



**Bayesian spatio-temporal and joint modelling of malaria and
anaemia among Nigerian children aged under five years,
including estimation of the effects of risk factors**

by

Jecinta Ugochukwu Ibeji

A thesis submitted in partial fulfillment for the

degree of Doctor of Philosophy in Statistics

in the

College of Agriculture, Engineering and Science

School of Mathematics, Statistics and Computer Science

October 2023

Dedication

To the Trinity

My dearest husband Dr. Collins U. Ibeji and my Jesus babies.

Declaration

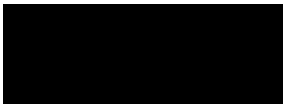
This study was done under the supervisor the supervision of Prof H. Mwambi at the School of Mathematics, Statistics and Computer Science, University of Kwa-Zulu Natal, Pietermaritzburg campus and Dr Abdul-Karim Iddrisu at the School of Science, Mathematica and Statistics, University of Energy and Natural Resources, Ghana. The work represents an original work by the author, and it has not been submitted in any form for any degree or any other qualification at this University or any other institution. Due acknowledgement is given where works of others have been quoted.



Author: Jecinta U. Ibeji

15 – 10 - 2023

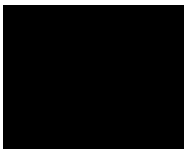
Date



Supervisor: Prof. Henry Mwambi

15-10-2023

Date



Co-supervisor: Dr. Abdul-Karim Iddrisu

15-10-2023

Date

List of Journals Publications

1. Jecinta Ibeji, Henry Mwambi and Abdul-Karim Iddrisu. “Spatial variation and risk factors of malaria and anaemia among children aged 0 to 59 months: A cross-sectional study of 2010 and 2015 datasets”. *Scientific reports* (2022): 12:11498-11513 (**Published**)
2. Jecinta Ibeji, Henry Mwambi and Abdul-Karim Iddrisu. “Bayesian spatio-temporal modeling and mapping of malaria and anaemia among children between 0 to 59 months in Nigeria”. *Malaria Journal* (2022) 21:311 (**Published**)
3. Jecinta Ibeji, Henry Mwambi, Abdul-Karim Iddrisu and Elphas Okango. “Spatial joint modelling of malaria and anaemia prevalence among Nigerian children under 5 using a semi-parametric approach”. *Heliyon* (2023). Under review.
4. Jecinta Ibeji, Henry Mwambi and Abdul-Karim Iddrisu, “Spatial pattern of childhood malaria and anaemia in Nigeria and their risk factors using 2010 and 2015 datasets”. Postgraduate Research and Innovation Symposium (PRIS) 2022 South Africa on 8 and 9 December 2022.

Abstract

Childhood mortality and morbidity in Nigeria have been linked to malaria and anaemia. This thesis focused on exploring the risk factors and the complexity of the relationship between malaria and anaemia in under 5 Nigerian children. Data from the 2010 and 2015 Nigeria Malaria Indicator Survey conducted by Demographic Health Survey were used. In 2010, the prevalence of malaria and anaemia was 48% and 72%, respectively, while in 2015, 27% and 68% were the respective prevalences of malaria and anaemia diseases. Machine learning-based exploratory classification methods were used to explain the relationship and patterns between the independent variables and the two dependent variables, namely malaria and anaemia. Decisions made by the public health body are centered on the administrative units (i.e., states) within the country. Therefore, the development of disease mapping and a brief overview of limiting assumptions and ways of tackling them was explained. Consequently, malaria and anaemia spatial variation for 2010 and 2015 was analyzed with the inclusion of their respective risk factors. A separate multivariate hierarchical Bayesian logistic model for each disease was adopted to investigate the spatial pattern of malaria and anaemia and adjust for the risk factors associated with each disease. Furthermore, a multilevel model analysis was applied to independently investigate the spatio-temporal distribution of malaria and anaemia. A joint model was further adopted to check for the relationship between malaria and anaemia and their common risk factors and relax the nonlinearity assumption. In the 2010 data, type of place of residence, mother's highest educational level, source of drinking water, type of toilet facility, child's sex, main floor material, and households that have electricity, radio, television, and water were significantly associated with malaria and anaemia. While in the 2015 data, the type of place of residence, source of drinking water, type of toilet

facility, households with radio, main roof material, wealth index, child's sex, and mother's highest educational level had a significant relationship with malaria and anaemia. The results from this study can guide policymakers to tailor-make effective interventions to reduce or prevent malaria and anaemia diseases. This will help adequately distribute limited state health system resources, such as personnel, funds and facilities within the country.

Keywords: Conditional autoregressive, structured spatial effect, spatial effect, spatio-temporal, spline smoothing, joint modeling, fully Bayesian, Information Criteria, Risk factors, Indicator Survey, Bayesian Hierarchical, Heterogeneity, Convolution.

Acknowledgements

To the Almighty God, Who speaks and stands by His word, all glory to You Lord. My God sent academic father and supervisor, Prof. Henry Mwambi, really, I am short of words but one thing I am very sure is this; My God will bless you indeed. My co-supervisor, Dr. Abdul-Karim Iddrisu, you are more like a brother. Thank you for always being patience and eager to listen and guide. Your kind is rare. My special thanks to my husband and my children for being supportive throughout this research programme. God bless and keep you all for Himself and for me. I am forever grateful to my precious parents, Oke Nze, Lolo Clement Anyaogu and my siblings thank you for your comfort. Thank you, Moses, Elphas and Danielle for being true friends. Thanks to everyone who has made this dream come through, God bless you all.

Contents

Declaration	iii
List of Journal Publications	iv
Abstract	v
Acknowledgement	vii
Contents	viii
List of Figures	xi
List of Tables	xiii
Abbreviation	xiv
Chapter 1: General Introduction	1
1.1 Overview.....	1
1.2 Mapping of disease.....	2
1.3 Models for mapping disease.....	3
1.4 Problem statement.....	8
1.5 Aim.....	10
1.6 Specific objectives.....	10
1.7 Significance of the study.....	11
1.8 Thesis outline.....	12
Chapter 2: Statistical methodology	14
2.1 Spatial modelling of the binary response outcome (malaria or anaemia)	14
2.1.1 Generalized Linear Models.....	14
2.1.2 Spatial Logistic Regression Models.....	15
2.1.2.1 Logistic Regression Models.....	16
2.1.3 Estimation of Parameter.....	17
2.2 Spatio-Temporal Modeling of Malaria or Anaemia.....	18

2.2.1	Spatio-temporal models.....	19
2.2.2	Specifications of prior.....	20
2.2.3	Parameters estimation.....	21
2.3	Spatial joint modeling of malaria and anaemia using parametric approach.....	21
2.3.1	The Penalized regression splines.....	23
2.3.2	Prior distributions.....	24
2.3.3	Estimation of parameter.....	25
2.4	Markov Chain Monte Carlo (MCMC).....	26
2.5	Integrated Nested Laplace Approximation (INLA).....	29

Chapter 3: Spatial variation and risk factors of malaria and anaemia among children

aged 0 to 59 months: A cross-sectional study of 2010 and 2015 datasets 32

3.1	Introduction.....	32
3.2	Data and method.....	35
3.2.1	Studied data.....	35
3.2.2	Dependent variables.....	37
3.2.3	Independent variables.....	39
3.3	Methods.....	42
3.3.1	Spatial models.....	42
3.4	Results.....	47
3.5	Non-linear effect of rural area on malaria and anaemia.....	57
3.6	Discussion.....	58

Chapter 4: Bayesian spatio-temporal modelling and mapping of malaria and anaemia

among children between 0 to 59 months in Nigeria 62

4.1	Introduction.....	62
4.2	Data description.....	65
4.3	Method.....	70
4.3.1	Prior specifications.....	73

4.4 Parameter estimation.....	74
4.5 Results.....	75
4.6 Discussion.....	83
Chapter 5: Spatial joint modelling of malaria and anaemia among Nigeria children	
under 5 years using a semi-parametric approach	85
5.1 Introduction.....	85
5.2 Methods.....	89
5.2.1 Data.....	89
5.3 Statistical methods.....	96
5.4 Estimation of parameter.....	99
5.5 The penalized regression splines.....	100
5.6 Distributions of prior	101
5.7 Distribution of posterior	101
5.8 Diagnostics of model	102
5.9 Data analysis.....	102
5.10 Results.....	105
5.10.1 Model comparison.....	105
5.11 Fixed effects.....	106
5.12 Malaria (2010 and 2015)	107
5.13 Anaemia (2010 and 2015)	108
5.14 Non-linear effects of age.....	110
5.15 Joint spatial effects of malaria and anaemia.....	112
5.15 Discussion.....	113
Chapter 6: Discussion and conclusion	117
Appendix A	119
Appendix B	125
Appendix C	129
References	138

List of Figures

3.1 The 37 states including the capital territory of Nigeria showing the 2010 prevalence of malaria and anaemia of under-5 years children.....	38
3.2 The 37 states including the capital territory of Nigeria showing the 2015 prevalence of malaria and anaemia of under-5 years children.....	39
3.3 The nonlinearity relationship between child's age in months and AOR of malaria and anaemia (2010).....	47
3.4 The nonlinearity relationship between child's age in months and AOR of malaria and anaemia (2015).....	48
3.5 Nigeria state-level estimated posterior mean of the structured spatial effect maps: a.) Mean, b.) Median, c.) 25% quantile and 97.5% quantile for 2010.....	52
3.6 Nigeria state-level estimated posterior mean of the unstructured spatial effect maps: a.) Mean, b.) Median, c.) 25% quantile and 97.5% quantile for 2010.....	53
3.7 Nigeria state-level estimated posterior mean of the structured spatial effect maps: a.) Mean, b.) Median, c.) 25% quantile and 97.5% quantile for 2015.....	56
3.8 Nigeria state-level estimated posterior mean of the unstructured spatial effect maps: a.) Mean, b.) Median, c.) 25% quantile and 97.5% quantile for 2015.....	57
3.9 Estimated significance of non-linear effect of rural area, the posterior mean and the 95% credible intervals.....	58
4.1 Location map of Nigeria showing the 6 geopolitical zones and their 37 states including the capital territory.....	66
4.2 Map of Nigeria showing state rates based on sampling weights of under 5 years old malaria prevalence.....	67
4.3 Map of Nigeria showing state rates based on sampling weights of under 5 years old anaemia prevalence.....	68
4.4 Maps displaying residual spatial effects of malaria in Nigeria for year 2010 and 2015 obtained Spatio-temporal interaction logistic regression model, i.e., Model 7.....	77
4.5 Depicting estimated posterior relative prevalence of malaria for the logistic regression best fitting model	78
4.6 Relationship between child's age in month and AOR of malaria and anaemia.....	81
4.7 Maps displaying residual spatial effects of anaemia in Nigeria for year 2010 and 2015 sprang from the Spatio-temporal interaction logistic regression model i.e., Model 6.....	82

4.8 Depicting estimated posterior relative prevalence of anaemia for the logistic regression best fitting model	82
5.1 Estimating mean of the non-linear effect of age (black) on malaria infection and the corresponding 95% credible interval (red and blue) for 2010 and 2015.....	111
5.2 Estimating mean of the non-linear effect of age (black) on anaemia infection and the corresponding 95% credible interval (red and blue) for 2010 and 2015.....	112
5.3 Residual spatial effect of state on malaria infection in 2010 and 2015.....	113
5.4 Residual spatial effect of state on anaemia infection in 2010 and 2015.....	113

List of Tables

3.1 Selected variable (risk factors) of malaria and anaemia among children aged 0 to 59 months.....	40
3.2 Percentage of malaria and anaemia prevalence with the p-value of children under the age of 5 years in 2010 and 2015 for each variable.....	41
3.3 Model comparisons for malaria and anaemia.....	49
3.4 95% credible interval and adjusted posterior odd ratios estimates (AOR) of malaria.....	51
3.5 95% credible interval and adjusted posterior odd ratios estimates (AOR) of anaemia.....	55
4.1 Exploratory data analysis.....	68
4.2 Summary of the model comparisons of malaria.....	75
4.3 Adjusted posterior odd ratios estimates (AOR) of malaria with 95% confidence interval.....	76
4.4 Summary of the model comparisons of anaemia.....	78
4.5 Adjusted posterior odd ratios estimates (AOR) of anaemia with 95% confidence interval.....	79
5.1 Description of all the variables used in the study.....	92
5.2 Sample size (n) (based on variables) and percentage of all the children tested for malaria and anaemia in 2010 and 2015.....	93
5.3 Univariate exploratory analysis of all the children tested for malaria and anaemia in 2010.....	94
5.4 Univariate exploratory analysis of all the children tested for malaria and anaemia in 2015.....	95
5.5 Summary of all constituents for each model used in the study.....	105
5.6 2010 and 2015 comparative fit statistics of malaria and anaemia.....	106
5.7 Adjusted posterior odds ratio estimates (AOR) and 95% credible interval (95%) of malaria and anaemia for 2010.....	108
5.8 Adjusted posterior odds ratio estimates (AOR) and 95% credible interval (95%CI) malaria and anaemia for 2015.....	110

Abbreviations

NMIS	Nigeria Malaria Indicator Survey
INLA	Integrated Nested Laplace Approach
MCMC	Markov Chain Monte Carlo
DIC	Deviance Information Criteria
GLMM	Generalized Linear Mixed Model
GLM	Generalized Linear Model
CAR	Conditional Auto-Regressive
RBM	Roll Back Malaria
LLNIs	Long-Lasting Insecticide-Impregnated Nets
GAM	Generalized Additive Model
BSAR	Binary Structured Additive Regression
EAs	Enumeration Areas
PSU	Primary Sampling Unit
GPS	Global Positioning System
NMEP	National Malaria Elimination Program
NPopC	National Population Commission
NBS	National Bureau of Statistics
NMCP	Nigeria Malaria Control Program
NPHC	National Population Housing Census

Chapter 1

Introduction

1.1 Overview

Malaria, which can be transmitted to humans from the bites of infected female *Anopheles mosquitoes*, is an acute febrile illness caused by *Plasmodium parasites* (WHO, 2021b). In 2020, about half of the world's population was at risk of malaria, with children under the age of 5 years mostly affected in sub-Saharan Africa in terms of mortality and morbidity (WHO, 2021b, Ugwu and Zewotir, 2018; Weiss et al., 2019; Gething et al., 2016). Malaria and anaemia have been known to be the major cause of mortality and morbidity among children in sub-Saharan Africa especially in Nigeria (Roberts et al., 2020; Ibeji et al., 2022; WHO, 2021a). Anaemia is defined as a decrease of functional haemoglobin in the blood which reduces the quantity of oxygen needed in the tissues and organs of the body, thus incapacitating their functionality (NNPC et al., 2012, NMC report, 2015). Notwithstanding that there are other causes of anaemia, the consensus is that in malaria endemic areas, malaria accounts for a large portion of anaemia prevalence in children under the age of 5 years. Recently, georeferenced data has been associated with several fields of study like agriculture, epidemiology, ecology, and public health. This geo-referencing is typically by areal referencing such as states, provinces, local governments, districts, counties, and other administrative units or point referencing such as longitudes and latitudes. Due to the emergence of this type of data, there is need to develop and apply spatial statistical methods in analyzing tergeographical correlated data. In this dissertation, hierarchical modelling of binary data, which has been extensively utilized in mapping disease, particularly the modelling of malaria and anaemia prevalence, was adopted to investigate the relationship between the two diseases among

children under the age of 5 and also to identify the risk factors of malaria and anaemia in Nigeria using spatial and spatiotemporal models to map the diseases.

1.2 Mapping of Disease

Mapping of disease enables the geographical distribution of a disease within a population to be represented virtually which serves as an important tool in revealing the source of an outbreak. Also, according to Okango et al., (2015) and Rezaeian et al., (2007), it is defined as the estimation and demonstration of the sum-up measures of health results by geographical location. This is because it is used to:

- Create hypotheses concerning disease and its risk factors
- Define the risk factors and change in diseases geographically
- Create maps and atlases of disease
- Identify the clustering of disease

Identification of the clustering disease, comprehending the risk factors and geographical variation or change of diseases and exact hypotheses about a disease can help policy makers in taking decisions on assessment of inequalities, distribution of public health resource and incorporating correct intervention strategies. The applications for discrete variation or difference of disease were discussed, and the lattice data (areal data) was considered as a type of spatial data. Also, Geostatistical data (point referenced data) and point pattern data are other types of spatial data was considered. Cressie, (1993) discussed to a large extent the modeling of geostatistical data types while modeling of the point pattern data types was discussed at length by (Diggle et al., 1998).

1.3 Models for mapping disease

The development and implementation of many diseases mapping models have been made easy because of the availability of limitless amount of georeferenced data and free statistical software. The modelling approach was mainly developed from the generalized linear model. Let y_{ik} denote the disease status of a child i in state k ; a child is said to be positive for the disease if $y_{ik} = 1$ and 0 otherwise. Therefore, a state or comparable structure comprises groups of observations that share common characteristics. The disease status y_{ik} is modelled as a Bernoulli random variable under the generalized linear model (GLM) framework. Explanatory variables can be added in this model which can be categorical or continuous. By assumption, categorical explanatory variables have linear effect on the dependent variable or response function while the continuous explanatory variable may not always have a linear relationship with the dependent variable or response function.

According to Lawson et al., (2003), explanatory variables do not take care of all the variation inherent in the dependent variable in most instances, therefore, the need to extend the GLM to generalized linear mixed model (GLMM) by adding random effects to account for unobserved effects. For each response unit i , the basic type of random effects model initiates an extra parameter μ_i into the linear predictor. The μ_i 's are independent and exchangeable by assumption. Also, they are often modeled under the assumption of normal distribution with mean zero and unknown variance σ^2 . The model is stated in a hierarchical method alongside two stages. Given the values of the random effects, which may be correlated or uncorrelated, the observed statuses in a group of observations that share common characteristics are conditionally independent, and the random effects distribution is stated in the second stage. A spatial covariance matrix can be a way through which the correlated random effects are introduced. It turns the random effects into a single vector

with a proper distribution with an actual mean and a spatial variance-covariance matrix. This spatial variance-covariance matrix comprises of parametric functions that define the covariance structure centered on any two units of study. The spatial covariance between two observations is determined by the distance between the two observations with regards to geostatistical data (Cressie, 1993; Diggle et al., 1998; Waller and Gotway, 2004; Okango et al., 2015). In the case of lattice data, the dissociation between the center point of any pair of regions (i.e., neighbourhood) can be determined by shared borders.

According to Waller and Gotway, (2004); Sherman, (2011); Gaetan and Guyon, (2010), the multivariate Gaussian distribution is majorly used for the joint distribution of random effects partly because of the ease of manipulation and computation. Normality assumption is not preferred here because some of the random effects may show skewness, fat tailness, multimodality etc. which may hide certain vital characteristics of between subjects and within subjects' variation. Notwithstanding, several research works have attempted to relax the normality assumption using parametric and non-parametric distributions (Ngesa et al., 2014), adopting Generalized Gaussian distribution to relax the normality assumption. They concluded that a good result could be obtained if there is a violation of the normality assumption depending on how low or high the peakedness of the data distribution was. It is known that heterogeneous data with asymmetric features have been modeled effectively using skew distributions and the distributions comprises of skew normal and skew-t (Azzalini, 2020) among others. Bayesian nonparametric spatial modeling methods for disease incidence and prevalence data incorporate the Dirichlet distribution (Ferguson, 1973) or its change and the polya trees procedures by Lavine, (1992) for random effects distribution.

Clayton and Kaldor, (1987) were the first to introduce spatially structured prior distribution for random effects with an estimate of the relative risk of a region as an interaction that connects the

local data and a weighted average of observations in the neighbourhood of that region using the empirical Bayes method. Besag et al., (1991) introduced the conditional autoregressive model (CAR) that was performed using Markov Chain Monte Carlo (MCMC) algorithm, and it is equivalent to Clayton and Kaldor, (1987) method. This description gives room for multivariate Gaussian models. At this point, the random effects conditional distribution in a region resulting from all the others is a weighted average of all other random effects. Different weighting schemes which are either data-driven or fixed have been used in some studies. According to Besag et al., (1991), the weights were assigned on the condition of if or not the regions shared similar boundaries, i.e., the weight is said to be 1 if they share a border and 0 otherwise. On the other hand, Best et al., (2001) utilized distance-based spatial weights. Still, preference was more for the adjacency-based model than the distance-based model on the ground of Deviance Information Criteria (DIC) as the model selection criterion. In the smoothing properties of the CAR model, substantial dissimilarities that depend on the specified neighbourhood structure were discovered by Earnest et al., (2007). However, these differences had effect on the ability of the models to predict the observed risk in an area. These results have considerable implications for all researchers using CAR models, given that the neighbourhood weight matrices selected may influence the outcomes of a study. Lu et al., (2007) created a Bayesian hierarchical model that estimates the spatial neighbourhood structure. Their method permitted other observed covariate information, including the data, to assist in the degree and the type of spatial smoothing.

The convolution model comprises of spatially unstructured and spatially structured random effects. The selection of a particular random effect is dependent on the prior information about the scope of dominant risk factors. Nevertheless, to avoid clustering or heterogeneity, it is essential to fairly assign the prior weight to the structured and unstructured components. This can be obtained by

ensuring that the spatially structured random effects conditional distribution has a standard deviation of 0.7 times the spatially unstructured random effects. Still, this submission is debatable (Bernardinelli et al., 1995). Another way of formulating a convolution model is by utilizing just one random intercept. Still, its variance-covariance matrix should comprise spatial and non-spatial components alongside a parameter controlling the spatial dependency (Leroux et al., 2000). Alternatively, a parametric bootstrap method by MacNab and Dean, (2002) and the hidden Markov method by Green and Richardson, (2002) can be used, though uncommonly used.

It is more definite to model diseases with similar risk factors collectively. Modelling disease outcomes jointly within a spatial statistical framework can give a better understanding of the disease's interplay both at the individual and state levels. The multivariate CAR method (Carlin and Banerjee, 2003; Gelfand and Vounatsou, 2003), the shared component method (MacNab, 2010) and the Multiple Membership Multiple Classification (MMMC) methods are the most commonly used methods for carrying out multiple disease modelling.

In the joint modelling of two diseases based on the shared component model, one component is common to both diseases while the other is explicit to one of the diseases. Both components are responsible for the unobserved (also known as latent) spatial variables that affect disease risk, which is not taken care of by the systematic component through the independent variables. Even though MCAR and MMMC seem similar, their slight difference is based on the way each method obtains spatial correlation. While the spatial association is obtained via a variance structure in MCAR, the spatial association for MMMC is obtained via a multiple membership relationship, and the random neighbourhood effects are not independent.

Epidemiologically, observing disease risks in space and time will be a significant gain because changes in the effects of the determinants with time and space can easily be seen. Also, it helps

unveil the most endemic regions and periods of disease outbreaks, including new determinants of these diseases and the ones that have been phased out based on the prevalent disease. This will help policymakers (i.e., lawmakers) to see whether there is a resultant effect from their intervention strategies over time and also if there is a need to incorporate new approaches. The reason for developing spatiotemporal models is to help in the passive investigation of diseases in space and time. Bernardinelli et al., (1995) assumed a Poisson regression model in the Generalized Linear Model (GLM) with the linear independent variables consisting of different terms for space and time and an additional period of space and time interaction effects that allows other temporal trends in different regions. For spatiotemporal modelling, the CAR model developed by Besag et al., (1991) was used by Waller et al., (1997). Their model permitted a distinct and non-spatial random effect every time.

Instead of presenting the spatial dependence through the random effects, allowance should be made for coefficients to differ via spatial domains resulting in spatially varying coefficients. This permits the correlation between the dependent variable and the covariates to vary by region in the spatial domain. By assigning the parameters of regression of the Bayesian autoregressive (BAR), the conditional autoregressive or simultaneous autoregressive (SAR) models, the covariates were permitted to differ spatially (Assunção, 2003). It was also reported that covariates could vary region-wise in the spatial domain (Hastie and Tibshirani, 1990; Gelfand and Vounatsou, 2003).

Thus far, the introduction of spatial dependence through random effects and varying spatial coefficients has been discussed at length. Further, it is worthy to mention that the extent of spatial autocorrelation amongst independently measured values observed in geographical space can also be introduced through observations (Okango et al., 2015). Auto-Poisson models (Besag,

1974; Griffith, 2002) and auto-logistic models (Hoeting et al., 2000) represent spatial dependence through observations.

The MCMC procedure has been used regularly for inference regarding the latent Gaussian models even though other methods like the expectation propagation (EP) by Minka, (2001) and the variational Bayes method (Hinton and Van Camp, 1993) have been adopted. According to (Rue et al., (2009), the disadvantage of the MCMC procedure is always the longer time of computation and convergency. Many researchers have used the integrated nested Laplace approximation (INLA) developed by Rue et al., (2009) to get similar results as MCMC but quicker. The following are the ternary stages of the INLA procedure.

1. Approximation of the hyper-parameters posterior in the data used in ascertaining the hyper-parameter grid values.
2. Approximation of the posterior marginal distributions in the data and the hyper-parameter grid values.
3. Performance of numerical integration of the product of the two approximations to find the posterior marginal of interest.

Time is better managed in this procedure than in MCMC, which jointly samples directly from the posterior distribution.

1.4 Problem statement

Recently, disease mapping has been widely used because of easy access to geo-referenced data and free software. Even with notable results from developed models for disease mapping, there are still some limiting assumptions, leading to incorrect results and interpretations. Usually, spatial, and non-spatial components are contained in these models' random effects. The data may

exhibit fat tailness, skewness and multimodality resulting in inappropriate use of the normality assumption used for the non-spatial component, distorting the results. As a result of this, it is necessary to look at flexible models that permit non normal random effects. Non-parametric models can be used instead of parametric models for the random effects. In this research work, the prevalence of malaria and anaemia in children under the age of 5 years was investigated using Nigeria Malaria Indicator Survey (NMIS) datasets. Developing suitable statistical methods will enable our understanding of the fundamental factors of malaria and anaemia in Nigeria and the geographical variations of malaria and anaemia prevalence in Nigeria of children under the age of 5 years and their changes over time. To avoid the normality assumption restriction, we adopted the non-parametric (also known as distribution-free) modelling of the random effects employing the mixture of Dirichlet and the mixture of Polya trees models. In addition, this work looked at several approaches to relaxing the linearity assumption on covariate effects with the application of the penalized regression splines and the random walk model.

It is incorrect to assume that similar stimuli can cause similar responses throughout the study region. Different responses across the study area can be obtained from similar stimuli in terms of spatial variation. And this may be due to attitudes, cultures, preferences, and climate, along with other causes predominant within an area. Hence, it will be rational to say that the effects of a child's risk factors are based on location. This is possible by letting the regression coefficient differ through space.

It is advisable and more effective to model diseases jointly if they share similar risk factors. By utilizing data from Nigeria Malaria Indicator Survey (NMIS) data from (NDHS), there are advantages on efficiency of estimates and precision. In this research work, we developed a spatial

and spatio-temporal model for malaria and anaemia and later did a joint modelling of the two diseases.

The effect of policymakers' involvement in initiating plans and ideas for eradicating a particular disease can be studied over time. Also, the progressive study of the evolution of the effect of a specific risk factor in space will be of an enormous value. Here, their effects were studied by developing spatial and spatiotemporal spatially varying coefficient models and applying them to the Nigeria Malaria Indicator Survey (NMIS) data.

1.5 Aim

This study's aim majors on investigating the risk factors and the difference in the geographical spread of malaria and anaemia in children under the age of 5 years using Bayesian hierarchical models.

1.6 Specific objectives

Specifically, the research objectives include the following:

1. To review statistical techniques for binary choice outcomes utilized in the mapping of disease.
2. To ascertain suitable spatial models for modelling the prevalence of malaria or anaemia at the state level using 2010 and 2015 NMIS data.
3. To ascertain suitable spatiotemporal models for modelling and mapping malaria or anaemia prevalence in Nigeria at the state level using 2010 and 2015 NMIS data.
4. To use semi parametric-joint models and apply them to the Nigeria Malaria-anaemia Indicator Survey (NMIS) data.

5. To relax the covariates stationary assumption and apply spatially varying covariates model to the data.

1.7 Significance of the study

At the beginning of 21st century, malaria was recognized as a priority global health issue. But this consciousness phased out between the 1960s and the late 1990s and resulted in a dramatic rise in malaria morbidity and mortality rate. In addition to malaria, anaemia is another global health issue, especially in areas of high malaria transmission due to the effect of malaria on the red blood cells. There is a need for continuous political commitment, extensive and predictable financing, and increased regional partnership to bring down this trend. Also, knowing the risk factors of malaria and anaemia will significantly help improve people's health, especially children. Furthermore, this research work will be advantageous to the populace, specifically government for policymaking and intervention. Nevertheless, the contributions of this research work should not be of interest only to healthcare or some organization desiring to develop models for tackling problems, especially in public health, but also to society and the government for intervention and policy making. Both in developed and developing countries, malaria and anaemia diseases are considered under epidemiology and biostatistics. Childhood malaria and anaemia diseases has been a global concern. Therefore, understanding the risk factors associated with malaria and anaemia infection can enhance appropriate solutions to mitigate their spread and initiate program developments. In addition, there is need to appreciate the spatial, spatio-temporal, and joint spatial distribution of these diseases. This research will present appropriate statistical methods and modelling techniques for the determination of the significant factors connected with malaria and anaemia in children. Thus, the importance of

this research work will be of interest to biostatistician, epidemiologist, public health policy makers and intervention.

1.8 Thesis outline

In this thesis, we develop models and methods for spatial and spatio-temporal analysis of disease. This thesis has six chapters. Chapter 1 is the general introduction of this research work that comprises the developments in disease mapping. Chapter 2 is an overview of the methods. Chapters 3 to 5 are the full research articles published or submitted in peer-reviewed journals. While chapter 6 contains the overall conclusions of this research work. Here are the contents of the chapters.

Chapter 1: This chapter is the general introduction of this research work which focuses on the developments in disease mapping for single and multiple diseases and the study's objectives.

Chapter 2: The second chapter is an overview of the methods used in this study.

Chapter 3: In this third chapter, we develop models for spatial variation and risk factors of malaria and anaemia among children aged 0 to 59 months: A cross-sectional study of 2010 and 2015 datasets. Model comparison was carried out on each dataset; the convolution and generalized linear mixed models gave the least Deviance Information Criteria (DIC) in 2010 for malaria and anaemia, respectively. While in 2015, the conditional auto-regressive (CAR) and convolution models showed the least DIC for malaria and anaemia, respectively.

Chapter 4: A spatio-temporal model was developed for Bayesian spatio-temporal modelling and mapping of malaria and anaemia among children between 0 to 59 months in Nigeria. For each of the diseases, model 7 [interaction with one random time effect (random walk)] and model 6 [interaction with an auto-regressive structure of order 1 or AR(1)] were used to model and map malaria and anaemia, respectively, because they have the least DIC.

Chapter 5: In the fourth chapter, we use the nonlinear relationship between malaria and anaemia of Nigerian children under age 5: A semi-parametric spatial joint modelling. The convolution model was used both 2010 and 2015 datasets, which captures both spatially structured and spatially unstructured random effects.

Chapter 6: The summary of this research work is contained in this chapter. The findings, conclusions and highlights of some topics that need further research are summarized here.

Chapter 2

Statistical methodology

2.1 Spatial modelling of the binary response outcome (malaria or anaemia)

Spatial data, sometimes called geospatial data, contains information about a specific location on the earth's surface. Incorporating spatial data within disease mapping in epidemiology, public health, and other research work on diseases has been a long practice. This disease mapping utilizes the hierarchical Bayesian methods, which permit overdispersion and spatial correlation. In this section, models that are usually employed on the binary outcome will be reviewed here. Most importantly, a flexible method for similar models permits capturing different covariates. The primary concern is to include spatial random effects comprising correlated and uncorrelated heterogeneity. And it is known that the group of models utilized for binary response outcomes are the generalized linear models (GLMs).

2.1.1 Generalized Linear Models

McCullagh and Nelder, (1989) introduced the generalized linear models (GLMs) as classes of models employed to extend the modelling of normal data to include non-normal data and, as such, can be used to model both normal and non-normal data. It is assumed that the observations y_i , $i = 1, \dots, n$ are independent with a distribution in the exponential family, and their probability density is expressed as;

$$f(y_i|\mu_i, \phi, \omega_i) = \exp\left(\frac{y_i\mu_i - b(\mu_i)}{\phi}\omega_i + c(y_i, \phi, \omega_i)\right), i = 1, \dots, n \quad (2.1)$$

With regards to the GLMs, the dispersion parameter is $\phi = 1$, and the weight for the observations is ω_i , while naturally, the parameter of the exponential family is μ_i . Also, $c(y_i, \phi, \omega_i)$ and $b(\mu_i)$ are specific dependent functions of the exponential family. Therefore, the linear predictor of the covariates x_i' is defined as;

$$\eta_i = x_i' \beta, i = 1, \dots, n,$$

and it has synergy with conditional mean $\theta_i = E(y_i|x_i, \beta)$ through a link function

$$g(\theta_i) = \eta_i = x_i' \beta, i = 1, \dots, n,$$

where the natural link function is $g(.)$. The link function has a number of choices. Thus, when $y_i \in \{0,1\}$, the link function of a Bernoulli distributed random variable is the logit link function. Then, a logit model is formed, which denotes the systematic logistic distribution function, thereby satisfying the link function GLMs component. The mean (y_i) determines the link between the structure and distribution assumption; in the same vein, it is distributionally assumed as;

$$E(y_i|x_i, \beta) = \theta_i = b'(\mu_i) \text{ and } Var(y_i|x_i, \beta) = b''(\mu_i)/\omega_i.$$

McCullagh and Nelder, (1989; Fahrmeir et al., (1994) and Agresti, (2007) discussed essentially the theory of GLMs. The special case of the univariate GLMs employed in modeling the binary response is discussed below.

2.1.2 Spatial Logistic Regression Models

Under Bayesian and frequentist framework, logistic regression models are mostly used to study the link between the binary response outcome and the covariates. Also, in disease mapping, dichotomous response data is modelled using logistic regression models to expound the geographic differences that occur in the data.

2.1.2.1 Logistic Regression Models

Let y_{ijk} represents a binary response for disease $k = 1, 2$ of child $j = 1, 2, \dots, n$ in state $i = 1, \dots, 37$, with response 0 or 1 i.e.,

$$y_{ijk} = \begin{cases} 1, & \text{If child } j \text{ in state } i \text{ has malaria (k=1) or is anaemic (k=2)} \\ 0, & \text{Otherwise} \end{cases}$$

By assumption, the response y_{ijk} is independent Bernoulli distributed that is

$$y_{ijk} \sim \text{Bernoulli}(\pi_{ijk}), \quad i = 1, \dots, 37, \quad j = 1, \dots, n, \quad k = 1(\text{Malaria}) \text{ or } 2(\text{Anaemia})$$

where unknown probabilities are $\pi_{ijk} = P(y_{ijk} = 1)$ and $1 - \pi_{ijk} = P(y_{ijk} = 0)$ with $E(y_{ijk}) = \pi_{ijk}$ relating to predictors through a logit link function as;

$$\text{logit}(\pi_{ijk}) = \log \left(\frac{P(y_{ijk}=1)}{1-P(y_{ijk}=1)} \right) = \eta_{ijk} = \mathbf{x}'_{ijk} \boldsymbol{\beta} \quad (2.2)$$

The categorical covariates are the vector $\mathbf{x}_{ijk} = (1, x_{ijk1}, \dots, x_{ijkp})'$ and the vector of the regression coefficients are $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$. This model permits a parametric type of categorical covariates. This section focuses on extending the linear predictor η_{ijk} of the logistic regression model in Equation (2.2) to cater for extra flexible forms. Therefore, by incorporating distinct types of covariates to increase the model complexity, the logistic regression model was first extended to take care of state-specific random effects by adding to the linear predictor in Equation (2.2) a geoaddivitive predictor. Thereby integrating the random effects in the model to capture additional variation. Hence, to capture unobserved influential factors that differ throughout the states, the model is expressed as follows;

$$\eta_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + f_{str}(s_i) \quad (2.3)$$

where $f_{str}(s_i)$ accounts for structured random effects. While the unstructured random effect is another alternative model, defined as;

$$\eta_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + f_{unstr}(s_i), \quad (2.4)$$

where f_{unstr} takes care of unobserved heterogeneity inside each state.

The convolution model comprises the random spatial components below;

$$\eta_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + f_{spat}(s_i) \quad (2.5)$$

where the spatial random effects consist of two components, i.e., $f_{spat} = f_{str}(s_i) + f_{unstr}(s_i)$.

Also, by assumption, both components have independent prior distributions (Besag et al., 1991).

The regression coefficients $\boldsymbol{\beta}$ are presumed to have diffuse priors ($\boldsymbol{\beta} \propto const$), the spatially unstructured random effects were assumed to follow an i.i.d Gaussian distribution and to model the spatially structured random effects, an intrinsic conditional autoregressive (iCAR) model was employed.

2.1.3 Estimation of Parameter

A fully Bayesian (FB) technique was adopted to obtain the parameter estimation. All unknown parameters in a fully Bayesian approach are taken as random variables and are given priors, and further, hyperpriors were assigned to hyperparameters. The estimated parameters are obtained using the Markov chain Monte Carlo (MCMC) simulation technique to sample from the posterior distribution. Nevertheless, the parameter estimation in this work was obtained using the integrated nested Laplace approximation (INLA). For all the models formulated earlier in this section,

$\vartheta = \{\{\beta\}, \{f_{str}(\cdot)\}, \{f_{unstr}(\cdot)\}\}$ are the latent Gaussian variables, and the hyperparameters are represented by a group of precision parameters $\psi = \{\zeta_{str}, \zeta_{unstr}\}$. Thus, the posterior distribution is defined as;

$$p(\vartheta, \psi | y) \propto L(y | \vartheta, \psi) p(\vartheta, \psi) \quad (2.6)$$

Conjugate gamma priors are assigned to the hyperparameters ζ_{str} and ζ_{unstr} , with $\zeta_{str} \sim \text{Gamma}(1, 0.00005)$ and $\zeta_{unstr} \sim \text{Gamma}(1, 0.00005)$. The changeability of structured and unstructured random effects is obtained respectively by $\sigma_{str}^2 = \frac{1}{\zeta_{str}}$ and $\sigma_{unstr}^2 = \frac{1}{\zeta_{unstr}}$. Additionally, there was an imposition of a sum to-zero constraint on both functions of spatial random effects for identification.

2.2 Spatio-Temporal Modeling of Malaria or Anaemia

Spatio-temporal modelling is employed when a reasonable number of spatially referenced health data is gathered over a period of time. This section introduces the extension of spatial modeling in disease mapping, i.e., spatio-temporal modeling, to model disease risk in space and time. Assuming y_{ijkt} is malaria or anaemia status of a child j in state $i: i = 1, \dots, 37$ during time $t: t = 1, 2$ and $k = 1$ (malaria) and 2 (anaemia). The binary response is defined as;

$$y_{ij1t} = \begin{cases} 1, \text{Malaria} \\ 0, \text{No malaria} \end{cases}$$

$$y_{ij2t} = \begin{cases} 1, \text{Anaemia} \\ 0, \text{No anaemia} \end{cases}$$

Let $y_{ijkt} \sim \text{Bernoulli}(\pi_{ijkt})$, where π_{ijkt} are unknown probabilities associated with the event probabilities of the models. This response outcome belongs to the exponential family distributions in univariate GLMs. The logistic regression model is defined as follows;

$$\text{logit}(\pi_{ijkt}) = \beta_0 + \eta_{ijkt} \quad (2.7)$$

The predictor $\eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta}$ has the covariate vector $\mathbf{x} = (x_{ijkt1}, \dots, x_{ijktq})'$, and the vector of regression coefficient $\boldsymbol{\beta}$ is expressed as $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)$ while the intercept in the model under the logit model is β_0 . The cohesive framework of the structured additive regression (STAR) models is employed to allow flexibility where an extension to include a more flexible additive predictor can be made to the classical predictor. Therefore, extension is done on structured additive predictor for spatiotemporal modelling as;

$$\eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{\text{spat}}(s_i) + f_{\text{year}}(t) + f_{it}(s_i, t) \quad (2.8)$$

where f_{spat} , f_{year} and f_{it} are functions applicable for space, year and space-year interaction, respectively. It is essential to know that these two spatial and temporal predictors are independent. The spatial components f_{spat} comprise of spatially structured f_{str} and unstructured f_{unstr} effects. Also, the random year effects f_{year} can be modeled as a first-order random walk or AR(1), and $f_{it}(s_i, r)$ represents the space-year interaction (DiMaggio, 2012).

2.2.1 Spatio-temporal models

Here, seven models under the spatiotemporal model of logistic regression models were studied.

These models are:

$$\text{Model 1: } \eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{\text{str}}(s_i) + f_{\text{unstr}}(s_i),$$

$$\text{Model 2: } \eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{\text{str}}(s_i) + f_{\text{unstr}}(s_i) + \beta_t,$$

$$\text{Model 3: } \eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{\text{str}}(s_i) + f_{\text{unstr}}(s_i) + f_{\text{year}}(t),$$

Model 4: $\eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{str}(s_i) + f_{unstr}(s_i) + f_{it}(s_i, t),$

Model 5: $\eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{str}(s_i) + f_{unstr}(s_i) + \beta_t + f_{it}(s_i, t),$

Model 6: $\eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{str}(s_i) + f_{unstr}(s_i) + f_{year}(t) + f_{it}(s_i, t),$

Model 7: $\eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{str}(s_i) + f_{unstr}(s_i) + f_{1year}(t) + f_{it}(s_i, t),$

In general,

- \mathbf{x}'_{ijkt} denotes the row vector of categorical covariates effects for child j in state i for disease k in time t .
- $\boldsymbol{\beta}$ is a vector of regression coefficients.
- β_t stands for the year-specific fixed effects.
- $f_{str}(s_i)$ and $f_{unstr}(s_i)$ are the respective structured and unstructured spatial components.
- f_{year} and f_{1year} are the smooth functions of the temporal random effects where the first term is the distinct temporal year effect, and the second term is a first order random walk effect capturing dependence between years.
- $f_{it}(s_i, t)$ represents the spatial-year interaction effect.

2.2.2 Specifications of prior

This chapter utilized the full Bayesian approach for the spatio-temporal logistic regression models. By assigning diffuse priors to the fixed effect and linear year trend, the *intrinsic* conditional autoregression (iCAR) model was employed to model the spatially structured random effects, and i.i.d Gaussian prior was used for spatially unstructured random effects. The temporal year random effects f_{1year} were modeled by a first order walk structure. But it should be noted

that in the models, distinct prior specifications for the temporally varying year random effects f_{year} were assigned. The penalized splines were assigned for the spatio-temporal logistic regression model to capture spatial year-specific effects.

2.2.3 Parameters estimation

A fully Bayesian procedure was utilized in the parameter estimation of the spatio-temporal logistic regression model. Therefore, all unknown parameters are random variables and are assigned appropriate prior distributions. The posterior distribution is expressed as;

$$p(\varphi, \omega|y) \propto L(y|\varphi, \omega)p(\varphi, \omega) \quad (2.9)$$

where the likelihood is $L(y|\varphi, \omega)$ and the prior distribution of the model is $p(\varphi, \omega)$. $\varphi = \{\{\beta\}, \{\beta_t\}, \{f_{str}(\cdot)\}, \{f_{unstr}(\cdot)\}, \{f_{year}(\cdot)\}, \{f_{1year}(\cdot)\}, \{f_{it}(\cdot)\}\}$ and the equivalent hyperparameters are given by $\omega = \{\varsigma_{str}, \varsigma_{unstr}, \varsigma_{year}, \varsigma_{1year}, \varsigma_{it}\}$. To enhance the speed of computation, all hyper-parameters priors were assigned conjugate gamma priors $Gamma(1, 0.00005)$. R-integrated nested Laplace approximation (INLA) package was used to estimate the parameters.

2.3 Spatial joint modeling of malaria and anaemia using parametric approach

Assuming the response variable y_{ijk} is the status (0/1) of disease k , where Malaria is $k = 1$ and Anaemia is $k = 2$, for child j in state $i: i = 1, \dots, 37$. The notation that child j in state i has malaria is $y_{ij1} = 1$ and zero otherwise and $y_{ij2} = 1$ if child j in state i has anaemia and zero otherwise. For this study, y_{ijk} follows a bivariate Bernoulli distribution with π_{ij1} and π_{ij2} being the respective

probabilities of a child j having malaria or anaemia. Therefore, the generalized linear model is defined as;

$$g_1(\pi_{ij1}) = \mathbf{x}_{ij1}\boldsymbol{\beta}_1 + \mathbf{z}_{ij1}\mathbf{r}_1 \quad (2.10)$$

$$g_2(\pi_{ij2}) = \mathbf{x}_{ij2}\boldsymbol{\beta}_2 + \mathbf{z}_{ij2}\mathbf{r}_2 \quad (2.11)$$

where the assumed vectors of fixed effects are $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, \mathbf{r}_1 and \mathbf{r}_2 represent the vectors of random effects, and the designed vectors for fixed effects and random effects are \mathbf{x}_{ij1} , \mathbf{x}_{ij2} , \mathbf{z}_{ij1} and \mathbf{z}_{ij2} respectively.

\mathbf{x}_{ijk} is a q dimensional vector of continuous independent variables for disease k given by

$\mathbf{x}_{ijk} = (x_{ij1}, x_{ij2}, \dots, x_{ijq})'$. Here, $q = 1$ since age is the only continuous variable in this study and \mathbf{z}_{ijk} is a v dimensional vector of categorical independent variables for disease k , while j is a child in state i .

The unknown $E(y_{ijl}) = \pi_{ijl}$ is connected to independent variable as:

$$g(\pi_{ij1}) = \mathbf{x}_{ij1}\boldsymbol{\beta}_1 + \mathbf{z}_{ij1}\mathbf{r}_1 \text{ for Malaria} \quad (2.12)$$

and

$$g(\pi_{ij2}) = \mathbf{x}_{ij2}\boldsymbol{\beta}_2 + \mathbf{z}_{ij2}\mathbf{r}_2 \text{ for Anaemia} \quad (2.13)$$

where the logit link function is represented by $g(\cdot)$, for the continuous independent variables, $\boldsymbol{\beta}$ is the q dimensional vector of regression coefficients, and \mathbf{r} (i.e., r_1 and r_2) denotes a v dimensional vector of regression coefficients for categorical independent variables. The non-linear effects of the continuous predictors and the spatial autocorrelation in the data were handled by

adopting the convolution model and the penalized regression spline approach under a semi-parametric model.

The highly restrictive linear predictor was relaxed by penalized regression spline approach to a more flexible semi-parametric predictor, expressed as:

$$g(\pi_{ij1}) = \sum_{t=1}^q f_t(x_{ijk}) + f_{spat}(s_{i1}) + \mathbf{z}'\mathbf{r}_1 \text{ for Malaria} \quad (2.14)$$

and

$$g(\pi_{ij2}) = \sum_{t=2}^q f_t(x_{ijk}) + f_{spat}(s_{i2}) + \mathbf{z}'\mathbf{r}_2 \text{ for Anaemia} \quad (2.15)$$

The non-linear twice differentiable smooth function for the continuous covariate is the function $f_t(\cdot)$ while the factor that takes care of the spatial effect of each state is $f_{spat}(\cdot)$. The convolution model was implemented in this study, assuming that the spatial effect can be disintegrated into a sum of two components: spatially structured and spatially unstructured, i.e., $f_{spat}(s_{il}) = f_{str}(s_{il}) + f_{unstr}(s_{il}), l = 1, 2$ (Ngesa et al., 2014; Manda and Leyland, 2007; Okango et al., 2015). Cultures, common cultural practices, climate etc., are called unobserved predictors inherent within the states or the relationship within the states which are under spatially unstructured random effects. Alternatively, any unobserved predictors which vary spatially across the states can be explained by spatially structured random effects. This is known as spatial autocorrelation, though technically, it is expressed as the dependence due to geographical nearness.

2.3.1 The Penalized regression splines

The approaches for assessing the smooth function $f_t(\cdot)$ have been discussed to a reasonable extent by some studies (Hastie et al., 2001; Fahrmeir et al., 1994; Hall and Patil, 1995). The Penalized regression spline suggested by Eilers and Marx was used in this study (Eilers and Marx, 1996).

Now, assuming that the polynomial spline is utilized, then approximating the effect of the continuous predictors will be necessary. They assumed a spline of degree d with K equally spaced knots $x_{q,min} = \varphi_{q1} < \varphi_{q2} \dots \varphi_{qk-1} < \varphi_{qK} = x_{q,max}$ giving:

$$f(x, \theta) = \vartheta_0 + \vartheta_1 x + \dots + \vartheta_q x^d + \sum_{k=1}^K b_k (x - \varphi_k)_+^d \quad (2.16)$$

where $\theta = (\vartheta_0, \vartheta_1, \dots, \vartheta_q, b_1, b_2, \dots, b_k)'$ and the $(\Delta - \Psi)_+$ is the same as $(\Delta - \Psi)$ if $(\Delta - \Psi)$ is positive and zero otherwise. Δ is the predictor variable, Ψ is the knot location and the subscript denotes the positive part of the argument.

In this study, a quadratic spline ($d = 2$) comprising of 20 knots was employed to guarantee flexibility and the k^{th} knot was expressed as the sample quantile of the continuous independent variables obtained by the probability equal to $\frac{k}{k+1}$. A roughness penalty $-\frac{1}{2}\alpha \int_{x_{min}}^{x_{max}} [f''(x)]^2 dx$ applied in the log likelihood function to prevent getting an unsmooth function which wriggles to a large extent as suggested by Green and Silverman, (1993) to avoid over fitting. This gives the penalized log-likelihood function expressed as;

$L = l(y, \theta, \delta) - \frac{1}{2}\alpha \int_{x_{min}}^{x_{max}} [f''(x)]^2 dx$, where α (smoothness parameter) balances flexibility and smoothness.

2.3.2 Prior distributions

A prior distribution for the spatially structured effects $f_{str}(s_i) = (f_{str}(s_{i1}), f_{str}(s_{i2}))'$ was used based on the nearby neighbor multivariate Gaussian Markov random field (GMRF), and is expressed as:

$$f_{str}(s_i, s_k) \sim MCAR(1, \Sigma),$$

where Σ is the covariance matrix capturing correlation.

Two regions are said to be neighbours if there is a common border between them, i.e., spatial-wise, otherwise, they are not. The univariate form of the Multivariate Conditional Autoregressive Model (MCAR) and conditions in which the conditional multivariate distributions distinctively ascertain the equivalent multivariate joint probability density function, respectively, were considered by Besag et al., (1991) and Mardia, 1988). From their results, Multivariate Conditional Autoregressive Model (MCAR) was developed (Carlin and Banerjee, 2003). By assumption, the unstructured spatial effects follow a Multivariate Gaussian prior i.e., $f_{unstr}(s_i, s_k)/\tau_{unstr}^2 \sim MVN(0, \tau_{unstr}^2)$. Inverse gamma distributions were allocated to the variance hyper-parameter as:

$$\tau_{str}^2 \sim IG(0.0001, 0.0001) \text{ and } \tau_{unstr}^2 \sim IG(0.0001, 0.0001),$$

while the fixed effects coefficients were provided with the following priors distributions:

$$\varrho_0, \varrho_1, \dots, \varrho_q \sim N(0, 10^6), \delta_1, \alpha_2, \dots, \alpha_v \sim N(0, 10^6), b_j \sim N(0, \tau_b^2) \quad \text{and} \\ \tau_b^2 \sim IG(0.0001, 0.0001), \beta_1, \beta_2 \sim N(0.01, 0.01) \text{ is the intercept}$$

2.3.3 Estimation of parameter

Here, the parameters estimation was carried out using a fully Bayesian approach, and suitable prior distributions were given, as discussed earlier. The prior distribution with the observed data is upgraded to obtain posterior distribution. Therefore, it is known as the distribution of the parameters after observing the data. The samples for Bayesian inference are given by posterior distribution. The problem of high dimensionality may be avoided in implementing Markov Chain

Monte Carlo (MCMC) by sampling from the posterior distribution repeatedly. At the same time, mean and median estimates are computed from the data sample summaries.

By assumption, let the Conditional Independence be in the middle of the response variable, and the hyperparameters, the posterior distribution for the Bernoulli model is defined as:

$$Q_{post}(\vartheta, \alpha, b, \tau^2 / y) \propto L(y / \vartheta, \alpha, b, \tau^2) Q_{pri}(\vartheta, \alpha, b, \tau^2) \quad (2.17)$$

$$\begin{aligned} &= \prod_i \prod_k L(y_{ik} / \theta, \alpha, \tau^2) \prod_{l=1}^q [Q(b_l / \tau_l^2) Q(\tau_l^2)] \\ &\quad \times \prod_k^v [Q(\delta_k / \tau_k^2) Q(\tau_k^2)] \times Q(f_{str} / \tau_{str}^2) Q(\tau_{str}^2) Q(f_{unstr} / \tau_{unstr}^2) Q(\tau_{unstr}^2) \end{aligned}$$

WinBUGS 14 was used to carry out all the analyses in this study (Spiegelhalter et al., 2007 and Okango et al., 2015). A choice of 20000 iterations of Markov Chain Monte Carlo (MCMC) were implemented to execute each model, while the 10000 iterations done earlier was removed to take care of the burn-in period. The other 10,000 were used to access the meeting point of the MCMC and parameter estimation.

2.4 Markov Chain Monte Carlo (MCMC)

Physicists at Los Alamos developed the MCMC techniques in the 1940's. Statistically, Markov Chain Monte Carlo (MCMC) techniques comprise a class of algorithms for sampling from a probability distribution, i.e., they are applied to approximate the posterior distribution of a parameter of interest by random sampling in a probabilistic space. Here are the procedures of MCMC techniques;

a) Recall that for a sequence of random variables $\{X_1, X_2, \dots, X_t\}$ on a discrete state space ω to be called a Markov chain of order 1,

$$Q(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_1 = x_1) = Q(X_t = x_t | X_{t-1} = x_{t-1}).$$

By restricting this Markov chain to time-homogenous, then,

$$Q(X_t = x_t | X_{t-1} = x_{t-1}) = Q(x_t | x_{t-1}) \in \mathbb{R}^{\omega \times \omega}$$

$$\text{Note: } [Q(X_{t+1} = x_t | X_{t-1} = x_{t-1})]_{(x_t, x_{t-1}) \in \omega} = Q^2$$

b) If there exist h such that $Q^h > 0$, then the Markov chain is called ergodic. Therefore, considering this condition, Markov chain is said to be equivalent to

i) Irreducible i.e., for all $x, \alpha \in \omega$, there exists $h(x: \alpha)$ such that $Q^{h(x, \alpha)}(x, \alpha) > 0$

ii) Aperiodic i.e., for all $x \in \omega$, (the greatest common divisor) $\text{GCD}\{h: Q^h(x, x) > 0\} = 1$.

c) The distribution of σ on ω is a stationary distribution of a Markov chain such that

$$\sigma(\alpha) = \sum_{x \in \omega} Q(\alpha | x) \sigma(x). \quad (2.18)$$

d) Then $d_{TV}(u, z)$ i.e., the total variation distance between two probability measures u, z on ω will be

$$\|u - z\|_{TV} := \max_{R \subset \omega} \{u(R) - z(R)\} = \frac{1}{2} \sum_{b \in \omega} |u(b) - z(b)|$$

The time until the total variation distance to σ is below ϵ is called the mixing time function, $T_{mix}(\epsilon)$.

$$T_{mix}(\epsilon) := \max_{x_0 \in \omega} \min\{t: \|Q^t(\cdot, x_0) - \sigma(\cdot)\|_{TV} \leq \epsilon\}$$

$$\sim \ln\left(\frac{1}{\epsilon}\right).$$

e) If there exists a probability distribution σ on ω such that

$$Q(x|\alpha)\sigma(\alpha) = Q(\alpha|x)\sigma(x), \quad (2.19)$$

then the Markov chain is reversible. If σ satisfies Equation (2.18), then it is called the stationary distribution of the Markov chain given that

$$\sum_x Q(\alpha|x)\sigma(x) = \sum_x Q(x|\alpha)\sigma(\alpha) = \sigma(\alpha)$$

which is Equation (2.18). When Equation (2.19) is at equilibrium, it means that the chain moves at the same rate from x to α as well as from α to x and this is the reason Equation (2.19) is called the detailed balance equation. Though satisfying the detailed balance equation is sufficient, but it is not a required condition for σ to be a stationary distribution because ergodicity is also necessary to justify that σ is a stationary distribution.

Usually, much focus is on scalar-valued functions of the parameter vector θ and let $\gamma(\theta)$ be that function. Assuming that there is an n MCMC samples from the stationary distribution then this will result to n samples of $\gamma(\theta)$, i.e.,

$$\{\gamma_1 := \gamma(\theta_1), \dots, \gamma_n := \gamma(\theta_n)\}.$$

Therefore, the sample mean will be $\gamma = n^{-1} \sum_{i=1}^n \gamma_i$. Also, the posterior intervals for $\gamma(\theta)$ can be calculated as follows;

- i) Let $L(\tau_1) := \tau_1$ be the lower sample quantile and $U(\tau_2) := \tau_2$ the upper sample quantile of $\gamma_1, \dots, \gamma_n$. Then $(L(\tau_1), U(\tau_2))$ is a $1 - (\tau_1 + \tau_2)$ posterior interval.

- ii) An equi-tailed $1 - \tau$ posterior interval is obtained if $\tau_1 = \tau_2 = \tau/2$.
- iii) The highest posterior density interval is solved numerically for τ_1 and τ_2 such that $\tau = \tau_1 + \tau_2$ and $U(\tau_2) - L(\tau_1)$ is minimized. Remember that if the posterior of $\gamma(\theta)$ is not unimodal, this interval could be a union of intervals.

2.5 Integrated Nested Laplace Approximation (INLA)

INLA is known to be more flexible for inference in the Bayesian framework than MCMC. This INLA technique was proposed recently by Rue et al. (2009). INLA is preferred because it gives similar results faster than MCMC. Nevertheless, in the INLA software, some vital models have been implemented.

In general, the idea of INLA is to approximate the posterior marginals for latent Gaussian models. This class of models is a subset of all the flexible and extensively used Bayesian additive regression models. Let the observed data points be α , x is the vector of all the latent Gaussian variables and θ represents the vector of hyperparameters. Considering the likelihood of observations n , let's assume the conditional independence α to be:

$$Q(\alpha|x, \theta) = \prod_{i=1}^n Q(\alpha_i|x_i, \theta) \quad (2.20)$$

and

$$Q(x|\theta) = \frac{1}{\sqrt{2\pi}} |G(\theta)|^{\frac{1}{2}} \exp\left(-\frac{1}{2} x' G(\theta) x\right) \quad (2.21)$$

is the density function of the latent effects while assuming a multivariate Gaussian prior on the mean of x to be zero and $G(\theta)$ as the precision matrix. At the same time, the matrix determinant is $|\cdot|$ (Blangiardo and Cameletti, 2015). According to the properties of x , they are conditionally independent in such a way that $G(\theta)$ is a sparse matrix that gives room for inference with Gaussian

Markov random fields (GMRFs). The joint posterior distribution for the latent Gaussian models in the work of Rue et al. (2009) concentrated on the estimation of the Equation given by the product of Equation (2.18), (2.19) and hyperparameter prior $Q(\theta)$ as

$$Q(x, \theta | \alpha) \propto Q(\theta)Q(x | \theta)Q(\alpha | x, \theta) \propto$$

$$Q(\theta) | G(\theta) |^{\frac{1}{2}} \exp \left(-\frac{1}{2} x' G(\theta) x + \sum_{i=1}^n \log(Q(\alpha_i | x_i, \theta)) \right) \quad (2.22)$$

INLA aims at approximating each element of the posterior marginals $Q(x_i | \alpha)$ and $Q(\theta_k | \alpha)$ parameter vectors. Recall that

$$Q(x_i | \alpha) = \int Q(x_i | \theta, \alpha) Q(\theta | \alpha) d\theta \quad (2.23)$$

is the posterior marginals of the latent Gaussian components and

$$Q(\theta_k | \alpha) = \int Q(\theta | \alpha) d\theta_{-k}, \quad k = 1, \dots, b, \quad (2.24)$$

is the hyperparameter vector for each element of θ . The subscript θ_{-k} is all the parameter elements outside current θ_k . Then, by applying Equations (2.21) and (2.22), the nested approximation of

$$\tilde{Q}(x_i | \alpha) = \int \tilde{Q}(x_i | \theta, \alpha) \tilde{Q}(\theta | \alpha) d\theta \quad (2.25)$$

and

$$\tilde{Q}(\theta_k | \alpha) = \int \tilde{Q}(\theta | \alpha) d\theta_{-k} \quad (2.26)$$

can be constructed and computed. Hence, the computational approximation of $Q(\theta | \alpha)$ and $Q(x_i | \theta, \alpha)$ in association with numerical integration methods to integrate out θ (the hyperparameter) is the key to approximating $Q(x_i | \alpha)$. Based on the Laplace approximation, INLA approach exploits the assumptions of the model to produce the posterior of interest using numerical approximation (Tierney and Kadane, 1986). This approximation is in three categories, and they are:

- a) Compute an approximation to the hyperparameters joint posterior for the components of vector of parameters $Q(\theta|\alpha)$ as

$$Q(\theta|\alpha) \propto \frac{Q(x|\theta, \alpha)}{Q(x|\theta, \alpha)} \Big|_{x=x^*(\theta)} \approx \frac{Q(x|\theta, \alpha)}{\tilde{Q}G(x|\theta, \alpha)} \Big|_{x=x^*(\theta)} =: \tilde{Q}(\theta|\alpha), \quad (2.27)$$

where the Gaussian approximation is $\tilde{Q}G(x|\theta, \alpha)$ to the complete conditional of x based on the Gaussian distribution, and for a given θ , x^* is the mode of the complete conditional of x .

- b) There is need for a good approximation to the conditional distribution $Q(x_i|\theta, \alpha)$. Rue et al. (2009) has developed an approximation that is better than the Gaussian approximation by rewriting $\theta = (\theta_k, \theta_{-k})$ then again utilized the Laplace approximation to obtain

$$\tilde{Q}LA(x_i|\theta, \alpha) \propto \frac{Q(x, \theta, y)}{\tilde{Q}GG(x_{-i}|x_i, \theta, \alpha)} \Big|_{x_{-i}=x_{-i}^*(x_i, \theta)}, \quad (2.28)$$

where the Gaussian approximation to the conditional distribution of $x_{-i}|x_i, \theta, \alpha$ is $\tilde{Q}GG(x_{-i}|x_i, \theta, \alpha)$ and for a given θ , the mode $x_{-i}^*(x_i, \theta)$ is where the overall distribution is centered on. $\tilde{Q}LA(x_i|\theta, \alpha)$ is the posterior distribution of Laplace approximation.

- c) The marginal posterior distributions for Equation (2.23) can be computed by numerical integration on the condition that $\tilde{Q}(\theta|\alpha)$ and $\tilde{Q}LA(x_i|\theta, \alpha)$ are obtained as in

$$\tilde{Q}(x_i|\alpha) \approx \sum_{m=1}^n \tilde{Q}(x_i|\theta_m \alpha) \tilde{Q}(\theta_m|\alpha) \Delta_m, \quad (2.29)$$

where θ_m represents a set of grid points corresponding to the set of weights Δ_m . Further, in Equation (2.26) the computation of $\tilde{Q}(\theta_k|\alpha)$ can be obtained by integrating out θ_{-k} from the approximation $\tilde{Q}(\theta|\alpha)$.

Chapter 3

Spatial variation and risk factors of malaria and anaemia among children aged 0 to 59 months: A cross-sectional study of 2010 and 2015 datasets.

3.1 Introduction

Malaria is an acute febrile illness caused by *Plasmodium parasites*, which are transmitted to humans via the bites of infected female *Anopheles mosquitoes* (WHO, 2021b). Nearly half of the world's population was at risk of malaria in 2020, where children under the age of 5 years were mostly affected in sub-Sahara Africa both in terms of mortality and morbidity (WHO, 2021b, Ugwu and Zewotir, 2018; Weiss et al., 2019; Gething et al., 2016). This has remained a major public health problem notwithstanding it could be treated, prevented, and cured (NMC report, 2015). Based on the latest projections from the United Nations data, Nigeria, as one of the countries in Western Africa, has a population of approximately 208 million (Olanrewaju et al., 2020). By this projection, Nigeria is termed the largest community in Africa. Africa bears over 80% of global malaria burden, and Nigeria has about 29% of this burden. Furthermore, Nigeria together with the Democratic Republic of Congo contributes approximately 40% of global burden (WHO, 2015a). In Nigeria, malaria is accountable for up to 60% of outpatients' visits and 30% of admissions. Also, malaria is believed to contribute to about 11% of maternal death, 25% of infant death, and 30% to under 5-year age mortality (NMC report, 2015). Therefore, the country contributes a big proportion of deaths of mostly children and pregnant women every year in the region (Adebayo et al., 2016). According to a recent report from WHO, Nigeria has 1.3 million cases of malaria out

of 3.5 million cases recorded worldwide by the highest burden countries (Chukwuekezie et al., 2020). Though WHO suggested that for global burden of disease to lessen, there is an urgent need for a speedy reduction in the incidence of the disease in high-burdened countries which unfortunately includes Nigeria which is still lagging (Chukwuekezie et al., 2020). Also, anaemia is another global public health problem that mostly affects children under 5 years of age and pregnant women (WHO, 2016). It is a condition in which the hemoglobin (Hb) in the blood is lesser than needed by the body for its optimal physiological functions (Roberts et al., 2020; Kinyoki et al., 2021). Anaemia is one of the complications seen in malaria infection, which contributes to the resulting morbidity and mortality. According to World Health Organization (WHO) report, between 2015 and 2018, anaemia prevalence in children under the age of 5 years with a positive rapid diagnostic test (RDT) was twice that of children with a negative RDT in 21 African countries with moderate to high malaria prevalence (WHO, 2020). Children who tested positive for malaria had 9% severe anaemia and 54% moderate anaemia. On the other hand, children who tested negative for malaria had only 1% severe anaemia and 31% moderate anaemia (WHO, 2020). Nigeria has been confirmed to be among the countries in Africa with the highest anaemia prevalence of under aged 5 years children (WHO, 2015b, Aregbeshola et al., 2021). The risk factors of anaemia in children are multifactorial and interrelated in a complex way (WHO, 2014). In high-income countries, iron deficiency is known to be the major reason for anaemia unlike in low and middle-income countries (LMIC) where another factor like malaria is a major contributor to childhood anaemia especially in Sub-Sahara Africa (SSA) (Roberts et al., 2020; White, 2018).

The success of any healthcare program intervention is determined by a comprehensive and adequate understanding of the multifaceted risks of the occurrence of diseases and death (Adebayo

et al., 2016). Until recently, data and information on malaria, anaemia and other childhood diseases in Nigeria has been from clinics and hospitals. But then, these data and information are just a small portion of the total cases because it has been proven that a large percent of caregivers are conversant with the symptoms of fever associated with malaria, as such, about 80% of fevers are taken care of at home (Adebayo et al., 2016; Akogun, 2008; Kofoed et al., 2004; Kazembe and Namangale, 2007). Thus, data from the hospital may not be sufficient for estimating the prevalence and risk factors of malaria and anaemia for appropriate program development (Battle et al., 2019). As a replacement, the cross-sectional data from the Malaria Indicator Survey (MIS) was collected to gather information on malaria and anaemia in children under the age of 5 years.

The Nigerian government via the National Control Program, in line with many other non-governmental bodies such as Roll Back Malaria (RBM) through the implementation of (2009–2013) malaria control strategic plan has continued to make obvious efforts in reducing the spread of malaria and associated child death. Also, malaria awareness programs through mass distribution of long-lasting insecticide-impregnated nets (LLINs) within the selected states of the country has been put in place. As a result from 2010 to 2015, there was a reduction in malaria prevalence from 52 to 45% through their effort (NMC report, 2015). Despite these intervention programs, there has been a continuous record of morbidity and mortality from malaria, anaemia and other infectious diseases for children under the age of 5 years in most developing countries (Ugwu and Zewotir, 2018; Gayawan et al., 2014). Therefore, preventing fatal outcomes in malaria cases requires identifying infection, accurate laboratory diagnosis, and prompt therapy (Kain et al., 1998; Akinbo et al., 2009). Geoaddivitive latent variable models for binary/ordinal indicators was adopted to analyze the influence of variables of different types on the morbidity among young children in Nigeria (Adebayo et al., 2016). Nevertheless, epidemiology studies focus on analyzing the

geographic variation for the severity of a disease. A Multilevel logit model accounting for sampling design were adopted to assess individual, household and community factors associated with malaria infection (Oguoma et al., 2020). On the other hand, Generalized Additive Model (GAM) has been applied to identify the important risk factors that influence the prevalence of childhood malaria in Nigeria (Ugwu and Zewotir, 2020a). A multivariable hierarchical Bayesian geospatial model with the inclusion of a spatial effect was used by Roberts et al., (2020) to investigate the spatial variation and risk factors of childhood anaemia in four sub-Saharan African countries. While Ugwu and Zewotir, (2020b) adopted a binary structured additive regression (BSAR) regression that incorporates a Markov Random Field (MRF) prior, with posterior parameters estimated via Bayesian framework to examine the spatial heterogeneity and determinants of childhood anaemia in Nigeria. Studies on the risk factors and relationship between these diseases is scanty, also, no study has compared the differences in their spread for two separate years. Therefore, this study is aimed at investigating the spatial pattern of malaria as well as anaemia and adjusting for risk factors associated with each disease using a separate multivariate hierarchical Bayesian logistic model for each disease.

3.2 Data and method

3.2.1 Studied data

This study used the 2010 and 2015 data collected in the Malaria Indicator Survey (MIS) carried out in Nigeria. In both years, the sampling frame was obtained from the 2006 Population and Housing Census of the Federal Republic of Nigeria that was conducted by the National Population Commission (NPC report, 2015, NNPC et al., 2012). Nigeria as a nation is divided into 37 states administratively including the capital territory and each state is divided into local government areas

(LGAs) then each LGA is further divided into localities. For convenience, each locality was subdivided into census enumeration areas (EAs). These EAs from the 2006 EA census frame were used to define the cluster (i.e. primary sampling unit (PSU)) (NMC report, 2015).

A two-stage probability sampling was assumed in 2010 and 2015. While in 2010, the two-stage cluster design has a total number of 240 clusters, 83 in the urban areas and 157 clusters in the rural areas. In the end, 239 clusters were used due to intercommunal uproar in one of the clusters. Approximately, a representative sample of 6000 households was selected for the survey, with a minimum target of 920 completed individual women's interviews per zone. This is for the first stage. In the second stage, by equal probability systematic sampling, an average of 26 households were selected in each cluster. All women from age 15–49 were interviewed, also, children from 6 to 59 months were tested for malaria and anaemia (NNPC et al., 2012). On the other, in 2015, a two-stage sampling was also carried out. In the first stage, 9 clusters (EAs) were selected from each state and the Federal Capital Territory (FCT). Each state was represented in the sample with a total of 333 clusters around the country, 138 in urban areas and 198 in rural areas. From each cluster, 25 households were selected in the second stage by equal probability systematic sampling. At the same time, all women aged 15–49 were interviewed, and all children aged 6–59 months were tested for malaria and anaemia (NMC report, 2015).

Furthermore, a comprehensive inventory of households was carried out in both years. In 2010, the mapping exercise was done from August to September while in 2015 it was carried out from June to July (NMC report, 2015, NPC, 2013). In addition, global positioning system (GPS) receivers were used by NPC listing enumerators to record the coordinates of the 2010 and 2015 NMIS sample clusters, respectively.

3.2.2 Dependent variables

The dependent variables used in this study are the malaria status (presence or absence) and anaemia status (presences of absences) variables. Malaria is a disease spread to a human through the bite of infected anopheles' mosquito. The presence of malaria antigens discharged from the parasitized red blood cells is detected by the malaria diagnostic test which is a form of immunochromatography test. Both rapid diagnostic and microscopy testing have been approved by World Health Organization (WHO) as procedures for malaria diagnosis. Even though microscopy is recognized as the standard approach for malaria diagnosis, the application is demanding. While microscopy requires an experienced microscopist, a good environment, time etc., RDTs do not need skilled personnel, specialized equipment, and long process (Ugwu and Zewotir, 2018).

Anaemia is a condition resulting from the decrease or dysfunctional red blood cells in the body. Iron deficiency is known as a common cause of anaemia but in developing countries, malaria as one of the infectious diseases is largely attributed to anaemia disease (NMC report, 2015). As recommended by WHO, children from age 6 to 59 months are said to be anaemic if the Hb concentration level is below 11.0 g/dl: those within age 5 to 11 years are anaemic if Hb level is below 11.5 g/dl and children from age 12 to 14 years are considered anaemic if Hb is below 12.0 g/dl (WHO, 2008). The cause of anaemia is partly dependent on the part of the world a child lives.

Timely diagnosis and immediate treatment have been advised by WHO as major strategies in managing malaria and anaemia and in reduction of high mortality in most prevalent regions. Considering the strong correlation between malaria infection and anaemia, both microscopy and Rapid Diagnostic Tests were approved for the diagnosis of the two diseases in field surveys (Adigun et al., 2015). In this work, the upshot of interest was basically on the result of malaria

rapid test and anaemic status as binary indicators of the presence of malaria and anaemia in a child's blood sample respectively, where 1 denotes the presence of malaria or anaemia and 0 otherwise. The yearly distribution of malaria and anaemia during the last two weeks before the interview is as follows; in 2010, 5056 and 5147 were tested for anaemia and malaria, where for anaemia, 3512 children tested negative and 1544 tested positive while for malaria, 2719 tested negative and 2428 tested positive. While in 2015, 6021 and 6025 were tested for anaemia and malaria respectively, this has in the record that 4062 tested negative and 1959 tested positive for anaemia, on the other hand, 3399 tested negative and 2626 tested positive for malaria.

Figures 3.1 and 3.2 present the map of Nigeria with its 37 states including the capital territory.

Each of the maps displays the prevalence of malaria and anaemia in 2010 and 2015, respectively.

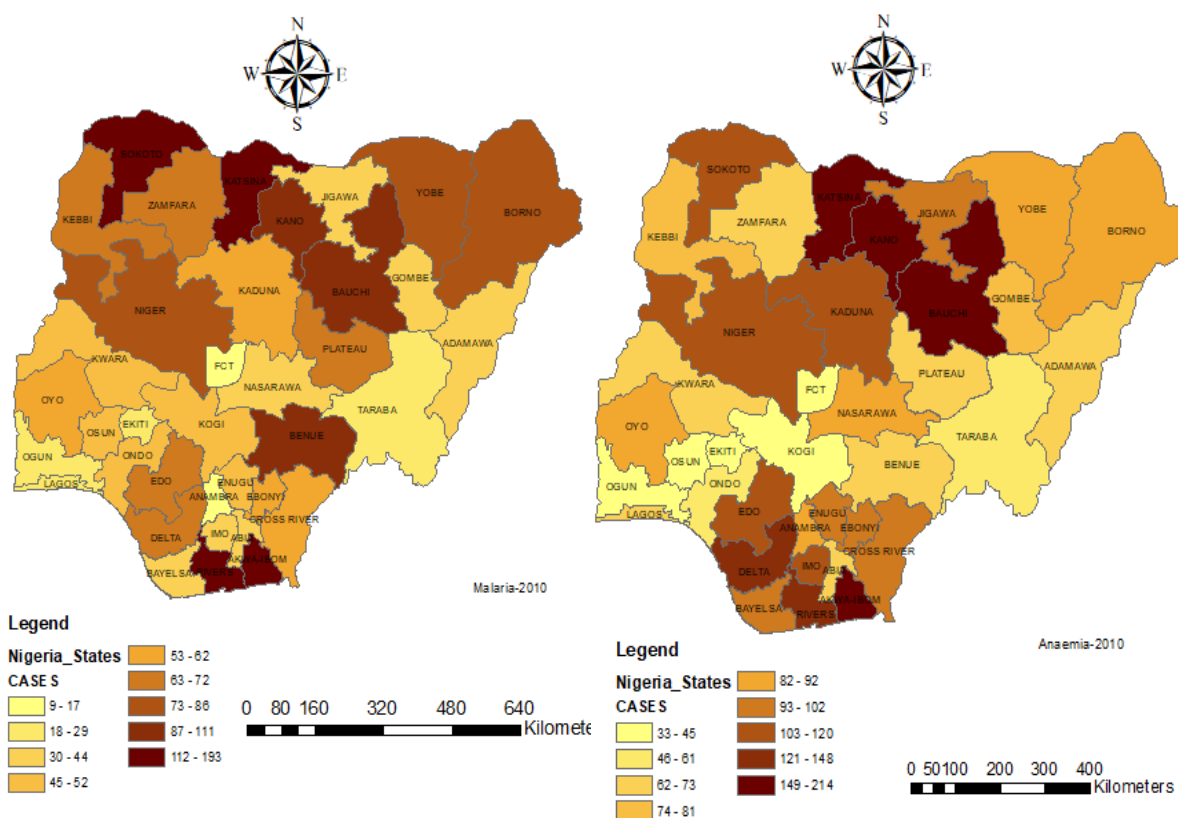


Figure 3.1: The 37 states including the capital territory of Nigeria showing the 2010 prevalence of malaria and anaemia of under- 5 years children

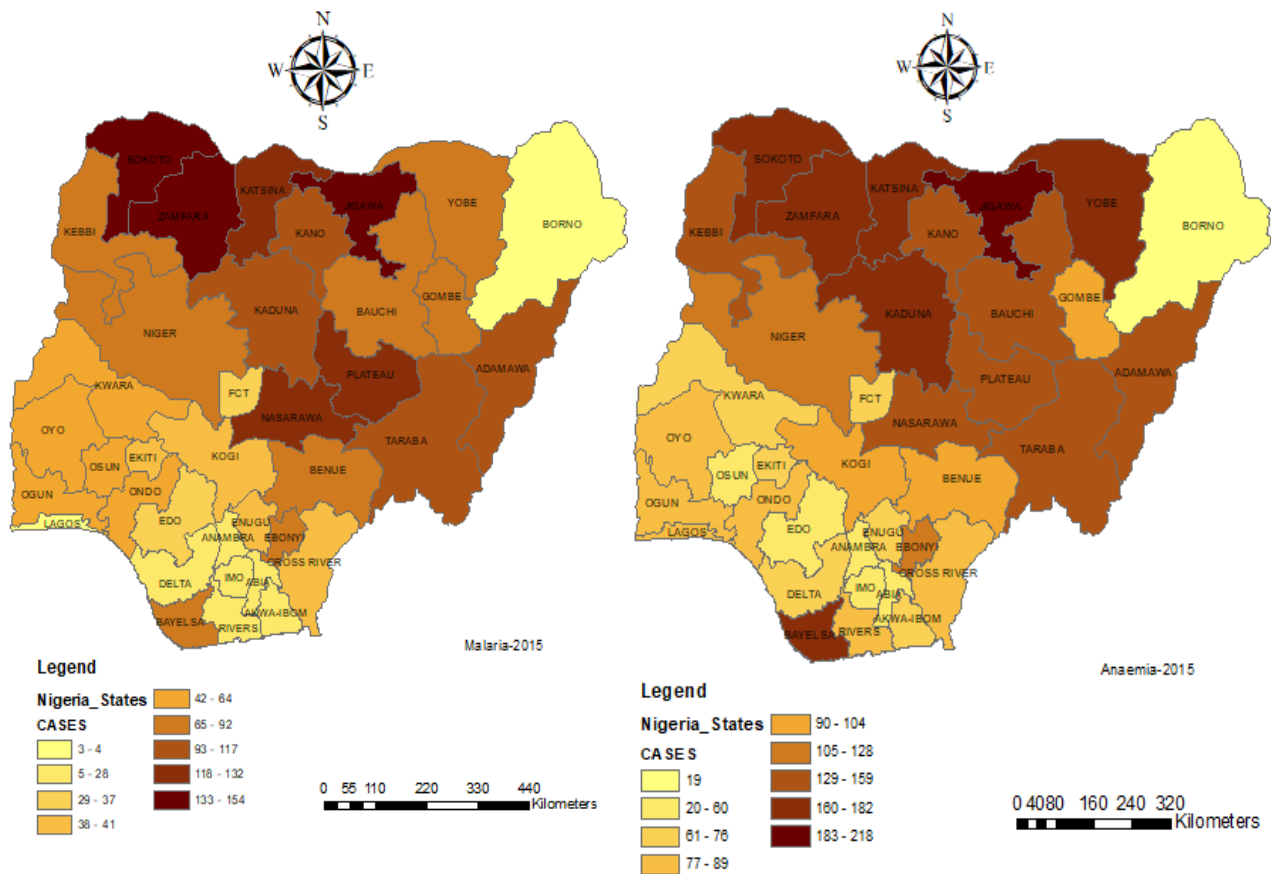


Figure 3.2: The 37 states including the capital territory of Nigeria showing the 2015 prevalence of malaria and anaemia of under- 5 years children.

3.2.3 Independent variables

In this study, the independent variables considered included some demographic, socio-economic and geographical variables which were based on the previous studies (Ugwu and Zewotir, 2018; Roberts et al., 2020; Ugwu and Zewotir, 2020c). These variables comprise of demographic variables (type of place of residence, child's age in months), socio-economic variables (mother's highest educational level, sex of the child, own a radio, own a television, wealth index, source of drinking water, type of toilet facility, presence of electricity, main floor material, main wall

material, main roof material) and geographical variables (state and region). The selected variables in Table 3.1 were based on DHS and MIS data sets as well as relevant literature.

Table 3.1: Selected variable (risk factors) of malaria and anaemia among children aged 0 to 59 months.

Demographic factors	Socio-economic factors	Geographical factors
Type of place of residence	Wealth index	State
Sex of the child	Mother's highest educational level	
Child's age in months	Has radio	
	Has Television	
	Source of drinking water	
	Type of toilet facility	
	Main roof material	
	Main floor material	
	Main wall material	
	Has electricity	

Table 3.2 presents the 2010 and 2015 percentage of children examined for malaria and anaemia. A respective individual data records of malaria and anaemia were constructed two weeks before the interview for 5147 and 5056 children, respectively within the age 0 to 5 years old in 2010 to ascertain those with fever. The same was done in 2015 with 6025 children for malaria and 6021 children for anaemia. In child's age in months, group 6 (age 51-59 months) had the highest record of the two diseases in both years and it was more in male children (2010: anaemia - 50.5%, malaria – 50.6%; 2015: anaemia – 50.4%, malaria – 50.4%). Most of these children lived in rural areas where there is no electricity, no television, no good toilet facilities, and the source of drinking water is not healthy. Regarding region, in both years, Northwest had the highest record of the two diseases with majority of illiterate mothers (2010: anaemia – 45.6%, malaria – 46.1%. 2015: anaemia – 43.5% and malaria – 43.5%) and low standard of living, with also poor building materials.

Table 3.2: Cross tabulation of proportions of malaria and anaemia prevalence of children under the age of 5 years for 2010 and 2015 HDHS data with Chi-square p-values of association for each variable.

2010 and 2015 HDHS data with Chi-square p-values of association for each variable.									
Individual characteristics		2010				2015			
		Malaria(%)	p-value	Anaemia(%)	p-value	Malaria(%)	p-value	Anaemia(%)	p-value
Factors									
Child's age in months			<0.001		<0.001		<0.001		<0.001
1		18.1		18.1		16.2		16.2	
2		13.7		13.8		15.5		15.5	
3		17.6		17.5		16.5		16.5	
4		17.2		17.1		18.1		18.1	
5		16.5		16.7		16.2		16.2	
6		16.9		16.9		17.6		17.6	
Sex			0.001		0.001		0.063		0.001
Male		50.6		50.5		50.4		50.4	
Female		49.4		49.5		49.6		49.6	
Region			0.004		<0.001		<0.001		0.145
North Central		17.3		17.5		20		20	
North East		17.7		17.9		19.2		19.2	
North West		24.5		23.3		25		25	
South East		12.4		12.6		9.9		9.8	
South South		18.7		19.1		12.8		12.9	
South West		9.4		9.6		13.1		13.1	
Electricity			<0.001		<0.001		<0.001		<0.001
No		52.8		52.9		52.1		52.1	
Yes		47.2		47.1		47.9		47.9	
Wealth index			<0.001		<0.001		<0.001		<0.001
Poor		36.9		36.7		40.2		40.2	
Middle		22.6		22.5		20.9		20.9	
Rich		40.5		40.8		38.9		38.9	
Mother's educational level			<0.001		<0.001		<0.001		<0.001
No education		46.1		45.6		43.5		43.5	
Primary		20.7		20.8		18.3		18.3	
Secondary		28.4		28.7		30.4		30.4	
Higher		4.8		4.9		7.7		7.7	
Has radio			<0.001		<0.001		<0.001		<0.001
No		28.7		28.6		39.8		39.8	
Yes		71.3		71.4		60.2		60.2	
Type of place of residence			<0.001		<0.001		<0.001		<0.001
Urban		27.3		27.3		33.9		33.9	
Rural		72.7		72.7		66.1		66.1	
Has television(ref=No)		59		58.9		54.2		54.2	

Yes	41	41.1	45.8	45.8
Source of drinking water	<0.001	0.324	0.064	<0.001
Tap/Other water	38.5	38.6	39.9	40.9
Well water	61.5	61.4	61.1	59.1
Type of toilet facility	<0.001	<0.001	<0.001	<0.001
Flush/other toilet	45.6	46.3	50	50
Pit toilet	54.4	53.7	50	50
Main roof material	<0.001	0.034	<0.001	<0.001
Wood/other	34.9	34.9	28.8	28.8
Zinc/Metal	65.1	65.1	71.2	71.2
Main floor material	<0.001	0.452	<0.001	<0.001
Earth/Other	49.2	48.9	44.2	44.2
Cement/Ceramics	50.8	51.1	55.8	55.8
Main wall material	<0.001	0.002	<0.001	<0.001
Wood/Other	56.2	55.9	39.5	39.5
Cement/Bricks	43.8	44.1	60.5	60.5

3.3 Methods

Here, we explained the Bayesian spatial models used to estimate the spread and the risk factors of malaria and anaemia in the 37 states including the Federal Capital in Nigeria.

3.3.1 Spatial model

Descriptive statistics approach was used to analyze the independent variables of the study sample in the form of descriptive table and simple percentages, while each of the dependent variables was described in the form of maps. For malaria or anaemia amongst children under the age of 5 years, their respective relationships with the independent variables were tested using the chi-square association analysis. Thereafter, a binary logistic regression was used to study the relationship between the independent variables and malaria or anaemia. A stepwise backward selection was done to pick the factors that have significant relationship with malaria or anaemia amongst children under 5 years. While for adjustment of clustering, common causes and sampling weights were

carried out using the already weighting factors constructed by measure DHS to compensate for unequal probability of selection, non-response and to make the key variables of interest conform to a known population distribution.

For the spatial study of malaria and anemia data, let y_{ijk} be a binary for disease k (malaria or anaemic status) of child $j, j = 1, 2, \dots, n_i$ where n_i is the number of children in state $i, i = 1, 2, \dots, 37$ that have malaria or anaemia. Then the binary response follows as:

$$y_{ijk} = \begin{cases} 1, & \text{If child } j \text{ in state } i \text{ has malaria (k=1) or is anaemic (k=2)} \\ 0, & \text{Otherwise.} \end{cases}$$

The y_{ijk} is assumed to follow a Bernoulli distribution with likelihood function defined as

$$y_{ijk} \sim \text{Bernoulli}(\theta_{ijk}), i = 1, \dots, 37, j = 1, \dots, n$$

where $\theta_{ijk} = P(y_{ijk} = 1)$ are unknown probabilities and $E(y_{ijk}) = \theta_{ijk}$ is related to predictor through a link function

$$\text{logit}(\theta_{ijk}) = \log\left(\frac{P(y_{ijk}=1)}{1-P(y_{ijk}=1)}\right) = \eta_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} \quad (3.1)$$

the vector $\mathbf{x}_{ijk} = (1, x_{ijk1}, \dots, x_{ijkp})'$ are categorical and continuous variables and the vector of regression coefficients is $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$. This model accepts only a parametric form of categorical variables. To account for more flexible approach, the linear predictor η_{ijk} of Equation (3.1) will be extended. Therefore, the model complexity is increased by including different forms of variables. The logistic regression model is then extended to give room for area-specific random effects by substituting the linear predictor η_{ik} in Equation (3.1) with a geoadditive predictor. These

random effects are put in the model to take care of extra variation. Consequently, to incorporate unobserved influential factors that change across the states, the structured random effects is accounted for by the model and it is defined as:

$$\eta_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + f_{str}(s_i) \quad (3.2)$$

On a general note, the spatial effects of an areal unit can be modelled using CAR model. Particularly in the second stage of hierarchical models, they are used to specify some classes of the model. In this work, CAR model is expressed as follows. Let $u = (u_1, \dots, u_n)$ be the vector of univariate random variables in relation with the observed spatial unit under study and represent $\{\partial(i): i = 1, \dots, n\}$ as the states sharing the same border with state i which means, for any j , $i = 1, \dots, n$, $i \in \partial(i)$ if only $j \in \partial(i)$ and $j \notin \partial(j)$ must be satisfied.

Therefore, suppose the conditional density of $u_i, j = 1, \dots, n$, follows the conditional normal variable defined as;

$$u_j/u_i, (j \neq i) \sim N(\mu_j + \sum_{i \in \delta(j)} c_{ij}(u_i - \mu_i), d_j^2), \quad j, i = 1, \dots, n \quad (3.3)$$

where u_i is the mean for state i , μ_j is the spatial trend at location j and $d_j^2 = \rho_u^2/\partial(j)$ is the conditional variance of the i th state, which depends on the number of neighbours. Therefore, the size of the variance for the current state is determined by the number of state neighbours. c_{ij} is the spatial dependence parameters for $j = 1, \dots, n$, such that $c_{jj} = 0$ for all i 's while ρ_u^2 denotes the variance parameter that controls the differences between spatial similarity. Particularly, the quantity c_{ij} captures spatial dependency. The matrix form of (3.3) is given by Cressie, (1993) as

$$u \sim N(\mu, B^{-1}L), \quad (3.4)$$

which denotes the joint distribution of u , where $B = (I - C)$, with $C = [c_{ij}]_{n \times n}$, $\mu = (\mu_1, \dots, \mu_n)$ and $L = \text{diag}(d_1^2, \dots, d_n^2)$ is an $n \times n$ diagonal matrix. (3.4) is correct if B can be inverted and $B^{-1}L$ is symmetric and the conditional constraints $c_{ij}d_j^2 = c_{ji}d_i^2$ for all $j \neq i$, and must be positive definite.

The elements of invertible matrix B is expressed as;

$$b_{(ij)} = \begin{cases} 1 & \text{for } j = i \\ -c_{ij} & \text{for } i \in \partial(j), \\ 0, & \text{otherwise.} \end{cases} \quad (3.5)$$

To get a valid joint distribution, the covariance matrix in Equation (3.4) must be symmetric and positive definite as mentioned above. Then the symmetric weighted adjacency matrix will be $W = (W_{ij})$, and set $c_{ij} = \emptyset W_{ij}$ where

$$W_{ij} = \begin{cases} 1 & \text{if } j \text{ and } i \text{ share the same border} \\ 0 & \text{otherwise,} \end{cases} \quad (3.6)$$

and the properness of the distribution is controlled by the parameter \emptyset .

Also, to account for unobserved heterogeneity within each state, the unstructured random effects is considered, and the model is expressed as:

$$\eta_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + f_{unstr}(s_i). \quad (3.7)$$

Besag, York and Mollie (BYM) model proposed by Besag et al., (1991) is the most popularly used tool under the spatial Bayesian hierarchical models for disease mapping. The BYM comprises of two random components, i.e., spatially structured u and spatially unstructured v components, which are included into the log-linear model for relative risk. By including of these random effects, the smoothing of the relative risk at state level is guaranteed. While spatially structured component u is correlated to neighbouring spatial units, the spatially unstructured component v is the

uncorrelated extra variation. Also, note that the vectors u and v have individual unit random effects $u_i (i = 1, \dots, n)$ and $v_i (i = 1, \dots, n)$ respectively.

Therefore, (3.3) is extended to convolution model by adding both structured and unstructured random effects as follows:

$$\eta_{ij} = x'_{ij}\beta + f_{str}(s_i) + f_{unstr}(s_i). \quad (3.8)$$

In (3.3), (3.4) and Equation (3.2), \mathbf{x}' is a k -dimensional row-vector of covariates with β as the corresponding vector of regression coefficients. u_j is the spatially structured random effect (correlated heterogeneity) and v_j is the spatially unstructured effect (uncorrelated heterogeneity).

On assumption, Besag et al., (1991) stated the that two random effects are independent and need a specification of independent priors. For the spatially unstructured v_j , the prior's distribution model is assumed to follow a normal distribution with a vector of mean $\mathbf{0}$ and a variance -covariance matrix $\sigma^2 \mathbf{I}$, where \mathbf{I} is the identity matrix and $\sigma^2 > 0$ is unknown. Following the argument of Besag et al., (1991), the prior of the spatial component is assumed to be represented by a Markov Gaussian field or conditional Gaussian autoregressive model. Therefore, excluding the i th state, let u_{-i} denote the vector of effects without i th state effect as μ_i/μ_{-i} . Then we assume that

$$u_i | u_{-i} \sim N\left(\frac{1}{n_i} \sum_{e \sim i} u_e, \frac{\tau_u^2}{n_i}\right), \quad (3.9)$$

where n_i is the number of neighbourhoods of state i , while $e \sim i$ are all units e neighbourhoods of state i and τ_u is the standard deviation parameter. Finally, the inverse gamma hyperpriors is assumed for the variance of the normal priors.

The posterior distributions of the parameters was estimated using Integrated Nested Laplace Approximation (INLA) (Chopin, 2009) in R. This is because it is a better approach compared to Markov Chain Monte Carlo sampling and approximate Bayesian inference (Martins et al., 2013). The Deviance Information Criteria (DIC) is calculated as; $DIC = \bar{D}(\theta) + Dp$, where \bar{D} is the posterior mean of the deviance that measures the goodness of fit while Dp is the effective number of parameters in the model. Based on the DIC , the final model was selected and the model with the smallest DIC was taken as a better fit (Spiegelhalter et al., 2002).

3.4 Results

Continuous variables with non-linear effect were studied; nonetheless, age in months was the only variable that showed a significant non-linear effect on the log-odds of a child's Malaria result and anaemic status as shown in Figures 3.3 and 3.4.

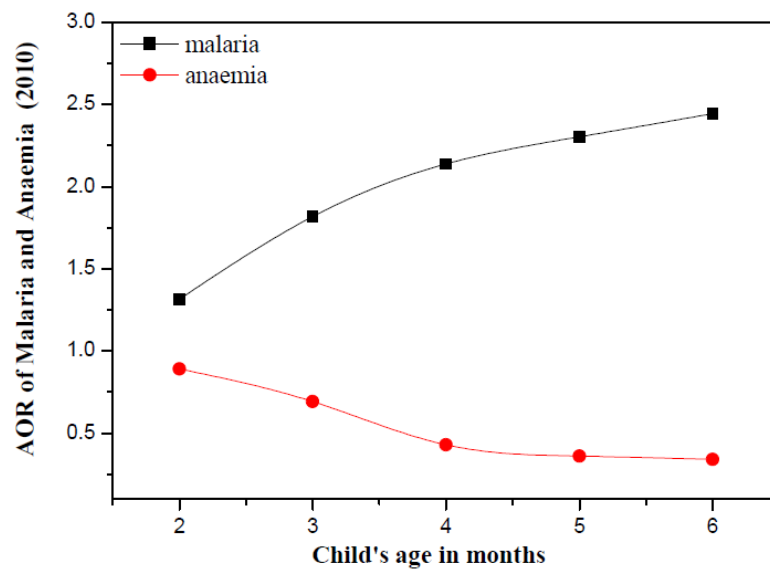


Figure 3.3: The nonlinearity relationship between child's age in months and AOR of malaria and anaemia (2010)

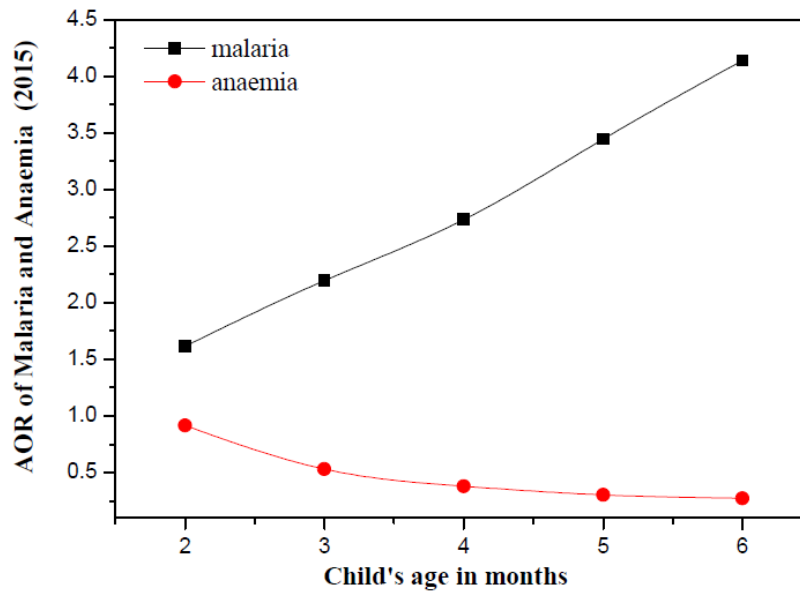


Figure 3.4: The nonlinearity relationship between child's age in months and AOR of malaria and anaemia (2015)

Therefore, this is the only non-linear effect incorporated in the fitted model, while the other independent variables were added as a linear fixed effect.

Table 3.3 contains the results of the DIC and effective number of parameters D_p of 2010 and 2015 for each of the fitted models of malaria and anaemia. For 2010, the results of malaria are based on model 4 because it gave the least DIC, and this model contains both the structured and unstructured spatial effects. On the other hand, the results of anaemia are based on model 2 because it gave the least DIC. In 2015, the results of malaria were based on model 3 because it gave the least DIC and the results of anaemia were based on model 4 (i.e., convolution model that comprises structured and unstructured spatial effect) because it gave the least DIC. Note that in both 2010 and 2015, there was not much difference in the estimates of the fixed effects in the four models; nevertheless, the variables differed significantly.

Table 3.3: Model comparisons for malaria and anaemia

	2010				2015			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
DIC(malaria)	5358.5	5014.26	5014.26	5014.01	5774.77	5773.55	5770.02	5770.23
DIC(anaemia)	4725.58	4542.03	4543.82	4542.15	5801.67	5793.63	5794.57	5793.43
Dp(malaria)	26.91	54.26	53.2	54.01	57.46	52.68	51	51.43
Dp(anaemia)	26.89	53.68	53.78	53.72	57.52	49.35	47.8	49.13

Table 3.4 presents the adjusted posterior odds ratio estimates (AOR) and 95% credible interval for the random effects included in the Bayesian hierarchical logistic regression model. In 2010, there was a significant increase in the odds of malaria for children in age groups 2(15 to 23 months), 3(24 to 32 months), 4(33 to 41 months), 5(42 to 50 months) and 6(51 to 59 months) relative to children in age group 1(6 to 14 months). There is a significant increase in the odds of malaria among children who have anaemia or reside in rural areas (Ugwu and Zewotir, 2018). On the other hand, the household that had electricity had significantly lower odds of malaria compared to a household with no electricity. In the same vein, the odds of malaria decrease significantly among those who use well water relative to those who use tap/other and those who use pit toilet relative to those who use flush/other toilet. Furthermore, lower odds of malaria were suggested for mothers with higher educational level (Secondary and Higher education) and wealth index (Rich). Female children had an increased odds of malaria than male children; however, these odds were not significant; the same applies to main wall material and households that had a radio. With respect to main roofing material, main floor material and households that had a television, there was no significant decrease in odds of malaria. While in 2015, the odds of malaria for children in age groups 2(15 to 23 months), 3(24 to 32 months), 4(33 to 41 months), 5(42 to 50 months) and 6(51 to 59 months) shows a significant increase relative to children in age group 1(6 to 14 months). Similarly, there is a significant increase in the odds of malaria among children who have anaemia

or live in rural areas with zinc/metal roof. On the other hand, the odds of malaria decrease significantly among those who use pit toilet relative to those who use flush/other toilet. Furthermore, the results suggested that children from mothers with higher educational level (Secondary and Higher education) and wealth index (Middle and Rich) have lower odds of malaria. Female children had higher odds of malaria than male children; however, these odds were not significant; the same applies to electricity, main floor material, main wall material and household that has a radio. There was a non-significant increase in odds of malaria with respect to source of drinking water and household that has television.

Table 3.4: 95% credible interval and adjusted posterior odd ratios estimates (AOR) of malaria

Variable	2010		2015	
	AOR	95% CI	AOR	95% CI
Intercept	0.468	(0.318, 0.685)	0.185	(0.133, 0.258)
Child's age in months(grouped)(ref=1)				
2	1.315*	(1.035, 1.670)	1.616*	(1.293, 2.020)
3	1.818*	(1.453, 2.275)	2.194*	(1.760, 2.738)
4	2.139*	(1.704, 2.688)	2.733*	(2.197, 3.405)
5	2.304*	(1.830, 2.905)	3.447*	(2.746, 4.334)
6	2.444*	(1.938, 3.086)	4.149*	(3.310, 5.211)
Has electricity(ref=No)				
Yes	0.773*	(0.631, 0.949)	0.773	(0.631, 1.064)
Source of drinking water(ref=Tap/Other water)				
Well water	0.810*	(0.696, 0.942)	1.122	(0.971, 1.295)
Type of toilet facility(ref=Flush/other toilet)				
Pit toilet	0.823*	(0.697, 0.973)	0.718*	(0.619, 0.833)
Anaemic status(ref=No)				
Yes	2.583*	(2.198, 3.039)	3.345*	(2.880, 3.890)
Wealth index(ref=Poor)				
Middle	0.887	(0.677, 1.163)	0.710*	(0.568, 0.887)
Rich	0.634*	(0.415, 0.967)	0.383*	(0.277, 0.529)
Sex(ref=Male)				
Female	1.093	(0.954, 1.252)	0.956	(0.842, 1.085)
Mother's educational level(ref=No education)				
Primary	1.062	(0.851, 1.324)	0.788*	(0.652, 0.953)
Secondary	0.684*	(0.538, 0.869)	0.790*	(0.646, 0.967)
Higher	0.529*	(0.355, 0.783)	0.389*	(0.268, 0.558)
Main roof material(ref=Wood/other)				
Zinc/Metal	0.88	(0.734, 1.055)	1.214*	(1.016, 1.450)
Main floor material(ref=Earth/Other)				
Concrete/Ceramics	0.86	(0.696, 1.063)	0.911	(0.769, 1.080)
Main wall material(ref=Wood/Other)				
Cement/Bricks	1.034	(0.808, 1.323)	0.968	(0.800, 1.171)
Has radio(ref=No)				
Yes	1.116	(0.948, 1.315)	0.902	(0.783, 1.039)
Type of place of residence(ref=Urban)				
Rural	1.348*	(1.117, 1.627)	1.889*	(1.568, 2.277)
Has television(ref=No)				
Yes	0.872	(0.698, 1.090)	1	(0.807, 1.240)

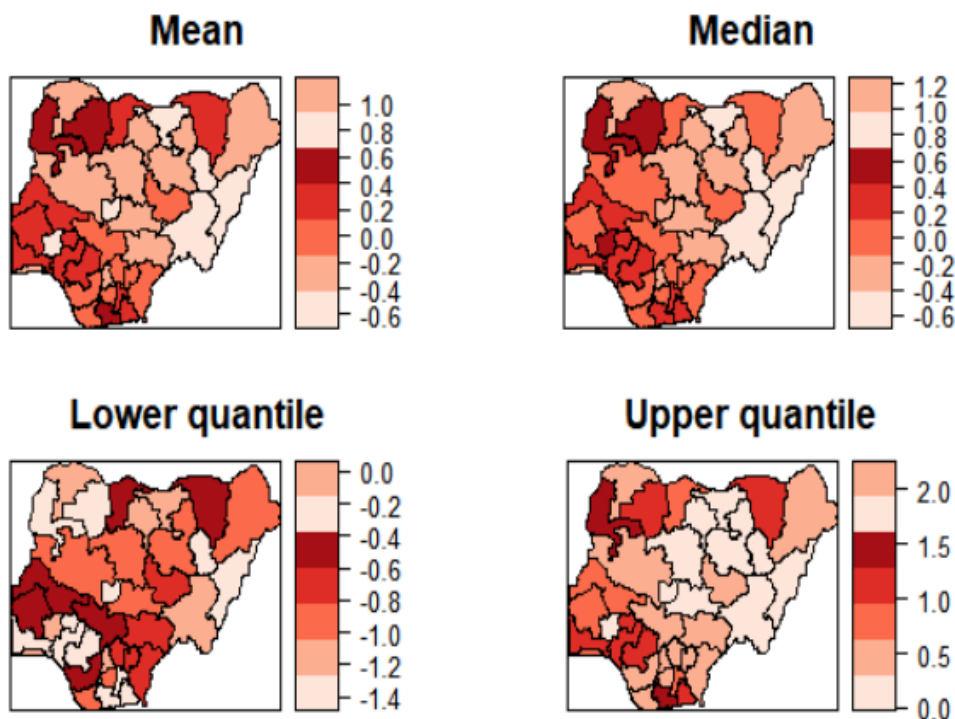


Figure 3.5: Nigeria state-level estimated posterior mean of the structured spatial effect maps: a.) Mean, b.) Median, c.) 25% quantile and 97.5% quantile for 2010

Figure 3.5 and Figure 3.6 show the estimated mean, median, 25% quantile and 95% quantile of the structured and unstructured spatial effects on the log-odds of malaria, respectively. Lower odds of malaria are associated to the eastern regions because they have a negative spatial effect, while higher odds of malaria are associated to the northern regions because they have a positive spatial effect. This could be attributed to several factors among them being that northern women are less educated and therefore are not well informed about this disease. Also, most northern women are rural dwellers and so lack the knowledge of malaria protection role of modern housing. In comparison, the structured spatial effect that ranged from -0.6 to 1.0 for mean, -0.6 to 1.2 for median, -1.4 to 0.0 for 25% quantile and 0.0 to 2.0 for 95% quantile is seen to be stronger than unstructured spatial effect that ranged from -0.8 to 0.8 for mean, -0.8 to 1.0 for median, -1.6 to 0.0 for 25% quantile and 0.0 to 1.8 for 95% quantile. From this, the structured spatially correlated

effect was stronger than the unstructured spatial effect, indicating that there is similarity in the effect a particular region has on the risk of malaria and her neighbouring regions. This shows that there is possibility of demographic, socio-economic, and geographical factors that go beyond boundaries of the regions playing a significant role in childhood malaria positivity. Also, the strong relationship between malaria and anaemia in Nigeria could be the result of the homogenous outcome of spatial effect on childhood malaria in the country, and this was explained by inclusion of child's anaemic status.

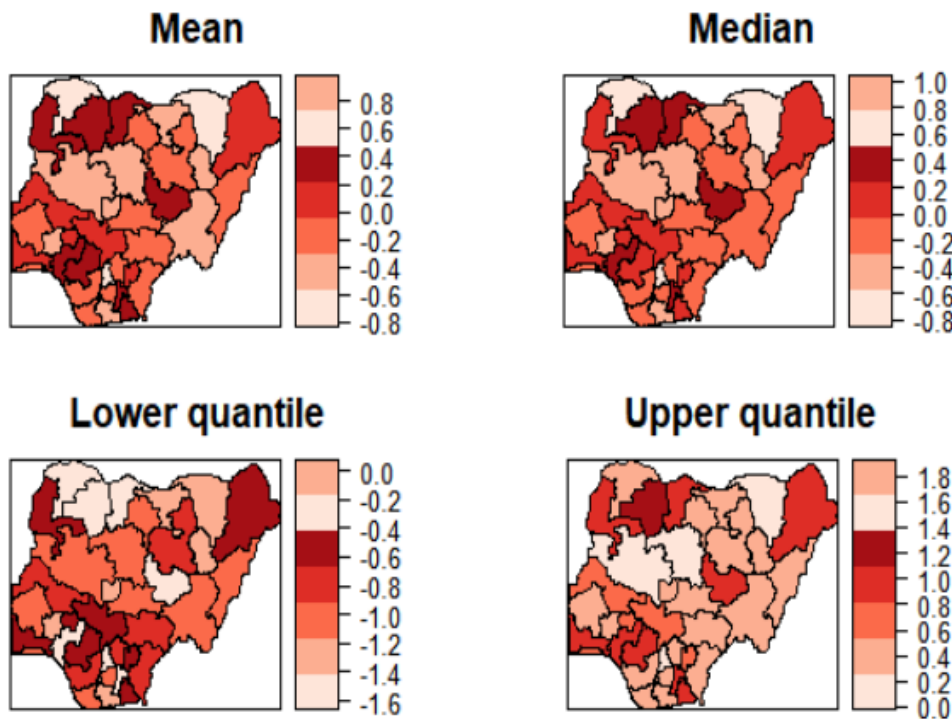


Figure 3.6: Nigeria state-level estimated posterior mean of the unstructured spatial effect maps: a.) Mean, b.) Median, c.) 25% quantile and 97.5% quantile for 2010

The adjusted posterior odds ratio estimates (AOR) and 95% credible interval for the random effects incorporated in the Bayesian hierarchical logistic regression model are presented in Table 3.5. In 2010, the odds of anaemia increase significantly for children in age groups 3(24 to 32 months), 4(33 to 41 months), 5(42 to 50 months) and 6(51 to 59 months) relative to children in age group

1(6 to 14 months). The odds of anaemia increase significantly among children who have malaria or live in rural areas. In the same vein, regarding wealth index, children from the middle and rich household had a significant increase in the odds of anaemia relative to children from a poor household. On the other hand, the household that have electricity had non significantly lesser odds of anaemia compared to a household with no electricity. There was a non-significant decrease in the odds of anaemia among those who use well water relative to those who use tap/other and those who use pit toilet relative to those who use flush/other toilet. Furthermore, mothers with higher educational level (Secondary and Higher education) had children with lower odds of anaemia but not significant. The odds of anaemia increased significantly for female children than male children. The same applies to households that had a radio and television. No significant decrease in the odds of anaemia was recorded with respect to main roof material, main floor material and main wall material.

Regarding 2015, the odds of anaemia decreased significantly for children in age groups 3(24 to 32 months), 4(33 to 41 months), 5(42 to 50 months) and 6(51 to 59 months) relative to children in age group 1(6 to 14 months). Also, the odds of anaemia among children who had malaria as well as living in rural areas and use well water and pit toilet increase significantly. On the other hand, regarding wealth index, electricity and concrete/ceramics floor, non-significant increase in the odds of anaemia was noted. The odds of anaemia among female children relative to male children and those who have television to those who do not have decreased non-significantly. With respect to main roof material, main floor material and main wall material and the household that have radio, the odds of anaemia did not decrease significantly.

Table 3.5: 95% credible interval and adjusted posterior odd ratios estimates (AOR) of anaemia

Variable	2010		2015	
	AOR	95% CI	AOR	95% CI
Intercept	4.184	(2.755, 6.396)	3.23	(2.338, 4.468)
Child's age in months(grouped)(ref=1)				
2	0.892	(0.682, 1.684)	0.914	(0.722, 1.158)
3	0.693*	(0.539, 0.890)	0.529*	(0.421, 0.664)
4	0.429*	(0.335, 0.549)	0.378*	(0.303, 0.470)
5	0.361*	(0.281, 0.463)	0.303*	(0.241, 0.380)
6	0.340*	(0.265, 0.435)	0.271*	(0.216, 0.340)
Has electricity(ref=No)				
Yes	0.98	(0.783, 1.227)	1.003	(0.825, 1.219)
Source of drinking water(ref=Tap/Other water)				
Well water	0.98	(0.832, 1.153)	1.205*	(1.049, 1.386)
Type of toilet facility(ref=Flush/other toilet)				
Pit toilet	0.961	(0.806, 1.145)	1.191*	(1.031, 1.377)
Result of malaria rapid test (ref=No)				
Yes	2.606*	(2.216, 3.068)	3.386*	(2.916, 3.936)
Wealth index(ref=Poor)				
Middle	1.465*	(1.093, 1.964)	1.069	(0.840, 1.361)
Rich	1.752*	(1.106, 2.778)	1.013	(0.728, 1.408)
Sex(ref=Male)				
Female	0.761*	(0.658, 0.877)	0.793*	(0.699, 0.901)
Mother's educational level(ref=No education)				
Primary	0.814	(0.642, 1.033)	0.877	(0.720, 1.068)
Secondary	0.87	(0.676, 1.119)	0.746*	(0.612, 0.910)
Higher	0.937	(0.633, 1.391)	0.578*	(0.436, 0.765)
Main roof material(ref=Wood/other)				
Zinc/Metal	0.904	(0.744, 1.099)	0.922	(0.769, 1.106)
Main floor material(ref=Earth/Other)				
Concrete/Ceramics	1.003	(0.802, 1.253)	1.105	(0.860, 1.198)
Main wall material(ref=Wood/Other)				
Cement/Bricks	0.881	(0.672, 1.152)	0.84	(0.685, 1.032)
Has radio(ref=No)				
Yes	0.659*	(0.548, 0.790)	0.877	(0.758, 1.014)
Type of place of residence(ref=Urban)				
Rural	1.455*	(1.201, 1.763)	1.440*	(1.205, 1.719)
Has television(ref=No)				
Yes	0.660*	(0.516, 0.836)	0.908*	(0.731, 1.128)

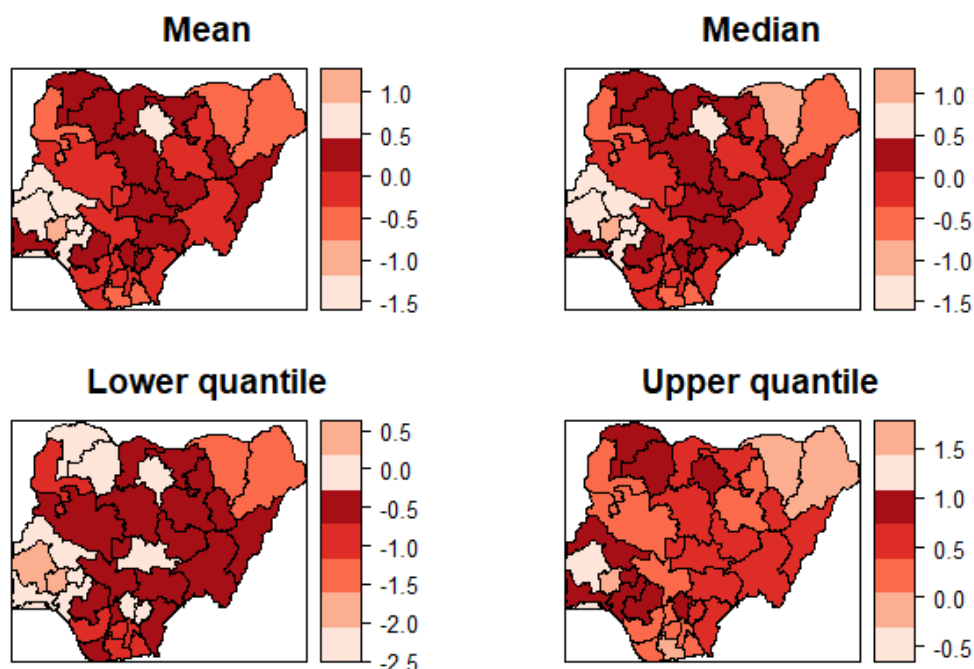


Figure 3.7: Nigeria state-level estimated posterior mean of the structured spatial effect maps: a.) Mean, b.) Median, c.) 25% quantile and 97.5% quantile for 2015

Figure 3.7 and Figure 3.8 show the estimated mean, median, 25% quantile and 95% quantile of the structured and unstructured spatial effects on the log-odds of anaemia, respectively. Lower odds of anaemia were attributed to the northeast regions because they have a negative spatial effect, while a higher odd of anaemia was ascribed to the core north regions because they have a positive spatial effect. This is because most northern women are less educated and therefore are not well informed about this disease. Also, most northern women are rural dwellers and so lack the knowledge of anaemia protection role of modern housing. In comparison, the structured spatial effect that ranged from -1.5 to 1.0 for the mean, -1.5 to 1.0 for the median, -2.5 to 0.5 for the 25% quantile and -0.5 to 1.5 for the 95% quantile is estimated to be stronger than unstructured spatial effect that ranged from -0.15 to 0.15 for the mean, -0.08 to 0.08 for the median, -0.8 to -0.2 for the 25% quantile and 0.1 to 0.8 for the 95% quantile. Like in the previous maps, there is similarity of factors that influences the risk of malaria which transcend boundaries among the neighbouring

regions. In addition, the strong correlation between malaria and anaemia may be because of the homogenous results of spatial effect on childhood anaemia. It was explained by adding a child's malaria status.

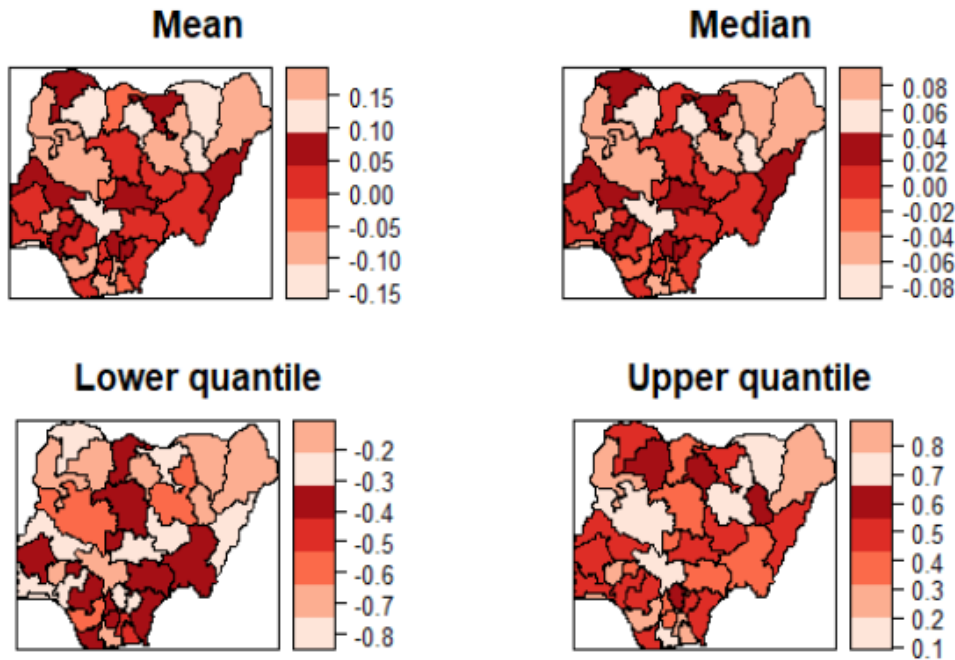


Figure 3.8: Nigeria state-level estimated posterior mean of the unstructured spatial effect maps: a.) Mean, b.) Median, c.) 25% quantile and 97.5% quantile for 2015

3.5: Non-linear effect of rural area on malaria and anaemia

Figure 3.9 shows the significance of non-linear effect of rural area on malaria and anaemia. It can be deduced from our results that the odds ratio increased for malaria in rural areas in both years but decreased slightly for anaemia. This is because as malaria infection increases, it get to a point where individuals develop a disease controlling immunity such that all those who have malaria are asymptomatic and this diminishes the prevalence of anaemia (White, 2018).

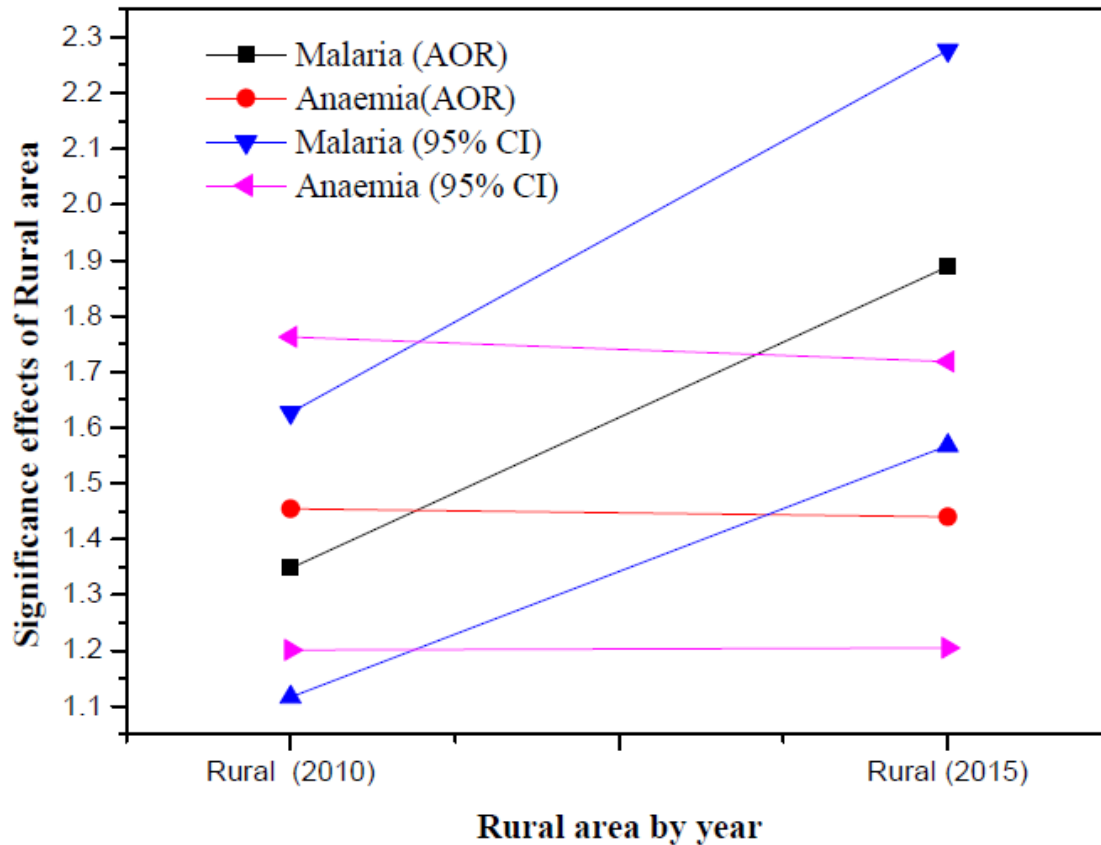


Figure 3.9: Estimated significance of non-linear effect of rural area, the posterior mean and the 95% credible intervals

3.6: Discussion

In this study, a hierarchical Bayesian logistics regression model was adopted to investigate the spatial variation, socio-economic, demographic, and geographical factors of malaria and anaemia in children under the age of 5 years in Nigeria using two datasets. The spatial effect maps were generated from the available data to identify the most endemic regions. The result from this study is in line with previous works (Ugwu and Zewotir, 2020b; Roberts et al., 2020; Gayawan et al., 2014; Magalhaes and Clements, 2011), that the risk of girls having malaria or anaemia is less than the risk of boys, and an increase in wealth index and mother's educational level decreases a child's risk. It can be inferred that the more educated an individual, the more aware they are, the more

understanding of health-related issues they have. Also, from the results, inefficient public health resource allocations and socio-economic conditions are strongly associated with geographical inequalities in health (Ugwu and Zewotir, 2020b). In the same vein, when individuals have high income, access to good health care and nutritional food sources the risk of the diseases is reduced. In both years, type of place of residence is significantly associated with the two diseases. Children in rural areas tend to have higher rates of the diseases than their contemporaries in Nigeria and Zambia (Gayawan et al., 2014; Riedel et al., 2010; Ugwu and Zewotir, 2020b). In addition, female children had lower risk of having anaemia in both years while for malaria, female children had non-significant higher odds in 2010 and lower odds in 2015 compared to male children. This could be attributed to the customary practice that male children are most preferred by their fathers in Nigeria especially in Igbo tribe than female children because they guarantee continuity of their family lineage while female children are married out of the family to start a new family (Milazzo, 2014). This could be the reason female children have the advantage of being breastfed for a longer period of time and also get better healthcare and nutrition, thereby improving their health status because they are left to be taken care of by their mothers (Aregbeshola et al., 2021). There is high possibility of denying the male children their mothers care which may result to blood loss from injury, etcetera and higher levels of parasitosis. This is in line with similar study by Gayawan et al., (2014); Pasricha et al., (2018a) and Ugwu and Zewotir, 2020b). As seen in Figure 3.6, there is a significant association between malaria and anaemia in the two years. This suggests that high rate of anaemia in this country is caused by malaria (Roberts et al., 2020; White, 2018). Children with a better source of drinking water are less affected by malaria and anaemia. It could be said that better source of drinking water ensures good health for children which increases their body immunity. In accordance, type of toilet facilities has significant association with malaria which in

turn causes anaemia. This is because when there is a poor toilet, there is all possibility for the facility to give a conducive environment for the breeding of mosquitoes which increases the risk of malaria parasites and then results in an increase anaemia cases (Roberts et al., 2020). In this study, wall, floor, and roof material were found not to be significantly associated with malaria or anaemia. But these factors have high correlation with malaria, for this reason, much of these factors on childhood anaemia will be accounted for if child's malaria status is included (Ugwu and Zewotir, 2018; Oguoma et al., 2020; Roberts et al., 2020). It can be seen from our results that the vulnerability of under-5 malaria infection increases with age while anaemia infection decreases with age. For malaria, the reason could be that children within 6 to 23 months are protected through maternal immunity (Adebayo et al., 2016; Ugwu and Zewotir, 2020b). On the other hand, for anaemia, a disease controlling immunity develops as the child grows, as such, nearly all malaria infections are asymptomatic by adolescent and adult, thus, there will be a decline in prevalence of anaemia (White, 2018).

The reason for considering more than one disease in two different years is to ascertain if both diseases are affected by the same demographical, socio-economic, or geographical factors. From this study, the effect of unstructured spatial correlation is seen to be moderately weak as contrasted to structured spatial effect, signifying that the prevalence of malaria and anaemia in each of the years is the same among neighbouring states. In 2010, malaria and anaemia diseases were denser in all Nigeria states than in 2015. Also, in 2015, there is an obvious increase in mother's educational level which is a major reason for this change.

Notwithstanding the differences in the two years considered, there were similarity as seen in Figures 3.1 and 3.2; both malaria and anaemia were predominant in North-West, North-Central and some part of South-East in the considered two datasets.

Chapter 4

Bayesian Spatio-temporal modelling and mapping of malaria and anaemia among children between 0 to 59 months in Nigeria

4.1 Introduction

Childhood malaria infection has been a major concern, especially in developing countries like Nigeria. A recent World Health Organization (WHO) report estimated 241 million malaria cases and 627000 deaths worldwide (WHO, 2021b; Steketee et al., 2021). Malaria infection rate dropped from 81% in 2000 to 59% in 2015 and 56% in 2016 but went up again to 59% in 2020 due to Covid-19 pandemic. Globally, 96% of malaria cases is majorly from 29 countries with Nigeria (27%) topping the list of 6 countries that contribute almost 55% of malaria cases (Organization, 2021b, Rahi et al., 2021). Children between the age of 0 to 59 months are most vulnerable with an estimate of 213 million to 228 million malaria cases between 2019 and 2020 and a mortality rate of 534000 to 602000 in the respective years; 80% of all malaria deaths are among children under the age 5 years (WHO, 2021b; Douglas et al., 2021). Like most other vector borne diseases, malaria is characterized by Spatio-temporal variations or changes due to demographic, socio-economic and geographical factors. These covariates can help determine the Spatio-temporal patterns of disease and recognize hotspots to aid efficient examination of disease, cost-effective allocation of resources and most importantly, effective disease control (Ahmad et al., 2017; Ra et al., 2012; Childs et al., 2006; Mabaso et al., 2006; Clements et al., 2009; Jana et al., 2022). The Spatio-temporal distribution of vector borne diseases is not only determined by environmental factors. Political and state borders are also major determinants because of their influence in the spatial

distribution and enactment of control and prevention programs (Jana et al., 2022). This can be explained by a Spatio-temporal study such as in Northern Nigeria and Northern Thailand, where there was a sharp difference in the malaria prevalence at Benue and Myanmar borders, respectively (Childs et al., 2006; Okunlola and Oyeyemi, 2019). Also, only environmental and biological factors cannot fully justify the differences in local diseases as claimed by Ra et al., (2012).

Anaemia is another disease that has become a public health challenge in Nigeria, especially amongst children aged under 5 years. It is a condition that arises as a result of the reduction of hemoglobin in the blood (Anaemias, 2017; Hailegebreal et al., 2021). Globally, more than 273 million children aged under 5 years are affected by anaemia (Hailegebreal et al., 2021). Sub-Sahara Africa is the most endemic region with about 53.8% childhood anaemia cases (Hailegebreal et al., 2021). According to WHO classification, anaemia is considered severe if its prevalence is 40% and above, moderate between 20% and 39.9%, and mild between 5% and 19.9% (Kassebaum and Collaborators, 2016; Hailegebreal et al., 2021). Anaemia has become major public health due to its prevalence and effect on children's health.

Pregnant women and children are most vulnerable to anaemia because of their high requirement of iron. Children between the age of 6 to 59 months are anaemic if their haemoglobin level is below 11g/dl. The major causes of anaemia in children are parasitic infection, dietary iron deficiency and inherited disorders but in malaria endemic region, malaria disease is the major cause (Leal et al., 2012; Zulkarnain, 2022).

The spread of malaria and anaemia in Nigeria has been a concern to researchers which has led to several studies such as (Okunlola and Oyeyemi, 2019; Okunlola et al., 2021; Hailegebreal et al., 2021; Akinyemi, 2021; Ozano, 2022; Borderon et al.). A quasi-experimental fixed-effect model was used to investigate the effect of malaria on the haemoglobin concentration of children under

5 years old in Burkina Faso (Starck et al., 2021). They concluded that there is a strong negative effect on haemoglobin levels among children from malaria infection. Furthermore, Ekvall, (2003) stated that anaemia caused by *Plasmodium falciparum* is a result of excess removal of nonparasitized red blood cells together with the destruction of parasitized red cells' immune system leading to malfunctioning of the bone marrow. Also, the main cause of mortality and morbidity in children who live in Kenyan malaria hotspots is *falciparum* malaria (Newton et al., 1997). There is a need to monitor the progress of malaria and anaemia eradication soon and use the available data to forecast cases in space and for quantifying the extent of the possible spread of these diseases.

The spatial pattern of diseases and exposures does not explain the temporal variation which is also important and interesting. Besag et al., (1991), introduced a spatial pattern (Besag-York-Mollie (BYM) model) which was extended by incorporating a linear time trend for interaction (Bernardinelli et al., 1995). Knorr-Held, (2000), included a non-parametric temporal trend in disease risk which comprises the time changing effect of predictors. Spatio-temporal models are mostly used in several fields of science (Mdakane, 2019) including disease surveillance studies. With the help of Bayesian hierarchical modelling framework, the implementation of these models is made possible. These models accommodate a composite and workable structure in space and time models, with Spatio-temporal interaction as the paramount feature. Here, our work is extended from the methods used by Bernardinelli et al., (1995), and Knorr-Held, (2000) for Spatio-temporal framework. By applying the multilevel model analysis (Mdakane, 2019), we investigate the Spatio-temporal distribution of malaria and anaemia and their associated risk factors using data from the Nigeria malaria indicator survey (NMIS) for 2010 and 2015.

4.2 Data description

This work obtained data from the Nigeria Malaria Indicator Survey (NMIS) which was carried out by the National Malaria Elimination Program (NMEP), National Population Commission (NPopC) and the National Bureau of Statistics (NBS). The data captures the surveys of 2010 and 2015 which were the first and second malaria indicator surveys conducted in Nigeria. The 2015 survey was put into action just a year after the development of the new national malaria strategic plan that covers 2014-2020 (NMC report, 2015). The two years were used because this survey is usually carried out every 5 years. To determine the risk of malaria or anaemia disease, two-stage sampling was carried out. Clusters were selected from each urban/rural strata in the first stage and systematic sampling were done for selection of households in the second stage. The data has 12,623 children under the age of 5 years old in total whereby 11,172 and 11,072 children were tested for malaria and anaemia out of 12,623, respectively. Finally, the sample size of 9,533 was used for analysis after removing the missing values and this is 75.5% of the original data. Figure 4.1 shows the map of Nigeria comprising 6 geopolitical zones and their 37 states including the capital territory. It further reveals the location of Nigeria in Africa. While the maps in Figure 4.2 and Figure 4.3 indicate the prevalence of malaria and anaemia in each state.

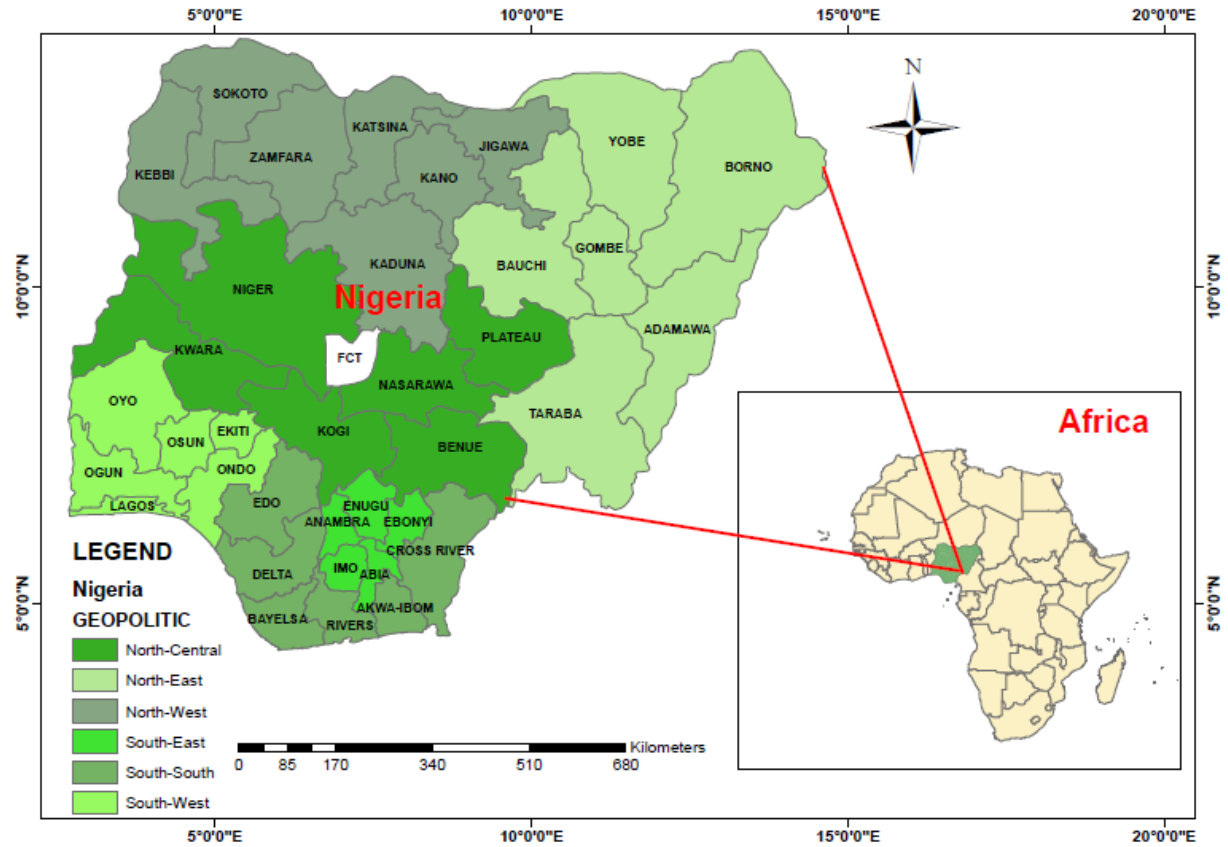


Figure 4.1: Location map of Nigeria showing the 6 geopolitical zones and their 37 states, including the capital territory Abuja.

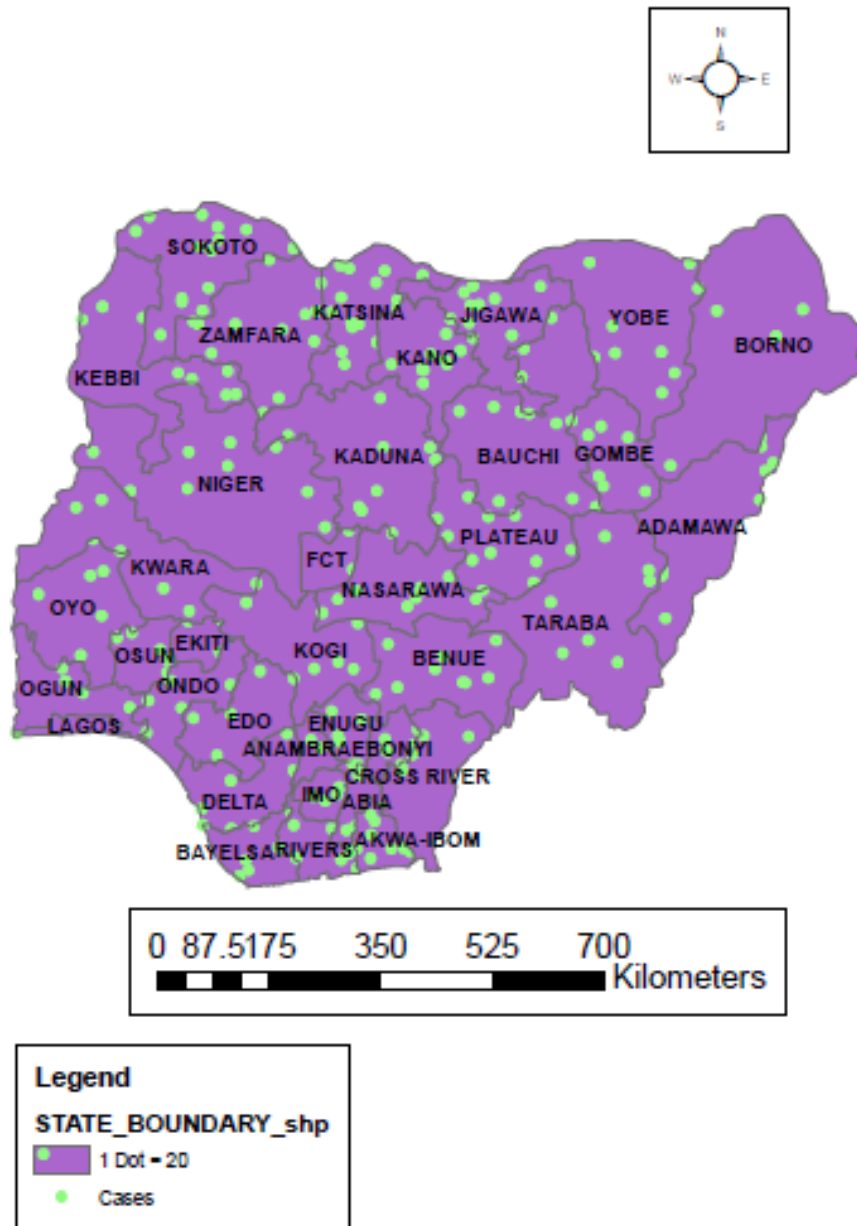


Figure 4.2: Map of Nigeria showing state rates based on sampling weights of under 5 years old malaria prevalence.

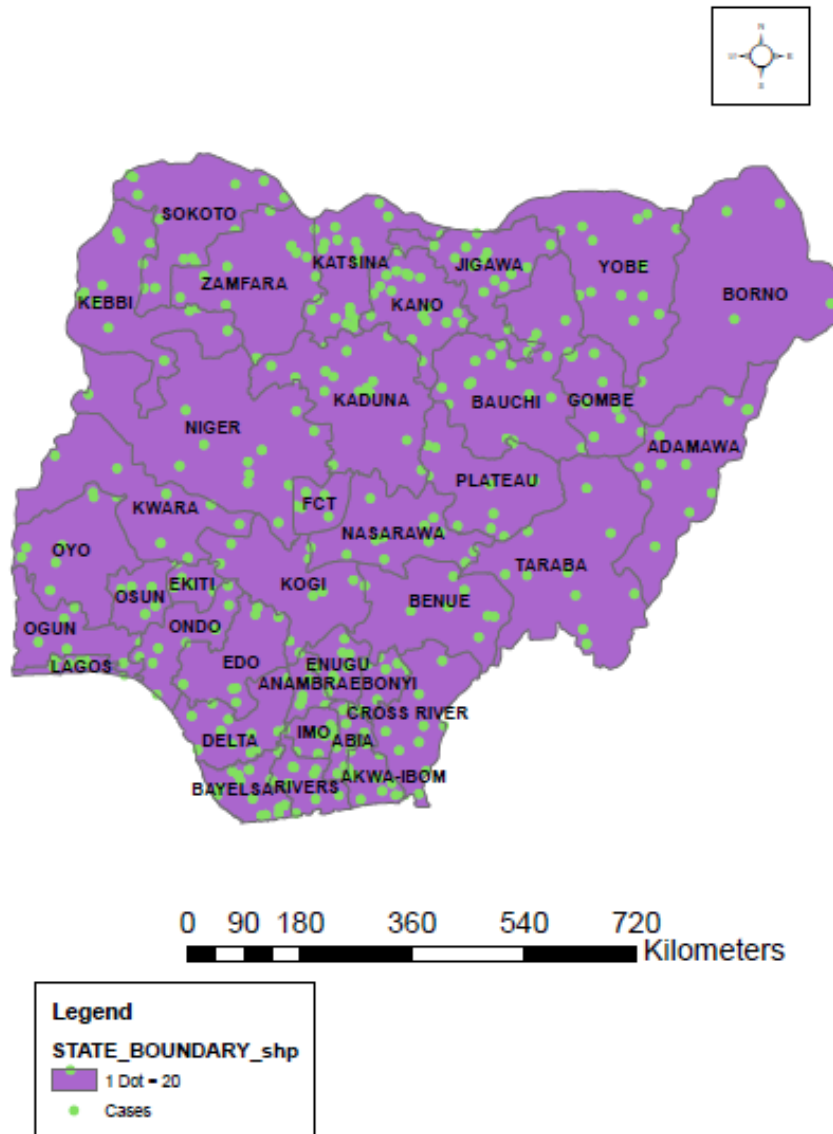


Figure 4.3: Map of Nigeria showing state rates based on sampling weights of under 5 years old anaemia prevalence.

In this study, the first dependent variable is the binary response from a child's RDT outcome while the second dependent variable is the binary response to the anaemic status of a child. For both dependent variables, 1 represents the presence of malaria or anaemia infection and 0 represents no presence of malaria infection or anaemia. The independent variables are the type of place of residence, source of drinking water, type of toilet facility, has electricity, has radio, has television,

main floor material, main wall material, main roof material, wealth index, child's age in months, sex, mother's highest educational level, state, and region. Table 4.1 contains the summary of all the variables used in this work.

Table 4.1: Exploratory data analysis

Variables	Category	Malaria		Anaemia	
		Positive(%)	Negative(%)	Positive(%)	Negative(%)
Type of place of residence	Urban	8.4	22.5	17.9	13
	Rural	36.9	32.2	50.5	18.6
Source of drinking water	Tap/Other water	17.5	22.3	26.6	13.2
	Well water	27.7	32.5	42	18.2
Wealth index	Poor	23.3	15.4	28.6	10
	Middle	10.8	10.9	15.5	6.2
	Rich	11.1	28.5	24.2	15.5
Region	North Central	8.6	10.1	11.2	7.5
	Northeast	7.8	10.7	11.6	7
	Northwest	13.9	10.8	19.1	5.1
	Southeast	3.7	7.4	7.3	3.8
	South South	6.7	8.9	12	3.8
	Southwest	4.4	7	7.3	4.3
Child's age in months	6-14 months	5.9	11.2	13.2	3.8
	15-23 months	5.8	8.9	11.2	3.5
	24-32 months	7.6	9.3	12	4.9
	33-41 months	8.4	9.2	11.4	6.2
	42-50 months	8.3	8.1	10.1	6.3
	51-59 months	9.3	8	17.4	6.8
Sex	Male	23.1	27.4	35.5	15
	Female	22.1	27.4	32.9	16.6
Electricity	No	30	22.4	38.6	13.9
	Yes	15.2	32.4	29.7	17.8
Television	No	31.7	24.8	41.6	14.7
	Yes	13.6	29.9	26.8	16.9
Radio	No	18	16.7	42.3	25.6
	Yes	27.2	38.1	22.5	9.2
Mother's educational level	No education	24.9	19.8	32.9	11.5
	Primary	9.2	10.2	13.5	6
	Secondary	9.6	9.9	18.9	10.8
	Higher	1.1	5.3	3.3	3.1
Main floor material	Earth/Other	25	21.4	33.3	13.1
	Cement/Ceramics	20.2	33.4	35.1	18.5

Main wall material	Wood/Other	28	22	36.2	13.6
	Cement/Bricks	17.2	32.8	32.2	18
Main roof material	Wood/Other	17.4	14.2	22.7	8.9
	Zinc/Metal	27.9	40.5	45.6	22.8
Type of toilet facility	Flush/Other toilet	22.7	25.4	32.8	15.7
	Pit toilet	22.5	29.4	35.5	16

4.3 Method

To begin with, we make use of a space-time model proposed by Bernardinelli et al., (1995) using Poisson distribution. The log risk ratio was defined as a linear function of time. Authors expressed the log risk ratio for area i ; $i = 1, \dots, I$ for time t ; $t = 1, \dots, T$ as

$$\log(\theta_{it}) = \eta_{it} = \mu + u_i + v_i + (\beta + \delta_i) \times t. \quad (4.1)$$

Following Besag et al., (1991) specification, $\phi = u_i + v_i$ are the spatial random effect (convolution model), where u_i 's are the structured variables and v_i 's are the unstructured variables, μ is the overall mean, β is the universal linear time trend effect and δ_i is the random effect for interaction between space and time. For the data to show the time trend, the parameters for the time trend were allocated unclear priors. In terms of the interaction random effect δ_i , the independent and identically distributed (i.i.d) Gaussian prior was adopted, though alternative prior specification can be given. Based on spatial models, the specification of priors for unstructured and structured spatial effects were described as in the strictly spatial models.

On the other hand, Knorr-Held, (2000) reformed the earlier method by disabling the parametric limitations. The authors adopted a Binomial distribution for the number of cases in country i ($i = 1, \dots, I$) at t th time ($t = 1, \dots, T$), while the log odds are expressed as;

$$\log\left(\frac{\pi_{it}}{1-\pi_{it}}\right) = \eta_{it} = \mu + u_i + v_i + \gamma_t + \nu_t + \delta_{it} \quad (4.2)$$

where γ_t , ν_t refers to the temporal random effects that take care of unnamed attributes of year t , and δ_{it} ; the interaction effects that take care of differences not described by the main effects. Intrinsic conditional autoregression (iCAR) and first order random walk structure were assigned to u_i and γ_t , while independent Gaussian priors were assigned to ν_i and ν_t . Based on the temporal effects interaction and spatial effects, the interaction δ_{it} was presumed to have four forms of prior inference.

Bernardinelli et al., (1995) and Knorr-Held, (2000) performed their parameter estimation under the fully Bayesian approach with the use of Markov chain Monte Carlo (MCMC) through Gibbs sampling techniques. But here, INLA approximation to fully Bayesian estimation was used. Therefore, the method used in this study is discussed as follows.

Let y_{ijkt} be disease k (malaria or anaemia status) of child j in state i : $i = 1, \dots, 37$ during year t : $t = 1, 2$. When the disease is malaria, $k = 1$ and the response outcome variable, which is a binary response, is defined as;

$$y_{ij1t} = \begin{cases} 1, \text{Malaria} \\ 0, \text{No malaria} \end{cases}$$

and $k = 2$ when the disease is anaemia and the response outcome variable, which is a binary response, is defined as;

$$y_{ij2t} = \begin{cases} 1, \text{Anaemia} \\ 0, \text{No anaemia} \end{cases}$$

where y_{ijkt} is the binary response outcome and it follows a Bernoulli distribution as

$y_{ijkt} \sim \text{Bernoulli}(\theta_{ijkt})$, where θ_{ijkt} are unspecified probabilities associated to the outcome probabilities of the models. The logistic regression model is expressed as;

$$\text{logit}(\theta_{ijkt}) = \beta_0 + \eta_{ijkt} \quad (4.3)$$

where β_0 is the model intercept and the linear predictor $\eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta}$ with covariate vector $\mathbf{x} = (x_{ijkt1}, \dots, x_{ijktq})'$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)$ is the vector regression coefficient. We employed the combined formulation of the structured additive regression to permit flexibility where the classical predictor can be expanded to a better flexible additive predictor. Therefore, the structured additive predictor is expanded to Spatio-temporal modelling as

$$\eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{\text{spat}}(s_i) + f_{\text{year}}(y) + f_{it}(s_i, t) \quad (4.4)$$

where f_{spat} , f_{year} , and f_{it} are respectively functions suitable for space, year and space-year interaction. The spatial components f_{spat} are disintegrated into two i.e., spatially unstructured f_{unstr} and spatially structured f_{str} effects. However, $f_{it}(s_i, t)$ denotes a space-year interaction, and f_{year} shows the random year effects, and this is modelled as a first-order random walk or AR(1) according to (DiMaggio, 2012).

To ascertain the efficiency of the estimators. we compared these seven models

$$\text{Model 1: } \eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{\text{str}}(s_i) + f_{\text{unstr}}(s_i)$$

$$\text{Model 2: } \eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{\text{str}}(s_i) + f_{\text{unstr}}(s_i) + \beta_t$$

$$\text{Model 3: } \eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{\text{str}}(s_i) + f_{\text{unstr}}(s_i) + f_{\text{year}}(t)$$

$$\text{Model 4: } \eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{\text{str}}(s_i) + f_{\text{unstr}}(s_i) + f_{it}(s_i, t)$$

$$\text{Model 5: } \eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{\text{str}}(s_i) + f_{\text{unstr}}(s_i) + \beta_t + f_{it}(s_i, t)$$

$$\text{Model 6: } \eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{\text{str}}(s_i) + f_{\text{unstr}}(s_i) + f_{\text{year}}(t) + f_{it}(s_i, t)$$

$$\text{Model 7: } \eta_{ijkt} = \mathbf{x}'_{ijkt}\boldsymbol{\beta} + f_{str}(s_i) + f_{unstr}(s_i) + f_{1year}(t) + f_{it}(s_i, t)$$

where,

- \mathbf{x}_{ijkt} signifies the vector of categorical variables effects for disease k of child j in state i during year t .
- $\boldsymbol{\beta}$ is a vector of regression coefficients.
- β_t means the year-specific fixed effects.
- $f_{unstr}(s_i)$ and $f_{str}(s_i)$ are the unstructured and structured random effects, respectively.
- f_{year} and f_{1year} show the smooth functions of the temporal random effects.
- $f_{it}(s_i, t)$ signifies the spatial-year interaction effect.

Model 1 takes care of the spatially structured random effects, and this accounts for unobserved significant factors that change spatially transversely over the states and spatially unstructured random effects that caters for undetected variables inside the states. Hereafter, by assuming a categorical variable, it will have a linear effect on malaria and anaemia. The temporal effect is not assumed by this model. Model 2 follow the same pattern but in addition assumes a linear year trend in β_t . On the other hand, Model 3 contains separable space and year random effect which takes care of the linear effect of categorical variables. Also, Model 4 and Model 1 are parallel to each other but additionally, Model 4 takes care of space and year interaction which captures differences that are not shown by the main effects. Regarding Model 5, the assumptions of linear effects of the categorical variables, spatial random effects, linear year trend and space year-interaction are made. While both Model 6 and Model 7 take on linear effects of categorical variables and spatial random effects of the location, space, and year interaction, they vary in prior assumptions of the temporal random year effects $f_{year}(t)$ and $f_{1year}(t)$. In other words, all models take on linear effects of categorical variables through the term \mathbf{x}'_{ijkt} .

4.3.1 Prior specifications

Here, for the Spatio-temporal logistic regression models, the full Bayesian approach was adopted. Diffuse priors were allocated to fixed effects and linear year trend, intrinsic conditional autoregressive (iCAR) was used to model the spatially structured random effects, while the independent and identically distributed (i.i.d) Gaussian prior was assigned spatially unstructured random effects. A first-order random walk was used to model the temporal year random effects f_{1year} . Nevertheless, it is interesting to note that varied prior specifications for the temporally changeable year effects f_{year} were given in the models and penalized spline was given in the Spatio-temporal logistic regression model. Also, independent penalized splines for the logistic independent first-order autoregressive model were adopted to model the spatial year-specific effects (interaction).

4.4 Parameter estimation

In this research, we discussed the procedure of parameter estimation of the Spatio-temporal logistic regression model of malaria and anaemia using the fully Bayesian approach. With due consideration, every unspecified parameter assumed a random variable and was given adequate prior distributions. The posterior of the priors is given as;

$$p(\varphi, \psi|y) \propto L(y|\varphi, \psi)p(\varphi, \psi) \quad (4.5)$$

where $L(y|\varphi, \psi)$ is the likelihood of the penalized spline and $p(\varphi, \psi)$ are the prior distributions of the Spatio-temporal logistic regression model. We represent the latent Gaussian field as

$\varphi = \{\{\beta\}, \{\beta_t\}, \{f_{str}(\cdot)\}, \{f_{unstr}(\cdot)\}, \{f_{year}(\cdot)\}, \{f_{1year}(\cdot)\}, \{f_{it}(\cdot)\}\}$ while the equivalent hyperparameters are shown as $\psi = \{q_{str}, q_{unstr}, q_{year}, q_{1year}, q_{it}\}$. Conjugate gamma priors

$\text{Gamma}(1, 0.00005)$ were allocated to all hyperparameters while the R-integrated nested Laplace approximation (INLA) package was used to estimate the parameters.

4.5 Results

Table 4.2 shows the model fit values for the Spatio-temporal logistic regression models of malaria which comprise Deviance Information Criteria (DIC) and the effective number of parameters (DP). Model 7 was chosen as a better model because it gave the least DIC value of 10819.24. Therefore, the presentation of results and interpretations are based on Model 7 which includes both linear and nonlinear effects as well as the Spatio-temporal effects.

Table 4.2: Summary of the model comparisons of malaria

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
DIC	11122.29	11110.03	11114.02	10821.05	10820.58	10821.11	10819.24
DP	52.31	53.40	54.27	86.24	86.14	86.19	85.06

Table 4.3 provides the adjusted posterior odds ratios estimates (AOR) and 95% confidence interval for the best fitting model mentioned above. Here, the results for significant covariates were discussed separately. Regarding child's age in months, the results showed an increase in the odds of malaria for all ages. In the same vein, there was a significant increase in the odds of malaria among children whose anaemic status is positive and those who live in rural area. On the other hand, odds of malaria decrease significantly with respect to wealth index and mother's educational level (Secondary and Higher). Furthermore, household with electricity had significantly lower odds of malaria compared to a household without electricity.

Table 4.3: Adjusted posterior odd ratios estimates (AOR) of malaria with 95% confidence interval

Malaria (Model7)		
Variables	AOR	95%CI
Child's age in months(grouped)(ref=1)		
2	1.452*	(1.235, 1.707)
3	1.979*	(1.692, 2.315)
4	2.399*	(2.051, 2.807)
5	2.787*	(2.372, 3.277)
6	3.221*	(2.742, 3.787)
Source of drinking water(ref=Tap/Other water)		
Well water	0.957	(0.863, 1.061)
Has electricity(ref=No)		
Yes	0.819*	(0.715, 0.939)
Main wall material(ref=Wood/Other)		
Cement/Bricks	0.943	(0.813, 1.094)
Main floor material(ref=Earth/Other)		
Concrete\Ceramics	0.919	(0.807, 1.047)
Main roof material(ref=Wood/Other)		
Zinc/Metal	1.004	(0.887, 1.137)
Anaemic status(ref=No)		
Yes	2.963*	(2.657, 3.306)
Wealth index(ref=Poor)		
Middle	0.758*	(0.641, 0.897)
Rich	0.468*	(0.365, 0.601)
Mother's educational level(ref=No education)		
Primary	0.915	(0.794, 1.056)
Secondary	0.748*	(0.642, 0.871)
Higher	0.442*	(0.339, 0.574)
Sex(ref=Male)		
Female	1.019	(0.930, 1.118)
Has radio(ref=No)		
Yes	0.983	(0.884, 1.093)
Type of place of residence(ref=Urban)		
Rural	1.637*	(1.436, 1.867)
Has television(ref=No)		
Yes	0.947	(0.814, 1.103)

* means significant

CI means Confidence Interval

Figure 4.4 presents the mapped estimated residual spatial effects of the year 2010 and year 2015.

The essence is to study how disease prevalence and risk factors change with time. These maps show unobserved spatial factors that are not captured in the survey or that captures the effects of

cultural patterns. From the Figure, there was an obvious spatial pattern change over the two years. Though higher concentrations in the two years are scattered, the states inside the Northeast, Northwest and Southwest regions have higher odds of 0.99 – 2.2 of malaria. In 2010, states in the north-east had higher odds of malaria but had lower odds of malaria later in 2015. This is to the states in the northwest.

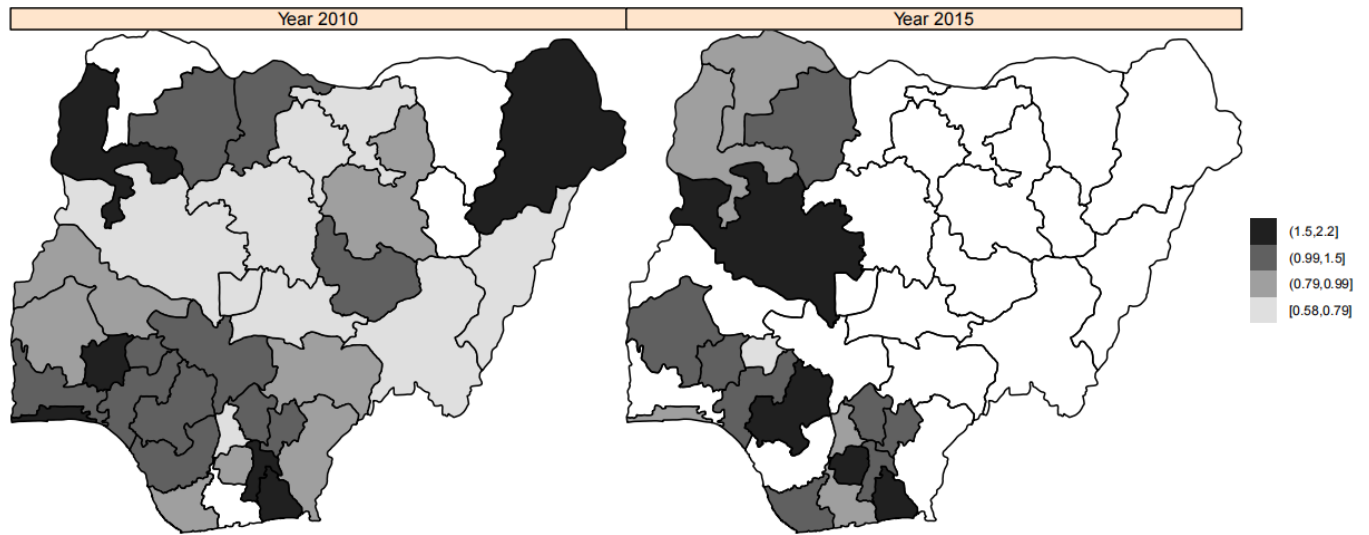


Figure 4.4: Maps displaying residual spatial effects of malaria in Nigeria for year 2010 and 2015 obtained from Spatio-temporal interaction logistic regression model, i.e., Model 7

Figure 4.5 shows the posterior relative risk of malaria. Here, there was an obvious change in the relative risk of malaria over the two years. This implies that there has been an increase in the relative risk of malaria from 2010 to 2015.

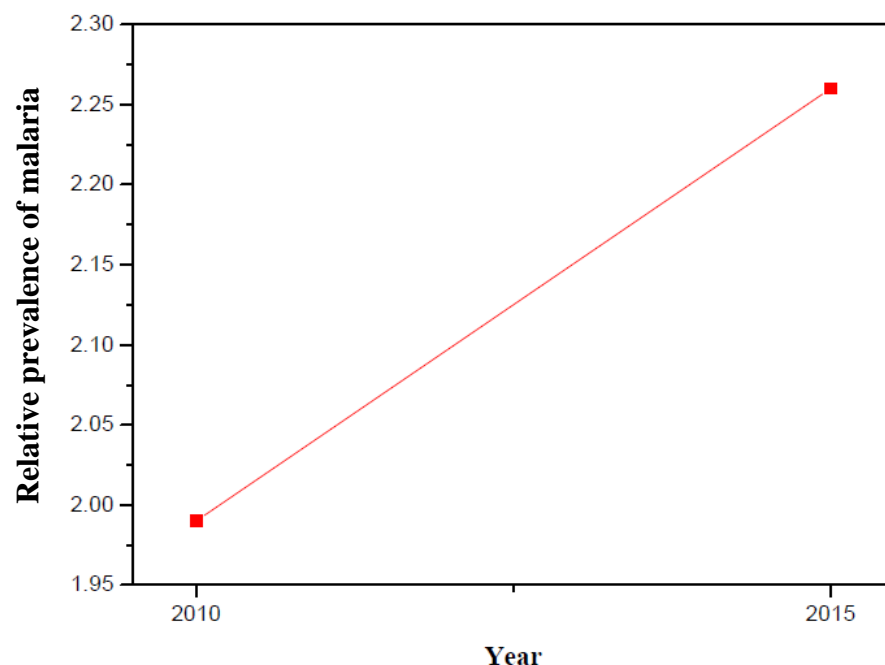


Figure 4.5: Depicting estimated posterior relative prevalence of malaria for the logistic regression best fitting model.

Table 4.4 contains the results of *DIC* and the effective number of parameters *DP* of the model fit values for the Spatio-temporal logistic regression models of anaemia. From the summary, Model 6 gave the least DIC value of 10330.78. Therefore, the interpretation and presentation of results were based on this model.

Table 4.4: Summary of the model comparisons of Anaemia

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
DIC	10472.33	10474.30	10472.42	10330.80	10330.85	10330.78	10330.87
DP	53.11	54.10	53.09	83.12	83.59	83.11	83.18

Table 4.5 presents the adjusted posterior odds ratios estimates (AOR) and 95% confidence interval for the best fitting model mentioned above. Discussion on the results was basically on the significant covariates. There was a significant decrease in odds of anaemia for children in age categories 3,4,5, and 6 but insignificant for age category 2. Also, the odds of anaemia decrease significantly with increasing mother's educational level. Furthermore, sex, household that has

radio and television had significantly lower odds of anaemia. On the other hand, with regards to child's age in months, the results showed that there was an increase in the odds of malaria but only significant for children in age 2, 3 and 6. While the odds of anaemia increased significantly for children whose malaria rapid test results were positive, their source of drinking water is well water with a moderate wealth index and live in a rural area as a type of place of residence compared to those who tested negative for malaria, drink tap water/other water, with a good wealth index and live in an urban area as a type of place of residence.

Table 4.5: Adjusted posterior odd ratios estimates (AOR) of Anaemia with 95% confidence interval

Malaria (Model7)		
Variables	AOR	95% CI
Child's age in months(grouped)(ref=1)		
2	0.910	(0.763, 1.086)
3	0.599*	(0.506, 0.708)
4	0.402*	(0.341, 0.473)
5	0.330*	(0.280, 0.390)
6	0.303*	(0.256, 0.358)
Source of drinking water(ref=Tap/Other water)		
Well water	1.134*	(1.021, 1.259)
Has electricity(ref=No)		
Yes	1.019	(0.881, 1.180)
Main wall material(ref=Wood/Other)		
Cement/Bricks	0.908	(0.774, 1.065)
Main floor material(ref=Earth/Other)		
Concrete\Ceramics	1.027	(0.900, 1.170)
Main roof material(ref=Wood/Other)		
Zinc/Metal	0.932	(0.818, 1.062)
Result of malaria rapid test(ref=No)		
Yes	2.985*	(2.676, 3.331)
Wealth index(ref=Poor)		
Middle	1.206*	(1.005, 1.448)
Rich	1.229	(0.946, 1.595)
Mother's educational level(ref=No education)		
Primary	0.849*	(0.730, 0.987)
Secondary	0.795*	(0.680, 0.928)
Higher	0.659*	(0.526, 0.825)

Sex(ref=Male)		
Female	0.784*	(0.713, 0.862)
Has radio(ref=No)		
Yes	0.778*	(0.695, 0.871)
Type of place of residence(ref=Urban)		
Rural	1.475*	(1.295, 1.679)
Has television(ref=No)		
Yes	0.789*	(0.673, 0.925)

* means significant

CI means Confidence Interval

Figure 4.6 is the graphical representation of child's age in months and the adjusted odd ratio (AOR) of malaria and anaemia. The relationship between child's age in months and the adjusted odd ratio of malaria increase significantly across the age group, while anaemia decreases significantly. This might be due to the stimulation of antimalarial immune defenses by malaria antigen in breast milk which reduces malaria risk in infants attributable to breastfeeding (Van Den Elsen et al., 2020). As children are weaned, they are vulnerable to malaria as they have lost maternal immunity and are yet to develop self-immunity against infection (Schumacher and Spinelli, 2012). As a result, there will be a decrease in anaemia infection because individuals at some point develop a disease-controlling immunity that makes them asymptomatic (White, 2018). This asymptotic means that there will be no apparent symptoms even when an individual is anaemic.

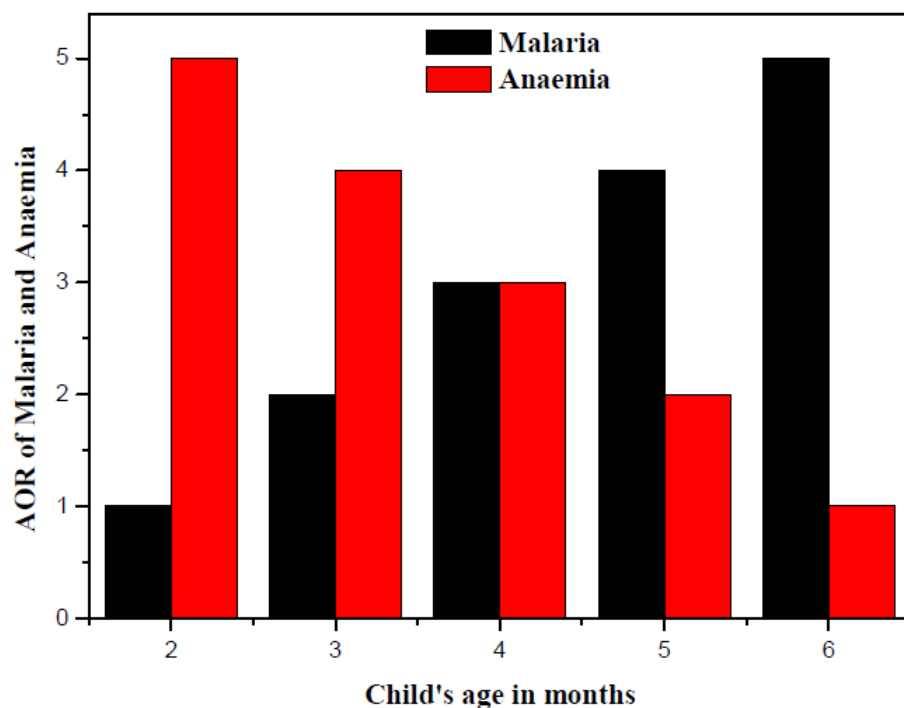


Figure 4.6: Relationship between child's age in months and AOR of malaria and anaemia

Figure 4.7 displays the mapped estimated adjusted posterior odds of residual spatial effects of the year 2010 and 2015 in Nigeria. In our first work, we focused on the evolution of the geographical variation of anaemia. The map in Figure 4.7 represents the estimated residual spatial effects for the two years. The colours for the regions are the same as described above. In both years, the South-South and partially South-West regions showed a higher concentration of anaemia with odds ratio of 0.90 to 2.2. Notwithstanding, in 2010, some part of North-Central region had a higher concentration of the same odds ratio. Other things seem similar to the previous figure.

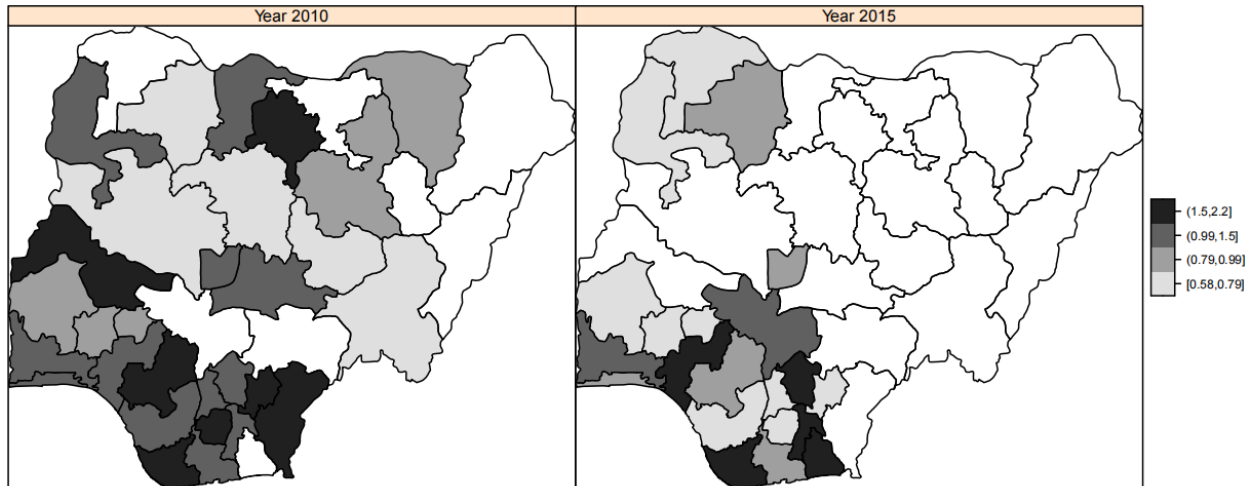


Figure 4.7: Maps displaying residual spatial effects of anaemia in Nigeria for year 2010 and 2015 sprang from the Spatio-temporal interaction logistic regression model i.e., Model 6

Figure 4.8 provides the estimated posterior relative risk of smooth function of anaemia. There was an increase in the relative risk of anaemia in the two years i.e., the prevalence of anaemia infection was high in 2015.

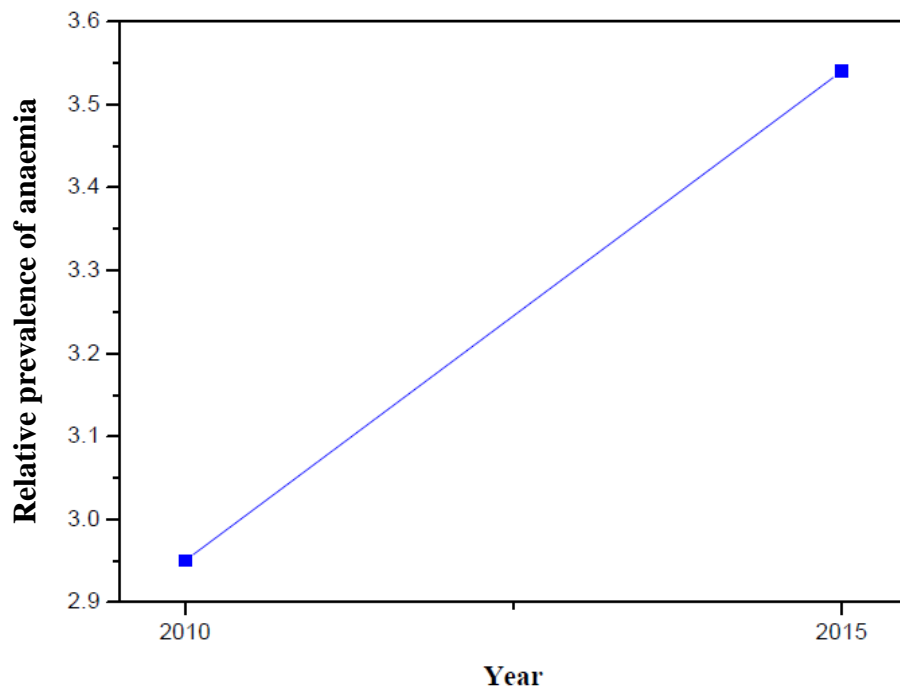


Figure 4.8: Depicting estimated posterior relative prevalence of anaemia for the logistic regression best fitting model.

4.6 Discussion

This research work applied Spatio-temporal models to investigate the relative risks and geographical variation of malaria and anaemia in Nigeria. This research work was carried out with the sole aim of developing and applying the appropriate statistical models to assess risk factors and geographical variations of malaria and anaemia. In addition, a unified framework of flexible models within Bayesian hierarchical modelling was applied to understand various factors associated with this discrete type of malaria and anaemia prevalence among children from 0 to 59 months in Nigeria. The models considered are an augmentation to classical models which include spatial and Spatio-temporal models for identification of geographical variation of year-specific effects. Logistic regression was applied to assess influential factors and state variation of malaria and anaemia prevalence. The structured additive modelling approach gives allowance for different kinds of predictors to be included in classical models in an additive manner by borrowing strengths from both parametric and non-parametric models. Integrated nested Laplace approximation was used to investigate the Spatio-temporal effect on childhood malaria and anaemia disease with the application to MIS (Malaria indicator survey) datasets in Nigeria. For each model, the Deviance Information Criteria (DIC) were compared, and the best model was used to fit malaria and anaemia data of Nigeria. Among the models considered, the Spatio-temporal interaction logistic regression model was chosen as the best model to fit malaria data while for anaemia, model 6 (interaction with one random time effect (autoregressive prior of order 1 (AR1))) was chosen. For both diseases, variation can be seen among the Nigeria states and clustering among states with high malaria and anaemia relative risk (RR). Child's age in month, main wall material, anaemic status, wealth index, mother's educational level and type of place of residence were significantly related to malaria and anaemia except for the source of drinking water, sex, and household that has radio and television

that were significantly related to only anaemia over the two years period, this is in line with Ugwu and Zewotir, (2020c) and Roberts et al., (2020). While anaemia is seen as a dominant risk of malaria, malaria as well is a major risk of anaemia (Weze et al., 2021; Rumisha et al., 2019). In the two years considered, we found out that the states within the northern, southern, and western regions have the higher prevalence of malaria and anaemia. The lowest prevalence of malaria and anaemia was seen in states within the eastern region. Also, we estimated year temporal effects on malaria and anaemia. For both diseases, the plots showed no obvious change in the spread of malaria and anaemia.

There are always limitations in every research. In this study, the main limitation is the number of years available for us to estimate the Spatio-temporal trends of malaria or anaemia. This matter poses a hinderance to investigating the trend of malaria or anaemia pandemic during the early years. Also, this study used secondary data from cross-sectional surveys which did not allow the causal relationships to be established. Furthermore, though iron deficiency is one of the major causes of anaemia, there was no information on iron levels in children. Notwithstanding the limitations, the strength of this study is in the use of individual-level malaria RDT results instead of indicators or estimates of malaria or anaemia.

Chapter 5

Spatial joint modelling of malaria and anaemia among Nigeria children under 5 using a semi-parametric approach

5.1 Introduction

World Health Organization (WHO) reports that a child dies of malaria every two minutes. Notwithstanding, many of these deaths can be prevented. Globally, 229 million malaria cases were recorded in 2019, which resulted in 558,000 deaths, of which 74% (416,000) were children under the age of 5 years (WHO, 2021b). This translates to approximately 750 daily deaths of children under the age of 5 years. Malaria is an urgent public health priority, especially in sub-Saharan Africa, where half of the world's population resides.

Malaria has been linked to anaemia infection due to its effect on the red blood cells. When malaria parasite enters the human blood after a bite from infectious mosquito this affects the red blood cells. The red blood cell gets ruptured at the end of the infection cycle and lowers the amount of the red cells which if not treated causes severe anaemia (Henrici et al., 2021). These two diseases have added to the mortality rate of children especially in Nigeria (Bamgbose, 2021; Ajakaye and Ibukunoluwa, 2020; Egbon, 2021). Notwithstanding, a recorded global improvement in fighting malaria and anaemia, the two diseases remain a health challenge more specifically in children from developing countries like Nigeria (Gaston et al., 2021; Adeyemi, 2021; Gayawan et al., 2022).

Anaemia is defined by World Health Organization (WHO) as the reduction of hemoglobin in the red blood cells which results in insufficient oxygen in the body tissue. It can be classified as mild,

moderate and severe in men, women and children (Pasricha et al., 2018b). Poor cognitive and motor development in children, and work capacity in adult which affects a nation's economic development can be associated with anaemia (WHO, 2008). A child within the age of 6 to 59 months is considered anaemic if the hemoglobin concentration is less than 11g/dl (Korenromp et al., 2004; WHO, 2015a; NNPC et al., 2010; Gaston et al., 2021). Nevertheless, the cut-off for hemoglobin concentration level of anaemia caused by malaria infection is less than 8g/dl (NNPC et al., 2010).

It was reported by World Health Organization (WHO) that the global prevalence of anaemia among children under the age of 5 years was 42%; the disease is more endemic in Africa, with 62.3% of children from 6 to 59 months being anaemic (Mbunga et al., 2021). According to 2010 and 2015 Malaria Indicator Survey (MIS), the prevalence of anaemia among children under the age of 5 years in Nigeria was about 72% and 68% respectively (NNPC et al., 2012; NMC report, 2015). The effect of anaemia on children is so detrimental that it impedes their mental and physical development which affects the socio-economics of the nation (Abegunde and Stanciole, 2006; Soares Magalhães and Clements, 2011; WHO, 2011; Gaston et al., 2018; Gaston et al., 2021).

Nutritional deficiencies such as iron deficiency, folate deficiency, insufficient Vitamin B12 and communicable diseases such as hookworm infection, HIV, malaria, and other parasitic infections are major causes of anaemia (Cappellini et al., 2020; Roche and Layrisse, 1966; Malizia et al., 2022; Gupta et al., 2022; Cao et al., 2022; Chao et al., 2022; Karagöl and Yiğit, 2022; Mrimi et al., 2022; Powers, 2022). Nevertheless, in the areas where malaria infection is most endemic, the major cause of anaemia is malaria (Agagliati et al., 2022; Keating et al., 2021; WHO, 2021b; Ibeji et al., 2022). Malaria is attributed to anaemia prevalence, and high mortality rate and morbidity in children under the age of 5 years (Sardar et al., 2021; Fischer, 2021; Gayawan et al., 2022; Okell

et al., 2022). Anaemia is said to be responsible for half of malaria related deaths in areas with high rate of malaria infection (Keating et al., 2021; Bahati et al., 2022; Papaioannou et al., 2019). From the above information, it is convincing that both diseases are related and that implies bringing malaria prevalence under control will reduce anaemia infection (Pasricha et al., 2018b; Cohee et al., 2020; White, 2018). According to McLean et al., (2009), over half of malaria reduction reduced 60% risk of anaemia infection. This is a confirmation of the relationship between malaria and anaemia, which can increase the death rate and cause more havoc in children if there is no timely intervention (Gaston et al., 2018; Gaston et al., 2021).

Several studies have proved that children are more vulnerable to both diseases, but studies on simultaneous modeling of both malaria and anaemia infection in children are scanty, especially in Nigeria and some African countries (Deribew et al., 2013; Adebayo et al., 2016; Petry et al., 2019; Gayawan et al., 2022; Egbon, 2021). Because children are a nation's future, there is need to give precedence to their health (Gaston et al., 2018, Gaston et al., 2021). Also, studies have been done on monitoring the trend of malaria and anaemia in Nigeria using national average to compare between countries (Obasohan et al., 2022; Oguoma et al., 2021). Even with the importance of these averages, there is a tendency that they can hide the rate at which the spread of malaria and anaemia varies among the administrative unit of a country which will hinder the practical application of intervention strategies at the lower administrative level.

The national prevalence rate of malaria and anaemia in Nigeria among children's population (6 – 59 months) was estimated to be 42% and 72% in 2010, respectively while in 2015 it was estimated to be 27% and 68%, respectively. In both years, there is a slight gender difference in the prevalence of malaria and anaemia but with a wide geographical variation. Male children had 50.5%, and female children had 49.5% prevalence of malaria in 2010 (Ibeji et al., 2022). In terms of anaemia

prevalence in the same year, male and female children prevalences were 50.4% and 49.6%, respectively. Concerning geographical variation, there is low malaria and anaemia prevalence in South-East Nigeria at 8% and 7.8%, respectively, while North-West Nigeria had the highest prevalence of 30.2% and 31.3%, respectively. In 2015, 50.6% and 49.4% was the prevalence of malaria for male and female children, respectively, while the prevalence of anaemia for male and female was 50.8% and 49.2%, respectively. Regarding geographical variation, South-Eastern region had 8.7% and 8.5% of malaria and anaemia prevalence, respectively, while the highest prevalence for both diseases was in the North-West region with 32.3% and 32.3%, respectively (NNPC et al., 2012; NMC report, 2015). According to Nigeria Malaria Indicator Survey report, age has nonlinear relationship with malaria and anaemia prevalence as reported by previous studies (Adebayo et al., 2016; NNPC et al., 2012; NMC report, 2015; Gaston et al., 2021).

Notwithstanding the assumption by many studies that there exists a linear relationship between the predictors and the response variable, on the contrary, the relationship between age and the response variable is nonlinear. Majorly, this research focuses on the association between malaria and anaemia, also between states under study and further captures the nonlinear relationship by using a spatial joint modelling. Additionally, using the same method proposed by Okango et al., (2015), we relaxed the linearity assumption from Neuhaus et al., (2009) and Chesnaye et al., (2020). We made a deliberate allowance for some covariates to have a nonlinear relationship with the response variable by applying the penalized regression spline to jointly model malaria and anaemia prevalence among Nigerian children.

5.2 Methods

5.2.1 Data

In this study, we used data from 2010 and 2015 Nigeria Malaria Indicator Survey to extract information on malaria and anaemia prevalence in the country (NNPC et al., 2012; NMC report, 2015). Data collection was done using a two-stage cluster design from October through December 2010, comprising 240 clusters, with 83 clusters from urban areas and 157 clusters from rural areas. In 2015, data collection took place between October and November 2015 with a nationally representative sample of more than 8000 households in 333 clusters, where 138 clusters are from urban areas and 195 clusters from rural areas. Each survey was commissioned by Nigeria Malaria Control Programme (NMCP) and implemented by National Malaria Elimination Program (NMEP), the National Population Commission (NPopC), the National Bureau of Statistics (NBS) and the malaria partnership in Nigeria. The survey included all women aged 15 to 49 years who were permanent occupants of the household or visitors in the household on the night before the survey of the 2015 NMIS sample were eligible to be interviewed. In addition, malaria and anaemia tests were run on every child from age 6 to 59 months. In both years, finger or heel prick blood samples were obtained to carry out on-the-spot testing for malaria and anaemia for children between the ages of 6 to 59 months. Also, to determine the presence of malaria parasitaemia, the same samples were used to prepare thick and thin blood smears that were read in the Department of Medical Microbiology and Parasitology at the University of Lagos, Nigeria. The Paracheck pf rapid diagnostic test, which tests for *P. falciparum*, was used for on-the-spot malaria testing. The test involves a loop applicator that comes in a sterile packet. A tiny blood fragment is captured on the applicator and placed on the device's well, giving the result in nearly 15 minutes (NMC report, 2015).

For the two years, the survey adopted a two-stage probability sampling method (NNPC et al., 2010, NMC report, 2015). The sampling frame for both years was the 2006 National Population and Housing Census (NPHC) of the Federal Republic of Nigeria. In the first stage, this was used to select the primary sampling unit. In 2010, the enumeration areas consisted of 240 clusters, 83 in urban areas and 157 in rural areas though 239 clusters were finally used because of inter-communal disturbance in a particular cluster. A complete listing of households was conducted within the 37 states, including the capital territory. A mapping exercise for each cluster was done from August to September 2010.

In the second stage, 26 households were selected on the average from each cluster by equal probability systematic sampling. And this includes all women aged 15 to 49 years who were either permanently residing in the house or visitors present in the house on the night before the survey were eligibly interviewed. Also, children from the age of 6 to 59 months were tested for malaria and anaemia. While in 2015, 333 clusters were selected from the whole country, comprising 138 in urban areas and 195 in rural areas, the 2010 NMIS covered clusters as mentioned earlier. Also, a complete listing of households was conducted, and a mapping exercise was done from June to July 2015. The list of households from the first stage was used in selecting households for the second stage. By equal probability systematic sampling, 25 households were selected in each cluster. All women aged 15 to 49 years who were permanent residents of the household or those who visited the household a night before the survey were interviewed. In addition, all children aged 6 to 59 months were tested for malaria and anaemia.

This study uses only 2010 and 2015 Nigeria Malaria Indicator Survey (NMIS) data because no data beyond 2015 have been released. In malaria and anaemia data, all predictors were categorical except age, which was considered continuous. An exploratory data analysis was done to ascertain

the relationship between the predictors and the dependent variables (Malaria and Anaemic status) using a univariate standard logistic regression model. These variables include mother's highest educational level, has radio, main floor material, wealth index, type of place of residence, sex, type of toilet facility, has electricity, has television, main wall material, main roof material, child's age in months, region, source of drinking water and state.

Table 5.1 contains an overview of all the variables used in this study, while Table 5.2 shows the percentages of children for each covariate used in the population under study. Over 60% of the population study who lived in rural areas were tested for malaria and anaemia in both years. Over 55% of children who drink well water did test for malaria and anaemia. Above 40% of those who use pit toilet were tested. Similarly, over 40% of children who live in a household with television, radio and electricity, and children whose mothers had no education and lived in a house made of earth/other floor material, wood/other wall, or roof material were also tested. Over 20% of children from middle-class homes were tested while less than 41% from rich homes were tested. In total, less than 51% of male children underwent malaria and anaemia tests.

Furthermore, Table 5.3 comprises the frequency distribution and percentages of all the children tested for malaria and anaemia with their related risk factors for 2010. The table used cross-tabulation to analyze the data and summarize the results. To check whether there was a significant relationship between the independent variables and each of the dependent variables, Pearson's chi-square test and p-values were used for both 2010 and 2015. The results from this exploratory analysis (Table 5.3) showed that all the predictors considered in this study are significantly related to malaria and anaemia infection in 2010, except for source of drinking water and child's sex, which is non-significantly associated with malaria. In addition, from Table 5.3, child's age in months was found to have a nonlinear relationship with the two diseases. From Table 5.4, only

child's sex had non-significant relationship with malaria, but all other predictors were significantly related with malaria and anaemia in 2015.

Table 5.1: Description of all the variables used in the study.

Variable	Description	Scales of measurement
Responses		
Malaria	Presence of malaria	Dichotomuos: yes or no*
anaemia	Presence of anaemia	Dichotomous: yes or no*
Predictors		
place of residence	Type of place of residence	Dichotomous: rural or urban*
drinking water	Source of drinking water	Dichotomous: well or tap/other*
toilet facility	type of toilet facility	Dichotomous: pit or flush/other*
electricity	Has electricity	Dichotomuos: yes or no*
Radio	Has radio	Dichotomuos: yes or no*
television	Has television	Dichotomuos: yes or no*
floor material	Main floor material	Dichotomous: cement/ceramics or earth/other*
wall material	Main wall material	Dichotomous: cement/bricks or wood/other*
roof material	Main roof material	Dichotomous: zinc/metal or wood/other*
wealth index	Wealth index	Categorical: poor*, middle or rich
age in months	Child's age in months	Continuous: lowest = 6, highest = 59
Sex	Sex	Dichotomous: male* or female
educational level	Mother's highest educational level	Categorical: no education*, primary, secondary, or higher
State	Nigerian region	37 states including the capital territory

* represent the reference category for categorical or dichotomous variables.

Table 5.2: Sample size (n) (based on variables) and percentage of all the children tested for malaria and anaemia in 2010 and 2015

Variables	Category	2010				2015			
		Malaria		Anaemia		Malaria		Anaemia	
		n	%	n	%	n	%	n	%
Type of place of residence	Urban	1406	27.3	1381	27.3	3983	33.9	2040	33.9
	Rural	3741	72.7	3675	72.7	2042	66.1	3981	63.1
Source of drinking water	Well water	3163	61.5	3097	61.4	3562	59.1	3558	59.1
	Tap/Other water	1976	38.5	1951	38.6	2463	40.9	2463	40.9
Type of toilet facility	Pit	2652	51.6	2570	50.9	2629	43.7	2629	43.7
	Flush/Other	2485	48.4	2476	49.1	3393	56.3	3389	56.3
Has electricity	Yes	2421	47.2	2370	47.1	2884	47.9	2883	47.9
	No	2706	52.8	2666	52.9	3141	52.1	3138	52.1
Has radio	Yes	3658	71.3	3596	71.4	3626	60.2	3623	60.2
	No	1471	28.7	1442	29.6	2399	39.8	2398	39.8
Has television	Yes	2099	41.0	2069	41.1	2760	45.8	2759	45.8
	No	3026	59.0	2965	59.9	3265	54.2	3262	54.2
Main floor material	Earth/Other	2530	49.2	2469	48.9	2663	44.2	2661	44.2
	Cement/Ceramics	2608	50.8	2578	51.1	3362	55.8	3360	55.8
Main wall material	Wood/Other	2887	56.2	2819	55.9	2502	41.5	2377	39.5
	Cement/Bricks	2252	43.8	2226	44.1	3523	58.5	3644	60.5
Main roof material	Wood/Other	1795	34.9	1762	34.9	1733	28.8	1732	28.8
	Zinc/Metal	3343	65.1	3285	65.1	4292	71.2	4289	71.2
Wealth index	Poor	1901	37.0	1856	36.8	2421	40.2	2419	40.2
	Middle	1160	22.5	1136	22.5	1259	20.9	1258	20.9
	Rich	2086	40.5	2064	40.7	2345	38.9	2344	38.9
Child's age in months	6 - 14 months	932	18.1	913	18.0	975	16.2	974	16.2
	15 -23 months	707	13.7	696	13.8	934	15.5	934	15.5
	24 - 32 months	904	17.6	885	17.5	992	16.5	991	16.5
	33 - 41 months	884	17.2	865	17.1	1089	18.1	1088	18.1
	42 - 50 months	851	16.5	845	16.7	974	16.1	974	16.2
	51 - 59 months	869	16.9	852	16.9	1061	17.6	1060	17.5
Sex	Male	2604	50.6	2554	50.5	3036	50.4	3035	50.4
	Female	2543	49.4	2502	49.5	2989	49.6	2986	49.6
Mother's highest educational level	No education	2007	46.1	1951	45.7	2305	43.5	2304	43.5
	Primary	903	20.7	891	20.8	971	18.3	971	18.3
	Secondary	1236	28.4	1227	28.7	1612	30.4	1610	30.4
	Higher	207	4.8	207	4.8	407	7.8	407	7.8

Table 5.3: Univariate exploratory analysis of all the children tested for malaria and anaemia in 2010

Variables	Category	Malaria			Anaemia		
		Yes	No	P-value	Yes	No	P-value
Type of place of residence	Urban	450(8.7)	956(18.6)	0.000	852(16.9)	529(10.5)	0.000
	Rural	1978(38.4)	1763(34.3)		2660(52.5)	1015(20.1)	
Source of drinking water	Well water	1459(28.3)	1704(33.2)	0.306	2160(42.8)	937(18.6)	0.000
	Tap/Other water	966(18.8)	1010(19.7)		1345(26.6)	606(12.0)	
Type of toilet facility	Pit	1199(23.3)	1453(28.3)	0.012	1787(35.4)	783(15.5)	0.000
	Flush/Other	1224(23.9)	1261(24.5)		1716(34.0)	760(15.1)	
Has electricity	Yes	891(17.4)	1530(29.8)	0.000	1582(31.4)	788(15.6)	0.000
	No	1526(29.8)	1180(23.0)		1914(38.1)	752(14.9)	
Has radio	Yes	1661(32.4)	1997(38.9)	0.000	2416(48.0)	1180(23.4)	0.000
	No	759(14.8)	712(13.9)		1082(21.5)	360(7.1)	
Has television	Yes	778(15.1)	1383(27.0)	0.000	1359(27.0)	710(14.1)	0.000
	No	1643(32.1)	1321(25.8)		2137(42.5)	828(16.4)	
Main floor material	Earth/Other	1336(26.0)	1194(23.2)	0.000	1698(33.6)	771(15.3)	0.000
	Cement/Ceramics	1088(21.2)	1520(29.6)		1806(35.8)	772(15.3)	
Main wall material	Wood/Other	1559(30.3)	1328(25.8)	0.000	1959(38.8)	863(17.1)	0.000
	Cement/Bricks	866(16.9)	1386(27.0)		1545(30.6)	681(13.5)	
Main roof material	Wood/Other	990(19.3)	805(15.7)	0.000	1240(24.7)	522(10.3)	0.000
	Zinc/Metal	1435(27.9)	1908(37.1)		2263(44.8)	1022(20.2)	
Wealth index	Poor	1090(21.2)	811(15.7)	0.000	1286(25.4)	570(11.3)	0.009
	Middle	609(11.8)	551(10.7)		837(16.6)	299(5.9)	
	Rich	729(14.2)	1357(26.4)		1389(27.4)	675(13.4)	
Child's age in months	6 - 23 months	655(12.7)	984(19.1)	0.000	1241(24.6)	368(7.3)	0.000
	24 -41 months	863(16.8)	925(18.0)		1223(24.2)	527(10.4)	
	42 - 59 months	910(17.7)	810(15.7)		1048(20.7)	649(12.8)	
Sex	Male	1224(23.8)	1380(26.8)	0.475	1832(36.2)	722(14.3)	0.005

Mother's highest educational level	Female	1204(23.4)	1339(26.0)		1680(33.2)	822(16.3)	
	No education	1079(24.8)	928(21.3)	0.000	1394(32.7)	557(13.0)	0.000
	Primary	457(10.5)	446(10.2)		622(14.5)	269(6.3)	
	Secondary	440(10.1)	796(18.3)		832(19.5)	395(9.2)	
	Higher	58(1.3)	149(3.5)		133(3.1)	74(1.7)	

Table 5.4: Univariate exploratory analysis of all the children tested for malaria and anaemia in 2015

Variables	Category	Malaria			Anaemia		
		Yes	No	P-value	Yes	No	P-value
Type of place of residence	Urban	487(8.1)	1555(25.8)	0.000	1128(18.8)	912(15.1)	0.000
	Rural	2139(35.5)	1844(30.6)		2934(48.7)	1047(17.4)	
Source of drinking water	Well water	1639(27.2)	1923(31.9)	0.000	2474(41.1)	1084(18.0)	0.000
	Tap/Other water	987(16.4)	1476(24.5)		1588(26.4)	875(14.5)	
Type of toilet facility	Pit	1182(19.6)	1447(24.0)	0.000	1872(31.1)	757(12.6)	0.000
	Flush/Other	1442(24.0)	1951(32.4)		2188(36.4)	1201(19.9)	
Has electricity	Yes	804(13.3)	1822(30.2)	0.000	1706(28.4)	1177(19.5)	0.000
	No	2080(34.6)	1319(21.9)		2356(39.1)	782(13.0)	
Has radio	Yes	1375(22.7)	1251(20.8)	0.000	2316(38.5)	1307(21.7)	0.000
	No	2251(37.4)	1148(19.1)		1746(29.0)	652(10.8)	
Has television	Yes	740(12.3)	1886(31.3)	0.000	1599(26.6)	1160(19.3)	0.000
	No	2020(33.5)	1379(22.9)		2463(40.9)	799(13.2)	
Main floor material	Earth/Other	1463(24.3)	1200(19.9)	0.000	1983(32.9)	678(11.3)	0.000
	Cement/Ceramics	1163(19.3)	2199(36.5)		2079(34.5)	1281(21.3)	
Main wall material	Wood/Other	1427(23.7)	952(15.8)	0.000	1847(30.7)	530(8.8)	0.000
	Cement/Bricks	1199(19.9)	2447(40.6)		2215(36.8)	1429(23.7)	
Main roof material	Wood/Other	945(15.7)	788(13.1)	0.000	1274(21.2)	458(7.6)	0.000
	Zinc/Metal	1681(27.9)	2611(43.3)		2788(46.3)	1501(24.9)	

Wealth index	Poor	1513(25.1)	908(15.1)	0.000	1883(31.3)	536(8.9)	0.000
	Middle	598(9.9)	661(11.0)		875(14.5)	383(6.4)	
	Rich	515(8.5)	1830(30.4)		1304(21.7)	1040(17.2)	
Child's age in months	6 - 23 months	646(10.8)	1263(21.0)	0.000	1466(24.3)	442(7.3)	0.000
	24 -41 months	929(15.4)	1152(19.1)		1369(22.7)	710(11.8)	
	42 - 59 months	1051(17.4)	984(16.3)		1227(20.4)	807(13.5)	
Sex	Male	1359(22.6)	1677(27.8)	0.063	2098(34.8)	937(15.6)	0.002
	Female	1267(21.0)	1722(28.6)		1964(32.6)	1022(17.0)	
Mother's highest educational level	No education	1322(25.0)	983(18.6)	0.000	1759(33.2)	545(10.3)	0.000
	Primary	430(8.1)	541(10.2)		668(12.6)	303(5.7)	
	Secondary	485(9.2)	1127(21.3)		976(18.4)	634(12.0)	
	Higher	50(0.9)	357(6.7)		183(3.5)	224(4.3)	

5.3 Statistical methods

Assuming the response variable y_{ijk} is the status (0/1) of disease k , where malaria is $k = 1$ and anaemia is $k = 2$, for child j in state i : $i = 1, \dots, 37$. The notation that child j in state i tests positive for malaria is $y_{ij1} = 1$ and zero otherwise and $y_{ij2} = 1$ if child j in state i test positive for anaemia and zero otherwise is used. For this study, y_{ijk} follows a bivariate Bernoulli distribution with π_{ij1} and π_{ij2} being the respective probability of a child j testing positive for malaria or anaemia. Therefore, the generalized linear model for each respective outcome is defined as;

$$g_1(\pi_{ij1}) = \beta_{01} + \mathbf{x}_{ij1}\boldsymbol{\beta}_1 + Z_{ij1}r_1 \quad (5.1)$$

$$g_2(\pi_{ij2}) = \beta_{02} + \mathbf{x}_{ij2}\boldsymbol{\beta}_2 + Z_{ij2}r_2 \quad (5.2)$$

where β_{01} and β_{02} are the respective model intercepts, the assumed vectors of fixed effects are β_1 and β_2 , $Z_{ij1}r_1$ and $Z_{ij2}r_2$ represent the Gaussian random field, and the designed vectors for fixed effects are x_{ij1} and x_{ij2} .

$x_{ijk} = (x_{ij1}, x_{ij2}, \dots, x_{ijq})'$ contains q continuous independent variables. Here, $q = 1$ (age) since age is the only continuous variable in this study and $v = 12$.

A multivariate Gaussian distribution $r \sim N(0, \Sigma)$ was assumed for the Gaussian random field where the covariance matrix is Σ . The constituents of Σ (covariance matrix) are expressed as a function of the marginal variance of the process σ_Z and the Matern correlation function Cor_M as shown below:

$$\Sigma_{ik} = \sigma_Z Cor_M(Z(r_i), Z(r_k)),$$

The Matern correlation function is defined as

$$Cor_M(z(r_i), z(r_k)) = \frac{2^{1-v}}{\Gamma(v)} (\mathbf{u} \parallel r_i - r_k \parallel)^v U_v(\mathbf{u} \parallel r_i - r_k \parallel) \quad (5.3)$$

$\parallel \cdot \parallel$ is the Euclidean distance while the modified Bessel function of second order is U_v , u and v are respectively scale parameter and smoothness parameter.

The inverse Gamma priors were given as u, v and σ_Z . By assumption, the study has weakly informative Gaussian prior $\beta \sim N(0, \tau_\beta^{-1})$ for fixed effect parameter β with small precision τ_β on identity matrix. In the data, the continuous predictor's non-linear effects and the spatial autocorrelation were handled by adopting the convolution model and the penalized regression spline approach under a semi-parametric model.

The highly restrictive linear predictor was relaxed by penalized regression spline approach to a better flexible semi-parametric predictor, expressed as:

$$g(\pi_{ij1}) = \beta_{01} + \sum_{k=1}^q f_k(x_{ijk}) + f_{spat}(s_{i1}) + Z^T s_i \text{ for malaria} \quad (5.4)$$

and

$$g(\pi_{ij2}) = \beta_{01} + \sum_{k=1}^q f_k(x_{ijk}) + f_{spat}(s_{i2}) + Z^T s_i \text{ for anaemia} \quad (5.5)$$

The non-linear twice differentiable smooth function for the continuous covariate is the function $f_t(.)$ while the factor that takes care of the spatial effect of each state is $f_{spat}(s_i)$. The convolution model was implemented in this study, and it assumes that the spatial effect consists of spatially structured and spatially unstructured components i.e., $f_{spat}(s_{ik}) = f_{str}(s_{ik}) + f_{unstr}(s_{ik}), k = 1, 2$ (Ngesa et al., 2014; Manda and Leyland, 2007; Okango et al., 2015). Cultures, common cultural practices, climate etc., are called unobserved predictors inherent within the states or the relationship within the states which are under spatially unstructured random effects. Alternatively, any unobserved predictors which varies spatially across the states can be explained by spatially structured random effects. This is known as spatial autocorrelation, though technically, it is expressed as the dependence (i.e., relationship between two nearby observations) because of geographical nearness .

Finally, in order to model malaria and anaemia jointly, a spatial shared component with one shared component common to malaria and anaemia was used. Based on the bivariate model which pools the two datasets together, the vectors connecting all observations for the two outcome variables were incorporated in

$$Y = \begin{bmatrix} y_{11} & NA \\ \vdots & \vdots \\ y_{n_i1} & NA \\ NA & y_{12} \\ \vdots & \vdots \\ NA & y_{n_i2} \end{bmatrix}$$

where n_{ik} is the number of observations for each outcome variable $k = 1, 2$. Therefore, by definition, the joint (bivariate) models are expressed as;

$$g(\pi_{ij1}) = \beta_{01} + \sum_{k=1}^q \beta_{j1}(x_{ij1}) + f_1(s_{i1}) + Z_1(s_i) \quad (5.6)$$

$$g(\pi_{ij2}) = \beta_{02} + \sum_{k=2}^q \beta_{j2}(x_{ij2}) + f_2(s_{i2}) + Z_2(s_i) + \tau h_1(s_i) \quad (5.7)$$

where vector \mathbf{x} of linear covariates is assigned to each outcome with corresponding regression parameters β_{jk} , s_{ik} is the vector of ages assumed to follow a random walk of order 1; the Gaussian random field shared by both outcomes is $h_1(s_i)$, while τ (interaction parameter) which links the two outcome variables also explains how well the structure captured in $h_1(s_i)$ is also inherent in $g(\pi_{ij2})$. In the same manner, prior distributions specified for univariate models were used for parameters and hyperparameters of the joint model.

5.4 Estimation of parameter

Estimation was done using a full Bayesian technique, and suitable prior distributions were given to the parameters. Basically, the joint posterior distribution is specified as;

$$p(\theta|y) \propto \pi(\theta).p(y|\theta)$$

where $\theta = (\theta_1, \dots, \theta_b)$ are the parameters of interest.

5.5 The Penalized regression splines

The approaches for assessing the smooth function $f_t(.)$ have been discussed to a reasonable extent by some studies (Hastie et al., 2001; Fahrmeir et al., 1994; Hall and Patil, 1995). The Penalized regression spline suggested by Eilers and Marx, (1996) was used in this study. Now, assuming the polynomial spline is utilized, then approximating the effect of the continuous predictors will be necessary. They assumed that using a spline that has a degree d with K knots of equal spacing,

$x_{q,min} = \varphi_{q1} < \varphi_{q2} \dots \varphi_{qk-1} < \varphi_{qK} = x_{q,max}$ giving:

$$f(x, \theta) = \vartheta_0 + \vartheta_1 x + \dots \vartheta_q x^d + \sum_{k=1}^k b_k (x - \vartheta_k)_+^d, \quad (5.8)$$

where $\theta = (\vartheta_0, \vartheta_1, \dots, \vartheta_q, b_1, b_2, \dots, b_k)'$ and $(\Delta - \Psi)_+$ is the same as $(\Delta - \Psi)$ if $(\Delta - \Psi)$ is positive(+) and zero(0) otherwise, the smooth function $f_t(.)$ can be estimated. Δ is the predictor variable, Ψ is the knot location and the subscript denotes the positive part of the argument.

In this study, a quadratic spline ($d = 2$) comprising of 20 knots was employed to guarantee flexibility and the k^{th} knot was expressed as the sample quantile of the continuous independent variables obtained by the probability equal to $\frac{k}{k+1}$. A roughness penalty $-\frac{1}{2} \alpha \int_{x_{min}}^{x_{max}} [f''(x)]^2 dx$ added to the log likelihood as a penalty term to prevent overfitting which wriggles to a large extent was suggested by Green and Silverman, (1993). This gives the penalized log-likelihood function defined as;

$L = l(y, \theta, \delta) - \frac{1}{2} \alpha \int_{x_{min}}^{x_{max}} [f''(x)]^2 dx$, where α , a smoothing parameter, balances smoothness and flexibility.

5.6 Distribution of prior

A multivariate Gaussian Markov random field (GMRF) prior was used to spatially structured random effects $f_{str}(s_i) = (f_{str}(s_{i1}), f_{str}(s_{i2}))^T$ based on nearby neighbour structure. The multivariate version of GMRF was defined by letting $X = (X_1, \dots, X_n)$ be a multivariate Gaussian random vector. The spatially structured random effects are assumed to follow a multivariate conditional autoregressive (MCAR) model, as expressed below;

$$f_{str}(s_i, s_k) \sim MCAR(\mathbf{1}, \Sigma)$$

where Σ denotes the covariance matrix causing the association.

5.7 Distribution of posterior

The posterior distribution is acquired by putting the observed data to update the prior distribution. Therefore, it can be referred to as parameter distribution after data observation. The samples for Bayesian inference are given by posterior distribution. The problem of high dimensionality is avoided by Markov Chain Monte Carlo (MCMC) because it can permit repeated sampling to be done directly from the posterior distribution. At the same time, the mean and median as estimates are computed from the data sample summaries.

By assumption, let the Conditional Independence (i.e., reductant observation) connects the dependent variable and the hyperparameters; for the Bernoulli model, the posterior distribution is defined as:

$$Q_{post}(\vartheta, \alpha, b, \tau^2 / y) \propto L(y / \vartheta, \alpha, b, \tau^2) Q_{pri}(\vartheta, \alpha, b, \tau^2) \quad (5.9)$$

$$\begin{aligned}
&= \prod_i \prod_k L(y_{ik} | \theta, \alpha, \tau^2) \prod_{l=1}^q [Q(b_l / \tau_l^2) Q(\tau_l^2)] \\
&\quad \times \prod_k^v [Q(\delta_k / \tau_k^2) Q(\tau_k^2)] \times Q(f_{str} / \tau_{str}^2) Q(\tau_{str}^2) Q(f_{unstr} / \tau_{unstr}^2) Q(\tau_{unstr}^2)
\end{aligned}$$

WinBUGS 14 was used to carry out all the analyses in this study (Spiegelhalter et al. 2003, Okango et al., 2015). We carried out 20000 iterations of Markov Chain Monte Carlo (MCMC) to execute each model while the 10000 iterations done earlier was removed to take care of the burn-in period. The other 10,000 were used to sample from the posterior distribution for the parameter estimation and Markov Chain Monte Carlo (MCMC) meeting point.

5.8 Diagnostics of model

The DIC, known as deviance information criteria proposed by Spiegelhalter et al. (2003), was used to compare models, and the model that had the least DIC was regarded as the best fitting. The value for DIC is obtained using the following formula: $DIC = \bar{D}(\theta) + pD$, $\bar{D}(\theta)$ being the posterior mean of the deviance utilized to measure the goodness of fit. At the same time, the effective number of parameters in the model which penalizes for the complexity of the model is pD . While model parsimony is denoted by small values of pD which is calculated by summing the intraclass correlation coefficients. Low value of $\bar{D}(\text{DIC})$ shows a better fit.

5.9 Data Analysis

This research investigated four models to ascertain the predictors' effects, some of the explanatory variables that are unobservable on the distribution and the correlation between childhood malaria and anaemia in Nigeria. Researchers have discussed at length on these models (generalized linear model, generalized additive model, multivariate CAR model and convolution model), basically on

their benefits over classical models (Wand et al., 2011; Kammann and Wand, 2003; Hastie and Tibshirani, 1995; Okango et al., 2015). The fitted four models in this study are:

Model 1: This model comprises fixed categorical predictors and one continuous variable. The latter variable is modelled using a non-linear smooth function. Using a non-smoothing prior to model age was supported by results from Montana et al., (2007); Johnson and Way, (2006). Here, the two diseases were modelled independently. Their respective models are expressed as:

$$\text{logit}(Q_{ik1}) = \beta_{01} + f(\text{age}) + z^T s_i \text{ for malaria}$$

and

$$\text{logit}(Q_{ik2}) = \beta_{02} + f(\text{age}) + z^T s_i \text{ for anaemia}$$

Model 2: This is an additive model that assumes linear fixed effects of categorical variables mentioned in Model 1, non-linear effect of child's age in months and spatially unstructured random effect which cover the unobserved predictors that are innate within the states. Here, multivariate normal distribution initiates the joint model. The model is defined as:

$$\text{logit}(Q_{ik1}) = \beta_{01} + f(\text{age}) + z^T s_i + f_{unstr}(s_{i1}) \text{ for malaria}$$

and

$$\text{logit}(Q_{ik2}) = \beta_{02} + f(\text{age}) + z^T s_i + f_{unstr}(s_{i2}) \text{ for anaemia.}$$

for the respective diseases.

Model 3: Two states are said to be neighbours if they share a boarder i.e., spatial wise but otherwise, they are not. The univariate type of MCAR and conditions that guide the conditional multivariate distributions to distinctively ascertain the equivalent multivariate joint probability

density function were considered respectively by Besag et al., (1991) and Mardia, (1988). Their results developed the Multivariate Conditional Autoregressive Model (MCAR) (Carlin and Banerjee, 2003). Assumedly, the unstructured spatial effects follow a Multivariate Gaussian prior, i.e., $f_{unstr}(s_i, s_k)/\tau_{unstr}^2 \sim MVN(0, \tau_{unstr}^2)$. The allocation of inverse gamma distributions to the variance hyperparameter is as follows:

$$\tau_{str}^2 \sim IG(0.0001, 0.0001) \text{ and } \tau_{unstr}^2 \sim IG(0.0001, 0.0001),$$

while the priors distributions below were allocated to the fixed effects coefficients:

$$\varrho_0, \varrho_1, \dots, \varrho_q \sim N(0, 10^6), \delta_1, \alpha_2, \dots, \alpha_v \sim N(0, 10^6), b_j \sim N(0, \tau_b^2) \quad \text{and} \\ \tau_b^2 \sim IG(0.0001, 0.0001), \beta_{01}, \beta_{02} \sim N(0.01, 0.01) \text{ is the intercept}$$

This model generally comprises model 1 and spatially structured random effect, which explains any unobserved predictors that vary spatially among states. The multivariate conditional autoregressive model initiates the joint modelling, and this is represented mathematically as:

$$\text{logit}(Q_{ik1}) = \beta_{01} + f(\text{age}) + z^T s_i + f_{str}(s_{i1}) \text{ for malaria}$$

and

$$\text{logit}(Q_{ik2}) = \beta_{02} + f(\text{age}) + z^T s_i + f_{str}(s_{i2}) \text{ for anaemia.}$$

Model 4: Usually, the joint modelling of two diseases through the random effects by a few likely methods such as the Multiple Membership Multiple Classification (MMMC), Multivariate Conditional Autoregressive Model (MCAR) and Multivariate Normal Distribution (MVN) methods. The random effects are divided into structured random effects that cater for unobserved covariates that differ spatially throughout the states and unobserved covariates inherent within the states that are taken care of by unstructured random effects under study. Therefore, this model is

an inclusion of the convolution model, the non-linear effects of child's age in months and the linear effects of the categorical predictors. Here, the multivariate normal distribution and the multivariate conditional autoregressive model initiate the joint modelling and it is expressed as:

$$\text{logit}(Q_{ik1}) = \beta_{01} + f(\text{age}) + z^T s_i + f_{unstr}(s_{i1}) + f_{str}(s_{i1}) \text{ for malaria}$$

and

$$\text{logit}(Q_{ik2}) = \beta_{02} + f(\text{age}) + z^T s_i + f_{unstr}(s_{i2}) + f_{str}(s_{i2}) \text{ for anaemia.}$$

Table 5.5: Summary of all constituents for each model used in the study

Models	Non-linear effect of child's age in months	Linear effects of categorical predictors	Spatially unstructured random effects	Spatially unstructured random effects
M1	P	P	NP	NP
M2	P	P	P	NP
M3	P	P	NP	P
M4	P	P	P	P

*P means Present

*NP means Not Present

5.10 Results

5.10.1 Model comparison

The total datasets sizes used in this study were 4240 children in 2010 and 5289 children in 2015, respectively. The summary of all constituents for each variable is shown in Table 5.5. Model 1 investigates the linear and non-linear effects of the predictors; model 1 was broadened by Model 2 to incorporate spatially unstructured random effects, while Model 3 broadens Model 1 by including spatially structured random effects. Lastly, Model 4 comprises the convolution model.

The diagnostics for the four models are presented in Table 5.6 and the model that has the least DIC is chosen as the best fit. Nevertheless, some research works maintained that two models cannot be

distinguished if their difference in DIC is 3, but a DIC value between 3 and 7 can be said to have a weak difference (Spiegelhalter et al., 2002; Kazembe et al., 2008). Based on this, Model 3 and Model 4 are not distinguishable for both years because their difference in DIC is below 3. Hence, results from Model 4 were used for discussion in 2010 and 2015 as it considers both spatially structured and unstructured random effects.

Table 5.6: 2010 and 2015 comparative fit statistics of malaria and anaemia

	2010				2015			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
DIC(Malaria)	5564.96	5125.11	5126.08	5124.01	6237.16	6021.47	6019.48	6045.56
pD(Malaria)	18.79	51.83	53.09	52.17	19.38	49.01	49.73	52.02
DIC(Anaemia)	4984.65	4668.52	4654.3	4655.31	6171.94	6052.75	6050.31	6053.2
pD(Anaemia)	19.29	52.43	52.03	52.35	19.45	47.01	46.85	48.02

5.11 Fixed effects

Table 5.5 and Table 5.6 provide the adjusted posterior odds ratio estimates (AOR) and 95% confidence intervals (CI) for the categorical predictors. The effects on malaria and anaemia are assumed linear and this is determined from model 4 for 2010 and 2015, respectively. The discussion will be based on the significant predictors.

For 2010, type of place of residence, mother's highest educational level, source of drinking water, type of toilet facility, child's sex, main floor material, households that have electricity, radio and television water were significantly associated with malaria and anaemia. While in 2015, type of place of residence, source of drinking water, type of toilet facility, households with radio, main roof material, wealth index, child's sex, and mother's highest educational level had significant relationship with malaria and anaemia.

5.12 Malaria (2010 and 2015)

In 2010, female children had increased odds of malaria [OR=1.276, 95%CI=(1.109, 1.482)] compared to their male counterparts. While for type of place of residence, the odds of malaria among children who dwell in urban areas decrease significantly. Also, the decrease in the odds of malaria among children that have radio [OR=0.675, 95%CI=(0.556, 0.8161)] and television [OR=0.64, 95%CI=(0.7008, 0.817)] in their houses was significant. Concerning 2015, type of place of residence (urban), source of drinking water (well water), and households with radio had lower odds of malaria. Alternatively, the odds of malaria increased significantly among female children and those from rich households in terms of wealth index. At the same time, there was a significant decrease in the odds of malaria regarding the mother's highest educational level. These are presented in Tables 5.7 and 5.8, respectively.

Table 5.7: Adjusted posterior odds ratio estimates (AOR) and 95% credible interval (95%CI) of malaria and anaemia for 2010

Covariates	Malaria		Anaemia	
	AOR	95%CI	AOR	95%CI
Type of place of residence (ref = Rural)				
Urban	0.652*	(0.544, 0.787)	0.713*	(0.592, 0.858)
Source of drinking water (ref = Tap/Other water)				
Well water	1.056	(0.899, 1.234)	1.246*	(1.071, 1.445)
Type of toilet facility (ref = Flush/Other)				
Pit toilet	1.085	(0.915, 1.293)	1.23*	(1.041, 1.446)
Has electricity (ref = No)				
Yes	0.927	(0.739, 1.171)	0.769*	(0.625, 0.947)
Has radio (ref = No)				
Yes	0.675*	(0.556, 0.8161)	1.051	(0.888, 1.241)
Has television (ref = No)				
Yes	0.64*	(0.7008, 0.817)	0.811	(0.6487, 1.017)
Main floor material (ref = Earth/Other)				
Concrete/Ceramics	1.034	(0.829, 1.293)	1.137	(0.926, 1.411)
Main wall material (ref = Wood/Other)				
Cement/Bricks	1.065	(0.809, 1.446)	1.151	(0.892, 1.504)
Main roof material (ref = Wood/Other)				
Zinc/Metal	1.143	(0.5856, 1.379)	1.151	(0.9692, 1.379)
Wealth index (ref = Poor)				
Middle	0.651	(0.405, 1.040)	1.344	(0.914, 2.117)
Rich	0.926	(0.666, 1.276)	1.295	(0.982, 1.782)
Sex (ref = Male)				
Female	1.276*	(1.109, 1.482)	0.956	(0.831, 1.092)
Mother's highest educational level (ref = No education)				
Primary	1.187	(0.815, 1.762)	1.912*	(1.252, 2.883)
Secondary	0.986	(0.687, 1.429)	1.966*	(1.319, 2.980)
Higher	0.966	(0.694, 1.384)	1.295	(0.869, 1.921)

5.13 Anaemia (2010 and 2015)

From Table 5.7 and Table 5.8, the results from 2010 shows that type of place of residence (urban) [OR=0.713, 95%CI=(0.592, 0.858)], households with electricity [OR=0.769, 95%CI=(0.625, 0.947)], and mother's highest educational level [higher; OR=1.295, 95%CI=(0.869, 1.921)] had a lower odd of anaemia. But source of drinking water (well water) [OR=1.246, 95%CI=(1.071,

1.445)] and type of toilet facility (pit toilet) [OR=1.23, 95%CI=(1.041, 1.446)] had higher odds of anaemia. In 2015, the odds of anaemia for the type of toilet facility (pit toilet) [OR=1.34, 95%CI=(1.151, 1.549)] increased significantly. Though the odds of anaemia for the wealth index (rich) significantly increased, other categories were insignificant. Regarding the type of place of residence (urban), source of drinking water (tap/other water), main roof material (zinc/metal) and mother's highest educational level had a significant decrease in the odds of anaemia.

Table 5.8: Adjusted posterior odds ratio estimates (AOR) and 95% credible interval (95%CI) of malaria and anaemia for 2015

Covariates	Malaria		Anaemia	
	AOR	95%CI	AOR	95%CI
Type of place of residence (ref = Rural)				
Urban	0.598*	(0.503, 0.711)	0.489*	(0.410, 0.584)
Source of drinking water (ref = Tap/Other water)				
Well water	0.819*	(0.715, 0.933)	0.839*	(0.730, 0.964)
Type of toilet facility (ref = Flush toilet)				
Pit /Other	0.9	(0.778, 1.030)	1.34*	(1.151, 1.549)
Has electricity (ref = No)				
Yes	0.987	(0.808, 1.199)	0.899	(0.747, 1.076)
Has radio (ref = No)				
Yes	0.849*	(0.736, 0.973)	0.872	(0.757, 1.004)
Has television (ref = No)				
Yes	0.914	(0.724, 1.145)	0.976	(0.790, 1.189)
Main floor material (ref = Earth/Other)				
Concrete/Ceramics	1.01	(0.857, 1.193)	1.105	(0.924, 1.321)
Main wall material (ref = Wood/Other)				
Cement/Bricks	1.199	(0.979, 1.455)	1.065	(0.888, 1.279)
Main roof material (ref = Wood/Other)				
Zinc/Metal	1.028	(0.855, 1.234)	0.841*	(0.705, 0.999)
Wealth index (ref = Poor)				
Middle	1.284	(0.919, 1.764)	1.215	(0.963, 1.532)
Rich	2.648*	(1.912, 3.561)	1.883*	(1.489, 2.349)
Sex (ref = Male)				
Female	1.283*	(1.132, 1.449)	1.098	(0.971, 1.239)
Mother's highest educational level (ref = No education)				
Primary	2.022*	(1.535, 2.672)	2.98*	(2.027, 4.293)
Secondary	1.648*	(1.247, 2.167)	2.314*	(1.563, 3.353)
Higher	1.436*	(1.125, 1.827)	2.238*	(1.591, 3.200)

5.14 Non-linear effects of age

The respective non-linear correlation connecting a child's age in month and malaria and a child's age in month and anaemia are shown in Figures 5.1 and 5.2, respectively. The smooth function based on the posterior mean and their equivalent 95% CI are shown in Figure 5.1 and Figure 5.2.

In both years, it can be seen from the Figures that the relationship between age and malaria or

anaemia infection is non-linear. For malaria, the spread of malaria infection with age was higher in children from aged 8 to 17 months in 2010 while in 2015, children from aged 6 to 14 months were mostly affected. The results in both years show a reduction in the spread of malaria infection for children within aged 20 months and above. On the other hand, children from aged 8 to 30 months were mostly affected by anaemia infection in 2010 and 2015. It can be concluded from the results that malaria and anaemia diseases are more in younger children than in older children. Reason being that as a child grows, it gets to a point where the child becomes asymptomatic to malaria thereby causing reduction in the spread of anaemia (White, 2018).

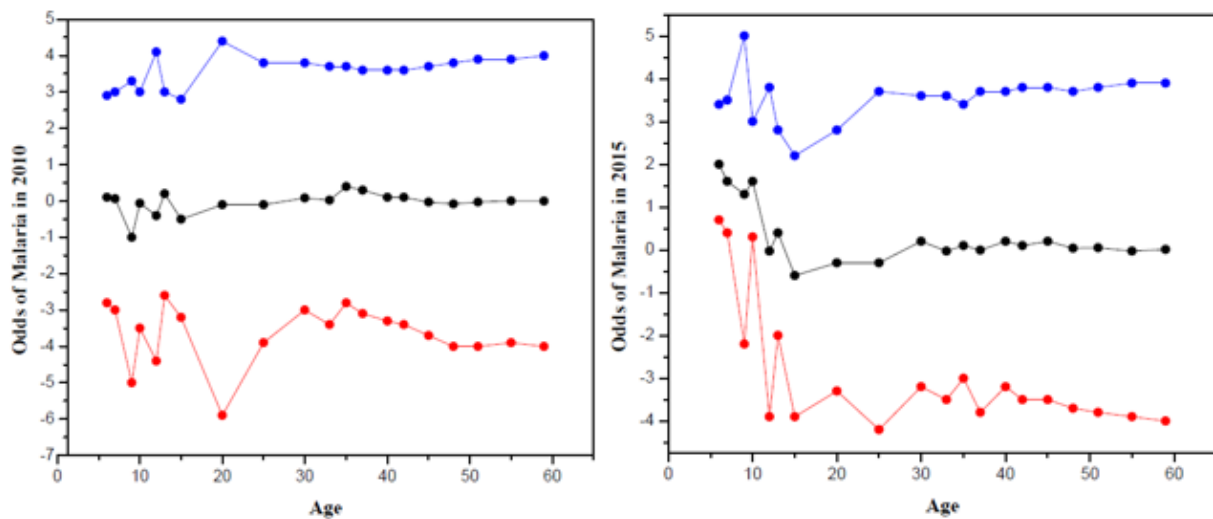


Figure 5.1. Estimated mean of the non-linear effect of age (black) on malaria infection and the corresponding 95% credible interval (red and blue) for 2010 and 2015.

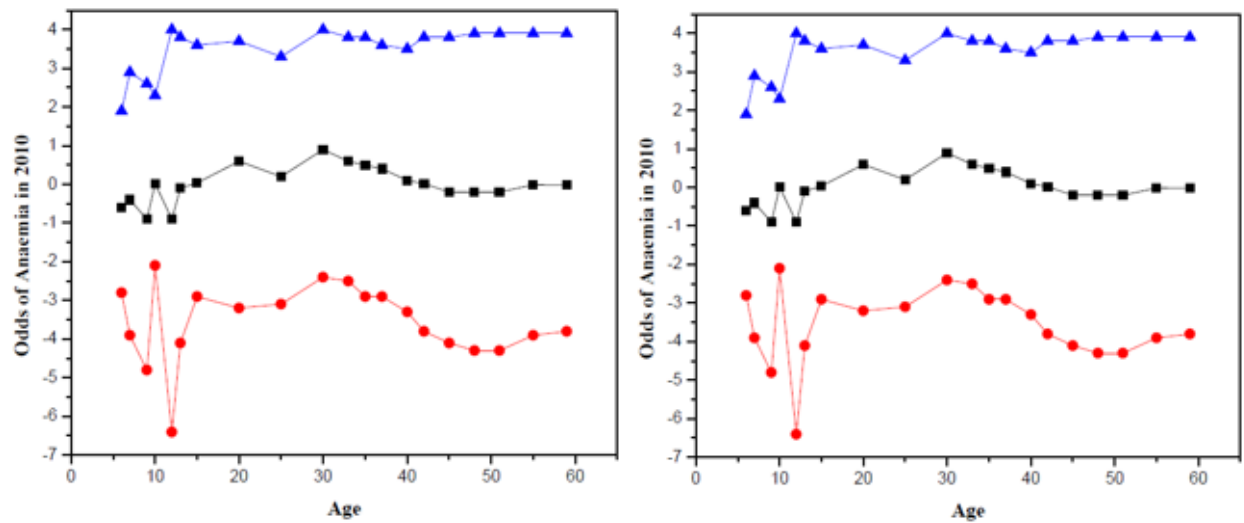


Figure 5.2. Estimated mean of the non-linear effect of age (black) on anaemia infection and the corresponding 95% credible interval (red and blue) for 2010 and 2015.

5.15 Joint spatial effects of malaria and anaemia

The spatial effects are presented in Figures 5.3 and 5.4 and they are based on Model 4. States that have dark red shading indicate high relationship between malaria and anaemia infection whereas light red shading show that the relationship between malaria and anaemia is low in 2010 and 2015. In 2010, there was a high prevalence of malaria in states within the North West, South West and part of middlebelt. While in 2015, the spread of malaria infection was more in North West, South West and South East. For anaemia infection, states in the North West, North Central, South South, South East and middle belt had high anaemia prevalence in 2010 while in 2015, North West, South South, South East and middle belt were most endemic with anaemia. From Figures 5.3 and 5.4, the spread of anaemia infection was more than malaria in 2010 as well as in 2015.

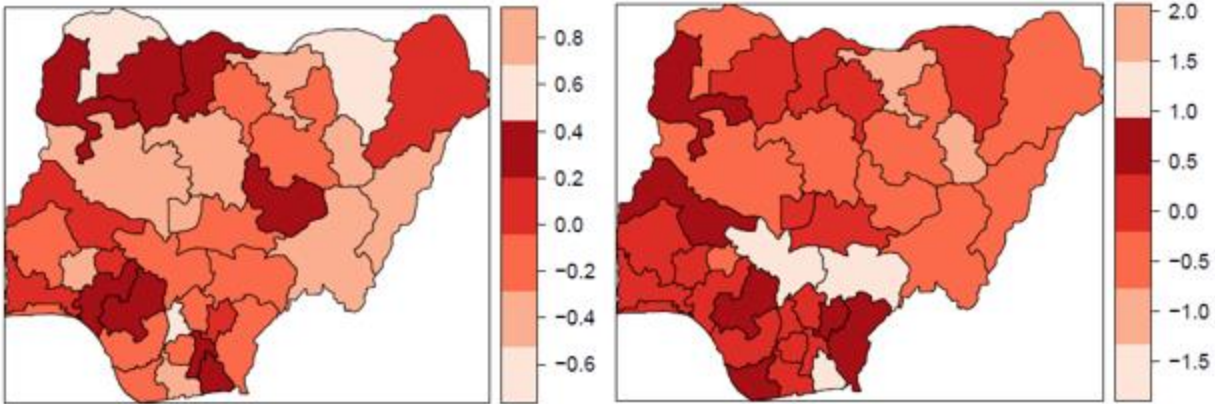


Figure 5.3. Residual spatial effect of state on malaria infection in 2010 and 2015

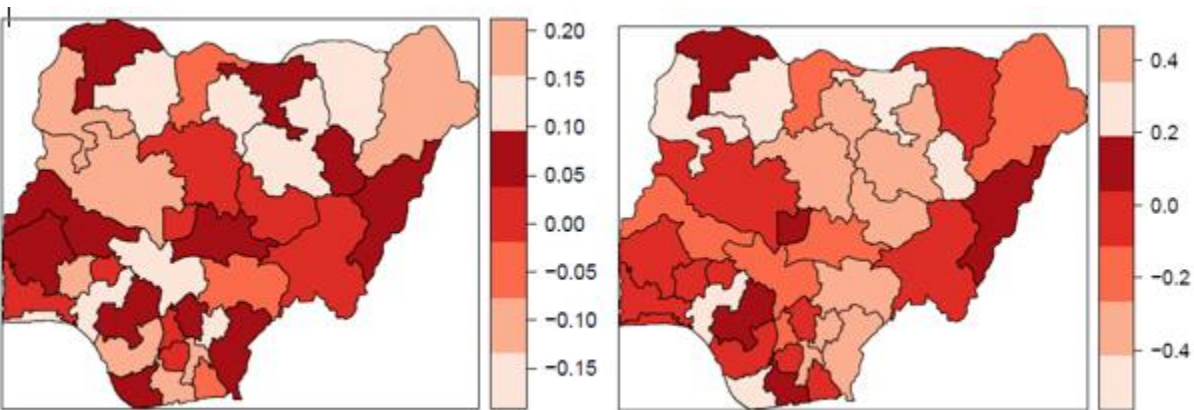


Figure 5.4. Residual spatial effect of state on anaemia infection in 2010 and 2015

5.16 Discussion

This study was carried out using a semi-parametric joint modelling under a fully Bayesian approach to observe the joint association of malaria and anaemia infection among Nigerian children from 6 to 59 months. Both diseases have been a major health challenge especially to children and there are similarities in their risk factors at various degrees among children and geographical locations. Specifically, the methods above were used to investigate the regional differences between malaria and anaemia risk factors and the relationship between the two diseases. This study was extended from the works of Kazembe et al., (2008) on the creation and penalized likelihood of B-splines and on semi parametrization regression by Wand et al., (2011). The penalized regression splines in a semi-parametric model exemplar were used to model the

non-linear effects by permitting variation spatially in the response variables. Due to the limiting and unrealistic assumption of the linearity amid the response variable and the predictors, disagreeing results in several instances cannot be avoided. According to Besag et al., (1991), semi parametric models are better because they are more flexible i.e., they have the ability to bring together both semi-parametric and parametric models thereby improving standard parametric models by reconnoitering the non-parametric domain at the same time ensuring that the linear structure is whole (Besag et al., 1991).

There was a non-linear relationship between age with malaria or anaemia. Though there were fluctuations in the likelihood of malaria infection with age for both 2010 and 2015, children within age 6 to 40 months were mostly infected by malaria. On the other hand, children from age 11 to 35 months were mostly affected by anaemia infection in both years. Notwithstanding, there was a slight increase in anaemia infection for 55 months old children. These results are inline with other studies (Adebayo et al., 2016; Gayawan et al., 2022; Egbon, 2021).

A Gaussian Markov Random Field (GMRF) modelled the spatially structured random effects, while a zero mean Gaussian process modelled the spatially unstructured random effects in the model (Besag et al., 1991; Kazembe et al., 2008). Johnson and Way, 2006 and Langford et al., (1999) have proposed Bayesian and non-Bayesian approaches for joint disease modeling. Because of the high complexity and intractability of maximum likelihood i.e., the frequentist approach, the workability in these models is not obtainable. Therefore, the Bayesian inference under MCMC methods is most needed here (Ngesa et al., 2014; Okango et al., 2015). Bayesian approach within the MCMC algorithm is easier to implement than frequentist approach. Bayesian approach permits complex and flexible hierarchical modeling and as well make provision for better estimates and predictions for several credible epidemiological hitches. Whereas parameters under the two approaches are estimated similarly, random effects variance estimates on a general note are

underestimated in the frequentist procedure in contrast to Bayesian approach (Manda and Leyland, 2007).

Type of place of residence had a significant relationship with malaria and anaemia infection among children in both years when controlled for other predictors. Children in rural areas were mostly affected with malaria and anaemia infection than children living in urban areas. This is in correspondent with other studies (Adebayo et al., 2016; Gaston et al., 2021). From our findings, malaria and anaemia were most prevalent in Kebbi, Sokoto, Zamfara, Benue, Ondo, Edo and Akwa-Ibom which are majorly rural. This can be a guide in state specific approach and campaign tactics that will facilitate the eradication of both diseases.

Also, Mother's highest educational level was significantly related to malaria and anaemia in both years. Children whose mothers are well educated were less affected by the two diseases compared to children from less educated or uneducated mothers. This could be attributed to the ability of mothers having access to educative material and programs that enlightens one on the prevention of these diseases. Unlike uneducated mothers, mothers who are educated are well informed and appreciative of health related issues (Roberts et al., 2020).

Furthermore, type of toilet facility was significantly related to anaemia in the different years. Children who use pit toilet were likely to be more anaemic than those who use flush toilet. This is because in pit toilet, there is a direct contact of the body and the heat from the pit. Also, there is danger of intestinal hook worm due to inadequate sanitation which can result to anaemia in infected children (Smith and Brooker, 2010; Roberts et al., 2020).

In the same vein, source of drinking water was found to be significantly related to the two diseases in both years. Where there is no proper drainage, there is high possibility of mosquito breeding which will in turn bite children and if not adequately taken care of can result to anaemia.

Malaria and anaemia were equally found to be highly and significantly correlated: . States that had high malaria prevalence also had high anaemia prevalence.

Chapter 6

Discussion, Conclusion and Recommendation

This research work aimed to understand and utilize already established statistical models for spatially mapping disease. These models, applied to malaria and anaemia, only took care of areal data. Here, we used 2010 and 2015 NMIS data.

In chapter 2, we explained in detail the methodology used in this work. While in chapter 3, we developed and applied appropriate statistical models to evaluate risk factors and variations in the spread of malaria or anaemia geographically. Also, an integrated framework of flexible models within the Bayesian hierarchical modelling framework was incorporated to ascertain the risk factors of malaria and anaemia in children between the ages of 6 to 59 months and to provide appropriate measures that will reduce the impacts of the diseases from a public health view. For the two diseases, children from rural areas and children whose mothers are illiterate are mostly affected. Control measures need to be used to monitor the spatial difference in malaria and anaemia in these regions. The odds of malaria and anaemia for children under the age of 5 increase with the child's age, mother's level of education, place of residence, source of drinking water, low income, type of toilet facility and presence of either of the diseases. Consideration should be given to contributing factors to spatial heterogeneity to focus on assessing local region-specific causes of anaemia and malaria in children.

Chapter 4 used the full Bayesian (FB) method under the Bayesian hierarchical modelling framework to model the prevalence of malaria and anaemia in Nigeria. Additionally, the Bayesian structured additive method was utilised to model malaria and anaemia risk factors. The result from

this study reveals a likely reduction in the prevalence of the two diseases if proper implementation of these commendations is done.

In chapter 5, malaria and anaemia were modelled by utilizing a semi-parametric joint model. Tangible information was obtained from a joint model on the association between malaria and anaemia. States that have a high rate of malaria have high rates of anaemia. In this chapter, the introduced model can only be used when one or more covariates have a non-linear relationship with the response variable and when there is the intention to model two or more disease outcomes jointly.

Spatial disease modelling as a subject cannot be addressed at once. Therefore, this study cannot be exhaustive. Up to now, most research works have modelled the covariate effect of the categorical covariates linearly by the normal prior. In future studies, consideration for modelling these effects should be done using a non-parametric fashion. It should be noted that boundary problems usually interfere with correct statistical parameter estimation in the spatial analysis due to boundaries cut off at two countries' borders with zero assumption of effect throughout the boundary. This is not always true because, most times, high prevalence at the border point may result from an outbreak. Future studies can focus on modelling the boundary effect.

Here, the research work puts into consideration a complete case analysis. A complete exclusion of children with missing entries was done. Incorporation of missing data analysis can be applied in further study to account for this deletion.

Differences are considered in the CAR model smoothing properties, though it is neighbours specific. Therefore, there are obvious consequences for CAR model users, given that selecting a

neighbourhood's weight matrices can remarkably affect the study's findings. Future studies can focus on the neighbourhood structure best for the CAR models.

In this study, the performance of non-parametric models employed for random effects was better than parametric models. Instead of the CAR model, a non-parametric spatial model can be developed and utilized. Though deviance information criteria (DIC) is the standard measurement method for model fit in Bayesian analysis, especially in mapping disease, it has some limitations. There is the problem of inconsistency; the model's effective number of parameters used for penalizing the complexity of the model is not unchanging to reparameterization. Future studies can improve the DIC or explore other model diagnostic techniques.

The findings from this work show that there will be likely a reduction in the spread of these diseases if commendations are adequately adhered. We recommend that the government and policy makers should focus on improving mother's education and standard of living. Also, pertaining to these diseases, there should be aggressive awareness on social television programs.

Appendix A

R codes for spatial model

```
#Packages required
library(readxl)
library(devtools)
library(foreign)
library(ggplot2)
require(RColorBrewer)
library(spdep)
require(INLA)
library(maptools)
library(rgdal)
library(tidyverse)
library(dplyr)
library(haven)

Besty<-read.csv(paste0(BASEDIR,file="Besty.csv"))
attach(Besty)

Map <- readOGR("C:/Users/Ibeji/Documents/NIG. STATE/STATE_BOUNDARY_shp.shp")
plot(Map)

nbnigeria<-poly2nb(Map)
nb2INLA(file="Map.graph",nbnigeria)

Given<-cbind(Besty,region=as.numeric(Besty$State1),
             region.struct=as.numeric(Besty$State1),
             region.unstruct=as.numeric(Besty$State1))
attach(Given)

library(survey)

nigeria.ana <- svydesign(ids = ~HV001, weights = HV005,
```

```

strata = HV022
, data = Besty, nest = T)

#GLM FOR MALARIA AND ANAEMIA

fit.glm.malaria<-with(Given,inla(MalariaresultRDT~factor(Agegroup)+factor(Electricity)+
factor(Watersource)+factor(Wallmaterial)+factor(Floormaterial)+
factor(Roofmaterial)+factor(Toilettype)+factor(Anaemicstatus)+
factor(Wealthindex)+factor(Mothereducvl)+factor(Kidsex)+
factor(Radio)+factor(Residence)+factor(Television)+factor(State1),
nigeria.ana,family="binomial",data=Given,control.compute=list(dic=1),
control.predictor = list(compute = TRUE), verbose = TRUE))

summary(fit.glm.malaria)

fit.glm.anaemia<-with(Given,inla(Anaemicstatus~factor(Agegroup)+factor(Electricity)+
factor(Watersource)+factor(Wallmaterial)+factor(Floormaterial)+
factor(Roofmaterial)+factor(Toilettype)+factor(MalariaresultRDT)+
factor(Wealthindex)+factor(Mothereducvl)+factor(Kidsex)+
factor(Roofmaterial)+factor(Floormaterial)+factor(Wallmaterial)+
factor(Radio)+factor(Residence)+factor(Television)+factor(State1),
nigeria.ana,family="binomial",data=Given,control.compute=list(dic=1),
control.predictor = list(compute = TRUE), verbose = TRUE))

summary(fit.glm.anaemia)

#GLMM FOR MALARIA AND ANAEMIA

inla.doc("iid")

names(inla.models())$prior

inla.models()$prior

inla.doc("priorname")

prior.prec<-list(prec = list(prior = "pc.prec", param = c(1,0.01)))

```

```

fit.glmm.malaria<-inla(MalariaresultRDT~factor(Agegroup)+factor(Electricity)+
  factor(Watersource)+factor(Wallmaterial)+factor(Floormaterial)+
  factor(Roofmaterial)+factor(Toilettype)+factor(Anaemicstatus)+
  factor(Wealthindex)+factor(Mothereducvl)+factor(Kidsex)+
  factor(Roofmaterial)+factor(Floormaterial)+factor(Wallmaterial)+
  factor(Radio)+factor(Residence)+factor(Television)+
  f(region.unstruct, model = "iid", hyper = prior.prec),
  family = "binomial", data = Besty, nigeria.ana,control.compute=list(dic=T),
  control.predictor = list(compute = TRUE), verbose = T)
summary(fit.glmm.malaria)

fit.glmm.anaemia<-inla(Anaemicstatus~factor(Agegroup)+factor(Electricity)+
  factor(Watersource)+factor(Wallmaterial)+factor(Floormaterial)+
  factor(Roofmaterial)+factor(Toilettype)+factor(MalariaresultRDT)+
  factor(Wealthindex)+factor(Mothereducvl)+factor(Kidsex)+
  factor(Roofmaterial)+factor(Floormaterial)+factor(Wallmaterial)+
  factor(Radio)+factor(Residence)+factor(Television)+
  f(region.unstruct, model = "iid", hyper = prior.prec),
  family = "binomial", data = Besty,
  nigeria.ana,control.compute=list(dic=T),control.predictor = list(compute = TRUE), verbose = T)
summary(fit.glmm.anaemia)

```

#CAR MODELS FOR MALARIA AND ANAEMIA

```

fit.malaria<-MalariaresultRDT~factor(Agegroup)+factor(Electricity)+factor(Watersource)+
  factor(Wallmaterial)+factor(Floormaterial)+factor(Roofmaterial)+
  factor(Toilettype)+factor(Anaemicstatus)+factor(Wealthindex)+
  factor(Mothereducvl)+factor(Kidsex)+factor(Roofmaterial)+
  factor(Floormaterial)+factor(Wallmaterial)+factor(Radio)+
  factor(Residence)+factor(Television)+f(region.struct, model = "besag",
  adjust.for.con.comp=TRUE, constr=TRUE, graph.file="nigeria.graph",
  scale.model = TRUE, hyper=prior.prec)

```



```
fit.car.malaria<-with(Given, inla( fit.malaria,family = "binomial", data = Given,
nigeria.ana,control.compute=list(dic=T), control.predictor = list(compute = TRUE), verbose = T))
summary(fit.car.malaria)
```

```
fit.anaemia<-Anaemicstatus~factor(Agegroup)+factor(Electricity)+factor(Watersource)+
factor(Wallmaterial)+factor(Floormaterial)+factor(Roofmaterial)+
factor(Toilettype)+factor(MalariaresultRDT)+factor(Wealthindex)+
factor(Mothereduclvl)+factor(Kidsex)+factor(Roofmaterial)+
factor(Floormaterial)+factor(Wallmaterial)+factor(Radio)+
factor(Residence)+factor(Television)+f(region.struct, model = "besag",
adjust.for.con.comp=TRUE, constr=TRUE, graph.file="nigeria.graph",
scale.model = TRUE, hyper=prior.prec)
```

```
fit.car.anaemia<-with(Given, inla( fit.anaemia,family = "binomial", data = Given,
nigeria.ana,control.compute=list(dic=T),
control.predictor = list(compute = TRUE), verbose = T))
summary(fit.car.anaemia)
```

#CONVOLUTION MODELS FOR MALARIA AND ANAEMIA

```
convolution.malaria<-MalariaresultRDT~factor(Agegroup)+factor(Electricity)
+factor(Watersource)+factor(Wallmaterial)+
factor(Floormaterial)+factor(Roofmaterial)+
factor(Toilettype)+factor(Anaemicstatus)+factor(Wealthindex)+
factor(Mothereduclvl)+factor(Kidsex)+factor(Roofmaterial)+
factor(Floormaterial)+factor(Wallmaterial)+factor(Radio)+
factor(Residence)+factor(Television)+ f(region.struct,
model = "besag",adjust.for.con.comp=TRUE, constr=TRUE,
graph.file="nigeria.graph",scale.model = TRUE, hyper = prior.prec)+
f(region.unstruct, model = "iid", hyper = prior.prec)
fit.convolution.malaria<- inla(convolution.malaria, family = "binomial",
data=Given, nigeria.ana,control.compute=list(dic=1,mlik=1,
```

```

      waic = 1), control.predictor = list(compute = TRUE), verbose = T)
summary(fit.convolution.malaria)

convolution.anaemia <- naemicstatus ~ factor(Agegroup) + factor(Electricity) + factor(Watersource) +
  factor(Wallmaterial) + factor(Floormaterial) + factor(Roofmaterial) +
  factor(Toiletttype) + factor(MalariaresultRDT) + factor(Wealthindex) +
  factor(Mothereducvl) + factor(Kidsex) + factor(Roofmaterial) +
  factor(Floormaterial) + factor(Wallmaterial) + factor(Radio) +
  factor(Residence) + factor(Television) + f(region.struct, model = "besag",
    adjust.for.con.comp = TRUE, constr = TRUE, graph.file = "nigeria.graph",
    scale.model = TRUE, hyper = prior.prec) +
  f(region.unstruct, model = "iid", hyper = prior.prec)

fit.convolution.anaemia <- inla(convolution.malaria, family = "binomial",
  data = Given, nigeria.ana, control.compute = list(dic = 1, mlik = 1,
    waic = 1), control.predictor = list(compute = TRUE), verbose = T)
summary(fit.convolution.anaemia)
exp(fit.convolution.anaemia$summary.fixed)

```

Appendix B

R codes for spatio-temporal model

```
rm(list=ls())

Given<-cbind(MD,region=as.numeric(MD$State1),
             region.struct=as.numeric(MD$State1),
             region.unstruct=as.numeric(MD$State1))

attach(Given)

library(survey)

nigeria.ana <- svydesign(ids = ~HV001, weights = HV005,
                       strata = HV022
                       , data = MDA1, nest = T)

attach(MDA1)

MDA$Year3<-as.numeric(MDA$Year)

Given$State2<-Given$State

Given$Year1<-Given$Year3

Given$Year2<-Given$Year3

#GLM FOR MALARIA AND ANAEMIA

formula1<-with(Given,inla(MalariaresultRDT~factor(Agegroup)+factor(Electricity)+
                        factor(Watersource)+factor(Wallmaterial)+factor(Floormaterial)+
                        factor(Roofmaterial)+factor(Anaemicstatus)+
                        factor(Wealthindex)+factor(MotherEduclvl)+factor(Kidsex)+
                        factor(Radio)+factor(Residence)+factor(Television)+factor(State1),
nigeria.ana,family="binomial",data=Given,control.compute=list(dic=1),
control.predictor = list(compute = TRUE), verbose = TRUE))

summary(formula1)

formula2<-MalariaresultRDT~factor(Agegroup)+factor(Electricity)+factor(Watersource)+
```

```

factor(Wallmaterial)+ factor(Floormaterial)+factor(Roofmaterial)+
factor(Anaemicstatus)+factor(Wealthindex)+
factor(MotherEducLvl)+factor(Kidsex)+factor(Roofmaterial)+
factor(Floormaterial)+factor(Wallmaterial)+factor(Radio)+
factor(Residence)+factor(Television)+f(State1,model="besag",
graph="Map")+f(State2,model="iid",graph="Map")+
as.factor(Year)#linear trend model

mod2<-inla(formula2,family="binomial",data=MDA,control.compute=list(dic=T,cpo=T),
verbose=F)
summary(mod2)

formula3<-MalariareultRDT~factor(Agegroup)+factor(Electricity)+factor(Watersource)+
factor(Wallmaterial)+ factor(Floormaterial)+factor(Roofmaterial)+
factor(Anaemicstatus)+factor(Wealthindex)+
factor(MotherEducLvl)+factor(Kidsex)+factor(Roofmaterial)+
factor(Floormaterial)+factor(Wallmaterial)+ factor(Radio)+
factor(Residence)+factor(Television)+f(State1,model="besag",
graph="Map")+f(State2,model="iid",graph="Map")+
f(Year,model="ar1")

mod3<-inla(formula3,family="binomial",data=MDA,control.compute=list(dic=T,
waic=T,cpo=T),verbose=F)
summary(mod3)

formula4<-MalariareultRDT~factor(Agegroup)+factor(Electricity)+factor(Watersource)
+factor(Wallmaterial)+ factor(Floormaterial)+factor(Roofmaterial)
+factor(Anaemicstatus)+factor(Wealthindex)+
factor(MotherEducLvl)+factor(Kidsex)+factor(Roofmaterial)+
factor(Floormaterial)+factor(Wallmaterial)+ factor(Radio)+
factor(Residence)+factor(Television)+
f(State1,model="besag",graph="Map")+
f(State2,model="iid",graph="Map")+

```

```

      f(State,model="iid",group      =      Year,control.group      =
list(model="ar1"),  adjust.for.con.comp =FALSE)#Interaction only

mod4<-
inla(formula4,family="binomial",data=MDA,control.compute=list(dic=T,cpo=T),verbose=F)

summary(mod4)

formula5<-MalariareultRDT~factor(Agegroup)+factor(Electricity)+factor(Watersource)+
      factor(Wallmaterial)+ factor(Floormaterial)+factor(Roofmaterial)+
      factor(Anaemicstatus)+factor(Wealthindex)+
      factor(MotherEduc1)+factor(Kidsex)+factor(Roofmaterial)
      +factor(Floormaterial)+factor(Wallmaterial)+ factor(Radio)+
      factor(Residence)+factor(Television)+
      f(State1,model="besag",graph="Map")+f(State2,model="iid",
      graph="Map")+ f(State,model="iid",group = Year,
      adjust.for.con.comp =FALSE)+as.factor(Year)

mod5<-inla(formula5,family="binomial",data=MDA,control.compute=list(dic=T,
      waic=T,cpo=T),verbose=F)

summary(mod5)

formula6<-Anaemicstatus~factor(Agegroup)+factor(Electricity)+factor(Watersource)
      +factor(Wallmaterial)+ factor(Floormaterial)+factor(Roofmaterial)+
      factor(MalariareultRDT)+factor(Wealthindex)+ factor(MotherEduc1)+
      +factor(Kidsex)+factor(Roofmaterial)+factor(Floormaterial)+
      factor(Wallmaterial)+ factor(Radio)+factor(Residence)+
      factor(Television)+ f(Year,model="ar1")+
      f(State,model="iid",group = Year,control.group = list(model="ar1"),
      adjust.for.con.comp =FALSE)

mod6<-
inla(formula6,family="binomial",data=MDA,control.compute=list(dic=T,cpo=T),verbose=F)

summary(mod6)

exp(mod6$summary.fixed)

```

```

formula7<-MalariareultRDT~factor(Agegroup)+factor(Electricity)+factor(Watersource)
      +factor(Wallmaterial)+factor(Floormaterial)+factor(Roofmaterial)
      +factor(Anaemicstatus)+factor(Wealthindex)+
      factor(MotherEduclvl)+factor(Kidsex)+factor(Roofmaterial)+
      factor(Floormaterial)+factor(Wallmaterial)+ factor(Radio)+
      factor(Residence)+factor(Television)+
f(State1,model="besag",graph="Map")+
f(State2,model="iid",graph="Map")+
f(Year,model="rw1")+f(State,model="iid",
group = Year,control.group = list(model="ar1"),
adjust.for.con.comp =FALSE)
mod7<-inla(formula7,family="binomial",data=MDA1,control.compute=list(dic=T),
      control.predictor = list(compute = TRUE), verbose = T)
summary(mod7)
exp(mod7$summary.fixed)

```

Appendix C

R codes for joint spatial model

```
#####Multivariate CAR model#####
```

```
model
```

```
{
```

```
#spline
```

```
for(i in 1: N)
```

```
{
```

```
for(l in 1:degree+1)
```

```
{
```

```
X[i,l]<-pow(Age[i],l-1)
```

```
}
```

```
}
```

```
for(i in 1: N)
```

```
{
```

```
for(k in 1:20)
```

```
{
```

```
u[i,k]<-(Age[i]-knot[k])*step(Age[i]-knot[k])
```

```
Z[i,k]<-pow(u[i,k],degree)
```

```
}
```

```
}
```

```
#likelihood
```

```
for(i in 1: N)
```

```
{
```

```
###No education=0,Primary=1,Secondary=2,Higher=3##
```

```
D.MotherEducvl1[i]<-equals(MotherEducvl[i],0)
```

```
D.MotherEducvl2[i]<-equals(MotherEducvl[i],1)
```

```
D.MotherEducvl3[i]<-equals(MotherEducvl[i],2)
```

```
###Wealthindex#####
```

```
###Poor=1,Middle=2,Rich=3#####
```

```
D.Wealthindex2[i]<-equals(Wealthindex[i],1)
```

```
D.Wealthindex3[i]<-equals(Wealthindex[i],2)
```

```
##Residence##
```

```
##Urban=1,Rural=2###
```

```
D.Residence[i]<-equals(Residence[i],1)
```

```
###Watersource###
```

```
###Tap/Other water=1,Well water=2###
```

```
D.Watersource[i]<-equals(Watersource[i],1)
```

```
###Kidsex###
```


###Male=1,Female=2###

D.Kidsex[i]<-equals(Kidsex[i],1)

##Electricity###

##No=0,Yes=1###

D.Electricity[i]<-equals(Electricity[i],1)

##Toiletttype##

###Flush/other toilet=1,Pit toilet=2###

D.Toiletttype[i]<-equals(Toiletttype[i],1)

##Radio##

###No=0,Yes=1###

D.Radio[i]<-equals(Radio[i],1)

##Television##

##No=0,Yes=1###

D.Television[i]<-equals(Television[i],1)

##Roofmaterial###

##Wood/other=1,Zinc/Metal=2###

D.Roofmaterial[i]<-equals(Roofmaterial[i],1)

##Floormaterial###

##Earth/Other=1,Concrete/Ceramics=2###

D.Floormaterial[i]<-equals(Floormaterial[i],1)

##Wallmaterial###

```
##Wood/Other=1,Cement/Bricks=2###
```

```
D.Wallmaterial[i]<-equals(Wallmaterial[i],1)
```

```
#for MalariareultRDT
```

```
MalariareultRDT[i]~dbern(p1[i])
```

```
p1[i]<-min(1,max(0,PMalariareultRDT[i]))
```

```
logit(PMalariareultRDT[i])<-beta1+edu1[1]*D.MotherEduc1v1[i]+edu1[2]*
```

```
D.MotherEduc1v2[i]+edu1[3]*D.MotherEduc1v3[i]+wealth1[2]*
```

```
D.Wealthindex2[i]+wealth1[3]*D.Wealthindex3[i]+res1*D.Residence[i]+
```

```
water1*D.Watersource[i]+sex1*D.Kidsex[i]+elect1*D.Electricity[i]+toilet1*D.Toilettype[i]+
```

```
rad1*D.Radio[i]+tel1*D.Television[i]+roof1*D.Roofmaterial[i]+floor1*D.Floormaterial[i]+
```

```
wall1*D.Wallmaterial[i]+U[STATE[i],1]+spline1[i]
```

```
spline1[i]<-inprod(b1[, Z[i, ])+inprod(betaS1[, X[i, ])
```

```
#for anaemicstatus
```

```
Anaemicstatus[i]~dbern(p2[i])
```

```
p2[i]<-min(0,max(1,PAnaemicstatus[i]))
```

```
logit(PAnaemicstatus[i])<-beta2+edu2[1]*D.MotherEduc1v1[i]+edu2[2]*
```

```
D.MotherEduc1v2[i]+edu2[3]*D.MotherEduc1v3[i]+wealth2[2]*D.Wealthindex2[i]+wealth2[3]
```

```
*D.Wealthindex3[i]+res2*D.Residence[i]+water2*D.Watersource[i]+sex2*D.Kidsex[i]+elect2*
```

```
D.Electricity[i]+toilet2*D.Toilettype[i]+rad2*D.Radio[i]+tel2*D.Television[i]
```

```
+roof2*D.Roofmaterial[i]+floor2*D.Floormaterial[i]+wall2*D.Wallmaterial[i]+
```

```
U[STATE[i],2]+spline2[i]
```

```
spline2[i]<-inprod(b2[, Z[i, ])+inprod(betaS2[, X[i, ])
```

```
}
```

```
#Anaemicstatus
```

```
edu2[4]<-0
```

```

wealth2[1]<-0

edu1[4]<-0
wealth1[1]<-0

#priors
beta2~dnorm(0.01,0.01)
beta1~dnorm(0.01,0.01)

#MotherEduclvl coefficients
for(j in 1: 3)
{
edu1[j]~dnorm(0.01,0.01)
edu2[j]~dnorm(0.01,0.01)
}

#Wealthindex coefficients
for(m in 2:3)
{
wealth1[m]~dnorm(0.01,0.01)
wealth2[m]~dnorm(0.01,0.01)
}

for(l in 1:degree+1)

{
betaS1[l]~dnorm(0,0.0001 )
}

```

```

#priorsplines
for(k in 1:20)
{
b1[k]~dnorm(0,taub1 )
}
taub1~dgamma(1000,0.001)
for(l in 1:degree+1)
{
betaS2[l]~dnorm(0,0.0001 )
}

```

```

#priorsplines
for(k in 1:20)
{
b2[k]~dnorm(0,taub2 )
}
taub2~dgamma(1000,0.001)

```

```

#ODDS ratios

```

```

#MotherEduclvl coefficients
for(j in 1:4)
{
ORedu1[j]<-exp(edu1[j])
ORedu2[j]<-exp(edu2[j])
}
#Wealthindex coefficients
for(k in 1:3)

```

```

{
ORwealth1[k]<-exp(wealth1[k])
ORwealth2[k]<-exp(wealth2[k])
}
ORres2~exp(res2)
ORwater2~exp(water2)
ORsex2~exp(sex2)
ORelect2~exp(elect2)
ORtoilet2~exp(toilet2)
ORfloor2~exp(floor2)
ORwall2~exp(wall2)
ORroof2~exp(roof2)
ORrad2~exp(rad2)
ORtel2~exp(tel2)
ORres1~exp(res1)
ORwater1~exp(water1)
ORsex1~exp(sex1)
ORElect1~exp(elect1)
ORtoilet1~exp(toilet1)
ORfloor1~exp(floor1)
ORwall1~exp(wall1)
ORroof1~exp(roof1)
ORrad1~exp(rad1)
ORTel1~exp(tel1)
#prior
for(i in 1: N)
{
for(j in 1: 37)

```

```

{
PM[j,i]<-(PMalariaresultRDT[i])*(equals(STATE[i],j))
PA[j,i]<-(PAnaemicstatus[i])*(equals(STATE[i],j))
}
}
for(j in 1: 37)
{
for(i in 1: N)
{
STA[j,i]<-equals(STATE[i],j)
}
number[j]<-sum(STA[j,])
PCM[j]<-sum(PM[j,])/number[j]
PCA[j]<-sum(PA[j,])/number[j]
}

#unstructured prior
for(i in 1:NSTATE)
{

U[i, 1:Ndiseases] ~ dmnorm(zero[], tau[ , ])

}

# Precision matrix of MV Normal
tau[1:Ndiseases, 1:Ndiseases] ~ dwish(Q[ , ], Ndiseases)
# Covariance matrix of MV Normal
sigma2.U[1:2, 1:2] <- inverse(tau[ , ])

```

```

sigma.U[1] <- sqrt(sigma2.U[1, 1])
sigma.U[2] <- sqrt(sigma2.U[2, 2])

# within-STATE correlation between unstructured component of variation in
#MalariaresultRDT and Anaemicstatus
corr.U <- sigma2.U[1, 2] / (sigma.U[1] * sigma.U[2])
}

#Data

#INITIALS

```

References

- Abegunde, D. and Stanciole, A. (2006). An estimation of the economic impact of chronic noncommunicable diseases in selected countries. *World Health Organization, Department of Chronic Diseases and Health Promotion*, 2006.
- Adebayo, S. B., Gayawan, E., Heumann, C. and Seiler, C. (2016). Joint modeling of Anaemia and Malaria in children under five in Nigeria. *Spatial and spatio-temporal epidemiology*, 17, 105-115.
- Adeyemi, E. O. (2021). Predictors of Malaria-Anemia Comorbidity among Under Five Children in Nigeria: A Cross Sectional Study.
- Adigun, A. B., Gajere, E. N., Oresanya, O. and Vounatsou, P. (2015). Malaria risk in Nigeria: Bayesian geostatistical modelling of 2010 malaria indicator survey data. *Malaria journal*, 14, 156.
- Agagliati, F., Silvestro, E., Denina, M., Garazzino, S., Mignone, F., Zoppo, M., Pruccoli, G. and Scolfaro, C. (2022). Imported malaria in children: A 13 years retrospective study. *Travel Medicine and Infectious Disease*, 46, 102273.
- Agresti, A. 2007. An Introduction to Categorical Data Analysis Second Edition Second edi., Hoboken, New Jersey, Canda: JohnWiley & Sons. Inc., Hoboken, New Je.
- Ahmad, F., Goparaju, L. and Qayum, A. (2017). Studying malaria epidemic for vulnerability zones: Multi-criteria approach of geospatial tools. *Journal of Geoscience and Environment Protection*, 5, 30-53.
- Ajakaye, O. G. and Ibukunoluwa, M. R. (2020). Prevalence and risk of malaria, anemia and malnutrition among children in IDPs camp in Edo State, Nigeria. *Parasite Epidemiology and Control*, 8, e00127.
- Akinbo, F. O., Omoregie, R., Mordi, R. and Okaka, C. E. (2009). Prevalence of malaria and Anemia among young children in a tertiary Hospital in Benin City, Edo state, Nigeria. *Fooyin Journal of Health Sciences*, 1, 81-84.
- Akinyemi, Y. C. (2021). Spatiotemporal Patterns and Determinants of Reproductive Health Care Utilization in Nigeria: 2008–2018. *Papers in Applied Geography*, 7, 453-481.
- Akogun, O. B. (2008). *A study on presumptive diagnosis and home management of childhood malaria among nomadic fulani in Demsa, Nigeria*. University of the Western Cape.
- Anaemias, W. N. (2017). Tools for effective prevention and control. *World Health Organization*, 1-83.
- Aregbeshola, B. S., Onifade, O. M. and Awuviry-Newton, K. 2021. Prevalence and correlates of anaemia among children aged 6 to 59 months in Nigeria. *World Nutrition*, 12, 58-74.
- Assunção, R. M. (2003). Space varying coefficient models for small area data. *Environmetrics: The official journal of the International Environmetrics Society*, 14, 453-473.
- Azzalini, A. (2020). Further Results on a Class of Distributions which Includes the Normal Ones—Looking Back. *Statistica*, 80, 161-175.
- Bahati, Y. L., Delanghe, J., Balaluka, G. B. and Philippé, J. (2022). Exploration of the relationship between anemia and iron and zinc deficiencies in children under 5 years of age living in the malaria endemic area of South Kivu/Democratic Republic of Congo. *Annals of Hematology*, 1-9.
- Bamgbose, T. (2021). *Analysis of Malaria-Induced Anemia and Gender Differences Among Children in Nigeria*. Walden University.

- Battle, K. E., Lucas, T. C., Nguyen, M., Howes, R. E., Nandi, A. K., Twohig, K. A., Pfeffer, D. A., Cameron, E., Rao, P. C. and Casey, D. (2019). Mapping the global endemicity and clinical burden of *Plasmodium vivax*, 2000–17: a spatial and temporal modelling study. *The Lancet*, 394, 332-343.
- Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. and Songini, M. (1995). Bayesian analysis of space—time variation in disease risk. *Statistics in medicine*, 14, 2433-2443.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36, 192-225.
- Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43, 1-20.
- Best, N., Cockings, S., Bennett, J., Wakefield, J. and Elliott, P. (2001). Ecological regression analysis of environmental benzene exposure and childhood leukaemia: sensitivity to data inaccuracies, geographical scale and ecological bias. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164, 155-174.
- Blangiardo, M. and Cameletti, M. (2015). Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley & Sons, New York.
- Borderon, M., Ludwig, A. and Brazeau, S. 2.5 Vulnerable Populations. https://geography.univie.ac.at/fileadmin/user_upload/a_population/News/2022_Brazeau-Odgen_Chapter-2.5. August, accessed 2022
- Cao, G., Wang, Y., Wu, Y., Jing, W., Liu, J. and Liu, M. (2022). Prevalence of anemia among people living with HIV: A systematic review and meta-analysis. *eClinicalMedicine*, 44, 101283.
- Cappellini, M., Musallam, K. and Taher, A. (2020). Iron deficiency anaemia revisited. *Journal of internal medicine*, 287, 153-170.
- Carlin, B. P. and Banerjee, S. Models for Spatio-Temporally Correlated Survival Data. Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting, 2003. Oxford University Press, 45.
- Chao, D. L., Oron, A. P., Chabot-Couture, G., Sopekan, A., Nnebe-Agumadu, U., Bates, I., Piel, F. B. and Nnodu, O. E. (2022). The contribution of malaria and sickle cell disease to anaemia among children aged 6 to 59 months in Nigeria: A secondary analysis using data from the 2018 Demographic and Health Survey. *medRxiv*.
- Chesnaye, N. C., Tripepi, G., Dekker, F. W., Zocalli, C., Zwinderman, A. H. and Jager, K. J. (2020). An introduction to joint models-applications in nephrology. *Clinical kidney journal*, 13(2), 143-149.
- Childs, D., Cattadori, I., Suwonkerd, W., Prajakwong, S. and Boots, M. (2006). Spatiotemporal patterns of malaria incidence in northern Thailand. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100, 623-631.
- Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 71, 319-392.
- Chukwuekezie, O., Nwosu, E., Nwangwu, U., Dogunro, F., Onwude, C., Agashi, N., Ezihe, E., Anioke, C., Anokwu, S. and Eloy, E. (2020). Resistance status of *Anopheles gambiae* (sl) to four commonly used insecticides for malaria vector control in South-East Nigeria. *Parasites & Vectors*, 13, 1-10.

- Clayton, D. and Kaldor, J. 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 671-681.
- Clements, A. C., Barnett, A. G., Cheng, Z. W., Snow, R. W. and Zhou, H. N. (2009). Space-time variation of malaria incidence in Yunnan province, China. *Malaria journal*, 8, 1-12.
- Cohee, L. M., Opondo, C., Clarke, S. E., Halliday, K. E., Cano, J., Shipper, A. G., Barger-Kamate, B., Djimde, A., Diarra, S. and Dokras, A. (2020). Preventive malaria treatment among school-aged children in sub-Saharan Africa: a systematic review and meta-analyses. *The Lancet Global Health*, 8, e1499-e1511.
- Cressie, N. 1993. Statistics for Spatial Data: Wiley Series in Probability and Statistics.(1993). *Google Scholar*
- Deribew, K., Tekeste, Z. and Petros, B. (2013). Urinary schistosomiasis and malaria associated anaemia in Ethiopia. *Asian Pacific journal of tropical biomedicine*, 3(4), 307-310.
- Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47, 299-350.
- Dimaggio, C. (2012). Spatial Epidemiology Notes. *Applications and Vignettes in R*.
- Douglas, N. M., Burkot, T. R. and Price, R. N. (2021). Malaria eradication revisited. *International journal of epidemiology*.
- Earnest, A., Morgan, G., Mengersen, K., Ryan, L., Summerhayes, R. and Beard, J. (2007). Evaluating the effect of neighbourhood weight matrices on smoothing properties of Conditional Autoregressive (CAR) models. *International journal of health geographics*, 6, 1-12.
- Egbon, O. (2021). Bayesian Analysis of the Joint Risk of Malaria and Anemia Among Under-age Ve Children in Nigeria.
- Eilers, P. H. And Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical science*, 11, 89-121.
- Ekvall, H. (2003). Malaria and anemia. *Current opinion in hematology*, 10, 108-114.
- Fahrmeir, L., Tutz, G., Hennevoel, W. and Salem, E. (1994). *Multivariate statistical modelling based on generalized linear models*, Springer.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 209-230.
- Fischer, P. R. (2021). Malaria and Anemia—Chemoprophylaxis Helps. *Infectious Disease Alert*, 40.
- Gaetan, C. and Guyon, X. (2010). *Spatial statistics and modeling*, Springer.
- Gaston, R. T., Ramroop, S. and Habyarimana, F. (2018). Determinants of factors associated with anemia among children under five years in Lesotho. *African Population Studies*, 32.
- Gaston, R. T., Ramroop, S. and Habyarimana, F. (2021). Joint modelling of malaria and anaemia in children less than five years of age in Malawi. *Heliyon*, 7, e06899.
- Gayawan, E., Arogundade, E. D. and Adebayo, S. B. (2014). Possible determinants and spatial patterns of anaemia among young children in Nigeria: a Bayesian semi-parametric modelling. *International health*, 6, 35-45.
- Gayawan, E., Egbon, O. A. and Adebayo, S. B. (2022). Spatial modelling of the joint burden of malaria and anaemia co-morbidity in children: A Bayesian geosadditive perspective. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 1-16.
- Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4, 11-15.

- Gething, P. W., Casey, D. C., Weiss, D. J., Bisanzio, D., Bhatt, S., Cameron, E., Battle, K. E., Dalrymple, U., Rozier, J. and Rao, P. C. (2016). Mapping *Plasmodium falciparum* mortality in Africa between 1990 and 2015. *New England Journal of Medicine*, 375, 2435-2445.
- Green, P. J. and Richardson, S. (2002). Hidden Markov models and disease mapping. *Journal of the American statistical association*, 97, 1055-1070.
- Green, P. J. and Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*, Crc Press.
- Griffith, D. A. (2002). A spatial filtering specification for the auto-Poisson model. *Statistics & probability letters*, 58, 245-251.
- Gupta, S., Haldar, P., Singh, A., Malhotra, S. and Kant, S. (2022). Prevalence of serum cobalamin and folate deficiency among children aged 6–59 months: A hospital-based cross-sectional study from Northern India. *Journal of Family Medicine and Primary Care*, 11, 1063-1069.
- Hailegebreel, S., Nigatu, A. M., Mekonnen, Z. A. and Endehabtu, B. F. (2021). Spatio-temporal distribution and associated factors of anaemia among children aged 6–59 months in Ethiopia: a spatial and multilevel analysis based on the EDHS 2005–2016. *BMJ open*, 11, e045544.
- Hall, P. and Patil, P. (1995). On wavelet methods for estimating smooth functions. *Bernoulli*, 41-58.
- Hastie, T. and Tibshirani, R. (1990). Generalized additive models, volume 43: CRC press.
- Hastie, T. and Tibshirani, R. (1995). Generalized additive models for medical research. *Statistical methods in medical research*, 4, 187-196.
- Hastie, T., Tibshirani, R. and Friedman, J. (2001). The elements of statistical learning. Springer series in statistics. *New York, NY, USA*.
- Henrici, R. C., Sautter, C. L., Bond, C., Opoka, R. O., Namazzi, R., Datta, D., Ware, R. E., Conroy, A. L. and John, C. C. (2021). Decreased parasite burden and altered host response in children with sickle cell anemia and severe anemia with malaria. *Blood advances*, 5, 4710-4720.
- Hinton, G. E. and Van Camp, D. 1993. Keeping the neural networks simple by minimizing the description length of the weights. Proceedings of the sixth annual conference on Computational learning theory, 5-13.
- Hoeting, J. A., Leecaster, M. and Bowden, D. (2000). An improved model for spatially correlated binary responses. *Journal of agricultural, biological, and environmental statistics*, 102-114.
- Ibeji, J. U., Mwambi, H. and Iddrisu, A.-K. (2022). Spatial variation and risk factors of malaria and anaemia among children aged 0 to 59 months: a cross-sectional study of 2010 and 2015 datasets. *Scientific Reports*, 12, 1-15.
- Jana, S., Fu, S. H., Gelband, H., Brown, P. and Jha, P. (2022). Spatio-temporal modelling of malaria mortality in India from 2004 to 2013 from the Million Death Study. *Malaria journal*, 21, 1-10.
- Johnson, K. and Way, A. (2006). Risk factors for HIV infection in a national adult population: evidence from the 2003 Kenya Demographic and Health Survey. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 42, 627-636.
- Kain, K. C., Harrington, M. A., Tennyson, S. and Keystone, J. S. (1998). Imported malaria: prospective analysis of problems in diagnosis and management. *Clinical infectious diseases*, 27, 142-149.

- Kammann, E. and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52, 1-18.
- Karagöl, C. and Yiğit, M. (2022). Evaluation of Clinical and Laboratory Findings and Diagnostic Difficulties in Children with Vitamin B12 Deficiency. *Pediatric Practice and Research*, 10, 1, 6-10.
- Kassebaum, N. J. and Collaborators, G. A. (2016). The global burden of anemia. *Hematology/oncology clinics of North America*, 30, 247-308.
- Kazembe, L. N., Chirwa, T. F., Simbeye, J. S. and Namangale, J. J. (2008). Applications of Bayesian approach in modelling risk of malaria-related hospital mortality. *BMC medical research methodology*, 8, 1-14.
- Kazembe, L. N. and Namangale, J. J. (2007). A Bayesian multinomial model to analyse spatial patterns of childhood co-morbidity in Malawi. *European journal of epidemiology*, 22, 545-556.
- Keating, E. M., Chiume, M., Fitzgerald, E., Mgusha, Y., Mvalo, T., Fino, N., Crouse, H. L., Eckerle, M., Gorman, K. and Ciccone, E. J. (2021). Blood transfusion and mortality in children with severe anaemia in a malaria-endemic region. *Paediatrics and International Child Health*, 41, 129-136.
- Kinyoki, D., Osgood-Zimmerman, A. E., Bhattacharjee, N. V., Kassebaum, N. J. and Hay, S. I. (2021). Anemia prevalence in women of reproductive age in low-and middle-income countries between 2000 and 2018. *Nature medicine*, 27, 1761-1782.
- Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in medicine*, 19, 2555-2567.
- Kofoed, P.-E., Rodrigues, A., C6, F., Hedegaard, K., Rombo, L. and Aaby, P. (2004). Which children come to the health centre for treatment of malaria? *Acta tropica*, 90, 17-22.
- Korenromp, E. L., Armstrong-Schellenberg, J. R., Williams, B. G., Nahlen, B. L. and Snow, R. W. (2004). Impact of malaria control on childhood anaemia in Africa—a quantitative review. *Tropical Medicine & International Health*, 9, 1050-1065.
- Langford, I. H., Leyland, A. H., Rasbash, J. and Goldstein, H. (1999). Multilevel modelling of the geographical distributions of diseases. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48, 253-268.
- Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modelling. *The annals of statistics*, 1222-1235.
- Lawson, A. B., Browne, W. J. and Rodeiro, C. L. V. (2003). *Disease mapping with WinBUGS and MLwiN*, John Wiley & Sons.
- Leal, L. P., Batista Filho, M., De Lira, P. I. C., Figueiroa, J. N. and Os6rio, M. M. (2012). Temporal trends and anaemia-associated factors in 6-to 59-month-old children in Northeast Brazil. *Public Health Nutrition*, 15, 1645-1652.
- Leroux, B. G., Lei, X. and Breslow, N. (2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence. *Statistical models in epidemiology, the environment, and clinical trials*. Springer.
- Lu, H., Reilly, C. S., Banerjee, S. and Carlin, B. P. (2007). Bayesian areal wombling via adjacency modeling. *Environmental and ecological statistics*, 14, 433-452.
- Mabaso, M. L., Vounatsou, P., Midzi, S., Da Silva, J. and Smith, T. (2006). Spatio-temporal analysis of the role of climate in inter-annual variation of malaria incidence in Zimbabwe. *International Journal of Health Geographics*, 5, 1-9.

- Macnab, Y. C. (2010). On Bayesian shared component disease mapping and ecological regression with errors in covariates. *Statistics in medicine*, 29, 1239-1249.
- Macnab, Y. C. and Dean, C. (2002). Spatio-temporal modelling of rates for the construction of disease maps. *Statistics in medicine*, 21, 347-358.
- Magalhaes, R. J. S. and Clements, A. C. (2011). Mapping the risk of anaemia in preschool-age children: the contribution of malnutrition, malaria, and helminth infections in West Africa. *PLoS Med*, 8, e1000438.
- Malizia, V., Giardina, F., De Vlas, S. J. and Coffeng, L. E. (2022). Appropriateness of the current parasitological control target for hookworm morbidity: a statistical analysis of individual-level data. *medRxiv*.
- Manda, S. O. and Leyland, A. (2007). An empirical comparison of maximum likelihood and Bayesian estimation methods for multivariate disease mapping: theory and methods. *South African Statistical Journal*, 41, 1-21.
- Mardia, K. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis*, 24, 265-284.
- Martins, T. G., Simpson, D., Lindgren, F. and Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis*, 67, 68-83.
- Mbunga, B. K., Mapatano, M. A., Strand, T. A., Gjengedal, E. L. F., Akilimali, P. Z. and Engebretsen, I. M. S. (2021). Prevalence of anemia, iron-deficiency anemia, and associated factors among children aged 1–5 years in the rural, malaria-endemic setting of Popokabaka, Democratic Republic of Congo: a cross-sectional study. *Nutrients*, 13, 1010.
- Mccullagh, P. and Nelder, J. A. (1989). Generalized linear models. Chapman and Hall. London, UK.
- Mclean, E., Cogswell, M., Egli, I., Wojdyla, D. and De Benoist, B. (2009). Worldwide prevalence of anaemia, WHO vitamin and mineral nutrition information system, 1993–2005. *Public health nutrition*, 12, 444-454.
- Mdakane, B. P. (2019). *Spatial and spatio-temporal modeling and mapping of self-reported health among individuals between the ages of 15-49 years in South Africa*.
- Milazzo, A. (2014). Son preference, fertility and family structure: Evidence from reproductive behavior among Nigerian women. *World Bank Policy Research Working Paper*.
- Minka, T. P. (2001). *A family of algorithms for approximate Bayesian inference*. Massachusetts Institute of Technology.
- Montana, L., Neuman, M. and Mishra, V. (2007). Spatial modeling of HIV prevalence in kenya.
- Mrimi, E. C., Palmeirim, M. S., Minja, E. G., Long, K. Z. and Keiser, J. (2022). Malnutrition, anemia, micronutrient deficiency and parasitic infections among schoolchildren in rural Tanzania. *PLoS neglected tropical diseases*, 16, e0010261.
- National Population Commission. (2010). National Malaria Control Programme, ICF International. *Nigeria malaria indicator survey*.
- Nigeria. National Population Commission, National Malaria Control Programme (Nigeria), ICF International (Firm) and MEASURE DHS (Program). (2012). *Nigeria Malaria Indicator Survey 2010*. NPC, NMCP and ICF International.
- National Population Commission. (2013). *Nigeria demographic and health survey 2013*. National Population Commission, ICF International.
- Neuhaus, A., Augustin, T., Heumann, C. and Daumer, D. (2009). A review on joint models in biometricals research. *Journal of Statistical Theory and Practice*, 3(4), 855-868.

- Newton, C., Warn, P., Winstanley, P., Peshu, N., Snow, R., Pasvol, G. and Marsh, K. (1997). Severe anaemia in children living in a malaria endemic area of Kenya. *Tropical medicine & international health*, 2, 165-178.
- Ngesa, O., Mwambi, H. and Achia, T. (2014). Bayesian spatial semi-parametric modeling of HIV variation in Kenya. *PloS one*, 9, e103299.
- NMC, P. 2015. Nigeria MIS final report. Federal Republic of Nigeria Abuja.
- Obasohan, P. E., Walters, S. J., Jacques, R. and Khatab, K. (2022). Individual, household and area predictors of anaemia among children aged 6–59 months in Nigeria. *Public Health in Practice*, 100229.
- Oguoma, V. M., Anyasodor, A. E., Adeleye, A. O., Eneanya, O. A. and Mbanefo, E. C. (2020). Multilevel modelling of the risk of malaria among children aged under five years in Nigeria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*.
- Oguoma, V. M., Anyasodor, A. E., Adeleye, A. O., Eneanya, O. A. and Mbanefo, E. C. (2021). Multilevel modelling of the risk of malaria among children aged under five years in Nigeria. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 115, 482-494.
- Okango, E., Mwambi, H., Ngesa, O. and Achia, T. (2015). Semi-parametric spatial joint modeling of HIV and HSV-2 among women in Kenya. *PloS one*, 10, e0135212.
- Okell, L. C., Kwambai, T., Dhabangi, A., Khairallah, C., Nkosi-Gondwe, T., Opoka, R., Mousa, A., Köhl, M.-J., Lucas, T. C. and Idro, R. (2022). Projected health impact of post-discharge malaria chemoprevention among children under the age of five years with severe malarial anaemia in Africa: a modelling analysis. *medRxiv*.
- Okunlola, O. A. and Oyeyemi, O. T. (2019). Spatio-temporal analysis of association between incidence of malaria and environmental predictors of malaria transmission in Nigeria. *Scientific reports*, 9, 1-11.
- Okunlola, O. A., Oyeyemi, O. T. and Lukman, A. F. (2021). Modeling the relationship between malaria prevalence and insecticide-treated bed net coverage in Nigeria using a Bayesian spatial generalized linear mixed model with a Leroux prior. *Epidemiology and Health*, 43.
- Olanrewaju, S., Olafioye, S. and Oguntade, E. (2020). Modelling Nigeria Population Growth: A Trend Analysis Approach. *International Journal of Innovative Science and Research Technology*, 5, 52-64.
- Organization, W. H. (2008). Worldwide prevalence of anaemia 1993-2005: WHO global database on anaemia.
- Organization, W. H. (2011). Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. World Health Organization.
- Organization, W. H. (2014). Global nutrition targets 2025: Stunting policy brief. World Health Organization.
- Organization, W. H. (2015a). *Investing to overcome the global impact of neglected tropical diseases: third WHO report on neglected tropical diseases 2015*, World Health Organization.
- Organization, W. H. (2015b). *World health statistics 2015*, World Health Organization.
- Organization, W. H. (2016). *Guideline daily iron supplementation in infants and children*, World Health Organization.
- Organization, W. H. (2020). World malaria report 2019. 2019. Reference Source. <https://www.who.int/malaria/publications/worldmalaria-report-2019/en>.

- Organization, W. H. (2021a). WHO malaria policy advisory group (MPAG) meeting: meeting report, April 2021.
- Organization, W. H. (2021b). World malaria report 2021.
- Ozano, K. (2022). Interventions Aimed at Preventing, Detecting and Treating Malaria, TB and HIV in Nigeria.
- Papaioannou, I., Utzinger, J. and Vounatsou, P. (2019). Malaria-anemia comorbidity prevalence as a measure of malaria-related deaths in sub-Saharan Africa. *Scientific reports*, 9, 1-9.
- Pasricha, S.-R., Armitage, A. E., Prentice, A. M. and Drakesmith, H. (2018a). Reducing anaemia in low income countries: control of infection is essential. *BMJ*, 362.
- Pasricha, S.-R., Colman, K., Centeno-Tablante, E., Garcia-Casal, M.-N. and Peña-Rosas, J.-P. (2018b). Revisiting WHO haemoglobin thresholds to define anaemia in clinical medicine and public health. *The Lancet Haematology*, 5, e60-e62.
- Petry, N., Jallow, B., Sawo, Y., Daboe, M. K., Barrow, S., Sarr, A., Pa, O. C. and Wirth, J. P. (2019). Micronutrient deficiencies, nutritional status and the determinants of anemia in children 0-59 months of age and non-pregnant women of reproductive age in the Gambia. *Nutrients*, 11(10), 2275.
- Powers, J. M. (2022). Nutritional anemias. *Lanzkowsky's Manual of Pediatric Hematology and Oncology*. Elsevier.
- Ra, P. K., Nathawat, M. S. Onagh, M. (2012). Application of multiple linear regression model through GIS and remote sensing for malaria mapping in Varanasi District, INDIA. *Health Science Journal*, 6, 731.
- Rahi, M., Das, P. and Sharma, A. (2021). Malaria elimination in India requires additional surveillance mechanisms. *Journal of Public Health*.
- Rezaeian, M., Dunn, G., St Leger, S. and Appleby, L. (2007). Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *Journal of Epidemiology & Community Health*, 61, 98-102.
- Riedel, N., Vounatsou, P., Miller, J. M., Gosoni, L., Chizema-Kawesha, E., Mukonka, V. and Steketee, R. W. (2010). Geographical patterns and predictors of malaria risk in Zambia: Bayesian geostatistical modelling of the 2006 Zambia national malaria indicator survey (ZMIS). *Malaria journal*, 9, 1-13.
- Roberts, D. J., Matthews, G., Snow, R. W., Zewotir, T. and Sartorius, B. (2020). Investigating the spatial variation and risk factors of childhood anaemia in four sub-Saharan African countries. *BMC Public Health*, 20, 126.
- Roche, M. and Layrisse, M. (1966). The nature and causes of "hookworm anaemia.". *American Journal of Tropical Medicine and Hygiene*, 15, 1031-102.
- Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71, 319-392.
- Rumisha, S. F., Shayo, E. H. and Mboera, L. E. (2019). Spatio-temporal prevalence of malaria and anaemia in relation to agro-ecosystems in Mvomero district, Tanzania. *Malaria journal*, 18, 1-14.
- Sardar, S., Abdurabu, M., Abdelhadi, A., Habib, M. B., Jamshaid, M. B., Hajjar, A. H., Ageila, M. A., Abdalla, T., Kartha, A. and Farooqui, K. (2021). Artesunate-induced hemolysis in severe complicated malaria—A diagnostic challenge: A case report and literature review of anemia in malaria. *IDCases*, 25, e01234.

- Schumacher, R.-F. and Spinelli, E. (2012). Malaria in children. *Mediterranean journal of hematology and infectious diseases*, 4.
- Sherman, M. (2011). *Spatial statistics and spatio-temporal data: covariance functions and directional properties*, John Wiley & Sons.
- Smith, J. L. and Brooker, S. (2010). Impact of hookworm infection and deworming on anaemia in non-pregnant populations: a systematic review. *Tropical medicine & international health*, 15, 776-795.
- Soares Magalhães, R. J. and Clements, A. C. 2011. Mapping the risk of anaemia in preschool-age children: the contribution of malnutrition, malaria, and helminth infections in West Africa. *PLoS medicine*, 8, e1000438.
- Spiegelhalter, D., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2003). Bayesian measures of model complexity and fit. *Quality control and applied statistics*, 48, 431-432.
- Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. WinBUGS user manual, Version 1.4. 3, (2007). URL <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>.
- 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, 583-639.
- Starck, T., Bulstra, C. A., Tinto, H., Rouamba, T., Sie, A., Jaenisch, T. and Bärnighausen, T. (2021). The effect of malaria on haemoglobin concentrations: a nationally representative household fixed-effects study of 17,599 children under 5 years of age in Burkina Faso. *Malaria journal*, 20, 1-10.
- Steketee, R. W., Choi, M., Linn, A., Florey, L., Murphy, M. and Panjabi, R. (2021). World Malaria Day 2021: Commemorating 15 Years of Contribution by the United States President's Malaria Initiative. *The American journal of tropical medicine and hygiene*, 104, 1955.
- Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association*, 81(393):82-86.
- Ugwu, C. L. J. and Zewotir, T. (2020a). Evaluating the Effects of Climate and Environmental Factors on Under-5 Children Malaria Spatial Distribution Using Generalized Additive Models (GAMs). *Journal of Epidemiology and Global Health*.
- Ugwu, C. L. J. and Zewotir, T. (2020b). Spatial distribution and sociodemographic risk factors of malaria in Nigerian children less than 5 years old. *Geospatial Health*, 15.
- Ugwu, C. L. J. and Zewotir, T. 2020c. Spatial heterogeneity and determinants of childhood anaemia in Nigeria.
- Ugwu, C. L. J. and Zewotir, T. T. (2018). Using mixed effects logistic regression models for complex survey data on malaria rapid diagnostic test results. *Malaria journal*, 17, 453.
- Van Den Elsen, L. W., Verhasselt, V. and Egwang, T. (2020). Malaria antigen shedding in the breast milk of mothers from a region with endemic malaria. *JAMA pediatrics*, 174, 297-298.
- Waller, L. A., Carlin, B. P., Xia, H. and Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical association*, 92, 607-617.
- Waller, L. A. and Gotway, C. A. (2004). *Applied spatial statistics for public health data*, John Wiley & Sons.
- Wand, H., Whitaker, C. and Ramjee, G. (2011). Geoadditive models to assess spatial variation of HIV infections among women in local communities of Durban, South Africa. *International Journal of Health Geographics*, 10, 1-9.

- Weiss, D. J., Lucas, T. C., Nguyen, M., Nandi, A. K., Bisanzio, D., Battle, K. E., Cameron, E., Twohig, K. A., Pfeffer, D. A. and Rozier, J. A. (2019). Mapping the global prevalence, incidence, and mortality of *Plasmodium falciparum*, 2000–17: a spatial and temporal modelling study. *The Lancet*, 394, 322-331.
- Weze, K., Abioye, A. I., Obiajunwa, C. and Omotayo, M. (2021). Spatio-temporal trends in anaemia among pregnant women, adolescents and preschool children in sub-Saharan Africa. *Public health nutrition*, 24, 3648-3661.
- White, N. J. 2018. Anaemia and malaria. *Malaria journal*, 17, 1-17.
- Zulkarnain, M. (2022). Plagiarism_Risk Factors for the Incidence of Anemia in Elementary-School Children Living in Malaria-Endemic Regions.