalence of anemia in women with asymptomatic malaria parasitemia at first antenatal care visit at the University of Calabar Teaching Hospital, Calabar, Nigeria. Int J Women's Health. 2010; **2**:229.
43. Hassan A, Mamman AI, Adaji S, Musa B, Kene S. Anemia and iron deficiency in pregnant women in Zaria, Nigeria. Sub-Saharan Afr J Med. 2014; **1**(1): 36.
44. Onoh RC, Ezeonu OP, Joannes UOU, Onoh TJP, Anozie BO, Onyebuchi AK. Anemia in pregnancy in Abakaliki, South East Nigeria. Int J Gynecol Obstet. 2014; **125**(3): 280-280. http://dx.doi.org/10.1016/j.ijgo.2014.01.004.
45. Anorlu RI, Oluwole AA, Abudu OO. Sociodemographic factors in anaemia in pregnancy at booking in Lagos, Nigeria. J Obstet Gynaecol. 2006; **26**(8): 773-76.
46. Buseri FI, Uko EK, Jeremiah ZA, Usanga EA. Prevalence and risk factors of anaemia among pregnant women in Nigeria. Open J Hematol. 2008; **2**: 1.
47. Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics. J R Stat Soc C-APPL. 1998. **47**(3): p. 299--350.
48. Elliot P, Wakefield JC, Best NG, Briggs DJ. Spatial epidemiology: methods and applications. Vol. 15. London: Oxford University Press Oxford; 2000.
49. Fahrmeir L, Lang S. Bayesian semiparametric regression analysis of multicategorical time-space data. Ann I Stat Math. 2001; **53**(1): 11-30.
50. NDHS. Demographic and health survey 2016: key indicators report. The DHS Program ICF; 2016. https://dhsprogram.com/what-we-do/survey/survey-display-528.cfm. Accessed 09 Nov 2019.
51. Warren JL, Perez-Heydrich C, Burgert CR, Emch ME. Influence of Demographic and Health Survey Point Displacements on Distance-Based Analyses. Spat demogr 2016. **4**(2): 155-73. https://doi.org/10.1007/s40980-015-0014-0.
52. Grillet ME. Factors associated with distribution of Anopheles aquasalis and Anopheles oswaldoi (Diptera: Culicidae) in a malarious area, northeastern Venezuela. J Med Entomol. 2000; **37**(2): 231-38. https://doi.org/10.1603/0022-2585-37.2.231.
53. Nkurunziza H, Gebhardt A, Pilz J. Bayesian modelling of the effect of climate on malaria in Burundi. Malaria J. 2010; **9**(1): 114. https://doi.org/10.1186/1475-2875-9-114.
54. Mayala B, and Fish TD, Eitelberg D, Dontamsetti T. The DHS Program Geospatial Covariate Datasets Manual. 2018, ockville, Maryland, USA: ICF.

55. Fahrmeir L, Kneib T, Lang S. Penalized structured additive regression for space-time data: a Bayesian perspective. Stat Sinica. 2004: 731-61.
56. Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. Ann I Stat Math. 1991; **43**(1): 1-20.
57. Kammann EE, Wand MP. Geoadditive models. J R Stat Soc: Series C (Applied Statistics). 2003; **52**(1): 1-18. https://doi.org/10.1111/1467-9876.00385.
58. Umlauf N, Adler D, Kneib T, Lang S, Zeileis A. Structured Additive RegressionModels: An r Interface to Bayesx. Journal of Statistical Software. 2015; 63:1-46.
59. Lang S, Brezger A. Bayesian P-splines. J Comput Graph Stat. 2004; **13**(1):183-212.
60. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. Stat Science. 1996; **11**(2): 89-102. https://www.jstor.org/stable/2246049.
61. Fahrmeir L, Lang S. Bayesian inference for generalized additive mixed models based on Markov random field priors. J R Stat Soc: Series C (Applied Statistics). 2001; **50**(2): 201-20. https://doi.org/10.1111/1467-9876.00229.
62. Besag J, Kooperberg C. On conditional and intrinsic autoregressions. Biometrika. 1995. **82**(4): 733-46. https://doi.org/10.1093/biomet/82.4.733.
63. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc: Series b (statistical methodology). 2009; **71**(2): 319-92.
64. Freni-Sterrantino A, Ventrucci M, Rue H. A note on intrinsic conditional autoregressive models for disconnected graphs. Spat spatio-temporal epidemiol. 2018; **26**: 25-34. https://doi.org/10.1016/j.sste.2018.04.002.
65. Klein N, Kneib T, Lang S, Sohn Alexander. Bayesian structured additive distributional regression with an application to regional income inequality in Germany. Ann Appl Stat. 2015; 9(2): 1024-52.
66. Zewotir T, Galpin JS. Influence diagnostics for linear mixed models. Journal Data Sci. 2005; **3**(2): 153-77.
67. Marx BD, Eilers PHC. Direct generalized additive modeling with penalized likelihood. Comput Stat Data An. 1998; **28**(2): 193-209.
68. Zewotir T, Galpin GS. Evaluation of linear mixed model case deletion diagnostic tools by Monte Carlo simulation. Commun Stat-Simul C. 2006; **35**(3): 645-82.
69. Brezger A, Kneib T, Lang S. Bayesx manuals. Ludwig-Maximilians-University, Munich, URL http://www. stat. uni-muenchen. de/\~{} bayesx}, 2005.
70. Belitz C, Brezger A, Kneib T, Lang S. BayesX-Software for Bayesian inference in structured additive regression models. URL http://www. stat. uni-muenchen. de/\~{} bayesx; 2009.
71. Belitz C, Lang S. Simultaneous selection of variables and smoothing parameters in structured additive regression models. Comput Stat Data An. 2008; **53**(1): 61-81. https://doi.org/10.1016/j.csda.2008.05.032
72. Sumbele IUN, Sama SO, Kimbi HK, Taiwe GS. Malaria, moderate to severe anaemia, and malarial anaemia in children at presentation to hospital in the Mount Cameroon area: a cross-sectional study. Anemia. 2016: **2016**.
73. Nambiema A, Robert A, Yaya I. Prevalence and risk factors of anemia in children aged from 6 to 59 months in Togo: analysis from Togo demographic and health survey data, 2013--2014. BMC Public Health. 2019. **19**(5): 215.
74. Simbauranga RH Kamugisha E, Hokororo A, Kidenya BR, Makani J.Prevalence and

factors associated with severe anaemia amongst under-five children hospitalized at Bugando Medical Centre, Mwanza, Tanzania. BMC hematology, 2015. **15**(1): p. 13.
75. Shell-Duncan B, McDade T. Cultural and environmental barriers to adequate iron intake among northern Kenyan schoolchildren. Food Nutrition bull. 2005; **26**(1): 39-48.
76. Parbey PA, Tarkang E, Manu E, Amu H, Ayanore MA, Aku FY, Ziema SA, Bosoka SA, Adjuik M, Kweku M. Risk Factors of Anaemia among Children under Five Years in the Hohoe Municipality, Ghana: A Case Control Study. Anemia, 2019: **2019**.
77. Huang Z, Jiang F, Li J, Jiang D, Xiao T, Zeng J. Prevalence and risk factors of anemia among children aged 6--23 months in Huaihua, Hunan Province. BMC public health, 2018. **18**(1): p. 1267. https://doi.org/10.1186/s12889-018-6207-x.
78. Walsh JF, Molyneux DH, Birley MH. Deforestation: effects on vector-borne disease. Parasitology; 1993; **106**(S1): S55-S75.
79. Ricotta EE, Frese SA, Choobwe C, Louis TA, Shiff CJ. Evaluating local vegetation cover as a risk factor for malaria transmission: a new analytical approach using ImageJ. Malaria J. 2014; **13**(1): 94. doi:10.1186/1475-2875-13-94.
80. Shimaponda-Mataa NM, Tembo-Mwase E, Gebreslasie M, Achia TNO, Mukaratirwa S. Modelling the influence of temperature and rainfall on malaria incidence in four endemic provinces of Zambia using semiparametric Poisson regression. Acta Tropica. 2017; **166**: 81-91. https://doi.org/10.1016/j.actatropica.2016.11.007.

Figure legends:

Figure 1. Survey data location. Childhood anaemia prevalence data: Locations where the survey dataset was collected based on the 2015 NMIS-DHS. The figure shows 37 states including FCT used as the geographical clusters in the analysis.

Figure 2. Nonlinear effects of child's age (a) and mother's age (b) on the log-odds of childhood anaemia with pointwise 80% and 95% credible intervals.

Figure 3. Nonlinear effects of cluster altitude (a), enhanced vegetation index (b), proximity to waters (c), population density (d) and Rainfall (e) on childhood malaria prevalence with pointwise 80% and 95% credible intervals.

Figure 4. Estimated posterior means of spatial random effect of childhood anaemia in Nigeria using Model II. (A) Spatially structured random effects, together with (B) spatially unstructured random effects. The figures were generated from the results of our analysis using shapefiles freely

obtained from the DHS Spatial Data Repository(https://spatialdata.dhsprogram.com/boundaries) and imported into ArcGIS software.

Tables:

Table 1. Summary of the environmental variables, sources of data and definition of variables

Covariate	Data sources	Definition
Enhanced Vegetation Index (EVI)	NDHS Spatial Analysis data	The average enhanced vegetation index obtained by measuring the density of green levels in the near-infrared and visible bands.
Proximity to Waters (Coast/Large Lakes)	NDHS Spatial Analysis data	The geodesic distance to either a lake or the coastline.
Rainfall	NDHS Spatial Analysis data	Annual mean rainfall (mm).
Population Density	NDHS Spatial Analysis data	Number of People/Km2.
Land surface temperature	NDHS Spatial Analysis data	Annual land environmental air temperature in Degrees Celsius.
Cluster altitude	NDHS Spatial Analysis data	Measure of surface altitude (m)

Table 2. Sociodemographic information of all the Nigerian children included in the study, based on the 2015 NMIS data (n=6065).

Characteristics		**Frequency (Percentage)**
Age of the children		
	6-12 months	721 (11.9%)
	13-24 months	1283 (21.1%)
	25-36 months	1309 (21.6%)
	37-48 months	1417 (23.3%)
	49-59 months	1341 (22.1%)
Sex of children	Male	3079 (50.7%)
	Female	2992 (49.3%)
Place of residence	Urban	2037 (33.6%)
	Rural	4033 (66.4%)
Household wealth quintiles	Lowest	1244 (20.5%)
	Second	1407 (23.2%)

	Middle	1175 (19.4%)
	Fourth	1115 (18.4%)
	Highest	1129 (18.6%)
Mother's education	No education	2423 (39.9%)
	Primary	948 (15.6%)
	Secondary	1568 (25.8%)
	More than secondary	412 (6.8%)

Table 3. Summary of model fit criteria for models with no spatial effect (Model I) and model spatial effects (Model II).

	Fit criterion			
Models	$-2\times$ ***log likelihood***	**Df**	**AIC**	**GCV**
Model I	-1440.855	45.8505	2973.41	1.11302
Model II	**-1419.262**	**61.0794**	**2960.69**	**1.10913**

Table 4.Regression coefficients, odds ratios and 95% credible intervals from the binary structured additive regression model for childhood anaemia in Nigeria based on the 2015 NMIS dataset.

Variables	**Posterior mean (95% Crl)**	**Odds ratio (OR)**
Intercept	0.055 (-0.562, 0.630)	1.056
Child has malaria (Ref=NO)	**1.000**	**1.000**
Yes	1.275 (1.047, 1.501)	3.577
Child has fever (Ref=No)	**1.000**	**1.000**
Yes	0.231 (0.047, 0.426)	1.260
Sex of the child (Ref=Male)	**1.000**	**1.000**
Female	-0.332 (-0.514, -0.150)	0.718
Residence (Ref=Urban)	**1.000**	**1.000**
Rural	0.353 (0.102, 0.594)	1.424

Mothers Educ. (Ref=Non)	**1.000**	**1.000**
Primary	-0.421 (-0.717, -0.122)	0.657
Secondary	-0.466 (-0.754, -0.172)	0.628
More than secondary	-0.769 (-1.155, -0.382)	0.463
Wealth quintile (Ref=Lowest)	**1.000**	**1.000**
Second	-0.279 (-0.554, 0.002)	0.757
Middle	-0.375 (-0.547, -0.205)	0.687
Fourth	-0.767 (-0.949, -0.585)	0.465
Highest	-1.394 (-1.612, -1.192)	0.248
Toilet (Ref=Improved)	**1.000**	**1.000**
Unimproved	0.215 (0.106, 0.421)	1.240
Ethnicity (Ref=Hausa/Fulani)	**1.000**	**1.000**
Yoruba	0.133 (-0.251, 0.518)	1.142
Igbo	-0.092 (-0.462, 0.276)	0.912
Others	-0.062 (-0.255, 0.131)	0.940
Land surface temperature	0.096 (0.019, 0.074)	1.101

Figures

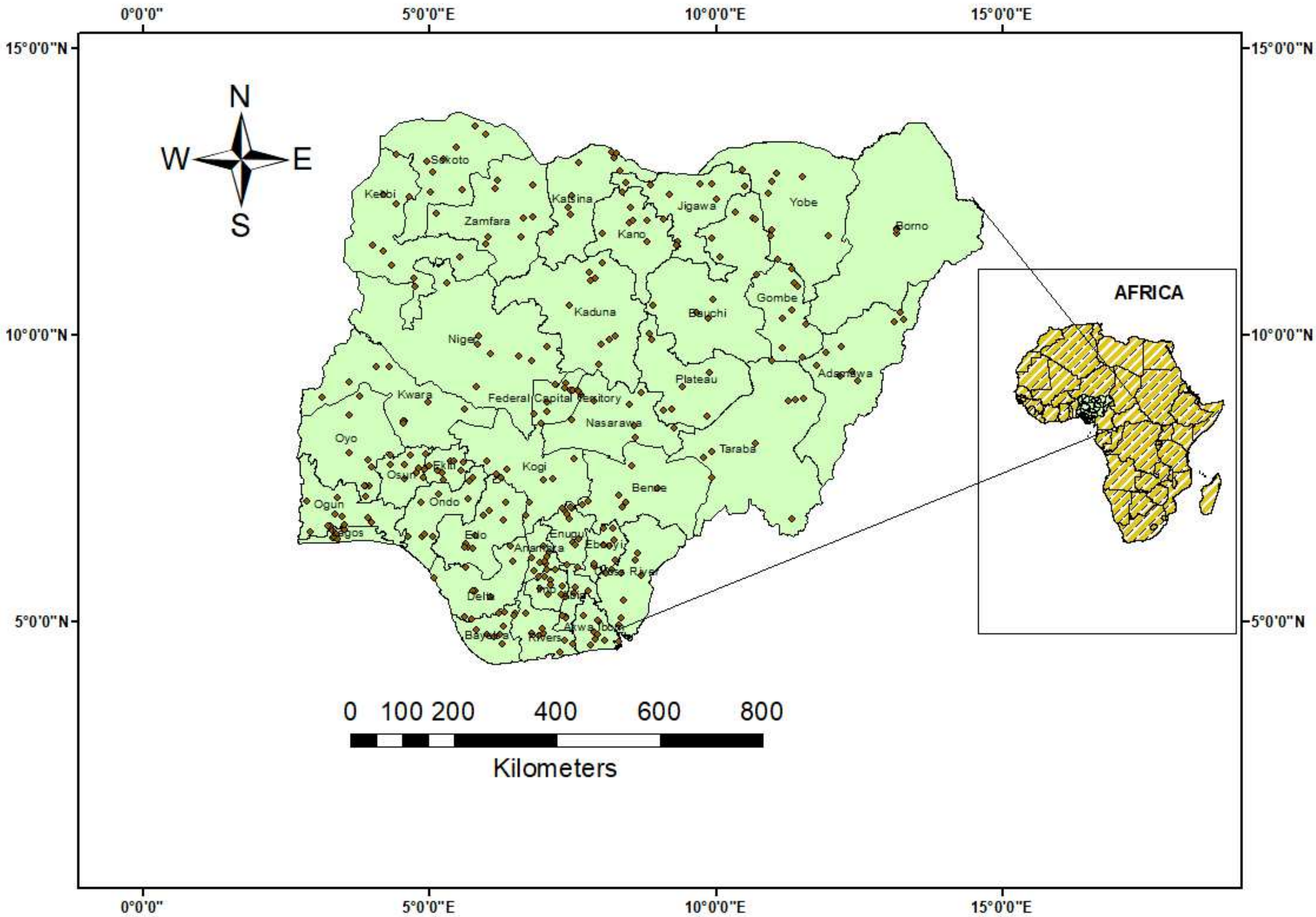

Figure 1

Survey data location. Childhood anaemia prevalence data: Locations where the survey dataset was collected based on the 2015 NMIS-DHS. The figure shows 37 states including FCT used as the geographical clusters in the analysis.

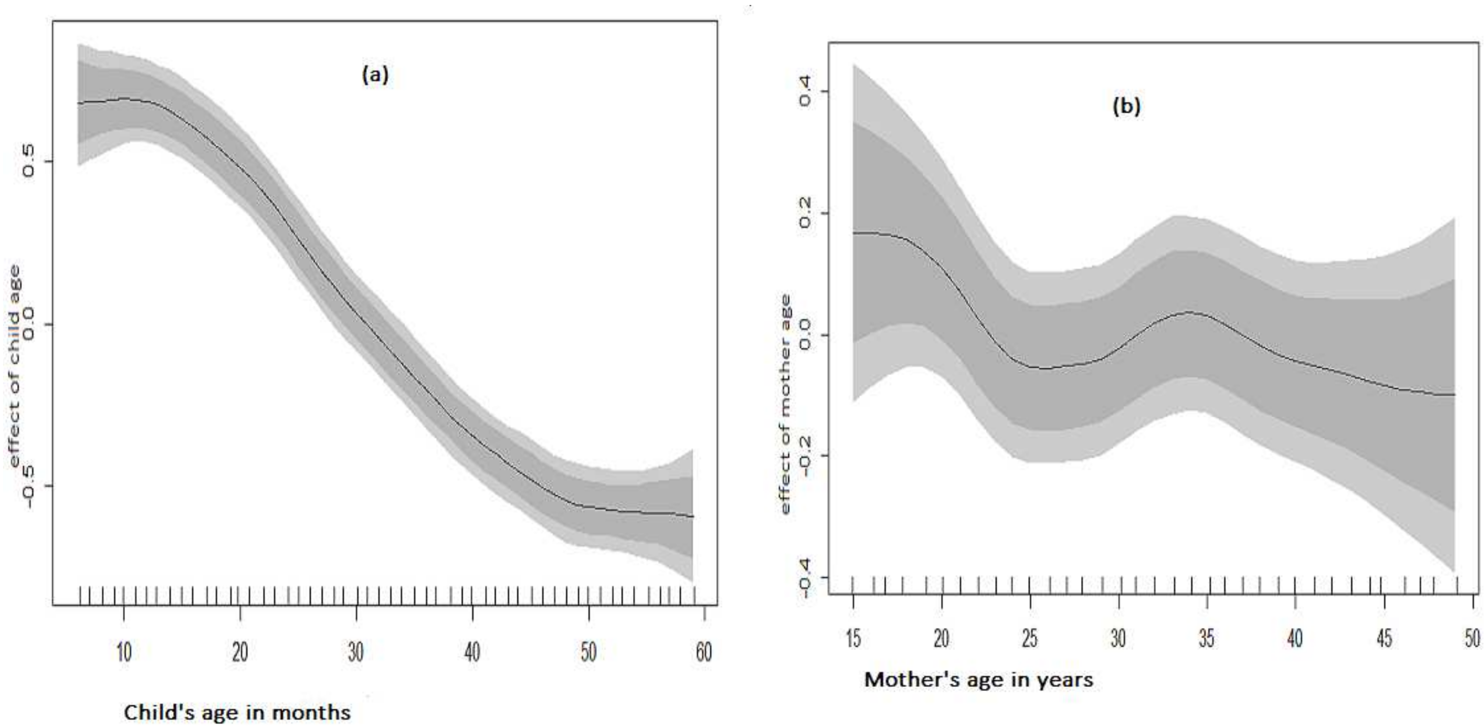

Figure 2

Nonlinear effects of child’s age (a) and mother’s age (b) on the log-odds of childhood anaemia with pointwise 80% and 95% credible intervals.

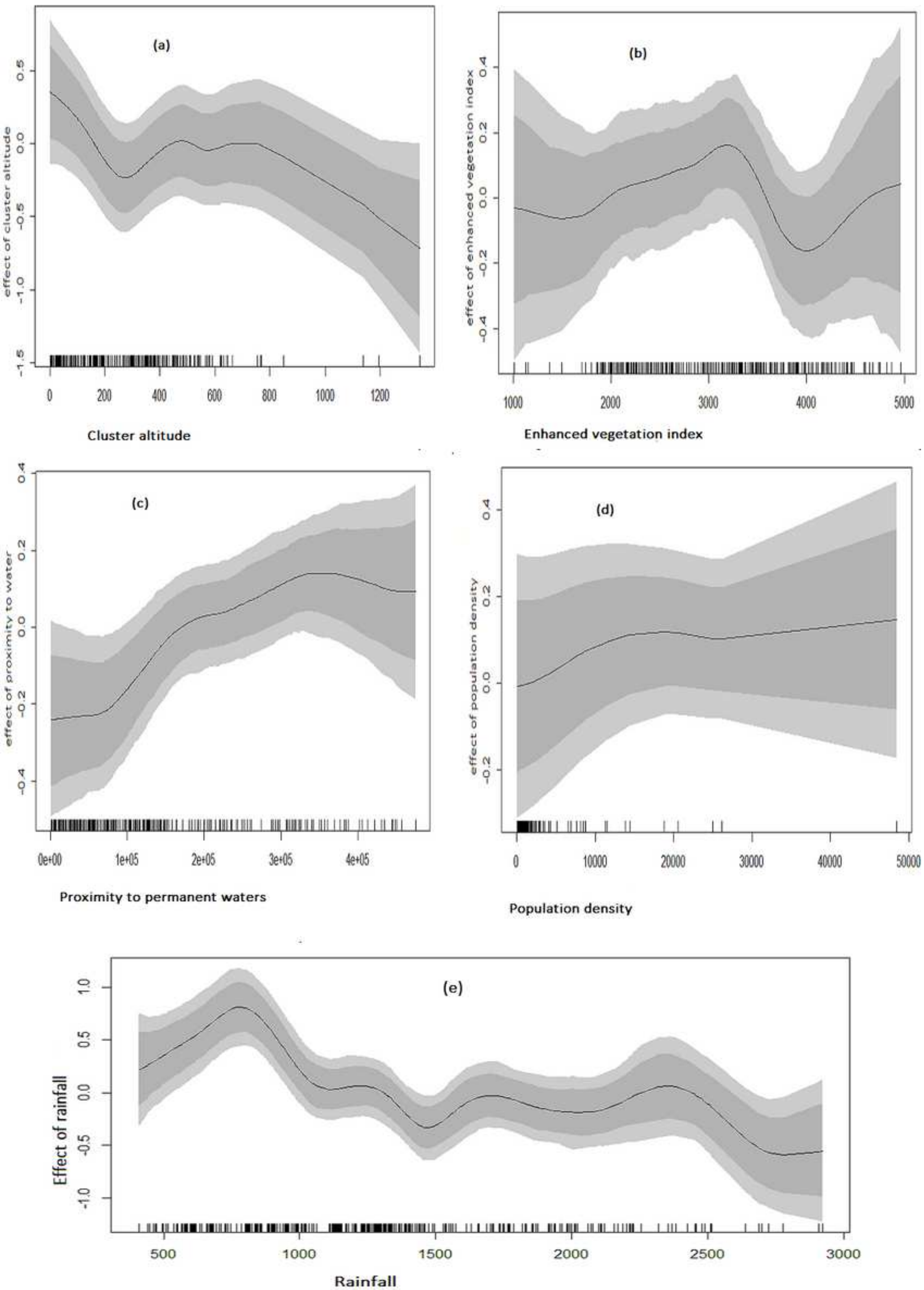

Figure 3

Nonlinear effects of cluster altitude (a), enhanced vegetation index (b), proximity to waters (c), population density (d) and Rainfall (e) on childhood malaria prevalence with pointwise 80% and 95% credible intervals.

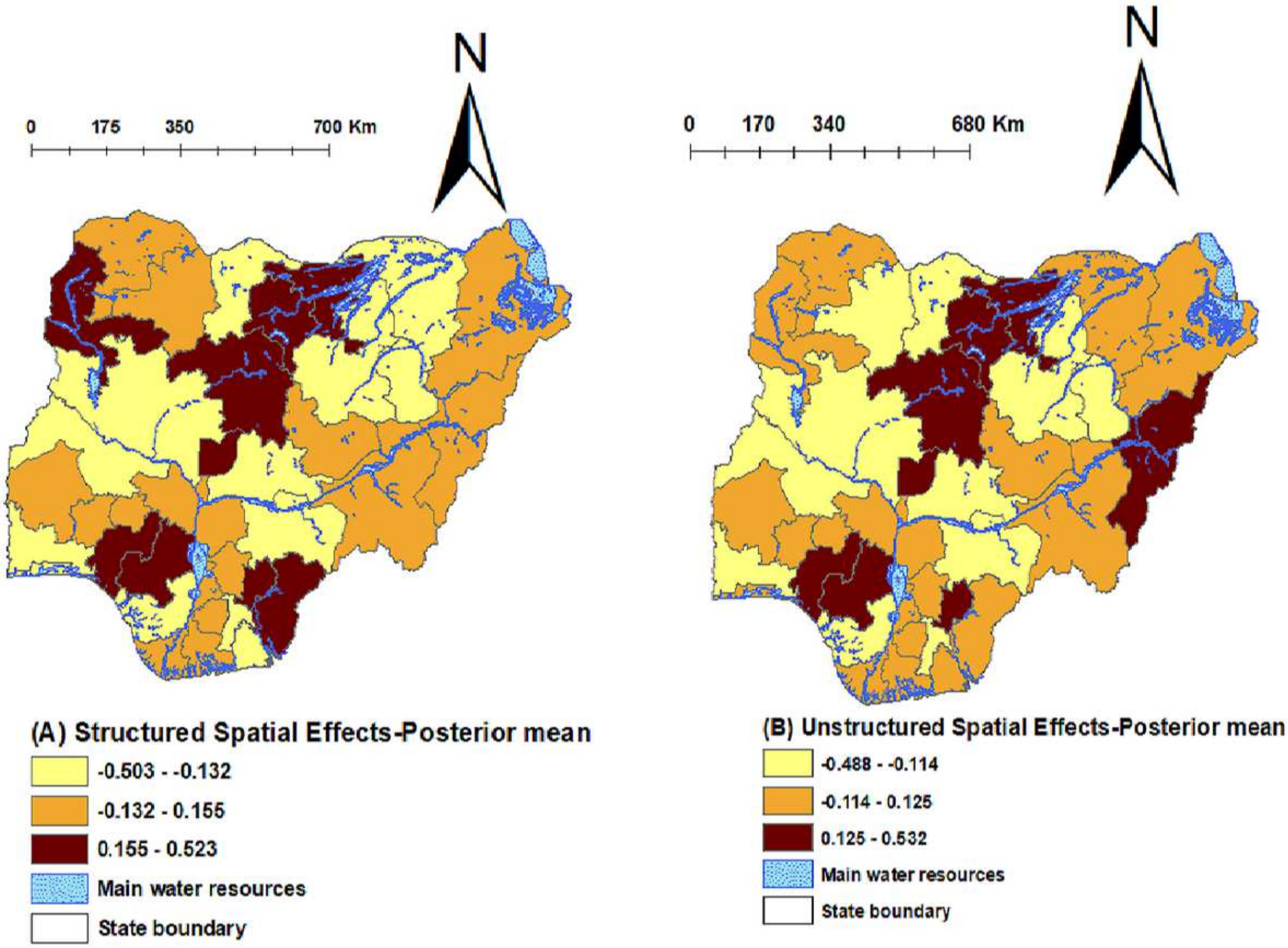

Figure 4

Estimated posterior means of spatial random effect of childhood anaemia in Nigeria using Model II. (A) Spatially structured random effects, together with (B) spatially unstructured random effects. The figures were generated from the results of our analysis using shapefiles freely obtained from the DHS Spatial Data Repository(https://spatialdata.dhsprogram.com/boundaries) and imported into ArcGIS software.